2021-06

VOT and F0 in the production and perception of Swahili obstruents: From the island to the coast to the inland region

Alsamaani, Mohammad

Alsamaani, M. (2021). VOT and F0 in the production and perception of Swahili obstruents: from the island to the coast to the inland region (Unpublished doctoral thesis). University of Calgary, Calgary, AB.
http://hdl.handle.net/1880/113605
doctoral thesis

University of Calgary graduate students retain copyright ownership and moral rights for their thesis. You may use this material in any way that is permitted by the Copyright Act or through licensing that has been assigned to the document. For uses that are not allowable under copyright legislation or licensing, you are required to seek permission.

Downloaded from PRISM: https://prism.ucalgary.ca
VOT and F0 in the production and perception of Swahili obstruents: From the island to the coast to the inland region

by

Mohammad A. Alsamaani

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN
PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN LINGUISTICS

CALGARY, ALBERTA

JUNE, 2021

© Mohammad A. Alsamaani 2021
Abstract
The status of aspiration in Swahili has received conflicting historical and linguistic accounts. To date, it is not fully understood if this laryngeal setting in the language’s four voiceless obstruents (/p/, /t/, /k/, and /tʃ/) is phonemic or allophonic. This dissertation analyzes the phonetic laryngeal variations in four Swahili varieties, which are spoken in East Africa as a first language (Zanzibar in Tanzania and Mombasa in Kenya) or as a second language (Iringa in Tanzania and Nairobi in Kenya). Two experiments, one in production and one in perception, examined the acoustic cues of voice-onset time (VOT) and fundamental frequency (F0) to investigate how speakers employ these language-specific details. A total of 98 participants (male and female) took part in these experiments.

This dissertation first explores statistically the productions of real words by all subjects. Linear mixed-effects models indicate that the phonetic cue that accounts for more variance in the data is VOT, with little significance found for F0. However, the VOT cue was not a significant dependent measure for three of the locations. Speakers from Zanzibar, both men and women, demonstrated that their dialect is most distinct from other dialects in that a) the language, as they speak it, has an aspiration contrast in minimal and non-minimal pairs; b) VOT measurements are different depending on word origin (long for English loanwords, intermediate for native Swahili words, and short for Arabic loanwords); and c) while females had significantly longer VOT durations in the other towns, the productions of both genders in Zanzibar were not statistically different.

Next, this dissertation analyzes statistically how the same subjects perceived and imitated modified VOTs and F0s (three levels each) of non-words. Imitation was found for VOT only among L1 speakers of the language, and a closer look revealed that the Zanzibar participants’ imitations were the strongest: that is, they mimicked all three levels with statistically significant accuracy. Mombasa participants, on the other hand, distinguished only
between Level 1 and the other two. By contrast, L2 speakers of Swahili from Nairobi and Iringa showed no difference across the three levels in either VOT or F0.

In short, only Zanzibar preserves produced VOT-based aspiration. The loss of aspiration along the coast and in the inland regions, which may be due to the linguistic influence of neighboring Bantu languages, is reflected by the abandonment of unstable orthographic systems that distinguished aspirated consonants in favour of the simplified orthography of later times.
Acknowledgements

All praise is due to God (Allah) for providing me with every means possible in order to finalize this project.

The rationale for this PhD dissertation would not have been possible without the inspiration from a real-life situation. Back in the days when I had just graduated from university and while embarking on a trip to South Africa, I heard the word *imali* ‘money’, which rang a bell in my linguistic system. A (Zulu) person in downtown Durban was asking people for some spare change by saying, “*givuna mali*” (at least, this what I heard/perceived). For a while I was thinking, and filled with enthusiasm about working on some linguistic aspects of Zulu loanwords, yet even after buying an old hard-copy Zulu dictionary, I could not find enough data to work on. Ever since, my focus has shifted from Zulu to Swahili, a Bantu language where the words *imali* (and *safari*) originated—borrowed a long time ago from Arabic.

Aside from this early motivation, I have tons of appreciation for many other influential people who helped this work get accomplished.

First and foremost, I would like to express my gratitude and deepest thanks to my supervisor Darin Flynn and co-supervisor Stephen Winters for their continuous support and valuable advice throughout my entire doctoral program at the University of Calgary. Without their professional guidance, constant help, and unlimited encouragement, this thesis would not have possibly been finished. Indeed, I was lucky enough to have Dr. Flynn, who gave me the opportunity to join the Linguistics division at the U of C and later helped me to construct the background section for this project. He was always available whenever I needed him. In addition, when things got tough during my experiments or I felt frustrated by statistics, Dr.
Winters’ support and encouragement helped me to overcome those hard moments. I am much indebted to these two individuals for their help, expertise, advice, care, and kindness.

I would also like to thank other members of my supervisory committee. Drs. Amanda Pounder and Angeliki Athanasopoulou have shaped many aspects of my dissertation through their constructive critiques. Gratitude is also extended to both the internal examiner, Dr. Ozouf Amedegnato, and the external examiner, Dr. Yoonjung Kang. Dr. Kang’s work will always inspire my future work on VOT and F0.

I am very fortunate to have been a student at the U of C, where I have worked with exceptional people during my PhD years. A big thank-you goes to every member of the Linguistics division, in particular to my respected professors from whom I learned a lot, as well as to the talented students whom I enjoyed talking with over the years. I am also thankful to all the other faculty members and academic staff in the department. A special thank-you goes to Biljana Arnautov, who was always there to help with any administrative tasks. I also should not forget to thank the graduate program director, Dr. Elizabeth Ritter, for her precious advice and heartfelt support.

In addition, I would like to thank all the people in the locations where I made my recordings. In Zanzibar, I would like to thank Sumait University and Drs. Adil Saad and Fowzi Mohamed Barrow; I am grateful to Fowzi for making every effort possible to help me feel at home. I am also thankful to my research assistant in Zanzibar, Juma Juma, and all of my Zanzibar informants. In Stone Town, the old district of Zanzibar city, I am glad that I got to know Saeed Ahmad, who was the first person to introduce me to the history of the island. In Mombasa, I would like to thank the Mombasa campus of Pwani University, Dr. Najya Mwenzangu, Dr. Rukiya Swaleh, my research assistant Swafiya Msellem, and all my Mombasa participants. At the University of Nairobi, I am grateful for the help I received.
from the Department of Kiswahili, in particular Professors Habwe Hamu, Iribe Mwangi, and Prisca Jerono, as well as my research assistant Bernard Kigen and all my Nairobi subjects. In Iringa, I am grateful to the University of Iringa, my research assistant Flavian Mkupasi, and all my Iringa informants.

My sincere appreciation also goes to my sponsor through all my graduate studies: my beloved country, Saudi Arabia, represented by Qassim University. To all my Saudi friends with whom I shared this journey, I am grateful, including my colleagues in the Department of English Language and Translation; in particular, I thank Drs. Bader Alharbi and Mohammad Aljutaily.

My family members have always been a source of encouragement; my appreciation for all of them is beyond words. I am grateful for the prayers I have always received from my parents, as well as for the support of my siblings, especially my brothers, Dr. Abdullah and Mr. Abdulsalam. I am in debt to my wife Nafilah Alajaji, my source of comfort and positivity. Also, I will not forget my kids, from whom I gain my happiness and strength: Sarah, Abdulaziz, and Maria.

Although the core written stages of my PhD dissertation were completed during the spread of Covid-19 and the resulting lockdown and isolation, many other turbulences, either personal or political, threatened its completion even before the pandemic. Thus, I am joyful that it finally is coming to its full, happy ending.
Dedication
To my parents, my wife, and my kids
Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgements ........................................................................................................... iv
Dedication ............................................................................................................................. vii
Table of Contents ................................................................................................................ viii
List of Figures ..................................................................................................................... xii
List of Tables ....................................................................................................................... xiii
List of Charts ....................................................................................................................... xiv
List of Maps ......................................................................................................................... xv
List of Abbreviations .......................................................................................................... xvi

Chapter 1: Introduction ........................................................................................................ 1
  1.1 The issue ..................................................................................................................... 1
  1.2 Aims of the study ....................................................................................................... 3
  1.3 Research questions ................................................................................................... 4
  1.4 Why Swahili? ............................................................................................................ 5
  1.5 The materials ........................................................................................................... 6
  1.6 Testing paradigm ....................................................................................................... 7
  1.7 Structure of this dissertation ...................................................................................... 8

Chapter 2: Background ......................................................................................................... 9
  2.1 Phonetic cues and laryngeal contrasts ...................................................................... 9
  2.2 Crosslinguistic background – VOT and F0 ............................................................... 15
  2.3 Crosslinguistic background – Sensitivity to VOT and F0 ........................................ 21
  2.4 Crosslinguistic background – Gender ..................................................................... 25
  2.5 Crosslinguistic background – Sociophonetics ........................................................ 31
  2.6 Background – debate re: loanword adaptation ......................................................... 33
  2.7 Crosslinguistic background – Prototype theory ....................................................... 35
  2.8 Background – The dilemma of Swahili orthography ................................................. 39
  2.9 Background – Introduction to the current study ...................................................... 50
  2.9.1 Background on Swahili/word choice .................................................................. 53
  2.9.2 Distribution of the word list ................................................................................ 55
  2.9.3 Background – Sounds under investigation ........................................................... 57
  2.9.4 Locations of interest ............................................................................................. 58
  2.9.5 Participants ........................................................................................................... 59
    2.9.5.1 L1 speakers (Zanzibar and Mombasa) ............................................................... 60
    2.9.5.2 L2 speakers (Iringa) ......................................................................................... 61
      2.9.5.2.1 Bena obstruents ......................................................................................... 61
    2.9.5.3 L2 speakers (Nairobi) .................................................................................... 61
      2.9.5.3.1 Kikuyu obstruents .................................................................................... 62
  2.10 Conclusion .................................................................................................................. 63

Chapter 3: Production Experiment ...................................................................................... 64
  3.1 Participants ................................................................................................................ 64
  3.2 Materials .................................................................................................................... 64
  3.3 Design ........................................................................................................................ 64
  3.4 Production procedure ............................................................................................... 65
3.5 Production coding and annotation methods ........................................65
3.6 Production results ........................................................................68
3.7 Linear mixed-effects analysis ..........................................................69
  3.7.1 VOT in Zanzibar ..................................................................69
    3.7.1.1 VOT/Phonation .........................................................70
    3.7.1.2 VOT/Loanword .........................................................72
  3.8 F0 in Zanzibar .........................................................................74
    3.8.1 F0/Phonation .................................................................74
    3.8.2 F0/Loanword .................................................................76
  3.9 Zanzibar interim summary .........................................................77
  3.10 VOT in Mombasa ..................................................................77
    3.10.1 VOT/Phonation ............................................................78
    3.10.2 VOT/Loanword ............................................................79
  3.11 F0 in Mombasa ..................................................................80
    3.11.1 F0/Phonation .............................................................81
    3.11.2 F0/Loanword .............................................................82
  3.12 Mombasa interim summary .....................................................83
  3.13 VOT in Nairobi ..................................................................84
    3.13.1 VOT/Phonation ............................................................84
    3.13.2 VOT/Loanword ............................................................85
  3.14 F0 in Nairobi ..................................................................87
    3.14.1 F0/Phonation .............................................................87
    3.14.2 F0/Loanword .............................................................88
  3.15 Nairobi interim summary .........................................................89
  3.16 VOT in Iringa ..................................................................90
    3.16.1 VOT/Phonation ............................................................90
    3.16.2 VOT/Loanword ............................................................91
  3.17 F0 in Iringa ..................................................................92
    3.17.1 F0/Phonation .............................................................92
    3.17.2 F0/Loanword .............................................................93
  3.18 Iringa interim summary .........................................................94
  3.19 Overall F0 of all towns’ words by Phonation ...............................95
    3.19.1 Overall F0 of all towns’ words by Phonation (Place of articulation) .........................................................................................................................97
    3.19.2 F0/Phonation summary ..................................................98
  3.20 Overall VOT of all towns’ words by Phonation ............................99
    3.20.1 Overall VOT of all towns’ words by Phonation (Place of articulation) .................................................................100
      3.20.1.1 VOT/Phonation of bilabials .......................................101
      3.20.1.2 VOT/Phonation of alveolars ...................................101
      3.20.1.3 VOT/Phonation of velars .......................................102
      3.20.1.4 VOT/Phonation of post-alveolars ...........................103
    3.20.2 VOT/Phonation summary ................................................104
  3.21 Overall Loanword/VOT differences ..........................................105
    3.21.1 Overall VOT of all towns’ words by Loanword (Place of articulation) .................................................................108
      3.21.1.1 VOT/Loanword of bilabials .....................................108
      3.21.1.2 VOT/Loanword of alveolars ...................................108
      3.21.1.3 VOT/Loanword of velars .......................................110
      3.21.1.4 VOT/Loanword of post-alveolars ...........................111
    3.21.2 VOT/Loanword summary ................................................112
  3.22 Evidence of VOT gender differences ........................................113
Chapter 4: Imitation Experiment .................................................. 117
4.1 Participants ........................................................................ 117
4.2 Materials ........................................................................... 117
4.3 Design .............................................................................. 117
4.4 Perception procedure .......................................................... 119
4.5 Perception annotation method ............................................. 120
4.6 Results of the imitation task .................................................. 121
4.7 Linear mixed-effects analysis ............................................... 122
  4.7.1 VOT variations in all towns .......................................... 122
  4.7.2 F0 variations in all towns .............................................. 124
  4.7.3 Summary of variations in imitated VOTs ......................... 125
  4.7.4 Summary of variations in imitated F0s ........................... 126
4.8 Ranges of F0 in all four towns ............................................. 127
4.9 Durations of VOT in all four towns ....................................... 129
  4.9.1 Durations of VOT by gender ........................................ 130
  4.9.2 Summary of VOT durations .......................................... 132
4.10 Differences in VOT duration for each obstruent .................... 133
  4.10.1 Differences in town Z by obstruent and gender .............. 133
  4.10.2 Differences in town M by obstruent and gender .......... 135
  4.10.3 Differences in town N by obstruent and gender .......... 136
  4.10.4 Differences in town I by obstruent and gender .......... 137
  4.10.5 Summary of VOT durations for each obstruent .......... 138
4.11 Between-gender VOT differences in all towns ...................... 138
  4.11.1 Summary of between-gender differences in VOT .......... 140
4.12 Gender differences in VOT duration when F0 is imitated ....... 141
4.13 A short summary of the imitation experiment .................... 143

Chapter 5: DISCUSSION .............................................................. 144
5.1 Production discussion .......................................................... 155
  5.1.1 Production discussion (aspiration contrast) ................. 156
  5.1.2 Production discussion (Loanword type) ....................... 163
5.2 Perception discussion (imitation) .......................................... 169
5.3 Closing remarks ................................................................... 175

Chapter 6: CONCLUSION .......................................................... 176
6.1 Summary of findings ............................................................ 176
  6.1.1 Summary of production .............................................. 179
  6.1.2 Summary of imitation ................................................ 182
6.2 Limitations ......................................................................... 183
6.3 Future directions ................................................................. 186

REFERENCES ............................................................................. 189
Appendix A ............................................................................... 200
Appendix A.1 Pemba Map (Zanzibar Northern Island) ............... 200
Appendix A.2 Unguja Map (Zanzibar Southern Island) ............. 200
Appendix A.3 Mombasa Map .................................................... 201
Appendix A.4 Iringa Map ........................................................... 201
Appendix A.5  Nairobi Map ........................................................................................................ 201
Appendix A.6  Arabic loanwords in Swahili ........................................................................ 202
Appendix A.7  Swahili native words and loanwords from English and Arabic ................ 203
Appendix B .................................................................................................................................. 205
Appendix B.1  F0 of All Towns (Bilabial*Gender) ............................................................ 205
Appendix B.2  F0 of All Towns (Alveolar*Gender) ............................................................ 205
Appendix B.3  F0 of All Towns (Velar*Gender) ............................................................... 206
Appendix B.4  F0 of All Towns (Post-Alveolar*Gender) .................................................. 206
Appendix B.5  VOT of All Towns by Loanword/p/ ............................................................. 207
Appendix B.6  VOT of All Towns by Loanword/t/ ............................................................. 207
Appendix B.7  VOT of All Towns by Loanword/k/ ............................................................. 208
Appendix B.8  VOT of All Towns by Loanword/ch/ ............................................................ 208
Appendix B.9  Gender Differences in Native words ............................................................ 209
Appendix B.10 Gender Differences in English Loanwords .............................................. 209
Appendix B.11 Gender Differences in Arabic Loanwords ................................................ 209
Appendix B.12 Aspiration Contrast in Town Z as per Subject .......................................... 210
Appendix B.13 Aspiration Contrast in Town M as per Subject .......................................... 210
Appendix B.14 Aspiration Contrast in Town N as per Subject .......................................... 211
Appendix B.15 Aspiration Contrast in Town I as per Subject .......................................... 211
Appendix B.16 Loanword Type in Town Z as per Subject .................................................. 212
Appendix B.17 Loanword Type in Town M as per Subject .................................................. 212
Appendix B.18 Loanword Type in Town N as per Subject .................................................. 213
Appendix B.19 Loanword Type in Town I as per Subject .................................................. 213
Appendix C .................................................................................................................................. 214
Appendix C.1  Level in Town Z as per Subject ................................................................. 214
Appendix C.2  Level in Town M as per Subject ................................................................. 214
Appendix C.3  Level in Town N as per Subject ................................................................. 215
Appendix C.4  Level in Town I as per Subject ................................................................. 215
Appendix D .................................................................................................................................. 216
Appendix D.1  Aspiration Contrast in Town Z as per Subject (F0) ..................................... 216
Appendix D.2  Aspiration Contrast in Town M as per Subject (F0) ..................................... 216
Appendix D.3  Aspiration Contrast in Town N as per Subject (F0) ..................................... 217
Appendix D.4  Aspiration Contrast in Town I as per Subject (F0) ..................................... 217
Appendix D.5  Loanword Type in Town Z as per Subject (F0) .......................................... 218
Appendix D.6  Loanword Type in Town M as per Subject (F0) .......................................... 218
Appendix D.7  Loanword Type in Town N as per Subject (F0) .......................................... 219
Appendix D.8  Loanword Type in Town I as per Subject (F0) .......................................... 219
Appendix E .................................................................................................................................. 220
Appendix E  Wordlist for other words in the data ............................................................. 220
List of Figures

Figure 2-1 PVI representation and measurement of voiceless alveolars (adapted from Schwarz et al., 2019: 119) .............................................................................................................. 17
Figure 2-2 PVI representation and measurement of voiced alveolars (adapted from Schwarz et al., 2019: 119) .............................................................................................................. 17
Figure 2-3 The shaded area represents the VOT of the word tai, meaning ‘tie’ (top) and ‘eagle’ (bottom), as produced by ZS8MPI ................................................................................. 19
Figure 2-4 The VOTs and F0s for /t/ of Speaker 1 (from Mombasa) ............................................. 50
Figure 2-5 The VOTs and F0s for /t/ of Speaker 2 (unspecified) ...................................................... 51
Figure 3-1 The second VOT repetition for /t/ in tembo, as produced by ZS11FPI ....................... 67
Figure 3-2 The distribution of F0 ranges in aspirated vs. unaspirated environments (pooled/overall) .......................................................................................................................... 96
Figure 3-3 The distribution of VOT ranges in aspirated vs. unaspirated environments (pooled/overall) .......................................................................................................................... 100
Figure 3-4 Line plots showing VOT (duration/length) ranges in aspirated vs. non-aspirated environments across all genders and towns (bilabials in A / alveolars in B) ........... 101
Figure 3-5 Line plots showing VOT (duration/length) ranges in aspirated vs. non-aspirated environments across all genders and towns (velars in A / post-alveolars in B) ....... 103
Figure 3-6 Differences in the VOT durations of native and borrowed words across towns Z, M, N, and I. ......................................................................................................................... 105
Figure 3-7 Between-gender differences in VOT duration, based on aspiration, in towns Z, M, N, and I. .......................................................................................................................... 114
Figure 3-8 Between-gender differences in VOT duration, based on non-aspiration, in towns Z, M, N, and I. ......................................................................................................................... 115
Figure 4-1 VOT stimuli construction for [t] in the nonce word tote at the three levels, from 15 to 35 to 55 ms ................................................................................................................. 118
Figure 4-2 F0 stimuli construction for [t] in the nonce word tote at the three levels, from 200 to 220 to 240 Hz ................................................................................................................. 119
Figure 4-3 The VOT of the non-word koke at 30 ms, the first repetition of word number 6 as produced by NS8MI ............................................................................................................... 120
Figure 4-4 Distribution of F0 ranges across the three levels and the four towns, divided by gender. ........................................................................................................................ 128
Figure 4-5 Distribution of VOT durations across the three levels and the four towns.............. 130
Figure 4-6 Distribution of VOT durations across the three levels and the four towns, divided by gender. ........................................................................................................................ 131
Figure 4-7 Within-gender VOT differences for the four words in town Z ..................................... 133
Figure 4-8 Within-gender VOT differences for the four words in town M ................................. 136
Figure 4-9 Within-gender VOT differences for the four words in town N, ................................. 137
Figure 4-10 Within-gender VOT differences for the four words in town I ................................. 137
Figure 4-11 Differences in VOT durations between the genders at the three levels, divided by town. ....................................................................................................................... 139
Figure 4-12 Within-gender differences in the VOT durations of imitated F0 tokens, divided by town. ........................................................................................................................ 142
List of Tables

Table 2-1 A survey of VOT distribution ................................................................. 28
Table 2-2 Phonemic vs. allophonic approximation ................................................... 35
Table 2-3 Taylor’s proposal for representing aspirated obstruents using Arabic characters ................................................................. 42
Table 2-4 Assorted examples of words/categories ....................................................... 55
Table 2-5 Selected Swahili obstruents ........................................................................ 58
Table 2-6 Obstruents of interest in Bena ..................................................................... 61
Table 2-7 Obstruents of interest in Kikuyu .................................................................. 63
Table 3-1 Phonation VOT coefficients from town Z, according to the last described model (Phonation + Place + Phonation*Place) ........................................................................... 71
Table 3-2 Loanword VOT coefficients from town Z, according to the best model as described above (Loanword + Place) ........................................................................... 73
Table 3-3 Phonation F0 coefficients from town Z, according to the last described model (Gender + Phonation + Gender*Phonation + Place) ........................................................................... 75
Table 3-4 Loanword F0 coefficients from town Z, according to the last described model (Gender + Loanword) ....................................................................................... 76
Table 3-5 Phonation VOT coefficients from town M, according to the last described model (Place + Gender + Place*Gender) ........................................................................... 78
Table 3-6 Loanword VOT coefficients from town M, according to the best model as described above (Place + Gender + Place*Gender) ........................................................................... 80
Table 3-7 Phonation F0 coefficients from town M, according to the last described model (Place + Gender + Place*Gender + Phonation) ........................................................................... 81
Table 3-8 Loanword F0 coefficients from town M, according to the last described model (Place + Gender + Place*Gender + Loanword) ........................................................................... 83
Table 3-9 Phonation VOT coefficients from town N, according to the last described model (Place + Gender + Place*Gender) ........................................................................... 85
Table 3-10 Loanword VOT coefficients from town N, according to the last described model (Place + Gender + Place*Gender) ........................................................................... 86
Table 3-11 Phonation F0 coefficients from town N, according to the last described model (Gender + Place + Gender*Place) ........................................................................... 87
Table 3-12 Loanword F0 coefficients from town N, according to the last described model (Gender + Place + Gender*Place) ........................................................................... 88
Table 3-13 Phonation VOT coefficients from town I, according to the last described model (Gender + Place + Gender*Place) ........................................................................... 90
Table 3-14 Loanword VOT coefficients from town I, according to the last described model (Gender + Place + Gender*Place) ........................................................................... 91
Table 3-15 Phonation F0 coefficients from town I, according to the last described model (Gender) ....................................................................................................................... 93
Table 3-16 Loanword F0 coefficients from town I, according to the last described model (Gender + Loanword + Gender*Loanword) ........................................................................... 93
Table 4-1 The levels of VOT and F0 for the target obstruents, as presented in the stimuli ......................................................................................................................... 117
Table 4-2 Imitation VOT coefficients from all towns, according to the last described model (Level*Town) ................................................................................................................. 123
Table 4-3 Imitation F0 coefficients from all towns, according to the last described model (Gender+Town) ................................................................................................................. 125
List of Charts

Chart 2-1 An example of prototypes for [t], along the y and x axes of F0 and VOT, respectively ..............................................................53
Chart 5-1 The overall distribution of the aspiration contrast in town Z’s obstruent VOTs
(each arrow targets the other cloud for the same place of articulation/gender). .....156
Chart 5-2 The overall distribution of the aspiration contrast in town M’s obstruent VOTs
(each arrow targets the other cloud for the same place of articulation/gender). .....160
Chart 5-3 The overall distribution of the aspiration contrast in town N’s obstruent VOTs
(each arrow targets the other cloud for the same place of articulation/gender). .....161
Chart 5-4 The overall distribution of the aspiration contrast in town I’s obstruent VOTs (each
arrow targets the other cloud for the same place of articulation/gender). ............162
Chart 5-5 The overall distribution of /p/ under the Loanword category (each arrow targets the
other cloud with the same color). .............................................................................165
Chart 5-6 The overall distribution of /t/ under the Loanword category (each arrow targets the
other cloud with the same color). .............................................................................166
Chart 5-7 The overall distribution of /k/ under the Loanword category (each arrow targets the
other cloud with the same color). .............................................................................167
Chart 5-8 The overall distribution of /tʃ/ under the Loanword category (each arrow targets the
other cloud with the same color). .............................................................................168
Chart 5-9 The overall distribution of F0s (for obstruents), as produced by female subjects. 170
Chart 5-10 The overall distribution of F0s (for obstruents), as produced by male subjects. 171
Chart 5-11 The overall distribution of VOTs, as produced by all subjects. .........................172
Chart 5-12 The overall distribution of VOT values for F0-based stimuli, as produced by all
subjects. ....................................................................................................................174
List of Maps

Map 2-1 The region of spoken Swahili..........................................................59
List of Abbreviations

VOT: Voice onset time
F0: Fundamental frequency or pitch
ms: Milliseconds
Hz: Hertz
L1: Speakers of the language as a first language
L2: Speakers of the language as a second language
Z: Town of Zanzibar
M: Town of Mombasa
N: Town of Nairobi
I: Town of Iringa
S: Swahili
E: English
A: Arabic
Chapter 1: Introduction

1.1 The issue

African languages with a limited number of speakers and influence are in flux for many reasons, including an increasing dominance of more powerful Bantu languages, which may compel minorities to communicate less and less in their mother tongues. Notably, Swahili is expanding into the territories of surrounding indigenous languages (Mous, 2003). Efforts to preserve languages are not limited to their speakers — libraries all over the world abound in dictionaries of African languages that are based on lexicons compiled by foreigners such as Western missionaries. This is the case for Swahili as well. Although much of the original motivation behind the documentation of Swahili was religious, speakers and linguists can still derive benefit from old scripts and use them to gain a deeper understanding of the language’s vocabulary and structure. In studying sound systems, however, we need to move beyond such written documentation. In particular, we can advance our knowledge of a specific language’s sound inventory by investigating its finest phonetic details.

An immediate difficulty when conducting a study of an African language is the lack of prior empirical data that can be used as points of comparison and contrast. Those who choose to work on an African language must often begin their studies from scratch. This holds true even for Swahili. It is the most widely spoken indigenous language in Africa, yet no extensive work has been done on it in terms of experimental linguistics. Consequently, a dissertation project on Swahili requires a large sample in order to be self-sufficient while developing a research scope that is concise and fueled by narrower concentrations.

Another challenge that arises when departing from commonly studied languages is choosing the starting point and other locations; it is not easy to design a base camp and a
journey to distinguish a priori which dialects are more innovative or conservative in their linguistic properties.

Another problem is that when an understudied language is documented by a small number of disparate linguists, the latter may propose conflicting judgments on very sensitive linguistic cues, which would require an intervention of speech science supported by statistics.

Such is the case in Swahili. To investigate the controversial and understudied issue of laryngeal contrast in Swahili, we combined production and perception to determine how aspiration is produced and perceived by monolingual and bilingual Swahili speakers in four towns. Previous studies (Engstrand & Lodhi, 1985; Frankl, 1991; Lodhi, 2003) have stated that a laryngeal contrast exist in voiceless obstruents, but these studies do not provide enough empirical evidence, and on some occasions, their theoretical accounts of the aspiration contrast are non-committal. Mohamed's (2001) Swahili grammar book does not mention any aspirated phonemes, which runs counter to Maddieson's (1984) judgment about Swahili: namely, that it has a three-way laryngeal contrast (see subsection 2.9.3). To shed further light on the issue, we gathered first-hand data from native Swahili speakers who pronounced real Swahili words, then imitated recordings of modified “phonetic prototype clouds” of four obstruents (/p/, /t/, /k/, and /tʃ/) in word-initial position; these two tasks were then followed by an identification task, which used native Swahili minimal pairs. The perception experiments were based on stimuli with controlled initial pitch and Voice-Onset Time (VOT). The production, imitation, and discrimination tasks were done in one sitting (session) in the summer of 2018, which lasted about 40-50 minutes per participant, and the results provide a sociophonetic account of the status of aspiration in Swahili. The geographical area I investigated spreads over two countries in East Africa, and the four towns were chosen as
locations of interest, to provide data from L1 and L2 speakers who speak different varieties of Swahili (see section 2.9).

1.2 *Aims of the study*

The first purpose of this study is to provide a description of the laryngeal contrast in the voiceless phonemic inventory of Swahili by examining people with different language backgrounds who speak it in various regions. How does the speech of first-language speakers differ from that of second-language speakers? Speakers of a language can claim to be native or non-native, and if we work with both groups, we can find out what linguistic details they either share or restrict. After reaching our conclusion, we will be able to report to the linguistic and scientific community about whether Swahili has a two-way or a three-way laryngeal contrast in its obstruents: that is, whether distinctive aspiration exists in the language, and if this acoustic feature does exist, to what extent it is declining or even vanishing.

Our second purpose is to determine which phonetic details are typically used by speakers of Swahili to represent the voiced/voiceless stops. To establish the primary cue or cues that speakers prefer when contrasting the phonemic inventory, we examined two phonetic details: temporal and spectral. The temporal parameter in our study is VOT, which traces the duration of a stop’s release before the vowel’s onset. The spectral parameter, on the other hand, is the Fundamental Frequency (F0) pitch level, which measures the height of the F0 in the post-stop vowel onset. It has been reported for other languages, such as Korean (Kwon, 2013; Lee et al., 2013), that phonetic distinctions between VOT and F0 are possible. In our study, these two dependent variables are examined in four distinct locations to see if different groups of speakers tend to articulate the two at the same level, or if they are heavily dependent on one cue over the other.
The third core interest of this study is to document which locations have speakers who exhibit the most sensitivity to the phonetic cues of Swahili phonemes. This can be determined by working with first- and second-language speakers in different geographical areas; as previously mentioned, we will test these variables across four different East African cities. We assume that, by the end of the study, we will be able to pinpoint the territories that exhibit significant variability in either VOT or F0.

The data for this study are based on native minimal pairs and many other words that were gathered and arranged according to their origins. Therefore, the fourth purpose of the study is not only to look into the contrast in phonation (aspirated vs. unaspirated), but also to shed some light on the difference, if any, between native words and loanwords. Swahili has its own native vocabulary, but there are also a large number of old (from Arabic) and new (from English) loanwords, which deserve an examination when it comes to the VOT/F0 parameters. Our hope is that our analysis will provide valuable evidence to linguists who are interested in the field of sociophonetics and historical linguistics, too.

Finally, it is common knowledge among phoneticians that F0 is higher on average for females and lower on average for males (Simpson, 2009); however, we do not have a clear picture regarding variabilities that may exist in the VOTs of men and women. Thus, the last purpose of this study is to compare and contrast the production and perception (imitation) of our participant groups, each of which will be divided equally between the two genders.

1.3 Research questions

There has been very little interest in studying the phonemic inventory of Swahili by means of a phonetic investigation aided by a quantitative analysis of contemporary spoken dialects. The Swahili stops are some of the phonemes that have received some attention in the literature as belonging to the three-way contrast of voiced, voiceless unaspirated and
voiceless aspirated (Engstrand & Lodhi, 1985; Moxley, 1992); however, the reports in both studies were limited and scarce. The limited nature of previous research means that the present writer can neither strongly confirm the laryngeal contrast’s existence nor confidently generalize any facts to a larger body of the Swahili-speaking population. As a result, we propose to answer the following questions in this thesis:

1) Does Swahili have a two-way or a three-way laryngeal contrast?

2) If a contrast exists, is it implemented to the same degree by all speakers of Swahili? Do first-language speakers differ from second-language speakers in this regard?

3) If a three-way laryngeal contrast exists, which phonetic detail (temporal or spectral) do speakers of the language use as a distinguishing factor, or do they potentially use both cues?

4) Are there gender differences in the production of VOT?

5) Do Swahili speakers mark native words and loanwords with varying VOT measurements?

6) If the production data suggest a laryngeal contrast, is the contrast also perceived by native Swahili listeners?

1.4 Why Swahili?

Swahili is a Bantu lingua franca for the vast majority of Africans in East Africa, and also a first language of long-time residents of the region who come from other backgrounds. It is spoken along the Indian Coast from southern Somalia to northern Mozambique, and as far as the Democratic Republic of Congo in the west (Wald, 2009). Before Swahili orthography was standardized into Roman characters by Johann Ludwig Krapf (1810–1881), the language was written in Arabic scripts (Samsom, 2015); therefore, it was documented for centuries before its alphabet changed, with ongoing preservation of its linguistics occurring since then.
Swahili has surpassed many other African languages in terms of the attention that it has received via bibliographic work on general topics. Citing earlier studies by himself and others, Baldi (2011) claims that many aspects of Swahili have received reasonable consideration, including its status as a borrowing language. He then proposes to bridge the gap by introducing Swahili as a donor language.

But there is still much work to do on African languages in comparison to the work that has been done on most popular languages around the world. When it comes to analyses in other fields of linguistics, we do not know how much still needs to be done for Swahili, but we do know that there is a need to revisit Swahili’s phonemic inventory of stops. Since Swahili is the predominant language in East Africa, which contains relics of long-ago contact with Arabic and more recent interactions with English, we believe that it is the ideal candidate for a deep instrumental analysis that will help us learn which part of East Africa has the most conservative speakers. In their book on Swahili, Nurse & Spear (1985: 61) reported that coastal (and island) varieties of Swahili were swamped by the dialects of mainland Tanzania, and many original characteristics were either lost or reduced as a result. In our sociophonetic study on Swahili as it is spoken today in various regions, it is our ultimate goal to see if the aspiration contrast, if any, is one of those phonological features that Nurse & Spear said are declining. This will ensure a further record of the language for specialists in the field and for many generations to come.

1.5 The materials

The individual words we studied were extracted, especially for this study, from old and new print dictionaries, online dictionaries, some early studies, and information provided by some Swahili experts and specialists (see subsection 2.9.1). Speakers of the language were consulted for further feedback on what words to keep or replace. The words we gathered
focused on word-initial stops and affricates, as well as post-nasal stops, for a total of 147 words. As the dissertation progressed and expanded, however, we had to minimize the data load to focus mainly on the voiceless stops and the voiceless affricate. Post-nasal voiceless stops and voiced stops were still included in the data, but the report on them will be saved for future studies.

1.6 *Testing paradigm*

I conducted three experiments to investigate our research questions. First, I collected production data by asking participants to look at pictures and then say the corresponding words twice. The purpose of this experiment was to test the phonetic realization of any contrast in phonation and offer further evidence of the phonetic cues employed to signal those contrasts. I expected our experimental evaluation of pitch and VOT to either support or oppose previous phonological claims about aspiration in Swahili. Second, I had the participants listen to modified VOTs and manipulated F0s of nonsense words and then try to imitate them as closely as possible (two imitations for each word). The aim of this experiment was to follow up on the phonetic analysis that I had conducted in the production task and see to what extent the speakers of Swahili are sensitive to fine details as presented in nonsense words. Third, I had our subjects discriminate four levels of spliced VOT and F0 durations, respectively, of native minimal pairs. The purpose of this experiment was to examine speakers’ ability to identify modified phonetic cues as presented in real words. Due to time and space limitations, the details of the discrimination task will be described in a future work. The participants were the same for each experiment and were recruited from four different locations. Each group had an equal number of males and females (12 each except Mombasa, 13 each).
1.7 Structure of this dissertation

Chapter 2 provides a theoretical background for the study and comments on the corpus of data and the collection procedures. The background section covers VOT/F0 contrasts and facts cross-linguistically, along with gender variability and sensitivity to VOT/F0 as reported for some non-African languages, as well as sociophonetic influences and the theory of phonetic prototypes. It discusses the historical account of the Swahili orthography and its influence on some of the language’s features. Then, the chapter offers further details on the methodology of the study, from the motivation behind choosing Swahili to the procedures used for the data preparation, collection, and analysis of the results of the three studies.

Chapter 3 presents the production data by looking at the production of all words’ VOT and F0. It also looks into words’ origins and gender differences. Chapter 4 examines the imitation data; it focuses on the imitation of nonsense Swahili-like words at three levels of VOT and F0, as well as on the way gender affects the responses in different towns. Chapter 5 discusses the results of both studies, with emphasis on the phonetic prototypes as shown in Chapters 3 and 4. For the production and imitation results, I created clouds for most phonetic cues in the form of visual charts to make for an easy reference when comparing and contrasting the results from the different towns. And, finally, Chapter 6 concludes and summarizes the study. It then comments on the limitations and suggests some future directions.
Chapter 2: Background

Before we can resolve the issue of phonemic aspiration in Swahili and to what extent it is widely spread across the spoken dialects of the language, some theoretical accounts have to be introduced. First, I will build an analysis of the VOT/F0 parameters by looking into measurements of voice-onset time and fundamental frequency crosslinguistically to see if speakers’ sensitivity to either cue is categorical and/or perceptual. After that, gender-based accounts of early studies on other languages will be discussed so that we can compare the VOT productions found elsewhere to those of Swahili-speaking men and women. Moreover, the sociophonetic variabilities may undergo different treatments from different users of the same language’s phonemic system, so it will be important to see if this match or mismatch is diachronic or synchronic. Furthermore, since Swahili as a borrowing language has taken a huge number of loanwords from Arabic and English, we also wonder if this created special cases for phonetic contrasts between native (Bantu) and adopted (non-Bantu) words and whether any adaptation is phonological or phonetic. Also, prototype theory will be presented as part of the background to help us understand how the best prototype of each sound may be categorized by participants in our four locations of interest. Finally, I will discuss how orthography posed some challenges for preserving certain linguistic details of Swahili. Then, with all of that theoretical background in place, I will discuss the main concern behind the current study and further comment on the methodology of the current study, including the selection of target words/sounds and locations of interest.

2.1 Phonetic cues and laryngeal contrasts

The original concepts of Chomsky & Halle (1968)’s phonological theory, which were intended to establish a basis for the Sound Pattern of English (SPE), attribute the contrasts
between phonemes to binary phonetic features, which Lisker & Abramson (1971: 774) critically describe as the “universal phonetic machinery”. Observing these properties in obstruents is facilitated by the presence (+) or absence (−) of: [voice], [tense], [heightened subglottal pressure], and [glottal constriction]. If we take bilabials (/bʱ/ → /b/ → /b̥/ → /p'/ → /p/ → /p̣/) as examples along the laryngeal dimension, ranging from aspirated voiced to highly aspirated voiceless, we see that the voiced /b/ is different from the voiceless /p'/. So, the first differentiating laryngeal mechanism here is the feature [±voice], according to which either the segment is vibrating, as in voicing [+voice], or no vibration is observed, as in voicelessness [–voice]. Alternatively, if the highly aspirated /p̣/ and the slightly aspirated /p/ are taken as examples of contrastive phonemes, then the second feature of interest is [±tense]. Voiceless phonemes with different degrees of aspiration require different gestures during the articulatory movements. The configuration of the articulators lasts longer in tenseness (highly aspirated phonemes) than in non-tenseness (moderately aspirated phonemes).

Furthermore, a voicing lead (where voicing persists for some time) requires the articulatory process to maintain a constant vibration of the vocal folds before the release of the (stop) sound. Therefore, some languages, including Hindi, show a phonemic contrast in

---

1 For more about the distribution of feature systems in Chomsky & Halle (1968), as well as the theorized feature specifications before and after them, see Mielke (2004: 153). In his comment on an ongoing dispute with reference to distinctive features, Kim (1965: 358) confirms that the “production of sound is a physiological activity, and the physiological activity is not as simple as a number of binary oppositions”.

2 As Chomsky & Halle (1968: 321-322) mention: “in Jakobson, Fant, & Halle (1963), the difference between plosives and affricates was characterized by means of the feature stridency. Plosives were characterized as nonstrident stops, affricates as strident stops”. However, for Chomsky and Halle, “plosives are stops with instantaneous (primary) releases, affricates are stops with delayed releases. The feature stridency can then be used to distinguish strident from nonstrident affricates”.

3 These bilabial phonemes are defined as follows: /bʱ/, voicing aspiration lead; /b̥/, voicing non-aspiration lead; /b/, coincident between voicing and non-voicing (extremely short lag), and according to Iverson & Salmon (2006), phonetically considered unvoiced (or passively voiced; see also Schwarz et al., 2019; Kirby & Ladd, 2019); /p'/, unaspirated voiceless (short lag; fortis/tense); /p/, slightly aspirated voiceless (moderate lag; lenis/non-tense), /p̣/, highly aspirated voiceless (long lag; tense). In this section, I will use these bilabials as referents (examples) for clarifying the behavior of stops in general.

4 Tenseness is found in both vowels and consonants: namely, in long vowels and (aspirated) consonants. Although it is very explicit in Chomsky & Halle (1968) that a long vowel is considered a tense vowel, they did not mention the exact phonetic cues for consonants. They did write, however, that “the differences between tense and lax consonants also involve a greater versus a lesser articulatory effort and duration” (p. 325, emphasis mine).
voiced stops, according to which they can be either aspirated or unaspirated. As stated by Chomsky & Halle (1968), the third laryngeal feature, [heightened subglottal pressure], is found only in aspirated voiced segments, not in the unaspirated ones: e.g. in /bʱ/, but not in /b/. This feature is also available in long voicing lags; in /pʰ/, voicing is delayed greatly, too.

However, there are times when the feature [voice] gets a tender prosody, resulting in an articulatory undershoot regarding voicing (e.g. common cases of English voiced /b̥/ vs. Spanish voiceless /p/: Lisker & Abramson, 1964; Dmitrieva et al., 2015). As such, the pulsing (voicing) coincides with the release, as in the voiced short lag of English, or there is an instant pulse after release, as in the Spanish voiceless short lag—which, according to Chomsky & Halle’s (1968) universal phonetic set, is assumed to be generated by the fourth feature, [glottis constriction], and could be affected by either the second feature, [tense], or the third feature, [high subglottal pressure]. That is to say, it is this laryngeal continuum, during the transition from voicing to non-voicing, whereby the overlap between some phonetic features gets its maximum complexity.

However, the SPE’s binary description of laryngeal contrasts was not very appealing to some early linguists⁶, and this motivated some others (Iverson & Salmons, 1995 and 2006; Lombardi, 1995; Honeybone, 2005), who, even years later, continued to propose some reformed categorization for contrasts in what is called “laryngeal realism”. Moving away from the primitive/abstract approach of Chomsky & Halle (1968), the general, “privative” realist model builds on the connection between the phonological representation and the phonetic cues. It posits that a language can be a true voicing language with [voice], an

---

⁵ The voiced stops of English are seen to have either a short lag (/b/) or lead voicing (/b/), depending on the production of informants (Lisker & Abramson, 1964; Dmitrieva et al., 2015).

⁶ Lisker & Abramson (1971) state: “This phonological ambiguity is in principle resolvable in the laboratory, provided the experimental phonetician has access to the deep phonetic facts and can verify just where in the course of production of the utterance there is a change in value for one or more of the phonetic features of the Chomsky-Halle universal set” (780).
aspirating language with [spread glottis (sg)], or a mix of both (Beckman et al., 2013). In both systems, the short duration lag, with either weak voicing or a weak burst, is left unspecified (Schwarz et al., 2019).

Schwarz et al. (2019) also reviewed how laryngeal realism reached the conclusion of dividing languages into either the voicing or the aspirating category. They stated that three “criteria” were tested by the allies of laryngeal realism. First is the need for “phonetic realization of segments” so that, if plosives are distinguished by pulsing (vibrating), then [voice] is specified; on the other hand, if duration is the differentiating phonetic aspect, then [spread] is specified. Second is a criterion that works on “diagnostics of control” and whether a specific acoustic cue is changing under some experimental conditions; one of these examined the speaking rates of voicing and aspirating languages, and it was found that the unspecified feature remained unchanged across different rates. Third, during the course of “phonological patterning”, it was seen that aspirating languages assimilate to their phonetic feature, [spread]; similarly, the assimilation of voicing languages is linked to [voice]. For this reason, Iverson & Salmon (2006) state that the unspecified feature attracts the state of “neutralization”; it is the feature that goes against the known laryngeal system of the language.

This realist assignment of features hinges on the laryngeal status of a language. When a two-way contrast includes aspirated obstruents, the language (such as English) gets specified for [sg], yet the sounds’ (un)voiced counterparts are assigned the non-feature category [∅]. However, if a true voicing language has a voicing lead for voiced obstruents, then the specification is for [voice]. For instance, the specification of languages such as Spanish, French, and Russian is linked to the feature [voice], and the plain voiceless obstruents are left unspecified. Moreover, languages with a three- or four-way distinction will
get specified for in two phonemes (in the former case) or three (in the latter). Representing a three-way contrast, Thai has voiceless aspirated [sg] and unaspirated [∅] stops alongside voiced stops [voice], so, as a result, it gets specified for [sg] and [voice] only. Then, some Indo-Aryan languages, including Hindi, have four plosives articulated in the laryngeal region (Beckman et al., 2013); these are voiced, voiceless unaspirated, voiced aspirated, and voiceless aspirated. Distinguishing these different phonemes is the feature [voice] for voiced unaspirated, as well as [sg] for aspirated voiceless. When combined, the two features [voice] and [sg] can specify aspirated voiced stops. In other words, in Hindi (and similar languages, such as Nepali: Schwarz et al., 2019), the phonological specification for true voicing, whether in the form /b/ or /bʱ/, is [voice], and if a sound is aspirated like /bʱ/ and /pʰ/, the feature is [sg].

However, even with this laryngeal realism and the re-evaluation of features, the outcomes of the phonetic data are still far more complex than classical distinctive features and/or the newer set of specified and unspecified segments. The case of Nepali, described above, is not as easy as it seems to be when specifying the language with either aspiration [sg] or voicing [voice]. Schwarz et al. (2019) examined the Nepali laryngeal distinction in both initial and medial position to test the strength of the two features. They found that the weight of feature distinction is quite different due to the effect of the phonological structure on the phonetic realization. In word-initial position, the results neatly presented the fact that the four Nepali segments are phonetically distinct from one another in duration (negative and/or positive VOT); this was further supported by soundwaves and spectrograms, which showed that the two voiceless series are specified as [sg], while the voiced ones are specified as [voice]. The medial (intervocalic) position, on the other hand, suggested some phonetic preference for the feature [voice] as being “stronger”. The piece of evidence for this was the
perceptual behavior of /p/ and /bʱ/, which tended to show more voicing and act like a pair from a voicing language, rather than an aspirating one.

Now, we know that [voice] is one of the influential features in both the classical (generative) and contemporary (realist) theories. Like many phonetic cues (e.g. voice quality and F0; see Schertz & Khan, 2020), voicing can be studied through tracing the durations of VOT (with mainly negative values in voiced segments, but positive ones in voiceless segments). Furthermore, VOT is one of the extensively studied acoustic details for understanding the variations of obstruents along the laryngeality continuum (Lisker & Abramson, 1964; Cho & Ladefoged, 1999; Ahn, 2018; among others). VOT, here, is taken to represent how complex it is to control for various durations during the process of laryngeal adjustment. Other phonetic cues suggest no less overlap when differentiating the contrast and/or change in configuration of the interacting perceptual gestures.

Adding to the facts regarding the variability of VOT in the languages of the world, Kakadelis (2018) studied different phonetic cues in varying phonological environments, including the VOTs of phrase-initial stops⁷. She chose three endangered languages⁸ with “no voicing distinction” (NVD), to see if there is a default pattern used by the three groups. Her findings are very complicated in terms of the voicing variations, not only among the three languages in her corpus, but also among the different phonological contexts in each one of them. All three languages were reported to have only one series of voiceless stops in their phonemic inventories, but the VOTs of the measured tokens exhibited a combination of positive values (a fine-grained phonetic detail of voiceless stops in aspirating languages) and negative values (a phonetic cue in voicing languages). Around 20% (28) of the 164 Bardi

---

⁷ Due to space limitations, I will report the findings regarding VOT behavior in initial position only.
⁸ These are Bardi, Arapaho, and Sierra Norte de Puebla Nahuatl, which are spoken in Australia, the USA, and Mexico, respectively (Kakadelis, 2018).
tokens had a negative VOT. In Arapaho, it was observed that 46% (38) of the 82 stops had pre-voicing (negative VOTs). However, the third language with NVD, Sierra Norte de Puebla Nahuatl, had the lowest percentage of negative VOTs, with only 6% (9) negative VOTs from a total of 148 in phrase-initial position. In her conclusion, Kakadelis (2018: 395) suggests that the different VOT patterns in the same category, in her data, were a result of the lack of competition and contrast and further states: “This suggests that the relationships found in languages with a voicing distinction are the result of deliberate adjustments to enhance a phonological contrast”.

At this point, we have to admit that, even with some unexplained and mysterious results of early studies, VOT is still one of the indispensable tools to help us develop and improve our knowledge of phonemes and the features [voice] and [sg]. Along with VOT, the use of the phonetic cue F0 was inspired by some evidence from early recordings (section 2.9 below) and studies (section 2.2 below), with regard to its persuasive position as a tool to help us learn more about the spectral frequency in voicing and aspirating languages alike.

2.2 **Crosslinguistic background – VOT and F0**

Phonetic VOT (Lisker & Abramson, 1964) cues have shown indiscutable variations in many languages, whether these languages have a voicing contrast as in Dutch and Hungarian (and Indonesian: Li et al., 2019; Batais, 2013), an aspiration contrast as in Cantonese (and Mandarin Chinese: Li et al., 2019), or contrasts based in both voicing and aspiration as in Hindi, Eastern Armenian, Thai, and Marathi (Lisker & Abramson, 1964). Ideally, languages

---

9 The variations in these languages’ VOTs are not surprising; there could be many sociolinguistic factors that have contributed to this level of interaction in voicing. Working on already-existing corpora compiled by different linguists, Kakadelis (2018) could not control for the sociolinguistic factors that were possibly motivating the data—such as bilingualism, gender, and age, to name a few. In a quite related case, Cho & Ladefoged (1999) could not find a legitimate reason why VOT in aspiration was seen to have three different durations (it cannot even be related to the distribution of the phonemic inventory of stops). For more about VOT patterns and the diversity of languages, see Cho & Ladefoged (1999).
that feature a three-way contrast using both voicing and aspiration would have three VOT lengths: a negative lead (a property of voiced plosives, with an exception for English), a short positive lag, and a long positive (reserved for voiceless stops), all located along the voice-onset time continuum. However, variabilities in voiceless plosives led Cho and Ladefoged (1999)\textsuperscript{10} to categorize languages with aspiration-related contrasts as “unaspirated” (30), “aspirated” (50), “slightly aspirated” (90), and “highly aspirated” (120). Therefore, the distribution of any such divisions is language-specific and depends mostly on the lengths of each language’s contrastive plosives.

It is further reported by Cho et al. (2019) that some languages exhibit a four- or five-way laryngeal contrast. Hindi and Urdu have a four-way contrast in their voiced and voiceless segments for each place of articulation (such as the bilabials: /p/, /pʰ/, /b/, /bʰ/). Furthermore, by adding the implosive [ɓ], Cho and Ladefoged (1999) claim that Sindhi qualifies as a language with a five-way contrast. Although the authors posit a challenge to voice-onset time to account for the (somewhat) overlapping contrast, they also call for investigation of other fine phonetic details for further control over the results. For this reason, Schwarz et al. (2019) propose an implementation of Pre-Vocalic Interval (PVI), following an earlier proposal reported by Berkson (2012).

Slightly different than VOT, which is the duration of the stop before the increased periodicities of the vowel, PVI (as Schwarz et al., 2019, define it) is a measurement of the interval from the release of the stop closure until the onset of the periodic noise in the following vowel. It begins from the burst of a voiceless segment (see Figure 2-1), but is delayed in voiced plosives until the end of pre-voicing; this marks the start of the breathy

\textsuperscript{10} Unlike Lisker & Abramson (1964), who found the highest aspiration (of 100) to be for /kʰ/ in Thai, the VOT of the velar /kʰ/ for one endangered language (Navajo) examined by Cho & Ladefoged (1999) was 154 ms.
duration before the vowel’s onset, with a longer PVI for a breathy voiced stop than for its non-breathy equivalent (see Figure 2-2)\textsuperscript{11}.

Berkson (2012) diverges from an earlier work by Mikuteit and Reetz (2007), which discussed an alternative way to account for the four-way laryngeal contrast in East Bengali. The focus of Mikuteit & Reetz’s study is intervocalic, and they argued that measuring VOT in medial positions has proven problematic. Therefore, after-closure time (ACT) and superimposed aspiration (SA) are crucial post-release intervals for them. They argue that aspiration extends to the vowel’s initial amplitude (pulsing), so that SA is the vowel’s early breathy portion and includes the aspiration. ACT, on the other hand, is like the positive VOT that is observed for voiceless stops, and they propose that it can be extended to those sounds’ voiced equivalents. Thus, ACT gives the measurements for both voiceless and voiced

\textsuperscript{11} The words reported in the figures are from Nepali (as in Schwarz et al, 2019: 119): tara ‘star’, thal ‘plate’, dal ‘lentils’, and d'ar ‘rice paddy’. Nepali has a four-way laryngeal contrast for bilabial, alveolar, retroflex, and velar sounds. The procedure carried out for the alveolars (in Figs. 1 and 2) was also used for measuring the other places of articulation.
segments as positive lags. The VOT voicing lead, as posited by Lisker and Abramson (1964, among others) is eliminated during the analysis of Mikuteit and Reetz (2007). Their measurements have shown similar conclusions for VOT (Cho & Ladefoged, 1999; Maddieson 1996; Lisker & Abramson, 1964), in which bilabials are short and velars, as most studies report, are longer than alveolar. Their general conclusion is that ACT is longer for voiceless sounds, whereas SA is longer for voiced sounds.

Yet, since our data for production and imitation present neither the voiced non-aspirate nor the voiced aspirate, we still must use the traditional VOT in measuring the produced tokens and in all manipulations of stimuli. The duration was measured from the release of the stop until the vowel’s onset (see Figure 2-3, representing similar criteria as those of PVI in Figure 2-1 for the homonymous word tai in two different meanings).

In addition to VOT, the maneuvering of the phonetic cues would target another variable in the study: namely, F0. The increase of fundamental frequency heights is manipulated in the portion that follows the ending point of VOT; it is the onset of the following vowel. As such, although other authors (Berkson, 2012; Mikuteit & Reetz, 2007; Schwarz et al., 2019) called attention to variabilities along the (VOT) durational continuum and made arguments about which technique is the best, our mission is to give attention to the ongoing VOT debate and place emphasis on the interaction between VOT and F0.
Figure 2-3 The shaded area represents the VOT of the word *tai*, meaning ‘tie’ (top) and ‘eagle’ (bottom), as produced by ZS8MPI (a male from the northern island – Town Z)

Even older than Lisker and Abramson (1964) study, which initiated linguists’ exploration of VOT, is House and Fairbanks (1953) study, which suggested that voicing has an effect on the F0 of the following vowel. As pioneers in this research, they showed that a voiced consonant would be followed by a lower average F0, while voiceless consonants would be followed by a higher F0. Later studies (Kingston, 2009; Kirby & Ladd, 2015, 2016) further supported the conclusion that voiceless and voiced segments work in correlation with high and low F0s, respectively. By analyzing /p/ and /b/, Kirby and Ladd (2015) also found that in both Italian and French, the longer the VOT lead is, the lower the F0 will be. While they failed to report the same for voiceless stops in French, Italian, on the other hand, showed that the longer the /p/ VOT, the lower is the F0 value.12

It is not unusual that a language should exhibit an F0/VOT contrast for different dialects. Two Korean dialects with a three-way contrast, perceptually analyzed by Lee et al. (2013), showed a trade relation between VOT and F0. A sample of forty-two participants (21 each from Seoul and South Kyungsang) performed an identification task using modified VOTs and F0s (12 each) for one token: *pul* ‘fire’. The result was that the Seoul dialect, which is a non-tonal language, relied on F0 to distinguish lenis responses, but it needed both VOT and F0 for aspirates. The other, tonal dialect (Kyungsang Korean), on the other hand, depended primarily on VOT and to a lesser degree on F0 to identify the tokens.

12 In their discussion of the results, Kirby & Ladd (2015) suggest that the absence of controlled intonation in their data led to the variation between Italian and French, which could be a product of the usual intonation in the two languages: falling for Italian but rising-falling for French.
Similarly, studies in language acquisition have demonstrated an ongoing phonetic relationship between VOT and F0. Data obtained from children have given evidence for implementation of phonetic cues while acquiring another language. Lee and Iverson (2012) report on Korean-English bilingual children, as well as Korean monolinguals and English monolinguals (thirty each, divided into two age groups of five and ten years old). The results revealed that the younger group (of bilinguals) depended on VOT (not F0), and their Korean stops tended to be lengthened more than their English stops. At the age of ten, bilinguals discriminated the stop categories of both languages more neatly, by using VOT and F0 in parallel. This result supports Flege (1991) argument that early bilinguals establish separate phonetic systems for their two languages. When asked to produce /t/, bilinguals in Flege’s study were able to articulate a shorter VOT for Spanish and a longer one for English. As the current study involves both monolinguals and bilinguals, it would be interesting to see how the bilinguals react to the different synthesized VOT/F0 tokens.

Furthermore, works on Indo-European languages have revealed that voicing contrasts can have an effect on F0. Kirby and Ladd (2016) examined the F0 (CF0) of four obstruents (/p-f/b-v/) and one sonorant (/m/). They found that in French and Italian, the CF0 phonetic gesture is lowered throughout the closure transition of voiced consonants but raised after voiceless consonants, giving evidence for true voicing in languages through a support system driven by primary phonetic cues. Kirby (2018) went a step further to test variation in languages with a three-way contrast: namely, tonal (Central Thai and Northern Vietnamese) and non-tonal (Khmer) languages spoken in Southeast Asia. The results revealed that aspiration in the three languages reflected a higher CF0, yet the non-aspirated sounds

---

13 The C stands for the effect of the consonant on the following F0.
presented some CF0 variabilities, which could be interpreted as individual- or language-specific.

To gain an overview of how laryngeal contrasts work in languages all over the world, Cho et al. (2019) investigated 11 studies involving 19 languages. Their final conclusion and suggestion was that “[w]e have seen further evidence that the phonetic and phonological nature of laryngeal contrast in voicing may never be complete without looking into other phonetic properties that may co-occur with or replace VOT” (Cho et al., 2019: 64). In our current study, I will consider VOT as the primary phonetic cue to be explored; however, I will use F0 as a secondary phonetic detail for supporting our analysis.

2.3 **Crosslinguistic background – Sensitivity to VOT and F0**

The scholarly written work on the sensitivity of humans to VOT has gotten further attention from researchers, particularly since the publication of Lisker and Abramson’s article (1964). Yet, despite the study’s valuable insight into variations in VOT, the work of these two phoneticians is limited to production. The variation in production of stops has also been attested to in a later study by Cho and Ladefoged (1999) on endangered languages. Due to the great variability in the VOT of voiceless stops across the languages they surveyed, Cho & Ladefoged as mentioned in the previous section, came up with four classifications: unaspirated, slightly aspirated, aspirated, and highly aspirated.

Therefore, if speakers of languages with a laryngeal contrast can articulate with different VOT durations to refer to non-identical lexical items, then we can test these speakers’ perceptual sensitivity to different VOT values. Furthermore, the perception test is not limited to languages with categorical aspiration; it has been reported that even subjects whose language’s laryngeal contrast is only phonetic have performed effectively in such an experiment. A claim posited by Caramazza et al. (1973), after examining mono- and bilingual
speakers of Canadian English and French, stated that, unlike what is found for English monolinguals, French monolinguals were insensitive to VOT duration. French and English bilinguals, on the other hand, presented a shift in their application of phonetic cues, with more control over production than perception.

Afterwards, studies that focused on the perception of phonetic cues showed how subjects reacted to different synthesized durations of the sounds (Allen & Miller, 2004; Eimas & Corbit, 1973; Iverson & Kuhl, 1995; Volaitis & Miller, 1992). Initially, a shift in speakers’ reactions to voiced and voiceless tokens is highly dependent on their level of exposure to the stimuli. In addition, stimuli near the phonetic boundary would compete to be adapted according to two overlapping VOT values. The final adaptation abides by a specific “linguistic feature detector” (Eimas & Corbit, 1973: 108).

Subsequent studies have expanded on this initial investigation and extended the analysis to link VOT duration with syllable duration. Thus, because of VOT sensitivity, values shift to be either higher or lower in parallel with the speaking rate—increased VOT with a slower speaking rate, and vice versa (Volaitis & Miller, 1992). Findings from Iverson and Kuhl’s study (1995), as mentioned above, support the idea of a “perceptual magnet” and suggest that prototypes form a cluster, in which a token that shows similarity with the “good” prototype is pulled towards the cluster, while tokens near the far edges of the cluster are perceived as “bad” prototypes. In other words, the prototype has a perceptual space, whereby a segment is processed during perception as being a good exemplar to a certain threshold level. Other segment tokens, which are perceived to be residing outside the category boundary, are not only bad prototypes but also different sounds.

14 This theory of phonetic prototypes will be explored in greater detail in a later section of this chapter (2.7).
Sensitivity to speakers’ specific VOTs is investigated in Allen and Miller (2004). In order to test the way listeners track phonetic variants, varying short/long values of two speakers’ VOTs for initial stops were presented to listeners in a training session. The purpose of this session was to accustom the listeners to the speakers’ voices, and it was followed by an identification task and then some feedback. The overall result after the test session was that the experience the two groups gained from the speakers during the training session significantly affected their responses on the test and extended their stored prototypic tokens to novel words. In other studies, similar to the results for VOT, discrimination of specific-talker phonetic categorizations was also detected in pitch.

Nazzi et al. (1998b) found that infants can successfully perceive F0 intonations with rising and falling pitch: French newborns showed an ability to distinguish between non-native (Japanese) words which varied in the pitch accent. This pattern is also observed with language pairs such as English and Japanese (Nazzi et al. 1998a). The results of Nazzi et al. (1998a) showed more variation than that indicated by Nazzi et al. (1998b) when the “rhythm class” was similar. Furthermore, when using filtered sentences, English cannot be contrasted with Dutch, as they share the same syllable pattern (stress-timed). A contrast between a group of stress-timed languages (English and Dutch) and a group of syllable-timed languages (Spanish and Italian) is possible. However, the contrast was not observed when the combination was mixed (English and Spanish vs. Dutch and Italian).

Working with adults, a study by Krishnan et al. (2010) contrasted speakers of two tonal languages (Mandarin and Thai), then compared them with a non-tonal language (English). The aim was to determine whether experience or lack of experience with tones could be transferred crosslinguistically. Subjects were exposed to a synthetic version of the Mandarin /yi/, representing six words with different tone levels: three for each tonal
language. Every token from the six had a different F0 contour, which is considered to be token-specific (voice fundamental frequency). The results of this study revealed an accuracy rate among Mandarin and Thai speakers regarding sensitivity to synthesized tone levels. Speakers of English, on the other hand, failed to differentiate between rising and falling pitch magnitudes. This effect is inherited from the features of their native language, English—a language with no highly active tone system.

Sensitivity towards a certain non-primary cue is hypothesized, in turn, to enhance a primary cue. In Kwon (2015), for example, one phonological category (such as aspiration) is perceived through a phonetically increased F0 height. For Korean listeners in the study, the F0 is a primary cue for the laryngeal contrast, whereas the VOT is a secondary cue which leads to a rough recognition of the contrast in the language. Therefore, when presented with a lengthened VOT, imitators would give VOT imitative responses accompanied by higher F0s, but raised F0s in the stimuli would not trigger a long VOT. That is to say, a modified VOT impacts the perception of the token under investigation and then extends its perceptual cue to influence the pitch, but on the other hand, manipulation of pitch has a minimal effect on lengthening or shorting the VOT. Kwon (2019) relates the mismatch between these two imitated cues to primacy vs. non-primacy. As a result, speakers’ high sensitivity to F0 has opened the door for the lower-level cue (VOT) to be perceived and then interpreted with reference to another phonetic property through a one-directional effect, from VOT to F0.

On the other hand, the test of perception done by Kwon (2013) on English speakers showed that sensitivity to one phonetic cue (VOT) was not strong enough to lead the participants to perceive the difference in F0 manipulations. Kwon tried to see if the Korean three-way contrast was identifiable by English listeners (n=122, University of Queensland students). However, since Korean is known for employing VOT (for fortis and aspirated) and
F0 (for fortis and lenis) to distinguish stops while English does not use F0 for this purpose, English speakers successfully contrasted what was based on VOT but not F0.

2.4 Crosslinguistic background – Gender

Gender, as an independent variable, is a crucial element in our study. It has been shown in the literature that gender shapes production of the fundamental frequency in that females have higher F0s than males. However, it is not clear whether VOT varies across genders. For this reason, one of the tasks of this study is to look into gender-based VOT variability, if there is any. It is, however, unknown whether samples collected for this sociophonetic endeavor would show any difference between Swahili as an L1 and an L2.

The only background we can refer to is some early literature which stated that gender enforces some sort of distinction in the produced VOT tokens. For example, it has been reported for languages other than Swahili (such as English) that females exhibit longer positive VOTs than males, but shorter negative durations (Ryalls et al., 1997). The participants in the Ryalls et al. study were male and female speakers of Caucasian or African-American English, who also exhibited dialectal differences in that the African-Americans had longer voiced stops than the Caucasians did (a similar observation was also reported by Herd (2020), who suggests a dialectal influence on VOT). Whiteside and Irving (1998) found that women (speaking British English) had longer positive VOT values but shorter VOTs for voiced stops, while Morris et al. (2008), working on American English, found that women’s VOTs were a bit longer than those of men, but not statistically different. Voiced stops in the Morris et al. study were very similar across the two genders, with a difference of approximately one millisecond (see also Robb et al., 2005)\textsuperscript{15}, who studied Caucasian

\textsuperscript{15} The study of Robb et al. (2005) was conducted under two conditions: a lab setting and a non-lab setting. In their study, voiceless stops yielded the same result in both settings.
speakers of American English and found that women’s voiceless stops were statistically different than those of men, but no significance was observed for any of the voiced stops). The lack of significance in the Morris et al. study is also supported by Oh (2019), who reported longer VOT values in voiceless and voiced stops for male speakers of American English, and this difference did not reach the level of significance, either.

Due to this evidence, we might expect gender-based variability in VOT to be a physiological property crosslinguistically, but in at least one case, an Indo-European language has shown measurement values that are the opposite of what is observed in many studies on English. In her analysis of the Serbian stops /p/, /t/, /k/, /b/, /d/, and /g/, both word-initially and in sentences, Sokolovic-Perovic (2012) revealed that males had significantly longer voiceless VOT values, while females had longer pre-voicing VOT values that did not reach the level of significance. A similar observation also holds true for Hungarian (a non-Indo-European language). In their analysis of Hungarian, Gósy and Ringen (2009) found that males’ positive VOTs were significantly longer for most stops (/k/ was not significantly different, but still longer for males). Voiced stops, on the other hand, were all significantly longer for females than for males.

In the Middle East, the Semitic language of Arabic has shown some variability in VOT across dialects and genders. Alshahwan (2015), who examined the Najdi and Bahraini dialects of Arabic, revealed that the voiceless stops /t/ and /k/ were longer for Saudis than for Bahrainis, but there was no difference for the voiced stops /b/ and /d/. And although there were no gender differences for voiced stops, males had shorter positive VOT durations (for most voiceless sounds) than females. The voiceless emphatic /tˁ/, on the other hand, yielded a different result, maintaining significantly longer VOT values for men rather than women.
Farther east, Li (2013), who compared standard Mandarin Chinese stops in various environments (/t/, /d/, /k/, and /g/ followed by /a/, /u/, or /i/), revealed gender differences in the production of stops: female speakers tended to have longer voiceless stops but shorter voiced stops. Regarding the different environments, Li found that the post-stop vowel had an effect on the length of /k/, /d/, and /g/ but not /t/, and it was also observed that all VOTs were significantly longer before /i/. The effect of the post-stop-vowel context on the VOTs of Mandarin matches some of the prior results for English (Morris et al., 2008) and Arabic (Alshahwan, 2015).

More evidence for the need to consider gender when examining VOT comes from (Peng et al., 2014), who looked into Taiwan Mandarin and Hakka\textsuperscript{16}. The two dialects’ unaspirated/aspirated stops behaved differently based on sex: men had statistically longer VOT values than women for all unaspirated stops, but women had longer measurements for the aspirated stops. When considering place of articulation, the authors confirm that statistical significance was found only for the bilabials of Mandarin, although their earlier study (Peng et al., 2014), also on Mandarin and Hakka) had shown significant statistical differences with respect to phonation and gender. In their 2014 study, Hakka failed to maintain the same significance pattern that was observed for Mandarin.

Oh (2011), working with 19 male and 19 female students from two Korean universities, examined the VOTs of fortis, lenis, and aspirated\textsuperscript{17} Korean sounds in a /CVn/ syllable structure. She observed that the males produced significantly longer VOTs for aspirated stops (both in isolation and in a phrase) and were able to set a distinct categorization boundary between the aspirated stops and the lenis ones; however, neither

\textsuperscript{16} This study concentrated on the voiceless laryngeal contrast before the three vowels /a/, /i/, and /u/.

\textsuperscript{17} “Fortis,” “lenis,” and “aspirated,” in the study, refer to a very short lag, an intermediate lag, and a very long lag, respectively.
observation was found for the females. Another categorical boundary, meanwhile (fortis vs. lenis), did not show any significant gender differences. Oh’s study concludes by questioning the theory that ascribes VOT gender differences to anatomical differences based on sex physiology (Oh, 2011: p. 65) and related lung volumes and laryngeal settings.

Other languages also exhibit significant gender-based variability in the production of VOTs. Ma et al. (2018), who worked on Mandarin, found that the gender difference was significant and applied to participants age 14 and older. After comparing two series of voiceless stops, Ma et al. concluded that the females had significantly longer aspirated stops than the males. The unaspirated stops /p/ and /t/ did not show any significant difference based on gender, but males did produce significantly longer unaspirated VOTs for the velar /k/. Furthermore, the Mandarin data revealed (as did Yu et al., 2015, who found longer VOTs in native English speakers between the ages of 8 and 11) that age and the level of aspiration also play a significant role in affecting VOT variability.

Table 2-1 below summarizes the results of the abovementioned studies, arranged by language and the types of VOT that were investigated.

<table>
<thead>
<tr>
<th>Language &amp; Authors</th>
<th>Voiceless VOT</th>
<th>Voiced VOT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>AE – Ryalls et al. (1997)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>AE – Robb et al. (2005)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>AE – Morris et al. (2008)</td>
<td>(N/S) long</td>
<td>short</td>
</tr>
<tr>
<td>AE – Herd (2020)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>AE – Oh (2019)</td>
<td>short</td>
<td>(N/S) long</td>
</tr>
<tr>
<td>BE – Whiteside &amp; Irving (1998)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>Language(s) &amp; Authors</td>
<td>Aspirated VOT</td>
<td>Unaspirated VOT</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Chi. – Li (2013)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>Chi. &amp; Hak. – Peng et al. (2009)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>Chi. &amp; Hak. – Peng et al. (2014)</td>
<td>(N/S) long</td>
<td>short</td>
</tr>
<tr>
<td>Chi. – Ma et al. (2018)</td>
<td>*long</td>
<td>short</td>
</tr>
<tr>
<td>Kor. – Oh (2011)</td>
<td>short</td>
<td>*long</td>
</tr>
</tbody>
</table>


VOTs are also seen to change over the course of time. Stuart-Smith et al. (2015), analyzing a variety of spoken Scottish English produced by women from different age groups, found that older speakers recorded in the 1970s (born in the 1890s) had a significantly short /p/ and other stops compared to recordings of younger speakers born in the 1960s. However, in similar recordings made in the 2000s, the elders (born in the 1920s) showed longer VOTs than the younger people (born in the 1990s). In addition to the fact that the voiceless values in the newer recordings were pretty long for all three voiceless stops, their /p/ and /t/ were significantly longer than those of the speakers recorded in the 1970s. With no further evidence obtained regarding the frequency of lexical items in their data and no effect detected from the following vowel’s height, the authors suggest that these VOTs are not related to age or an inherited articulatory or physiological property, but rather to a sociolinguistic factor such as a dialect shift between a vernacular and a standard. Stuart-
Smith et al. (2015) further speculate that the recordings’ speech style (a mix of formal and informal speech) might have played a role in these dissimilar observations of the same age group three decades before. Their study was meant to provide a sense of what effect age might have on VOT, yet the focus of the present review is gender differences and how they may influence the shape of VOT representations. In our current study, we controlled for the age variable by targeting a much less extreme age gap (working with mostly university students).

It is always clear-cut to understand the anatomy and physiology behind the distinctive fundamental frequency (F0) of males and females: on average, F0 is higher for females and lower for males. The issue of VOT, however, is quite a mysterious one. Although some linguists suggest that some sort of physiological property is behind the differing VOTs of males and females (Koenig, 2000; Whiteside et al., 2004; Yu et al., 2015), this kind of assumption cannot be applied crosslinguistically. A number of linguists (Herd, 2020; Li, 2013; Oh, 2011; Robb et al., 2005; Simpson, 2009; Sokolovic-Perovic, 2012) also call for a deep consideration of the possible sociophonetic/linguistic factors behind the variability that exists in VOT across different populations.

Most studies which discussed VOT as being affected by gender and/or age have focused mostly on production, where it is always the case that the alveolar is longer than the bilabial and the velar is longer than the alveolar (Brinca et al., 2016; Chodroff et al., 2019; Lisker & Abramson, 1964). To my knowledge, however, there has never been a lengthy study on an African language, with many participants, that looked into the imitation of VOT and F0 contrasts by shedding some light on both perception and production. Hopefully, by having an equal number of males and females in each chosen location, we can get some answers from Africa to explain bits and pieces of these overwhelming VOT variabilities.
2.5 **Crosslinguistic background – Sociophonetics**

Sociolinguistics has focused on the ways in which spoken languages are linguistically variable, and the works of Labov (1966; 1972, and many others), whether on the production of consonants or on that of vowels, have influenced the literature in this field in many ways. Since then, linguists have been investigating the varieties and variables of sounds in what is called “sociophonetics.” Thomas (2001) is one study, motivated by Labov’s approach, that discussed the acoustic properties of English as spoken in North America and some parts of the Caribbean. His groups (192 subjects in all) were divided ethnically and geographically. In each group, he examined unequal numbers of participants; his justification for having some groups be smaller than the others was that other linguists had already discussed the productions of those smaller groups exhaustively. Among his contributions is the analysis of a 1940s recording of former African-American slaves. By analyzing some contemporary speakers in comparison with older speakers, Thomas’ study shows a diachronic shift in the English vocalic inventory.

Moreover, the geography of spoken English has been seen to employ dissimilar phonetic cues for the realization of voicing in coda position (e.g., /Vt/, /Vs/ vs. /Vd/, /Vz/). In the United States, Jacewicz et al. (2007) have reported a consistent shift among the three major varieties of spoken English in North America: the pre-voiced vowels were longest in the Southern variety, intermediate in the Midland, and then shorter in the Inland North. In all regions, vowels were always longer before voiced consonants than they were before voiceless consonants. In addition, the diphthong /ai/ had the greatest length of all the vowels that preceded a voiced consonant. According to Jacewicz et al. (2007: 379), the interaction between vowel duration and gender in the three regional areas was not significant, which led the authors to call for revisiting the judgments of older studies on the role of women in sound change and to take a second look at gender-based variation in the production of vowels.
Furthermore, in their examination of the Scottish Vowel Length Rule (SVLR)\textsuperscript{18}, Scobbie et al. (1999) revealed that the effect of the SVLR is observed for the high vowels /i/ and /u/, as well as for the diphthong /ai/. An extensive analysis of /ai/, they claim, has also shown that the diphthong (and the SVLR) are being influenced by non-Scottish English dialects.

Studies on the status of laryngeals have found that contrasts are present in all positions within a word: initial, medial, and final. The aspiration of consonants in word-initial position is documented for languages such as English, and it is also found in unrelated languages. The Tibetan language, as studied by Hill (2007), has historical aspiration in word-initial and post-nasal positions, but word-medial positions are unaspirated\textsuperscript{19}. Standard Tibetan orthography still spells this laryngeal contrast with inconsistent aspiration, but on the other hand, loanwords from Chinese and words from some other dialects were assigned an unaspirated feature (ibid.).

Other languages, however, possess different sociophonetic mechanisms. Tyrolean, a variety of spoken German, had aspiration that was restricted to word-medial position. Vietti et al. (2018) investigation of this dialect considered all the possible contexts for aspiration and, interestingly, voiceless stops were neutralized in word-initial and word-final positions. According to the authors, the word-final neutralization is expected since it is a part of general German phonetics, but the word-initial neutralization of laryngeal contrasts was surprising and behaved unequally: bilabials neutralized the most, followed by velars, and then alveolars. We also note that Smith (2013) observed a different order for final devoicing in Toronto English; in his data, the most devoicing occurs in velars (26%), then bilabials (23%), and then alveolars (18%). When we consider the data for the present project, we see that stop-

\textsuperscript{18} That is basically the duration of the vowel as conditioned by the following consonant, mostly a stop and/or a fricative.

\textsuperscript{19} The post-nasal stops are still at the beginning of a word; therefore, the “word-medial” rule wouldn’t apply even though the nasal is—strictly speaking—the first sound in the word.
final neutralization is not possible since all the words have open final syllables. The medial position is a potential place for examining the voiceless behavior in our data; however, since our focus is on stops and affricates in word-initial position, non-initial stop locations will be put aside for now, along with vowel effects on the target stops.

2.6 **Background – debate re: loanword adaptation**

When previous scholars have considered the question of whether loanword adaptation is phonetic or phonological, they have usually found that phonetic details give the best account for any resulting alternations. In one of the published works (Ito & Kenstowicz, 2013) that investigate Japanese loanwords in Korean, analyses of recordings that had been made prior to 1940 revealed that a cue such as VOT measurements gave evidence for the specific VOT profile of the old period. In their data from the old period, the VOTs of lax and aspirated Korean stops (tense = 11; lax = 38; aspirated = 115)\textsuperscript{20} were categorically different from those of voiceless stops in Japanese (voiced = -93; voiceless = 17), but over time, the VOT differences between the lax and aspirated stops were reduced. When compared with further published data from the contemporary period for the two languages, the adaptation appeared to be rather phonetic, since the VOTs have shifted to resemble close values along the VOT line.

On the other hand, Desmeules-Trudel & Paradis’s (2012) analysis of English loanwords in Thai led them to the conclusion that these loanwords are adapted in a way that ignores the phonetic contrast of English stops because the bilinguals who presumably initiated the borrowings know that English aspiration is not categorical. In a corpus of interviews that feature Thai speakers using English words with aspirated stops, 58% of stops

---

\textsuperscript{20} These would meet the same categories that are referred to as “fortis”, “lenis” and “aspirated” in the discussion of other studies on Korean (see Oh, 2011: 61).
which are aspirated in English were realized as aspirated in the Thai speakers’ pronunciation; on the other hand, 40.5% of stops which are aspirated in English were realized as unaspirated by the Thai speakers. Desmeules-Trudel & Paradis suspected that some Thai speakers who were familiar with English were realizing the English stops according to English phonetics. Therefore, in a second phase of their investigation, they excluded Thais who had knowledge of English, and as a result, the percentage of English aspirated stops which were realized as unaspirated increased drastically (90%). That is to say, tracking VOT in loanword adaptations has demonstrated practical utility in determining the processes that govern the nature of adaptation.

In order to go further with our contrastive analysis of the sound system of Swahili (with regard to stops and some affricates), we need to deepen our understanding of the status of Swahili aspiration, implosives and word origin. Unlike with Arabic and English, scholars are still not quite sure what to stress when talking about the phonological categorization of Swahili. For some linguists, Swahili makes a phonological distinction between plosives and implosives (Hayward et al., 1989; Kharusi, 1994; Polomé, 1967; Tucker & Ashton, 1942). Others emphasize the contrast between aspiration and non-aspiration (Engstrand & Lodhi, 1985; Frankl, 1991; Lodhi, 2004). According to Engstrand & Lodhi (1985) personal examination, the aspiration is strictly for native words, not loanwords. So, in near-homonyms like karo, the variant with aspiration [kʰaro] has a purely Bantu origin; however, the unaspirated [karo] is a loanword\(^{21}\). However, my measurements of an audio file from the UCLA Phonetics Lab, which contained a recording made in 1973 of a female speaker from Mombasa, have supported the idea that aspirated/unaspirated contrasts are even found for pure Swahili minimal pairs\(^{22}\).

\(^{21}\) Lodhi stressed this assumption in an email I received in July 2017.

\(^{22}\) http://archive.phonetics.ucla.edu/Language/SWH/swh.html.

34
My initial investigation of the relation between Arabic and Swahili started by testing the adaptation of word-initial stops and whether they were borrowed phonetically or phonologically (Alsamaani, 2017). The basic idea is that, if the acoustic/perceptual cues (such as presence or absence of aspiration) of the donor language are transferred faithfully to the borrowing language, then the adaptation is very much linked to the prominence of the phonetic details. Some adaptation, as is claimed by bilinguals, follows the phonological account during the process of borrowing, so that phonetic cues constitute non-contrastive features, and as a result, they play no direct role in phoneme-to-phoneme mappings. For example, Table 2-2 gives a sense of how the adaptation of stops may proceed either categorically (i.e. phonologically) or perceptually (i.e. phonetically).

Table 2-2 Phonemic vs. allophonic approximation

<table>
<thead>
<tr>
<th>Phonetic Adaptation</th>
<th>Phonemic Cue → Phonemic Cue</th>
</tr>
</thead>
<tbody>
<tr>
<td>[VOICELESS] → [VOICELESS]</td>
<td>(VOICELESS) / (VOICED) → (VOICELESS) / (VOICED)</td>
</tr>
<tr>
<td>Phonological Adaptation</td>
<td>Phoneme → Phoneme</td>
</tr>
<tr>
<td>[VOICELESS] → [VOICED]</td>
<td></td>
</tr>
</tbody>
</table>

Building on what has been said about the nature of adaptation, I hesitate to support either theory. For instance, if Arabic has phonetically aspirated stops just like English, then the question is: do Swahili speakers ignore or preserve the phonetic details of Arabic and English? This question is raised because it is believed that Swahili speakers are perceptive to such cues.

2.7 Crosslinguistic background – Prototype theory

The theory of phonetic prototypes was first introduced in the field of psychology in order to reflect the cognitive representation of categories. One of the pioneers of this theory was Rosch, whose study (1975) demonstrated that subjects would use the name of a basic color as the “best” prototype of the category, whereas a less representative color in the same general
category is considered a “poorer” prototype. In a follow-up study, Rosch (1983) added that a subject’s reaction time to the model (representing the prototype) is slower or faster according to his or her familiarity with the category being tested. This hypothesis is based on experiments done by Rosch and Mervis (1975) with reference to “family resemblance.” That is to say, a prototypic member is the one that shares the most attributes with those of the prototype but correlates less with contrasting categories. Rosch & Mervis’ interest was not in drawing a clear-cut line between attributes of a prototype and those of a non-prototype; rather, as they proposed, “there is a principle of the structure of stimulus sets, family resemblances, which can be shown to underlie category prototype structure” (Rosch & Mervis, 1975: 576).

In support of the theory that prototypes have a varying distribution crosslinguistically, (Taylor, 2003: 11-12) disagrees with the idea that familiarity is a factor in creating categorical entities. He gives an example of Russian as having twelve terms for light blue and dark blue. On the other hand, Zulu (a language spoken in South Africa), like most other Bantu languages, lacks this basic color naming that separates blue from green focally; in order for Zulus to distinguish between the two colors, luhlaza (which refers to both) is embedded into a phrase to show the intended type of color. Taylor posits that fewer terms for basic colors would allow a wider color variation for each category: in other words, the categorization of colors has “a center and a periphery,” and prototypes for representing a certain color have a gradation in their typicality and resemblance to the best exemplar. He further believes that the names of colors are less arbitrary because, although the color terms are independent and can be contrasted in the color space, the category’s center is not contrasted (Taylor, 2003: 14).
To extend the prototype model to linguistics, Samuel (1982) worked on phonetic prototypes of /ga/ and /ka/, synthesized in a 3-ms length along the voice-onset time (VOT) continuum from 0 to 120 ms. In his main experiment, subjects (n=10 women who are native speakers of English) were asked to discriminate between the prototypes and non-prototypes on either end. The results indicated that adaptation is greater when the tokens are near the range of the subject’s prototype; however, the subject’s prototype location is constrained by individual variations observed in the study, as is evident from the VOT ranges of the voiced velar (3-33 ms) and the voiceless velar (87-117 ms). The gap between the values of /g/ and /k/ is the boundary range of 45-75 ms. Samuel’s (1982) results cast doubt on an earlier study by Miller (1977), who reported tighter VOT measurements (n=8) which, as a sequence, made it logical for her to suggest that the level of adaptation was stable. Samuel refuted Miller’s claim due to the possibility of having participants who happened to have obtained adjacent VOT values in their laryngeals.

Phonetic prototypes are also at the center of Kuhl’s (1991) question about the forces behind the formation of speech categories. She argued for the possibility of having an internal category cluster of prototypes which are dependent on the level of experience, and to provide evidence for this, she conducted experiments on adults, infants, and monkeys. The selected prototypes were elicitations from male speakers pronouncing the vowel /i/. The “best” prototype was based on an earlier study conducted by Grieser and Kuhl (1989), while the “poor” /i/, which is located on the periphery but still perceived as /i/, was selected as a non-prototype. In simple terms, the study had 32 variants located on 4 orbits, which were distributed on 8 vectors around the most prototypic variant. Additionally, since the results of Kuhl’s study were unconcerned with category boundaries, the internal structure of categories was perceived by humans (which, as Kuhl claimed, is due to the “perceptual magnet” triggered by the prototype) but not by animals, demonstrating a correlation between the
degree of “goodness” and perception. This is a property that is found among human beings but currently unobservable in nonhumans, a phenomenon that is in line with the prediction of Eimas & Corbit, who wrote in regard to VOT: “The apparent universality of this phenomenon suggests that it is a manifestation of the basic structure of the human brain” (Eimas & Corbit, 1973: 101). Later, the results from a study by Iverson and Kuhl (1995) revealed that participants could identify poor prototypes but almost failed to discriminate between prototypes that were clustered around the center.

Building on earlier proposals by Goldinger (1998); (Goldinger, 2000), who claimed that perceived linguistic words (of created phonological traces) can be stored in the memory until they are activated again, Nielsen (2011) extended the model of retained clusters of similar tokens to the phonemic level (i.e. /p/ and /k/). The main results of the two experiments she performed showed an asymmetrical relation between reduced and lengthened VOTs. Nielsen argued that the subjects, native speakers of American English (n=25), imitated artificially-extended VOTs but did not imitate shortened VOTs due to the possibility of overlapping the categorical boundary of short VOTs between voiced and voiceless segments. Also, the subjects imitated novel tokens. The study showed that imitation is not limited to phonemes, but extends to other phonetic features. It is worth studying, however, whether different categories of laryngeal prototypes are perceived and categorized distinctly.

As has been proven in many studies, phonetic cues are crucial in order to account for the variability of segments in a language, particularly obstruent consonants. Thus, our purpose in this study is to follow up on the observations of previous studies in regard to VOT and F0 laryngeal contrasts. This study also aims to target different places of articulation: bilabial (/p/), alveolar (/t/), velar (/k/), and the post-alveolar affricate (/tʃ/). Aside from the pure production test that is based on picture-naming, it is a listener-turned-speaker task that
uses imitation to test speakers’ sensitivity to synthesized tokens that have been designed for this project. Each token used in this experiment has been modified from a recording of a model talker by manipulating (and synthesizing) the VOT and pitch values.

2.8 Background – The dilemma of Swahili orthography

There is a serious debate which comes from the uncertainty that is shown by some Swahili linguists with regard to the exact region in which aspiration is active. Polomé (1967: 37-40), for example, assigns the language both a broader and narrower a set of phonemes: what he calls “maximal” and “minimal” sounds, respectively. Nevertheless, not all speakers of Swahili have a solid grasp of the minimal set, except for first-language speakers residing in Zanzibar and along the “Mrima coast”, which is the coast that faces the islands of Pemba and Unguja. However, Polomé (1967) claimed that the minimal set is well contrasted by people from Mombasa and well preserved by residents of Zanzibar in both the north and the south, with some reported decline among the people from the middle of the island. Yet, Frankl (1991) went a bit further and argued that not only have Zanzibar speakers lost the aspiration distinction, but another feature (alveolar vs. dental stops) has also disappeared from the island. Based on this, it is not a surprise that we find assertions in the literature that the loss of Swahili features is reflected in the orthography (Beck, 1960; Polomé, 1994; Engstrand & Lodhi, 1985; Harries, 1953; Knappert, 1971; Raia, 2014; Wójtowicz, 2016) So, why did this confusion arise in the language, threatening the representation of important linguistic features which were encoded by the orthographic system? This is a situation that needs to be explained.

23 This refers to the laryngeal contrast in voiceless obstruents. According to Polomé (1967:37-39), Swahili speakers would know the minimal set category if they distinguish between minimal pairs based on the presence or absence of aspiration, so that /tando/ ‘fungus’ is different from /tʰando/ ‘swarm’.
The alphabet of Swahili went through two stages; in both, foreign (Arabic and Roman, respectively) characters were adopted to record the language’s literature and legacy. In the early literature of Swahili, as mentioned in Topan (2006), the Arabic Swahili scripts were influenced by the relationship that the people of East Africa (the coast and the islands, in particular) had with Arabia and the religion of Islam. In early 1970, Jan Knappert (an expert Western linguist working on Swahili) confirmed this influence: that the literature of Swahili was and is very Islamic, even during and after the standardization of the language into Roman characters. The strong connection between Islam and the people of East Africa, which led speakers to embrace Arabic, is also confirmed by Marshall (2015), who states that the Arab traveler Ibn Battuta observed in 1331 that the majority of Swahili people are Muslims. For this reason, as stated in Harries (1953), Arabic was established for Swahili die-stamping as early as the thirteenth century, while known and documented Swahili-language work that uses an Arabic writing system appeared as late as the mid-eighteenth century.24

Given that long history of contact between Arabia and the east coast of Africa, it is no surprise that, of the 3,006 loanwords included in Johnson’s dictionary (1939), 2,534 came from Arabic (Lodhi, 2000: 8).25 Lodhi then adds that Swahili-language work in Arabic scripts goes back to 165226 and that the Hamziyah poem by Sheikh Aidarus bin Uthaima about the life of the prophet Mohammad is the oldest surviving piece of Swahili literature. Yet Arabic does not have the three-way laryngeal contrast which Swahili purportedly has, and no convention in the big donor alphabet can account for this distinction. The old Swahili in Arabic scripts, as Polomé (1994) put simply, lacked an aspiration distinction due to the

24 Citing Eastman (1983:8), Mdee (1999) also refers to the thirteenth century as the beginning of the time in which Swahili was written in Arabic scripts. Marshall (2015) confirms that the language is believed to have been found in a written form before the seventeenth century, but nothing seems to have survived.
25 Citing Nurse and Hinnebusch (1993: 321), Lodhi (2000: 97) writes that, “in the last three centuries” (presumably from the early eighteenth to the late twentieth century), most of the Arabic borrowings in Swahili came from Omani Arabic.
26 In this same year, the Omani fleet invaded parts of East Africa (Lodhi 2000: 62).
absence of adequate diacritics. Furthermore, as described in Harries (1953), the Arabic-style Swahili writing system is not enough for a Swahili person who is interested in knowing bits and pieces of the language’s features. For Harries, the reasons for this are that the nasal-stop sequence was not ideally described, glides were excluded, and one symbol in the alphabet could represent two phonemes. Phonologically speaking (Omar & Frankl, 1997), the alphabets of Swahili do not have a common convention to cover the wide variety (and vast feature set) of spoken Swahili: neither when the Arabic characters were used nor after the transition to the Western alphabet.

Whether this claim is ultimately supported or refuted, we need to have a better understanding of what the writing system of Swahili was like before the introduction of the Roman characters. As I found out in my readings, more than one linguist adopted diacritics of Arabic to represent features of Swahili, and one of those who were vigilant about Swahili aspiration was Taylor27. Harries (1953: 22) argues that Taylor’s scripts offer evidence that the characters of Arabic, when modified or revised, can in fact account for the phonemic contrast of aspiration: namely, Taylor took advantage of the Arabic geminate and dot diacritics to accommodate Swahili’s laryngeal contrasts. Aspirated obstruents (which Harries believed were followed by an apostrophe in the Roman system: /p’/, /t’/, /k’/, ch’/28) were represented using Arabic characters as in Table 3. The /p’/ character is a modified version of the Arabic /b/, which comes with one dot only. Placing the geminate diacritic, the shaddah symbol (•), under the letter is not authentic Arabic style because the shaddah always appears on top of every letter that represents a long consonant.

27 Taylor (1856-1927) was a British national and Swahili scholar based in Mombasa; see Frankl (1999).
28 Polomé (1975: 4) went a bit further and claimed that the Roman orthographic symbol <h> was assigned to bhara ‘piece’ because aspiration is reported in some voiced stops as well.
Table 2-3 Taylor’s proposal for representing aspirated obstruents using Arabic characters

\[
\begin{array}{|c|c|c|c|}
\hline
\text{/p'/} & \text{ه} & \text{/t'/} & \text{تّ} \\
\hline
\text{/k'/} & \text{كّ} & \text{/ch'/} & \text{خّ} \\
\hline
\end{array}
\]

The three dots under /p/ are pretty much an inspiration from Arabic and Persian scripts, as Omar and Frankl (1997) point out. In their article, they present some Spellings proposed by Taylor’s aide (Mwalimu Sikujua), who introduced new methods that were unheard of in Arabic, such as adding four dots above the Arabic /t/ and /d/. Sikujua, moreover, was able to capture the aspiration distinction by placing the symbol for /h/ (هـ) on top of the /t/’s four dots (تّ). The possibility of reforming the Arabic-style Swahili writing system was raised by some Swahili scholars, such as Sheikh Al-Amin Al-Mazrūʿī, the judge of Mombasa, in his 1940 letter to another Swahili scholar, J.W.T. Allen; however, the letter did not find open minds (Omar & Frankl, 1997: 59). Knappert (1971: 26) was in line with Muallim Yahya Ali Omar, who embraced the modified Arabic orthography and argued, for example, in favor of writing paa as (پها) for the aspirated meaning ‘gazelle’ and as (پا) for the unaspirated meaning ‘roof’, which should have been unified and spread across early writers. The proposal of Omar and Frankl (1997: 61) offered many ideas, most of them adapted from early linguists. Their proposed alphabet includes four different symbols for /t/, in which the influence of Urdu is very evident: /t/ (ت), /th/ (تّ), /dh/ (تّ), and /dh/ (تّ). However, Raia (2014) states that before the alphabet was Romanized, writing Swahili using Arabic graphemes was not a common convention, and on many occasions, writers had to

\[ \text{29} \these Arabic characters are as represented in Harries (1953), in my handwriting. \]

\[ \text{30} \in addition to having knowledge of Arabic, Al-Mazrūʿī knew the writing systems of Persian and Urdu. \]

\[ \text{31} \the first author of Omar & Frankl (1997). \]

\[ \text{32} \text{This was a 20th-century effort in favor of reform of Arabic characters.} \]
come up with their own visionary spelling for documenting and introducing Swahili to either Swahili people or the outside world.

With this chaos regarding orthographic misrepresentation of aspirated phonemes, the general pattern was basically to ignore aspiration (Harries, 1953), which led to questions about what the correct production of such sounds actually was. This confusion led Harries to quote from *Swahili Tales* by Steere (1870), “[i]t is absolutely necessary to have a good idea of what you are to read before you can read it all” (p.26). Harries (1953: 35) then states, “It is evident, therefore, that for a proper interpretation of Swahili texts in Arabic scripts, speech reference is essential”. This is exactly what should be reported whenever a claim is raised regarding the presence or loss of aspiration. Those who have interest in the language’s representation need to know the specific dialect or location where it is spoken when speech reference is mentioned.

The eventual transition of Swahili writing from Arabic to Roman style started even when Arabs were very economically and socially influential in East Africa. According to Wójtowicz (2014) and Omar and Frankl (1997), the documentation of Swahili in Roman alphabets was introduced by Dr. Ludwig Krapf around the mid-1800s, with more terms entering the language which were not available in the old written documents. Marshall (2015) mentions that, with more colonizing power on the part of the British, parts of East Africa became a British protectorate in 1895, which led to more need for a unified orthography. It was then the idea of Rivers-Smith, the first British Director of Education in early 20th-century Zanzibar, to introduce Roman characters officially for better communication among the public, the ruling Arabs, and their protectors, the British (Marshall, 2015: 22-23).

---

33 According to Lodhi (2000: 51), a mass immigration of Arabs and Persians to East Africa occurred in the 1820s, during the rule of Said bin Sultan Al-Busaidi of Oman.
According to Mdee (1999: 127), the Romanized letters were made standard in three stages. First, in 1925, a model spelling was set up, which received the endorsement of a spelling convention in 1928. Later, in 1930, an Inter-Territorial Language Committee was created to ensure the implementation of the new, minimized/reduced writing system (see also Polomé, 1975). Johnson’s dictionary (1939) was then approved by the Committee (see also Wójtowicz, 2016). Marshall (2015) states that the Committee examined the Swahili of four East African dependencies (Kenya, Tanganyika (mainland Tanzania), Uganda, and Zanzibar) and chose the dialect of Zanzibar as the standard. After World War II (1945), East Africans were appointed to higher posts in the Committee for the first time, but Polomé (1975) reports that with the British in charge, the progress of Swahili slowed for some time. In later years, the British were again in favor of standardizing and supporting Swahili, and the Committee was first led by East Africans in 1969.

Mdee (1999: 126-131) also reports that, before the Zanzibar dialect was selected as the standard, there were ten different Swahili dictionaries written between 1870 and 1928. The <kh> spelling was used to represent the Arabic uvular fricative /χ/, which was mastered by those with knowledge of Arabic; later, <h> was used instead. According to Mdee (1999), those previous dictionaries disregarded aspiration and double consonants, and so did Johnson’s, for two reasons: to match the common usage and simplify the spellings. Furthermore, dictionaries compiled after Johnson’s (at least three of them) did not follow

---

34 Marshall (2015: 31-34) also reports that Oscar Watkins, one of Kenya’s three officials who were assigned to choose the standard during the era of colonization, was in support of choosing the Zanzibar KiUnguja dialect because it was the one used in courts and in many dictionaries and books (KiMvita was the next candidate). Later, Johnson (1939) endorsed this recommendation, and the Zanzibar dialect was chosen.

35 As mentioned by Marshall (2015: 6), Ali Mazrui (1998) claims that the European colonizers, at some stage, believed that the African languages were inferior to their own.

36 The <kh> spelling is used in Johnson’s dictionary (1939: 184) to represent /χ/, which he claimed to be used by those who were influenced by Arabic. However, Johnson’s preferred method was to drop the <k> and preserve the <h>, in order to match the standard orthography as adopted by the Committee of which Johnson himself was the secretary.

37 The three dictionaries are: Rechenbach (1968), TUKI (1981) and Feeley (1990), see Mdee (1999: 129-130).
the Committee’s decision and conventions literally: although geminate spellings were widely omitted, variants of the same word still existed in later lexicographic work. Since literature is read in schools by embracing the local varieties, many writers committed to using their dialects in writing; and so, the conventions were taken lightly even by lexicographers. Johnson’s Swahili-English dictionary (1939) was different from his Swahili-Swahili dictionary (1935) due to the latter’s many additional variants.

For Lodhi (2000: 20), the fact is that Swahili dictionaries have left out some important linguistic facts and “have serious deficiencies”. Furthermore, Swahili as spoken by native speakers did not always match proper prescriptive use, and the standard dictionaries which were supposed to present “pure” Swahili failed to preserve a portion of Swahili linguistics. Along these lines, Wójtowicz (2016) mentions that the standardization omitted the double letters for geminates (see Harries, 1953: 35); and (Knappert, 1971: 29; uma ‘fork’ is different than umma ‘community’), and for many variants, only one form was adopted. Moreover, Knappert (1971), speaking about the effect of orthography on appropriate poetic representation, listed nine shortcomings, of which some are mentioned here. First, aspiration was ignored. Second, there was no distinction between dentals and alveolars. Third, there was no distinction between emphatics and alveolars. Fourth, there was no distinction between short and long vowels. Fifth, <ch> was a spelling for two phonemes. Sixth, the conventions for vowels violated the prosody of poetry. Seventh, the addition of <th> was not necessary.

As described by Beck (1960: 53), missionaries were required to study certain aspects of their language of interest. A number of Germans who went to work in Africa attended courses at Hamburg University and specialized in African languages, and upon their arrival in Africa, they had to keep working on the lexicographical aspect of their mission in

38 This Swahili-Swahili dictionary is written in Swahili for speakers of Swahili and Swahili learners.
consultation with their mentors back home. It is worth mentioning, however, that scholars whose work was not purely lexicographical were often more precise than lexicographers. As discussed in Frankl (1991), Taylor, who was a British missionary, noted that aspiration not only was a core aspect of the language as spoken in Mombasa (KiMvita), but also had great importance for the academic community. Frankl (1991) adds that Taylor was first introduced to aspiration by Edward Steere, who was based in Zanzibar, and to Steere, Frankl believes, belongs the first documented source about aspiration, which dates back to 1870. So in my opinion, it is not fully fair to attribute the loss of aspiration in orthography to all Western experts on Swahili39. However, the effort to identify those who were responsible for excluding aspiration from Swahili orthography as we know it today will lead to another long story.

As Frankl (1991: 366) notes, one of Taylor’s accounts in the late 1800s, which was supported in later years by Harries (1953), was that the plosives with apostrophes are spoken with a stronger puff of air and are different from their plain counterparts. On the one hand, speculation is ongoing among linguists to try to specify where in the language the contrast originates: while Polomé (1967) links aspiration to “stress”, Taylor claims that the relation between a word and aspiration is “a philological n”, which was lost before voiceless stops (Frankl, 1991: 368). On the other hand, Polomé (1994) attributes the marginalization of aspiration to the number of minimal pairs that show the contrast; these are very few and almost disappearing in big cities40. Moreover, one of Frankl (1991: 369-372) reasons for the loss of the contrast is that aspiration is not a property of minimal pairs, but can be found in singletons, as well as loanwords from Arabic and English (k’urasa ‘pages’ and p’ini ‘pin’,

39 Other Westerners who referenced aspiration, as cited in Frankl (1991: 367-368), were Meinhof (1899), Burt (1910), and Damman (1937).

40 In contrast to Polomé’s (1994) account, Frankl (1991: 369) who referenced a work from 1910, wrote: “there are some four dozen examples in Mrs BURT’s Swahili Grammar, pp. 149-151”.

46
respectively). It does not matter, however, where in the lexicon the laryngeal contrast is happening, as long as “aspiration is a predominant pronunciation feature in a majority of the native Swahili population” (Engstrand & Lodhi, 1985: 180). For this reason, these two scholars restrict the absence of a contrast to non-native speakers.

Engstrand and Lodhi (1985) also add that following the written literature of scholars is not always a promising method for tracing language-specific features. They cite a poet, Shaaban Robert of Tanga, as an example of a writer who tended to reduce his work to meet the standardized policy that had been imposed on the Swahili people by foreigners. Those who worked on planning a unified writing system for the Swahili language overlooked the ideas of Taylor, who had managed to show how aspiration can be represented in Arabic (Harries, 1953) and/or Roman (Frankl, 1999)41 alphabets42. Consequently, this underestimation of aspiration, and the decision to ignore it, were based on the ruling that had been issued in the 1930s by the Swahili Inter-Territorial Committee (Polomé, 1994). What is more, as maintained by Mdee (1999), is that, due to the Europeans’ high level of dependence on their mother tongues, the spelling of Swahili was subjected to the judgment of non-native speakers, which resulted in a mismatch in the orthography of Swahili words. This fact leads Mdee (1999: 132) to write that the “[s]tandardization of a language is an arbitrary decision”; so, the reasons why one dialect was made the standard, why the spelling had to be simplified, and why only one form was accepted even if many are used were arbitrary as well.

The major spoken Swahili dialects are KiUnguja in Zanzibar, KiMvita in Mombasa, KiAmu in Lamu, and KiNgwana in the eastern part of the DRC (Marshall, 2015)43. Harries

---

41 Taylor’s aspirated Roman characters included /pʰ/, /tʰ/, /kʰ/, and /chʰ/; see Frankl (1999: 172).
42 The first known published work that combined Arabic and Roman scripts was by Taylor, in 1893. Based on the Swahili of Mombasa, it featured Arabic characters on the left and Romanized ones on the right. For the Roman characters, he added symbols to differentiate between aspirated and unaspirated plosives (Omar & Frankl, 1997: 58).
43 Lodhi (2000: 33, citing Nurse and Hinnebusch, 1993) mentioned fourteen main dialects for Swahili. For Marshall (2015), however, the four are the leading ones.
(1953) states that *KiAmu* is the source of traditional Swahili linguistic heritage, and writers as far south as Zanzibar would use this dialect to introduce their work. Nonetheless, the ten dictionaries that were written between 1870 and 1928 were based on the dialect of Zanzibar (Mdee, 1999).\(^4^4\) No reference has been made to *KiAmu*, as far as I am aware, as being “pure” or holding any special status with respect to reflecting the language’s linguistic features, nor has any source used the dictionaries to show the aspiration contrast\(^4^5\). One reason for this could be the incorrect judgments of these lexicographers when it comes to the laryngeal distinction among the obstruent phonemes; native speakers should have been consulted during the preparation of these dictionaries. Omar and Frankl (1997: 60-61) state that none of the alphabets were able to bring the dialects of Swahili together, which necessitates reforming the writing system to reflect phonemic variations in the written form. They further assert that “there can be absolutely no correct speaking of the Swahili language as spoken by the Swahili people, and no scholarly writing of the language, unless a sensible distinction in the sound is indicated” (p. 61). Thus far, we have been discussing this debate from the perspective of orthography, but it extended even to the structure of the lexicon.

There is no doubt, as mentioned in Broomfield (1931), that what is called “Swahili” is of an authentic Bantu origin. The language’s name and lexicon could have been different before the overwhelming contact with speakers of Arabic, but Arabic loanwords still constitute a great portion of the language. Therefore, the question is: are borrowings from Arabic still important in the language? As Pastor Roehl believes, Arabic words are known more along the coast than on the mainland. In a response to Roehl’s New Testament in Swahili, which avoided as many Arabic words as possible, Broomfield (1931: 77-82) is not

\(^{44}\) Lodhi (2000: 38) also posited that the prose of the last 140 years may be based on the dialect of Zanzibar because of the Western works that were introduced into Swahili, and into education in particular.

\(^{45}\) Of the fourteen dictionaries surveyed by Wójtowicz (2016: 65), only one cited an interest in showing the pronunciation; this one was written in Polish and published in 1966.
in favor of this approach for many reasons. First, Roehl could not get rid of all the Arabic words and ended up using many of them. Second, it should be noted that, without Arabic, Swahili as a language could have been in a slower progress. Third, it should be stressed that the native Swahili people are those living along the African east coast with Arab-African blood, and the contact between the donor and borrower languages was enforced by the necessities of everyday life. Fourth, although Arabic words featured different pronunciations, they were absorbed into Swahili (making Swahili a Bantu language with an Arabic element). Fifth, the exclusion of semantically sufficient Arabic words sometimes left Roehl with no native Bantu ones that had the same linguistic power. An example that Broomfield cites for that last point is *injili* ‘Gospel’, which was eliminated for the sake of a native compound: *Utume Mwema* ‘good sending’ (p. 84). So that arbitrariness of choice/standardization has influenced the language on different linguistic levels, from ignoring the little fine details to the underestimation of some powerful lexical items.

To sum up, it has been seen that Arabic had a great impact on Swahili through both trade and religion; that is why the old system of Swahili orthography used Arabic characters. The abundant words of Arabic origin in the language are another factor that made it hard for missionaries and colonizers to isolate the Swahili people from their heritage. As far as the spelling is concerned—whether in Arabic or Roman alphabets—if we can consider the phonetic details a portion of the Swahili heritage, then the non-native decision-makers from around a century ago should have had no right to call for simplifying the orthography or to impose a forced unity among many spoken dialects. Orthographic unity may be less essential now that the era of colonization is over. Every dialect, from *KiUnguja* to *KiMvita*, has its

---

46 Because of the long contact between the east coast Swahili people and speakers of Arabic, some Arabic phonemes were borrowed too (Polomé 1967: 38).

47 Marshall (2015: 15) stated that, with Swahili being a lingua franca in the East African region, Swahili activists relied on their own language to promote anti-colonization movements and agendas that sought to eliminate any need to depend on the colonizers’ language and/or culture.
unique features that are worth studying and preserving. Finally, the current paper will use IPA symbols and portray the feature of aspiration where it is called for.

2.9 Background – Introduction to the current study

The current study is focused on the Swahili language; this language was chosen because of the ways in which VOT and pitch are believed to apply when contrasting its aspirated and unaspirated obstruents. Our concern is centered on the issue of overlapping phonetic VOT/pitch implementations, as observed in the productions of two Swahili speakers from several decades ago\(^48\). Preserved recordings from 1973 (in the UCLA Phonetics Lab Archive) showed a trade relation between VOT (Speaker 1) and F0 (Speaker 2) during the pronunciation of the word *tembo*, which means ‘elephant’ when aspirated but ‘alcoholic drink’ when not aspirated. An example of this contrast is given in the figures (Figure 2-4 and Figure 2-5) below.

![Figure 2-4 The VOTs and F0s for t/tʰ of Speaker 1 (from Mombasa)](image)

\(^{48}\) These are the two Swahili informants: Speaker 1 [swh_word-list_1973_02/03/04] and Speaker 2 [swh_word-list_1973_05], whose recordings are found in the UCLA Phonetics Archive: http://archive.phonetics.ucla.edu/Language/SWH/swh.html
To echo the earlier research questions once more, I stress that the current study is related to the body of work in linguistics with regard to Swahili’s sound inventory system as having a three-way laryngeal contrast. Do Swahili speakers have a contrast in VOT values only? Or do speakers also demonstrate a sort of pitch (tone) phenomenon to differentiate laryngeals? Is this contrast common to all Swahili speakers, or do we need to isolate those who consider it a mother tongue from other groups? And does exposure to a second language influence the listener-turned-speaker’s imitations?

This study will look into the cut-off point between aspirated and unaspirated utterances during the production of real words (Experiment 1), or when native speakers listen to and imitate synthetic speech of prototypes as employed at different VOT/pitch values (Experiment 2). The production task is mainly a straightforward process, in which participants’ job is to look at a picture and then pronounce the relevant word. For imitation, on the other hand, participants have to listen and then produce the target tokens. Three levels of manipulation are implemented to see if speakers, whether native or not, would create three distinct categories. The three levels of height and length are also intended to see if speakers of the language resort to dividing the three levels of height/length into aspirated, neutral, and unaspirated variants, as well as to find out which of the three VOT/F0 modified tokens is imitated most accurately. The focus of both the production experiment and the imitation experiment is voiceless obstruents in word-initial position. The study will investigate the produced laryngeal contrast, as well as the question of whether phonetic details of laryngeals...
are automatically shadowed or if the imitation is dialect-specific and cannot be extended to other dialects (due to some sociophonetic factors).

If the phonemic categorization is accurate based on the three-way contrast that was claimed by early linguists, the expectation for production is that two separate VOT lengths and two unequal F0 heights will be created for the voiceless obstruents. That means the clouds of the unaspirated voiceless consonants /p/, /t/, /k/, and /tʃ/ are theorized to start at 25, 35, 45, and 55 ms, respectively, with a 10-ms interval in between. Then, the levels for the aspirated voiceless obstruents would begin for the bilabial at 75 ms and increase by 10 as you move to alveolar, velar, and post-alveolar. Yet, the final result would determine the accurate categorization of these phonemes: specifically, if they generate two separate clouds per phoneme (suggesting that the contrast is retained), or if the two assumed phonemes (for each place of articulation) are actually a product of one phonetic representation and collapsed under only one cloud (so that the contrast is lost).

If aspiration is part of the language system, we hypothesize that our participants will comply with the phonetic experience they have obtained from their spoken varieties of Swahili. Sensitivity to one or more cues would result in an accurate production performance for real words and an approximate level of imitation for nonce words (in Experiment 2). This level of imitation doesn’t need to match the prototype that the original stimuli were based on; however, the listener-turned-speaker is expected to create three distinct phonetic categories for the three levels.

This means that the values for the imitated obstruents would shift, as a result of the level of sensitivity, to the left or right, which creates a unique prototype cloud for each level. Each dot (in red or blue, in Chart 2-1) forms a cloud representing the token’s level. If sensitivity towards the levels is low, then the three hypothesized categorizations are going to
be collapsed and merged into one or two categories; whether the three levels are different or similar will depend on the participant’s perceptual capacity and strength when approaching the task.

The study will rely on statistical analysis to judge whether these three levels are statistically different or fail to reach the level of significance. The typical prototype for each sound’s six phonetic details in perception will closely resemble the plotted chart for [t] (Chart 2-1 below). Other obstruents are expected to have some categories of this kind, with similar F0s but different VOTs. We should add here that the chart’s F0 heights, at 200 Hz or higher, are due to our having a female model talker. Males in our experiment, on the other hand, are expected to have F0 values below 200 Hz (a property that is known crosslinguistically). See Table 4-1 in the “4.3 Design” section for more on the length and height measurements for [p], [k], and [tʃ] that we expect to observe in the imitation task.

![Prototype Categorization of [t]](chart2-1.png)

**Chart 2-1**  An example of prototypes for [t], along the y and x axes of F0 and VOT, respectively

2.9.1  **Background on Swahili/word choice:**

Prior to the current study, there has been no clear-cut account of the real status of aspiration in native Swahili words. Although it has been claimed that the language features a laryngeal contrast, some scholars (Frankl, 1991; Polomé, 1967) believe that the aspiration is
geographically dependent on the region where the speaker lives—declining in the south (as in Zanzibar) but becoming more robust in the north (as in Mombasa). In line with Frankl (1991) and Polomé (1967); Wald (2009) also maintains that the aspiration has almost been lost in the south. However, Engstrand and Lodhi (1985) posit a symmetrical relation in aspiration between the north and the south and suggest that it is a characteristic found among native (conservative) speakers of Swahili.

Furthermore, H. Hussein (personal communication, September 28, 2017)\(^49\) firmly believes that the aspiration is preserved, but only among those who are “native Kiswahili Speakers.” Hussein affirms that aspiration is contrastive in the language, but that the minimal pairs which show this laryngeal contrast are few in number. On the other hand, another linguist from Kenya, M. Bosire (personal communication, October 1, 2017)\(^50\) confirms that he belongs to the school of thought which contends that Swahili has lost contrastive properties of aspiration (and that aspiration is no more than a redundant feature in the language). He further says, “I neither hear the aspiration nor do I make it.”

The Swahili words used in this study (see Appendix A-7) were collected from three dictionaries (Johnson, 1939; Awde, 2011; Kirkeby, 2000) and then compared and contrasted with the entries for the same words in two online dictionaries\(^51\). For the purpose of the analysis, it was crucial to focus on commonly-used words, and in order to decide on any word, I first had to look for it in Awde’s dictionary. Second, I made sure that the meaning is

\(^{49}\) Dr. Hussein Ali Hussein is one of Dr. Lodhi’s former students and did his PhD dissertation (on Arabic loanwords in Swahili) under Dr. Lodhi’s supervision at the State University of Zanzibar (SUZA).

\(^{50}\) Mokaya Bosire’s dissertation at the University at Albany, State University of New York, was about “Sheng,” a vernacular creole (from Swahili and English) that is spoken in and around Nairobi, Kenya. He is introduced in some talks as a native speaker of Swahili. Another scholar, Iribe Mwangi, a Swahili linguist at the University of Nairobi, affirms (personal communication, November 23, 2017) that, while Nairobi has mostly second-language speakers, Mombasa has a large population of native Swahili speakers. Mwangi, in line with Bosire, doesn’t support the theory that aspiration is still active in Swahili.

consistent in the two online dictionaries (or at least that the word exists in one of them).

Third, I consulted Kirkeby’s dictionary by looking up the English meaning (because that is the way this dictionary works) and making sure that the chosen Swahili word is one of the synonyms (at least one of the first three entries) that represent the English word. Words that are not found in Kirkeby’s dictionary are written in red; however, they still exist in at least two of the other dictionaries. After I decided on the words that would be included for the analysis, two Swahili speakers (one from Kenya and one from Tanzania) were recruited to speak the words from the pictorial images. This helped me to change, fix, and modify some of the pictures, as well as the corresponding clues. Later, a third Swahili speaker was consulted to double-check the spellings and the comprehensibility of the clues that had been prepared for the final PowerPoint stage presentation.

2.9.2 Distribution of the word list

The word list contains a total of 147 words in four categories (see Table 2-4; the target obstruents are in bold). The first category is a collection of 31 words that are meant to show the “laryngeal contrast” in the Swahili language. This category was gathered to address the issues raised by the UCLA recordings and some examples from Polomé (1967) and Engstrand & Lodhi (1985). Hussein (mentioned earlier) generously provided a short list of aspirated native words, as well as some minimal pairs, which I also used in this category. All words in this category are native Swahili words.

Table 2-4 Assorted examples of words/categories\(^52\)

<table>
<thead>
<tr>
<th>Category</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Category</strong></td>
<td><strong>Target</strong></td>
</tr>
<tr>
<td>(Swahili words)</td>
<td><strong>paa</strong> /pa:/ ‘roof’</td>
</tr>
<tr>
<td></td>
<td><strong>tembo</strong> /tembo/ ‘alcohol’</td>
</tr>
<tr>
<td></td>
<td><strong>kaa</strong> /ka:/ ‘charcoal’</td>
</tr>
<tr>
<td></td>
<td><strong>chungu</strong> /tʃuŋgu/ ‘clay pot’</td>
</tr>
<tr>
<td></td>
<td><strong>paa</strong> /pʰa:/ ‘gazelle’</td>
</tr>
<tr>
<td></td>
<td><strong>tembo</strong> /tʰembo/ ‘elephant’</td>
</tr>
<tr>
<td></td>
<td><strong>kaa</strong> /kʰa:/ ‘land crab’</td>
</tr>
<tr>
<td></td>
<td><strong>chungu</strong> /tʃʰuŋgu/ ‘black ant’</td>
</tr>
</tbody>
</table>

\(^52\) For the full lists of target and filler words, see Appendices A.7 and E.
The second category consists of mostly English loanwords. There are 32 borrowings from English, along with 16 native Swahili words that start with the voiced phonemes (/b/, /d/, /ɡ/, and /dʒ/; 4 each) and 4 more native examples for the voiceless /tʃ/. The native Swahili words were added because they contain sounds that are completely absent from the “laryngeal contrast” group. The voiced stops are very much needed to compare and contrast the English and Arabic loanwords with one another and with the native Swahili words. They are also useful for helping us recognize the difference between word-initial voiced stops and post-nasal voiced stops.53

The third category is a modified version of the word list from my first project, which exclusively looked into loanwords from Arabic. I had to eliminate some words from the old version (Appendix A-6) because it was near-impossible to match abstract nouns with pictures. Then, some of the other words in the old list were found to be less common, so these were excluded as well. Additionally, as Arabic alone (which misses the voiceless obstruents

<table>
<thead>
<tr>
<th>Second Category (English loanwords)</th>
<th>Filler</th>
<th>bundi /bundi/ ‘owl’</th>
<th>duma /duma/ ‘cheetah’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>gogo /ɡogo/ ‘log’</td>
<td>jengo /dʒengo/ ‘building’</td>
</tr>
<tr>
<td>Target</td>
<td>picha /piʃa/ ‘picture’</td>
<td>tai /tai/ ‘tie’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>keki /keki/ ‘cake’</td>
<td>chaki /tʃaki/ ‘chalk’</td>
<td></td>
</tr>
<tr>
<td>Filler</td>
<td>betri /betri/ ‘battery’</td>
<td>dereva /dereva/ ‘driver’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gita /ɡita/ ‘guitar’</td>
<td>jela /dʒela/ ‘jail’</td>
<td></td>
</tr>
<tr>
<td>Third Category (Arabic loanwords)</td>
<td>Target</td>
<td>tarehe from /tārī/ ‘date’</td>
<td>tabibu from /tˤabīb/ ‘doctor’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kitabu from /kitāb/ ‘book’</td>
<td>kafili from /qufīl/ ‘lock’</td>
</tr>
<tr>
<td>Filler</td>
<td>bahari from /bahr/ ‘sea’</td>
<td>duka from /dukkān/ ‘store’</td>
<td></td>
</tr>
<tr>
<td>Fourth Category (Swahili post-nasals)</td>
<td>Filler</td>
<td>mbuzi /mbuzi/ ‘goat’</td>
<td>mdalasini /mdalasini/ ‘cinnamon’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mgenu /mgeni/ ‘guest’</td>
<td>ndege /ndege/ ‘airplane/bird’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ngoma /ŋoma/ ‘drum’</td>
<td>mpiro /mpira/ ‘ball/football’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mão /mto/ ‘river’</td>
<td>mkanda /mkanda/ ‘belt’</td>
</tr>
</tbody>
</table>

53 The data for voiced stops and post-nasal stops were collected and annotated, but ultimately excluded from the analysis for several reasons (such as a narrowing of the study’s scope and some phonetic variations that were observed in the voiceless stops).
/p/ and /tʃ/ in its inventory system) cannot provide the opportunity to study sounds in a fully stop-to-stop environment, English loanwords could bridge this gap in the previous analysis.

Finally, the *fourth* category contains post-nasal voiced and voiceless stops (25 and 14 words, respectively). As can be seen from this list, more examples are available in the language for nasal-voiced stop sequences than for nasal-voiceless stop sequences. It is not clear why Swahili has more voiced stops than voiceless stops after nasal sounds. One plausible historical reason (see footnote 55) could be that aspiration has been introduced to compensate for the continuous loss of the initial nasal in nasal-voiceless stop combinations; see the next section below for more details on this hypothesis.

2.9.3 *Background – Sounds under investigation*

The stops are the main consonant group for studying the allophonic/phonemic contrast in Swahili loanword adaptation. However, Kharusi (1994: 59), who adopted Polomé’s (1967) consonant inventory of Swahili, added that aspiration in Swahili is not specific to stops. The post-alveolar affricate /tʃ/ is also contrastive and comes in minimal pairs such as *chungu* /tʃungu/ ‘earthen cooking pot’ and *chungu* /tʃhunugu/ ‘black ant’ (see Table 2-5).

In addition, it has been claimed historically (Wald, 2009; Contini-Morava, 1997) that aspiration in Swahili is a product of the loss of nasality before a voiceless stop. Thus, to aid in understanding the patterns of laryngeality, post-nasal voiceless stops (/mp/, /mt/, /mk/, and /nt/) were included in the word list. Unfortunately, due to the priority that was ultimately

---

54 One possible synchronic reason is that it’s articulatorily easier not to switch voicing in the middle of a consonant pair (from Dr. Winters’ comments on the first draft of this paper). According to Pater (2001), the avoidance of voiceless obstruent following a nasal is a crosslinguistically common restriction: *NC.

55 The nasality has not yet been lost in our data, but it was reported in Wald (2009: 888) that the words in the Swahili minimal pair *kaa* ‘land crab’ and *kʰaa* ‘charcoal’ are etymologically different because the latter goes back to Proto-Bantu *nkədd* and the loss of nasality in *nk*- words initially yielded kʰ. A similar transformation was claimed by Contini-Morava (1997).
given to obstruents in initial position, the analysis of post-nasal stops had to be saved for a future work.

Table 2-5 Selected Swahili obstruents (target sounds are bolded)\textsuperscript{56}

<table>
<thead>
<tr>
<th></th>
<th>voiceless</th>
<th>/p/, (/p\textsuperscript{b}/)</th>
<th>/t/, (/t\textsuperscript{b}/)</th>
<th>/k/, (/k\textsuperscript{b}/)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosives</td>
<td>voiced</td>
<td>/b/, (/b\textsuperscript{b}/)</td>
<td>/d/, (/d\textsuperscript{b}/)</td>
<td>/g/, (/g\textsuperscript{b}/)</td>
</tr>
<tr>
<td></td>
<td>voiced</td>
<td>/\textsuperscript{t}/</td>
<td>/\textsuperscript{t}/</td>
<td></td>
</tr>
<tr>
<td>Affricates</td>
<td>voiced</td>
<td>/\textsuperscript{d}/</td>
<td>/\textsuperscript{d}/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>voiced</td>
<td>/\textsuperscript{z}/</td>
<td>/\textsuperscript{z}/</td>
<td></td>
</tr>
<tr>
<td>Post-nasal plosives</td>
<td>voiced</td>
<td>/mp/</td>
<td>/mt/</td>
<td>/mk/</td>
</tr>
<tr>
<td></td>
<td>voiced</td>
<td>/mb/</td>
<td>/md/</td>
<td>/mg/</td>
</tr>
</tbody>
</table>

2.9.4 Locations of interest

Swahili is a Bantu language spoken in East Africa (Spear, 2000), whose domain extends from southern Somalia to northern Mozambique and from eastern Congo (Kinshasa) to the many scattered islands east of Tanzania and Kenya. It is a medium of communication for over 82 million people, but a first language for only 16 million (Ethnologue, 2019). The locations (in Map 2-1, labeled in red) were chosen to test, examine, and comment on the everyday dialects of Swahili as spoken by a younger generation of Swahili people in East Africa. The locations range from an island with bodies of water that restrict the residents’ movement (Zanzibar) to a coastal location where people have much more contact with the inland population but less interaction with the surrounding islands (Mombasa), followed by two mainland locations where people pretty much have access to other mainland populations from any direction but would have even less interaction with people on the islands (Nairobi and Iringa).

\textsuperscript{56} This is a phonemic summary of the prescribed stops and affricates in the literature. Disputed phonemes are shown in parentheses. For more information about early studies, see section 2.6 and the first two pages of Chapter 5.
2.9.5 *Participants*

Data collection took place in four different locations at which native Kenyans or Tanzanians, respectively, were recruited and recorded. These locations were Sumait University (on the island of Zanzibar), the University of Iringa, the campus of Pwani University in Mombasa, and the University of Nairobi. We selected these locations in order to obtain fair and balanced data and investigate whether there are any variations in the Swahili dialects from the island to the coast and then going deep inland. When recruiting participants for the data collection, we considered the need for equal gender ratios to test for any difference based on sex. Since it was nearly impossible to locate a sound-attenuated booth, every possible effort was made to assure minimal disturbances from background noise. Recording sessions had, on many occasions, to pause when interrupted by a moving vehicle, an ambulance siren, or a bird in a nearby tree. With all these difficulties well in hand, the data were gathered according to plan, and the distribution of participants was as follows:

---

57 A modified map adapted from jw.org under “God’s Name Made Known in Swahili”.

59
2.9.5.1 **LI speakers (Zanzibar and Mombasa)**

a. Zanzibar (see Appendices A-1 and A-2) has 24 subjects: 12 males and 12 females, with an average age of 22. The Zanzibar subjects speak Swahili as their first language. Considered to be one of the areas where the Swahili language was born, and featuring a “predominant” (Roehl, 1930) and “standard” (Wald, 2009) dialect, Zanzibar (consisting of Unguja and Pemba)\(^{58}\) provides participants who speak coastal Swahili. These speakers are believed to provide a sample of conservative linguistic data, as they live on the island with limited access to mainland Tanzania. Although it is open for tourism, Pemba Island—where most of our subjects come from—has a smaller population than Unguja (the main island) and is also poorer and underdeveloped, with fewer tourists visiting it. Unguja is the place where the history, the culture, and the old trading hub originated, long before the Zanzibar revolution took place in 1964. Furthermore, the importance of Zanzibar as a link between the Indian Ocean and the mainland has contributed to its status as the cradle of the Swahili language.

b. Mombasa (see Appendix A-3) has 26 subjects: 13 males and 13 females, with an average age of 28.5. In this study, Mombasa represents Kenyan coastal Swahili, which is spoken as a first language and regarded as the second main dialect of Swahili (Roehl, 1930). The city’s major dialect is *KiMvita*, and sub-dialects in the area include *Kingare* and *Chijomvu*. Another dialect of Kenyan coastal Swahili is *KiAmu*, which is spoken in and around the island of Lamu. There is one participant in this study who represents *KiAmu*. *Giriama* is another language that is spoken along the Kenyan coast; however, none of our Mombasa participants reported any knowledge of this language. All the

---

\(^{58}\)“The term Zanzibarite is usually taken to mean ‘an inhabitant of Zanzibar City,’ while other people on the islands, including Pemba, are called Zanzibaris” (Engstrand & Lodhi, 1985: 180).
participants from Mombasa speak Swahili as L1 and English as L2, and three women reported Digo (KiDigo) as a second language along with English.

2.9.5.2 **L2 speakers (Iringa)**

Iringa (see Appendix A-4) has 24 subjects: 12 males (but 11 in the imitation task) and 12 females, with an average age of 33. Hehe, which is the major language in Iringa, “shares a significantly great lexical similarity of 65% with Bena, 59% with Pangwa, 50% with Kinga and 48% with Vwanji” (Ethnologue, 2019). As opposed to coastal Swahili subjects, whose pronunciation is highly affected by the different Swahili dialects and sub-dialects, Iringa subjects use pronunciation that is almost 100% influenced by their first language.

2.9.5.2.1 **Bena obstruents**

There is no accessible source for Kihehe (Hehe)\(^{59}\) to shed light on the language’s obstruents; however, an alternative language with great similarity to Hehe is Bena. As reported in Morrison (2011), Bena follows the general two-way contrast pattern (table 2-6) in its inventory of plosives (/p/, /b/, /t/, /d/, /k/, /ɡ/); however, unlike Swahili, Bena has only one voiceless affricate (/ts/).

<table>
<thead>
<tr>
<th>Table 2-6 Obstruents of interest in Bena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosives</td>
</tr>
<tr>
<td>Affricate</td>
</tr>
</tbody>
</table>

2.9.5.3 **L2 speakers (Nairobi)**

Nairobi (see Appendix A-5) has 24 subjects: 12 males and 12 females, with an average age of 21.5. As is the case for Iringa in Tanzania, subjects from Nairobi (sampled from

\(^{59}\) There is a PhD dissertation titled “A comparative study of Ki-Hehe, Ki-Bena and Ki-Pangwa” by Mdemu, C (1995). Unfortunately, access to Mdemu’s work was not possible before the submission of the current work.
the University of Nairobi) represent a form of mainland Swahili. Just as in Iringa, Swahili in Nairobi is non-native and its pronunciation by speakers is highly affected by the predominant native language, which in this case is Kikuyu and its numerous dialects. Some participants who come from central Nairobi might also be affected in their pronunciation by Sheng60, a rapidly growing language among Nairobi city youth which is claimed to be a mix of Swahili and English (Bosire, 2008). Sheng was also spoken by some old-school speakers prior to its popularity and expansion (Bosire, 2008). It is strange that none of our speakers from Nairobi reported any knowledge of Sheng, but all of them mentioned in their consent forms that Kikuyu is their first language while Swahili and English are second languages. So, Sheng might be one extra language that has been accommodated by the participants’ knowledge of Swahili and English.

2.9.5.3.1 Kikuyu obstruents

Based on the works that describe Kikuyu, or “Gĩkũyũ” (Iribemwangi & Karũrũ, 2012; Englebretson & Wa-Ngatho, 2015), the language’s phonemic inventory is quite different from that of Swahili in many aspects. In reference to plosives and affricates, Englebretson & Wa-Ngatho (2015) state that there are no aspirated phonemes. In addition, the voiceless unaspirated obstruents lack the voiceless bilabial /p/ and the voiceless post-alveolar /tʃ/, and all voiced obstruents (/b/, /d/, /ɡ/, and /dʒ/) are in post-nasal position, represented as /mb/, /nd/, /nɡ/, and /ndʒ/. A bit different than what Englebretson & Wa-Ngatho (2015) posit, Iribemwangi & Karũrũ (2012) describe every voiced obstruent as being preceded by a unique nasal (/mb/, /nd/, /ŋɡ/), and according

---

60 See footnote 50.
to them, the post-nasal post-alveolar is actually a post-nasal palatal /ɲɟ/ (written below in parentheses). So, the two studies report the obstruents as in Table 2-7.

<table>
<thead>
<tr>
<th>Obstruents of interest in Kikuyu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosives</td>
</tr>
<tr>
<td>/mbl/</td>
</tr>
<tr>
<td>/t/, /d/</td>
</tr>
<tr>
<td>/k/, /ɡ/ (ʼɡ/)</td>
</tr>
<tr>
<td>Affricates</td>
</tr>
<tr>
<td>/dʒ/ (ʼʒ/)</td>
</tr>
</tbody>
</table>

2.10 Conclusion

To sum up, this chapter laid a theoretical background for understanding the phonemic variations in Swahili obstruents. The tools we used for this analysis are VOT and F0. We have revealed how these two phonetic cues were constructed and measured, and in some cases, how people tended to be sensitive to differences in them. Moreover, similar to the average F0 differences that are observed between men and women, women have typically had different VOT measurements than what is observed for men. Furthermore, different areas where the same language is spoken could show different sociophonetic patterns; for example, as we saw above in the discussion of Arabic loanwords in Swahili, speakers who have mastery of both languages might preserve the original Arabic pronunciation instead of using native Swahili phonology. On many occasions, this would lead to unequal prototype clouds based on the specific phonetic details that were under analysis. We have also seen that orthography is an important imperfect reflection of laryngeal features of a language. And, finally, the last subsections went over specific information regarding the data collection process and the recruitment of subjects.
Chapter 3: Production Experiment

3.1 Participants

For this study, I recruited 98 participants from different Swahili backgrounds (L1 and L2) and regions (one island, one coastal, and two inland), with a mean age of 26. In their history, none of the participants reported any history of speech, language or hearing impairments. (See section 2.8.5 for more information about the four locations where the participants were recruited.)

3.2 Materials

The stimuli included a total of 147 (66 target and 81 filler) words starting with obstruent consonants or post-nasal stops (voiced or voiceless). The native Swahili words in the obstruent category all began with <p>, <t>, <k>, or <ch>, with two minimal pairs and two non-minimal pairs included, which represented the potential contrast between aspiration and non-aspiration (/p/-/pʰ/, /t/-/tʰ/, /k/-/kʰ/, /tʃ/-/tʃʰ/). Other target words were borrowings from English, with four words per obstruent (/p/, /t/, /k/, and /tʃ/). However, since Arabic loanwords do not have the voiceless bilabial stop /p/ or the voiceless post-alveolar affricate /tʃ/, only words beginning with the voiceless alveolar stop /t/ or the voiceless velar stop /k/ were included (six and nine words, respectively). For more information on the target words, see Appendix A-7; for the filler words, see Appendix E.

3.3 Design

Pictures were found to represent the chosen words, and these were randomized and presented in a PowerPoint slide show, which each participant watched as they participated in the experiment. During the slideshow, each picture was paused for 7 seconds to allow the participant to speak the corresponding word twice. Whenever possible, a target word was followed by a filler word, and no target word was ever pronounced following another word.
from the same list: that is to say, a word from a minimal pair such as [paka]-[pʰaka] was not pronounced before or after any other word from the laryngeal contrast group. Also, words from the loanword groups were set apart from their group counterparts during the picture-naming task. The same randomized order of pictures was used for all participants.

3.4 Production procedure

For production, as mentioned above, data were elicited from participants via a picture-naming task that used a list of loanwords and native Swahili words (see Appendix A-7). For words with many synonyms and those which were assumed to be vague, clues were given. The recordings were made in Praat (Boersma & Weenink, 2018) using a microphone connected to a MacBook Air. In order to make sure that every informant was within his or her comfort zone, no more than one participant was allowed into a recording session at a time. The recording sessions took place in rooms or library offices on one of the university campuses that I visited for this purpose.

This elicitation task was intended to give some sense of the laryngeal distinctiveness of voiceless plosives and affricates\textsuperscript{61} so that we could see whether the contrasts and word origins—if observed in production—were discerned by all of the participants or dependent on factors such as a speaker’s L1 and/or regional location. This, in turn, would shed some light on the status of sociophonetic effects, particularly those of a speaker’s origin and proficiency level in Swahili.

3.5 Production coding and annotation methods

After I had the participants produce the whole word list by naming the pictures shown on the screen, every participant’s audio clip was assigned a specific filing code. In order to preserve the participants’ privacy and further protect their identities, the coding formula for each

\textsuperscript{61} Just as observed cross-linguistically in stops where closure is longer in aspiration than non-aspiration, the frication of aspirated affricates is produced with a longer duration.
experiment started with the first letter of the participant’s location (Z, M, N, or I), followed by the participant’s position in the recording session (1 to 12; this was done twice, once for men and once for women). Each gender group was then given a Gender identification letter (M for males and F for females). Since I conducted more than one experiment, the production experiment was assigned the capital letter P (e.g. ZS1MP), while the imitation experiment was identified by the uppercase letter I (e.g. ZS1MI). The production experiment worked on a combination of isolated words (partially analyzed here) and words in carrier sentences (not a part of this project), so that the last letter in the code for classifying the production data was either I (for words in isolation) or S (for words in sentences). That is, every participant had his/her own audio files for each portion of the different experiments.

To prepare for the annotation of the produced words, every audio file was opened up in Praat, and then a Matrix file was created; this Matrix file was needed for extracting the F0 at the vowel onset when the annotations were complete. The F0, indicated at the place where the vertical and horizontal dotted red lines intersect to mark the beginning of the vowel’s voicing, is calculated in Hz (see Figure 3-1). The annotating was then initiated by generating a Text Grid for the whole audio folder (all words), which was saved as a .wav extension for each informant. The Text Grid was defined to have three tiered layers: VOT, vowel, and word. In the VOT tier (Tier 1), the boundaries were drawn to reflect the closure of voicing by assigning the designation $b$ to the beginning of the closure and $v$ to the release of the closure.
Then, Tier 2 had boundaries that corresponded to the initial state of the vowel’s onset until the end of the voicing before the start of the next segment (which is the nasal /m/ in the provided example). The second tier was specified by the uppercase letter V, which stands for “vowel” and is meant to show the vowel’s duration when all the measurements are retrieved. Tier 3, on the other hand, presents the boundaries for the whole word, from the initiation of the utterance to the final specific segment. The boundaries of Tier 3 can give a sense of the whole word’s duration if this is needed for contrasting aspiration with non-aspiration (yet, the whole duration of the word was not a concern in the current study). This tier is the place where the word itself is written out, followed by “1” for the first repetition or “2” for the second repetition.

All the words in the audio files were annotated in this manner, except for the fact that the first tier would have v preceding b if the first segment was voiced; this revealed the duration of voicing (not the closure). The change in ordering between b and v was crucial so that the Praat script could differentiate the voiced obstruents from the voiceless ones once the annotations were complete and ready for the numbers to be extracted and opened up in a Microsoft Excel spreadsheet. The boundaries were inserted when we heard the word, figured out the VOT dimension (Tier 1) where the beginning of the voicing or closure predicts the start of the word, decided on the length of the following vowel’s voicing (Tier 2), and then

Figure 3-1 The second VOT repetition for \( t/t^h \) in tembo, as produced by ZS11FPI (a female informant from the southern island – Town Z)
finally set an end for the whole word’s duration (Tier 3). It is worth mentioning that all the boundaries were inserted manually for each repetition, resulting in total of around 29,000 annotated tokens, of which only 13,000 are reported in this dissertation.

3.6 Production results

Since little is known about VOT and F0 in the different varieties of Swahili, a linear mixed-effects model was used as a statistical tool to look into the measures of these phonetic cues and their explanatory factors. For evaluating the results of the production study, mixed-effects models were built for each dialect variety that was obtained from the participants in the four towns (Z, M, N, and I). The models were created in R 3.5.2 (R Core Team, 2012) with the lmer function from the lme4 package (Bates et al., 2015) and the summary function from the lmerTest package (Kuznetsova et al., 2017). The box plots for t-tests and ANOVAs were, in turn, plotted using the ggpubr package (Kassambara, 2020), and line graphs were generated using the ggplot2 package (Wickham, 2016).

All of the models included subject and word as random effects. The first step was to build a null model (with the abovementioned random effects) and then test whether single-factor models could do better at explaining the variance in VOT/F0 than that null model could. Predictors which did not significantly improve upon the null model were eliminated from the models one by one until only significant predictors remained; these were then reported in the final model. In order to validate the VOT/F0 interactions between factors, and to account for the variance in the data, different fixed predictors were examined. By having VOT and F0 as dependent measures, the following fixed explanatory factors were considered during the process of determining the models’ structure: Phonation (aspiration, non-
aspiration), Place (bilabial, alveolar, velar, post-alveolar), Gender (male, female), and Loanword (native\textsuperscript{62}, English, Arabic).

No data points were eliminated from this analysis unless I was sure that there was an error (for instance, values of zero for F0, since it is known that F0 cannot be zero). The process I followed across all locations and stimulus types involved testing each single-factor model against the null model to see if it significantly improved the amount of variance in the data that it could account for; this helped to explain which models yielded a significant \( p \)-value and which models did not significantly account for any variance in the dependent measures.

First, I will develop mixed-effects models for the VOT results. Once the interactions between the various models and VOT in each town are understood, a similar procedure will be presented for the F0 data to examine how they would interact with other mixed-effects models based on F0. Town Z will be analyzed first, and this analysis will be followed by separate analyses for towns M, N, and I, respectively. Later, a further analysis will be done to see where the variations are coming from.

3.7 \textit{Linear mixed-effects analysis}

3.7.1 \textit{VOT in Zanzibar}

The VOTs from town Z are analyzed with respect to two different sets of data. On the one hand, the whole data set was examined when I considered the Phonation contrast; on the other hand, only the unaspirated data set was used as a subset when I looked into the estimates for loanword predictors.

\textsuperscript{62} Native words are not a type of loanword, but just a value for the \textit{Loanword} category in the mixed-effects analysis.
3.7.1.1 VOT/Phonation

In evaluating Phonation, both aspirated and unaspirated data points for voiceless consonants were used and included in the models. At this stage, no elimination of any data point for VOT was considered.

The first single-factor model I tested aimed to see if the Gender factor had any significant effect on the VOTs of town Z speakers. This model showed that Gender, when tested against the null model, was not significant ($\chi^2 (1) = 1.0817; p = 0.2983$), and that the males and females from town Z exhibited VOT values that were very slightly, but not significantly, different. Specifically, females’ VOTs (which come first in the levels of Gender estimates that R reports) had an intercept of 54 ms while males went a bit higher by around 4.50 ms, which was not a large enough difference to reach the level of significance.

After I analyzed the Gender factor, the Phonation factor was examined to test the aspiration/non-aspiration contrast. The Phonation model proved to be significantly better than the null model ($\chi^2 (1) = 54.7; p < .001$). This model estimated that aspiration had an intercept of 82 ms while non-aspiration lowered VOT by 36 ms.

I then looked into the Place factor to see if it could account for any variance more effectively than the null model could. Just like Phonation, the Place model did show an effect on the produced VOT ($\chi^2 (3) = 34.99; p < .001$). However, not all between-Place effects were significant; the only observed significance was between the alveolar intercept of 46 ms and the post-alveolar VOT value, which was 33 ms higher. From the $\chi^2$ values of both Phonation (54.7) and Place (34.99), I confirm that Phonation accounts for more variance in the data from town Z than Place does. Rather than differences in place of articulation, phonation type is, in effect, the primary factor influencing the VOT that is produced in each initial segment in Swahili.
For further testing of the models, I independently added two more predictor slopes to Phonation: namely, Gender and Place. Although the combination of Phonation + Gender did not improve the model ($\chi^2 (1) = 1.0956; p = .2952$), the combination of Phonation + Place did yield a significant result ($\chi^2 (3) = 67.718; p < .001$), which indicates that Phonation + Place predicts the VOT values in the data better than the Phonation factor alone does. To improve the Phonation + Place model further, I took the final step of adding the interaction of Phonation by Place to the model to test whether the interaction of those two factors (Phonation*Place) would also improve the model, and this combination did turn out to make the model better ($\chi^2 (3) = 12.618; p = 0.005539$). This model’s output is illustrated in Table 3-1.\textsuperscript{63}

<table>
<thead>
<tr>
<th>Random effects:</th>
<th></th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>Name</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Word (Intercept)</td>
<td>254.1</td>
<td>15.94</td>
</tr>
<tr>
<td></td>
<td>Subject (Intercept)</td>
<td>189.7</td>
<td>18.48</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>431.5</td>
<td>20.77</td>
</tr>
<tr>
<td>Number of obs</td>
<td>3165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>groups: Word</td>
<td>134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) |
|----------------|----------|------------|----|---------|----------|
| (Intercept)    | 72.213   | 5.925      | 181.189 | 12.188 | < 2e-16  *** |
| Phonationunaspired | -34.420 | 6.234      | 164.412 | -5.522 | 1.29e-07  *** |
| Placebilabial  | 3.971    | 7.600      | 145.298 | 0.522 | 0.60212  |
| Placepost-alveolar | 22.199 | 8.834      | 143.636 | 2.673 | 0.00847  ** |
| Placevelar     | 18.716   | 8.932      | 143.489 | 2.030 | 0.02119  * |
| Phonationunaspired:Placebilabial | -19.680 | 9.219 | 146.428 | -2.135 | 0.03445  * |
| Phonationunaspired:Placepost-alveolar | 13.588 | 9.232 | 145.178 | 1.472 | 0.14323  |
| Phonationunaspired:Placevelar | -9.631 | 8.991 | 145.852 | -1.071 | 0.28587  |

Table 3-1 Phonation VOT coefficients from town Z, according to the last described model (Phonation + Place + Phonation*Place)

Based on Table 3-1, it is evident (for Phonation) that the town Z speakers produced longer VOTs when the obstruents were aspirated (72 ms) and shorter VOTs when they were unaspirated (38 ms). The estimates for Place also show an impact on the VOTs of town Z speakers. The post-alveolar place of articulation estimate\textsuperscript{64} revealed that the VOT for both

\textsuperscript{63} Formula for the final model’s interaction: VOT ~ Phonation + Place + Phonation * Place + (1 | Subject) + (1 | Word). The null baseline model used the same formula, but with the bold text excluded. It is worth mentioning that when interaction or interactional is stated in the results, it always refers to a similar syntax as in the above formula.

\textsuperscript{64} “Post-alveolar” refers to the affricates /tʃ/ and /tʃh/, which consistently have longer VOTs.
aspirated and unaspirated post-alveolar tokens increased by 22 ms overall, while the unaspirated post-alveolars increased in VOT by 14 ms over the alveolar stops, in general. This was expected, as one might assume that town Z speakers would produce longer VOTs when stops are aspirated. This pattern is also observed in the VOT decrease of non-aspiration.

Finally, the interaction between Place and Phonation has yielded set orders that are quite different. For the unaspirated obstruents, I have this hierarchy for VOT: Post-Alveolar > Alveolar > Velar > Bilabial; however, for the overall VOTs of obstruents, I have Post-Alveolar > Velar > Bilabial > Alveolar. In both hierarchies, it is somewhat unexpected that they do not follow the general pattern: Post-Alveolar > Velar > Alveolar > Bilabial.

### 3.7.1.2 VOT/Loanword

In the Loanword category, the mixed-effects model was only applied to the unaspirated data points. It is believed that, if the phonation contrast (aspiration in particular) was part of the data, then the aspiration may produce higher VOT values for native words at the expense of loanwords, which were assigned to an unaspirated category. Therefore, the Loanword category covers loanwords from Arabic and English, as well as native (unaspirated) words to contrast with the loanwords.

Again, with Loanword, I constructed a formula to test each model against the null model as a way to account for any variance in this category. The Loanword (Arabic vs. English vs. Swahili) model showed that VOT is produced differently based on a word’s

---

65 This just means that Post-alveolars had a longer VOT than Alveolars, Velars and Bilabials.
66 Based on the predictor estimates for the Loanword category when analyzing the whole data set (as used for Phonation): that is, by looking at Loanword as the only predictor for VOT (Loanword model against null model). Due to aspiration, and unlike what is shown in Table 3-2, native-word VOT values increased by 32 ms:
origin ($\chi^2 (2) = 16.38; p = 0.000277$). Unlike Place ($\chi^2 (3) = 57.477; p < .001$), Gender was not an improvement upon the null model ($\chi^2 (1) = 1.5467; p = 0.2136$). When I added Gender to the Loanword model, no significance was observed ($\chi^2 (1) = 1.5499; p = 0.2131$). Place + Loanword, on the other hand, did improve the model with very high significance ($\chi^2 (3) = 73.732; p < .001$), as can be seen in Table 3-2. Further testing was done to examine the interaction between Loanword and Place, but this model did not offer a better account of any interacting significance ($\chi^2 (4) = 3.6478; p = 0.4558$), not even when Gender was added to it ($\chi^2 (1) = 1.5803; p = 0.2087$).

Table 3-2 Loanword VOT coefficients from town Z, according to the best model as described above (Loanword + Place)

| Fixed effects: Estimate Std. Error df t value Pr(>|t|) |
|---------------|---------------------|--|-----------------|-----------------|
| Loanwordenglish | 27.203             | 4.472 | 97.853         | 6.103           | 2.12e-08 ***    |
| Loanwordnative | 12.963             | 4.453 | 98.592         | 2.931           | 0.00445 **      |
| Placebilabial   | -25.082            | 5.385 | 100.330        | -4.658          | 9.8e-06 ***     |
| Placepost-alveolar | 29.145           | 4.821 | 98.709         | 6.046           | 2.66e-08 ***    |
| Placevelar      | 11.200             | 4.037 | 98.894         | 2.774           | 0.00662 **      |

The Loanword + Place model showed that town Z participants had a lower VOT for Arabic words (by 26 ms). Native and English VOTs were increased by about 13 and 27 ms, respectively, and all the measurements that I observed for the three categories are significantly different from one another. Therefore, town Z speakers would use specific phonetic cues to signal words, depending on their origin. Based on the VOT of Arabic loanwords as an intercept of 26 ms, the slopes for place of articulation in Loanword VOTs are reduced for Bilabials (-25 ms), a bit higher for Alveolars, higher still for Velars, and the highest increase is for Post-Alveolars (around 29 ms), thus following this hierarchy: Post-Alveolar > Velar > Alveolar > Bilabal.
In sum, we have seen so far that both the Phonation contrast and a Loanword origin influence the VOT phonetic cue in town Z. The distinction of aspiration, as well as that of the different words’ origins, is well-maintained through the VOT dimension. To find out whether this result also applies to the F0 phonetic cue, I structured a linear mixed-effects model to account for any meaningful variation in the level of F0 in the data.

3.8 **F0 in Zanzibar**

F0 in town Z (and all the other towns) was analyzed after excluding the 0 Hz values in the original data. In order to figure out if F0 is a cue for the Phonation contrast and the Loanword type, two models, similar to those that I developed for VOT, were created for F0. As before, I used both the aspirated and unaspirated stop data for the Phonation analysis, but only unaspirated data points for the Loanword analysis. The same procedure that I employed for VOT was performed on the F0 data.

3.8.1 **F0/Phonation**

Three models were built by adding one variable (Gender, Phonation, or Place) to the F0 null model. Gender, as a fixed-effects model, was highly significant ($\chi^2 (1) = 32.009; p < .001$). Phonation was also a significant factor ($\chi^2 (1) = 4.5075; p = 0.03375$), as was Place ($\chi^2 (3) = 11.148; p = 0.01095$). Since Gender was the model with the highest $\chi^2$ value, it was selected first, before the other two variables were added to the model. The Gender + Place model was significant ($\chi^2 (3) = 11.147; p = 0.01095$), and the Gender + Phonation model reached the level of significance, too ($\chi^2 (1) = 4.5113; p = 0.03367$). I then built a model with the interactions (Gender + Place + [Gender*Place]) and compared it to the model with Gender and Place as fixed effects; however, it was not significant ($\chi^2 (3) = 2.6573; p = 0.4475$). To test the interaction more, I chose to test another fixed-effects model that interacts with the Phonation variable (Gender + Phonation + [Gender*Phonation]) and compared it to Gender
and Phonation as fixed effects, and this interaction was statistically significant ($\chi^2 (1) = 5.0918; p = 0.02404$). Since this last model was significant, I went a step further and added Place to the interaction of Gender and Phonation, and this was an improvement that yielded the final F0 model for town Z ($\chi^2 (4) = 18.102; p = 0.001179$), as shown in Table 3-3.

Table 3-3 Phonation F0 coefficients from town Z, according to the last described model (Gender + Phonation + Gender*Phonation + Place)

| Fixed effects                        | Estimate | Std. Error | df | t value | Pr(>|t|) |
|--------------------------------------|----------|------------|----|---------|----------|
| (Intercept)                          | 233.705  | 6.778      | 28.848 | 34.481  | < 2e-16  *** |
| GenderMale                           | -76.779  | 9.185      | 24.366 | -8.359  | 1.28e-08 *** |
| Phonationunaspirited                 | -5.694   | 1.839      | 161.134 | -3.096  | 0.00231 ** |
| Placebilabial                        | -1.460   | 2.342      | 131.920 | -0.627  | 0.53154  |
| Placepost-alveolar                   | 6.259    | 2.172      | 130.823 | 2.882   | 0.00462 ** |
| Placealveolar                        | 3.563    | 1.994      | 130.729 | 1.786   | 0.07634  |
| GenderMale:Phonationunaspirated      | 2.475    | 1.087      | 2768.947 | 2.277   | 0.02283 * |

With a high degree of significance, men’s F0s are 77 Hz lower on average than those of women. Unaspirated F0s are decreased significantly, by 6 Hz. As was expected with VOT, the place of articulation hierarchy showed lower values for bilabials, increased values for velars just slightly above the non-significance level, and even higher values with significance for post-alveolars, following this hierarchy: Post-Alveolar > Velar > Alveolar > Bilabial. The last estimate in the table shows the significant result that I obtained from male participants producing the unaspirated F0 tokens, which had a slightly higher F0 (by 2 Hz).

When I consider Phonation as a whole and look at its low estimate values, it does not seem that F0 is as reliable as VOT is for distinguishing aspiration in Swahili as spoken in town Z. F0 is not as salient a cue in this analysis, since most of the F0 differences can be explained by the Gender factor.
3.8.2  **F0/Loanword**

The same procedures that I employed above were repeated for the Loanword category. Three different explanatory models (with one variable each: Gender, Loanword, or Place) were tested against the null model. Gender proved to be of very high significance ($\chi^2 (1) = 30.906; p < .001$), and the Loanword model also reached the level of significance ($\chi^2 (2) = 10.976; p = 0.004137$), but Place failed to show any significance ($\chi^2 (3) = 4.1442; p = 0.2463$). Since Place had proved to be of no significant difference, I moved on to develop a model based on only the first two variables, Gender and Loanword. These two predictors together, when tested against the Gender variable, proved to be significant ($\chi^2 (2) = 10.981; p = 0.004126$). I then compared the interaction of Gender and Loanword with the two (Gender + Loanword) variables, but no improvement was observed in this interaction ($\chi^2 (2) = 4.9716; p = 0.08326$). Also, no significance was seen when Place was added to the interaction of Gender and Loanword ($\chi^2 (5) = 11.922; p = 0.07351$). Since neither of the last two models were significant, I picked the Gender + Loanword model as my final model for F0 in the Loanword category of town Z; see Table 3-4.

| Fixed effects:          | Estimate | Std. Error | df  | t value | Pr(>|t|) |
|------------------------|----------|------------|-----|---------|----------|
| (Intercept)            | 226.842  | 6.844      | 27303| 33.144  | < 2e-16  ***|
| GenderMale             | -74.365  | 9.369      | 23997| -7.938  | 3.63e-08 ***|
| Loanwordenglish        | 8.055    | 2.426      | 96249| 3.328   | 0.00027 ** |
| Loanwordnative         | 2.591    | 2.356      | 96953| 1.100   | 0.27411  |

Table 3-4 Loanword F0 coefficients from town Z, according to the last described model (Gender + Loanword)

In Table 3-4, it can be seen that male speakers’ F0s for loanwords were 74 Hz lower than those of women; they were also 3 Hz lower than the observed male F0 (77 Hz) in Table 3-3. In addition, the intercept is much higher when the Phonation contrast is included in the
data (234 Hz, also from Figure 3-4); however, the intercept in the Loanword category was 7 Hz lower, at 227 Hz. Following the trend that we saw in the VOTs of English loanwords in Table 3-2, which were significantly higher than those of Arabic and Swahili words, English loanwords are also seen to obtain a significantly higher F0 value (by 8 Hz).

3.9 Zanzibar interim summary

It can be confirmed so far that town Z speakers tend to use the VOT phonetic cue over the F0 one to signal the Phonation contrast in Swahili. This is quite clear from Table 3-1 and Table 3-2: the Phonation contrast is very visible in the former, and the Loanword distinction is also seen to play a big role in the latter. The significant reduction in unaspirated stops’ VOTs, relative to the aspirated VOTs, and the statistically different VOT dimensions as per word origin would support the theory that town Z speakers are dependent on VOT as a reliable cue to help them differentiate between different lexical items.

F0, on the other hand, has shown some variabilities. The results do confirm that men and women have two separate F0 pitch ranges, where on average women have higher F0 values than men. Comparatively speaking, however, the low level of significance and the smaller number of interactions could mean that F0 is possibly not very interesting or meaningful as a cue to obstruent contrasts when we interpret the data from town Z.

3.10 VOT in Mombasa

A similar procedure to those described above was applied to analyze the acoustic cues in the recorded productions from town M. First, an analysis of the influence of Phonation and Loanword on VOT was performed. After that, the influence of both of those categories on F0 was also conducted.
3.10.1 \textit{VOT/Phonation}

Using the whole set of VOT data, models based on Gender, Place, and Phonation were each tested independently against the null model. The Gender model was significant ($\chi^2(1) = 7.22; p = 0.00721$), but bigger significance was seen in the Place model ($\chi^2(3) = 273.02; p < .001$). Phonation, however, did not show any significance ($\chi^2(1) = 0.0212; p = 0.8842$).

Since it was found that Place had the highest $\chi^2$ values in the data, which means that it accounts for more variance, I added the combination of Place and Gender and then tested it against the Place model; this combination proved to be significant ($\chi^2(1) = 7.2859; p = 0.00695$). Conversely, Place and Phonation against Gender were not significant ($\chi^2(1) = 1.9966; p = 0.1577$), but the interaction of Place and Gender was highly significant ($\chi^2(3) = 241.95; p < .001$). A further addition of Phonation to the Place/Gender interaction was not significant ($\chi^2(1) = 2.011; p = 0.1562$), so I went with Place and Gender as the final model, as shown in Table 3-5.

<table>
<thead>
<tr>
<th>Random effects:</th>
<th>Variance Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word (Intercept)</td>
<td>33.06</td>
</tr>
<tr>
<td>Subject (Intercept)</td>
<td>148.58</td>
</tr>
<tr>
<td>Residual</td>
<td>237.68</td>
</tr>
<tr>
<td>Number of obs: 3430; groups: Word, 133; Subject, 26</td>
<td></td>
</tr>
</tbody>
</table>

| Fixed effects:           | Estimate Std. Error df t value Pr(>|t|) |
|--------------------------|----------------------------------------|
| (Intercept)              | 58.280 3.584 31.988 14.829 3.24e-15 *** |
| Placebilabial            | -25.710 1.885 194.455 -13.636 < 2e-16 *** |
| Placepost-alveolar       | 39.117 1.768 194.498 22.129 < 2e-16 *** |
| Placevelar               | 6.353 1.624 194.564 3.911 0.000127 *** |
| GenderMale               | -15.790 4.866 27.697 -3.232 0.000171 *** |
| Placebilabial:GenderMale | -15.690 1.295 3273.617 9.840 < 2e-16 *** |
| Placepost-alveolar:GenderMale | -10.090 1.494 3273.583 -6.745 1.88e-11 *** |
| Placevelar:GenderMale    | 4.076 1.373 3273.733 2.968 0.003038 ** |

**Table 3-5 Phonation VOT coefficients from town M, according to the last described model (Place + Gender + Place*Gender)**

In Table 3-5, the interaction of Place and Gender shows that the overall lowest VOT value was for the bilabials; their VOT decreased by 24 ms from the women’s intercept of 50 ms to be around 26 ms. Post-Alveolars’ and velars’ VOTs, meanwhile, were increased by 39 and 6 ms, respectively. This nicely follows the general place of articulation hierarchy: Post-Alveolar > Velar > Alveolar > Bilabial. The results also confirm that the males have
decreased their VOT values (16 ms) relative to those of the females. Furthermore, the interaction between Gender and Place showed that the males (as seen in the bottom three fixed effects) tended to maintain an increased VOT for bilabials (15 ms), but decreased ones for post-alveolars (by about 10 ms). Velars, meanwhile, are increased by only 4 ms. The males, therefore, exhibited a usual hierarchy pattern that follows this order: Post-Alveolar > Velar > Alveolar > Bilabial.

In sum, there is no evidence so far that VOT played any role in the variety of Swahili spoken in town M. The explanatory factors of Place and Gender proved that the significance in town M is seen only in that the places of articulation have different VOT lengths and that females’ VOTs are higher than males’.

3.10.2 VOT/Loanword

In the Loanword category for town M (unaspirated VOT data only), a procedure similar to that for the VOT/Phonation of town M was conducted. Independently, the models of Place, Gender, and Loanword were tested against the null model. The highest amount of variance was accounted for by the Place model (χ² (3) = 189.1; p < .001), followed by a significant effect for the Gender model (χ² (1) = 7.6766; p = 0.005594), while the Loanword model showed no significance (χ² (2) = 3.708; p = 0.1566). Since Place was the model that accounted for the most variance in town M’s category of Loanword, I added Gender to Place as a two-factor model and tested it against the single model of Place; this combination turned out to be statistically significant (χ² (1) = 7.7658; p = 0.005325). Testing the combination of Gender and Loanword against Place yielded no significance (χ² (2) = 1.9021; p = 0.3863); however, the interaction of Place and Gender was highly significant (χ² (3) = 160.69; p < .001) and was chosen as the final model (as shown in Table 3-6). Adding Loanword to the
The final model did not reach the level of significance (χ² (2) = 1.8882; p = 0.389), so that variable was ultimately left out.

```
Random effects:
Groups   Name     Variance Std.Dev.
Word     (Intercept) 41.02   6.405
Subject  (Intercept) 134.13 11.582
Residual            233.39 15.277
Number of obs: 2549, groups:  Word, 100;  Subject, 26

Fixed effects:  Estimate  Std. Error     df t value  Pr(>|t|)
(Intercept)        49.568      3.520  35.903  14.083     < 2e-16 ***
Place LiuBol       -27.220      2.506 136.031  -10.860     < 2e-16 ***
Placepost-Alveolar  40.492      2.180 136.383   18.578     < 2e-16 ***
PlaceVelar         5.936       1.950 136.581    3.043       0.00281 **
GenderMale         -14.735      4.681  28.629   -3.145       0.00386 **
PlaceVelar:GenderMale 14.217     1.963 2426.183   7.242      5.92e-13 ***
Placepost-Alveolar:GenderMale -11.827     1.785 2426.298  -6.468      1.20e-10 ***
PlaceVelar:GenderMale   4.880      1.527 2426.575   3.237       0.00147 **
```

**Table 3-6 Loanword VOT coefficients from town M, according to the best model as described above (Place + Gender + Place*Gender)**

A pattern similar to that observed in town M’s VOT/Phonation category was seen in its VOT/Loanword category: namely, that women had higher VOT durations (50 ms) than men because the latter decreased theirs by 15 ms. The general hierarchy regarding places of articulation was maintained in the overall data pool (in the three fixed effects above GenderMale); furthermore, this hierarchy was maintained when only male responses were taken into consideration (in the three fixed effects below GenderMale). The male responses followed the order Post-Alveolar > Velar > Alveolar > Bilabial, which is considered a common order that goes according to the normal pattern.

### 3.11 *F0 in Mombasa*

Following the same method that was used for the F0s from town Z, only the non-zero data points for F0 were examined for town M. Similarly, the Phonation category again looked at the whole data set, but only the unaspirated tokens were analyzed for the Loanword category.
3.11.1 **F0/Phonation**

The three independent explanatory factors Gender, Phonation, and Place were tested against the null model, and all three were statistically significant: Gender ($\chi^2 (1) = 39.549; p < .001$), Phonation ($\chi^2 (1) = 8.7046; p = 0.003174$), and Place ($\chi^2 (3) = 15.377; p = 0.001521$). The predictor combinations Gender + Place and Gender + Phonation were both significant when tested against the Gender model ($\chi^2 (3) = 15.382; p = 0.001517$ and $\chi^2 (1) = 8.7071; p = 0.00317$, respectively). The interaction of Gender and Place was also very significant ($\chi^2 (3) = 30.212; p = 1.245e-06$), as was the interaction of Gender and Phonation ($\chi^2 (1) = 13.835; p = 0.0001996$). Since the interaction of Gender and Place had the highest $\chi^2$ value, the Phonation predictor was added to this model to see if the two-factor model would account for more deviance in the data, which it did ($\chi^2 (1) = 11.761; p = 0.000605$). This model was then reported as the final model, as shown in Table 3-7.

<table>
<thead>
<tr>
<th>Random effects:</th>
<th>Variance Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word (Intercept)</td>
<td>170.5</td>
</tr>
<tr>
<td>Subject (Intercept)</td>
<td>747.5</td>
</tr>
<tr>
<td>Residual</td>
<td>252.1 15.88</td>
</tr>
<tr>
<td>Number of obs: 3357, groups: 133; Subject, 26</td>
<td></td>
</tr>
</tbody>
</table>

### Fixed effects:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 286.9935  | 8.1928  | 35.0599  | 35.030 < 2e-16 *** |
| GenderMale | -107.1650 | 10.7747 | 26.3572  | -9.946 2.00e-10 *** |
| Placebilabial | -6.7399  | 3.6646  | 146.3762 | -1.838 0.068078 ** |
| Placepost-alveolar | 10.0569  | 3.4905  | 146.3871 | 2.960 0.003591 *** |
| Placevelar | 0.2796   | 3.1207  | 146.3599 | 0.000 0.928723 |
| Phonationunaspirated | -9.3463  | 2.6835  | 138.7081 | -3.483 0.000064 *** |
| GenderMale:Placebilabial | 4.5644   | 1.6607  | 320.4005 | 2.749 0.000020 *** |
| GenderMale:Placepost-alveolar | 1.6760   | 1.5547  | 320.4721 | 1.078 0.281126 |
| GenderMale:Placevelar | 7.4135   | 1.4288  | 320.6756 | 5.187 2.26e-07 *** |

**Table 3-7 Phonation F0 coefficients from town M, according to the last described model (Place + Gender + Place*Gender + Phonation)**

As can be seen in Table 3-7, men decreased their F0 values by 107 Hz from the women’s intercept, which was at 287 Hz. The overall place of articulation hierarchy for F0 was higher for post-alveolars, similar lengths for velars and alveolars, and the lowest values for bilabials. The unaspirated tokens were also reduced significantly, by 10 Hz. When we
look at the interaction between Gender and Place for male participants, we notice that their place of articulation hierarchy is similar to the overall one. Here, males’ post-alveolars had the highest F0s, followed by velars, alveolars, and then bilabials.

3.11.2 F0/Loanword

The F0/Loanword category accounted for less variation in town M than what we have seen in the F0/Phonation section, for the first three tests in particular. Here, Place was not an improvement upon the null model ($\chi^2 (3) = 6.3919; p = 0.09403$). Gender and Loanword, on the other hand, revealed significant values ($\chi^2 (1) = 39.552; p < .001$ and $\chi^2 (2) = 9.7667; p = 0.007572$, respectively). Since Gender had the highest $\chi^2$ value, it was combined with Loanword (Gender + Loanword) as a two-factor model, which turned out to be significant ($\chi^2 (2) = 9.7729; p = 0.007548$); however, the combination of Gender and Place was not significant ($\chi^2 (3) = 6.3967; p = 0.09383$). Going further and testing the interaction between Gender and Loanword (Gender*Loanword) yielded no significance ($\chi^2 (2) = 4.8061; p = 0.09044$), yet, the interaction between Gender and Place (Gender*Place) revealed a model which proved to be of very strong statistical significance ($\chi^2 (3) = 24.185; p < .001$). I then added Loanword to the last model which featured the interaction of Gender by Place, and because it was significant ($\chi^2 (2) = 11.859; p = 0.00266$), it was reported as the final model, as shown in Table 3-8.

```
Random effects:
  Groups   Name    Variance Std.Dev.    
  Word (Intercept) 192.9       13.89    
  Subject (Intercept) 728.8      27.00    
  Residual          254.4       15.95    
Number of obs: 2485, groups:  Word, 100; Subject, 26

Fixed effects:     Estimate  Std. Error   df  t value  Pr(>|t|)   
(Intercept) 270.358     8.223     32 32.878   2e-16  ***  
GenderMale   -186.248    10.655    26 17.643   9.971   1.87e-10  ***  
Placebilabial -108.497    5.086     7 21.234   0.00037  
Placepost-alveolar 4.191     4.430     9 0.946    0.346265  
Placevelar   2.132      3.704     9 0.576    0.566833  
Loanwordbilabial 14.307     4.021     9 3.558    0.000078  
Loanwordpost-alveolar 8.972     4.882     9 1.824    0.072786  
Loanwordvelar 7.175      2.070    24 3.451    0.000567  ***  
GenderMole*Placebilabial 1.993     1.802     9 1.106    0.268815  
GenderMole*Placepost-alveolar 6.972     3.613    24 1.930    0.0664-05  ***  
GenderMole*Placevelar  
```
Based on Table 3-8, we notice that the average F0 of male participants is 106 Hz less than the female F0 estimate, and the overall Bilabial value is reduced by 10 Hz. Although the other places of articulation did not show any statistical significance, they still show a pattern in which the highest value is for post-alveolars, then a lesser one for velars, followed by alveolars, and the largest decrease is for the bilabials. However, if only men were taken into consideration, then the hierarchy would be a bit different: namely, we would see that the combined interaction of Gender and Place changes the usual F0 pattern slightly. Velars had the biggest increase, a bit less for post-alveolar, and bilabials were increased by only 7 Hz from the overall bilabial intercept. Furthermore, in the Loanword category, English and native words were given significantly higher pitch values than words of Arabic origin (by 14 and 8 Hz, respectively).

3.12 Mombasa interim summary

The data I obtained regarding the VOTs of town M participants revealed that Phonation is not an explanatory factor for the variety of Swahili that is spoken in this town; the only statistically significant slopes in the VOT/Phonation category were found in Gender and Place. For Gender, men were seen to have lower VOT durations than women. Moreover, the overall Place pattern followed the general cross-linguistic VOT order nicely: going from Post-Alveolar to Velar to Alveolar, then to Bilabial. Aslo when only males were looked at, this hierarchy was maintained in that Post-alveolar remained the most increased cue (the longest VOT) while Bilabials were the most decreased (the shortest VOT).

The VOT/Loanword data, which examined the unaspirated tokens only, matched the Phonation results in many respects. Town M subjects did not produce different VOTs based on the words’ origin; however, Place and Gender did matter and showed some significance.
Here, it is confirmed once again that the men had shorter VOTs than the women. And the VOT place of articulation pattern, with respect to both the overall and the gender-based utterances, was similar to the hierarchy that was seen in the Phonation data for town M. So, neither Phonation nor Loanword factors were able to explain any variation in the VOT data.

F0, on the other hand, showed some significant variation, based on both the Phonation and the Loanword categories. In both data sets, men’s F0s were lower than the women’s by about 107 Hz. The overall place of articulation hierarchy also matched quite well: higher for Post-Alveolars and lower for Bilabials. Nonetheless, the Gender/Place interaction was different in the one category when only men were taken into consideration. In male responses to Loanword category, velars were higher (Velar > post-alveolar).

Finally, participants in town M were seen to maintain a clearer distinction in F0 than those in town Z, based on the Phonation and Loanword categories. Non-aspiration was lowered by about 9 Hz, and both English and native words were significantly increased relative to the Arabic ones: by 14 and 8 Hz, respectively. So far, it is obvious that, unlike VOT, the F0 cue is a phonetic working mechanism for town M, but it is not as strong as VOT is in town Z.

3.13 VOT in Nairobi

For the first set of data from town N, the procedure that I conducted for the previous two towns’ data was followed in order to analyze the variation in the Nairobi speakers’ VOT productions.

3.13.1 VOT/Phonation

Phonation in town N did not seem to show any improvement upon the null model ($\chi^2 (1) = 2.1217; p = 0.1452$), but Gender and Place did reach the level of significance ($\chi^2 (1) = \ldots$).
16.452; \( p < .001 \) and \( \chi^2 (3) = 247.75; p < .001, \) respectively). Place and Gender together, when tested against the Place predictor which had yielded a higher \( \chi^2 \) value, was also significant \( \chi^2 (1) = 16.657; p < .001 \), but the combination of Place and Phonation was not \( \chi^2 (1) = 0.3152; \ p = 0.5745 \). Going a step further, an examination of the interaction between Place and Gender (Place*Gender) demonstrated an improvement upon the combination of Place + Gender \( \chi^2 (3) = 149.08; p < .001 \), and this final model is reported below in Table 3-9.

Table 3-9 Phonation VOT coefficients from town N, according to the last described model (Place + Gender + Place*Gender)

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| Intercept | 51.168 | 2.855 | 17.922 | <2e-16 *** |
| Placepost- alveolar | 42.253 | 2.127 | 19.886 | <2e-16 *** |
| Placevelar | 12.321 | 1.948 | 6.323 | 1.94e-09 *** |
| GenderMale | -19.848 | 3.613 | -5.591 | 8.73e-08 *** |
| Placebilabial:GenderMale | 10.842 | 1.472 | 7.307 | 9.47e-11 *** |
| Placepost-alveolar:GenderMale | -6.389 | 1.375 | -4.645 | 3.55e-06 *** |

The estimates in Table 3-9 confirm a decrease in VOT values among male participants by 20 ms. The overall place of articulation hierarchy is along the lines of the general pattern, as it follows this order statistically: Post-Alveolar > Velar > Alveolar > Bilabial. Also, the interaction of Place and Gender, as observed in male speakers, posited a similar pattern by keeping the positioning of place of articulation: Post-Alveolar > Velar > Alveolar > Bilabial. The category of most importance here, Phonation, was not statistically significant, and therefore it was not reported among the explanatory coefficients in the table.

3.13.2 VOT/Loanword

Town N’s VOT/Loanword statistics presented a result that is comparable to that of the Phonation category above. Loanword, when tested against the null model, did not have any
statistically significant difference ($\chi^2 (2) = 4.8943; \ p = 0.08654$), but Place and Gender, on the other hand, had bigger significance and stronger $\chi^2$ values ($\chi^2 (3) = 157.34; \ p < .001$ and $\chi^2 (1) = 15.325; \ p < .001$, respectively). Place + Gender, when tested against Place, were also statistically explanatory factors ($\chi^2 (1) = 15.519; \ p < .001$), but Place + Loanword failed to show any statistical probability ($\chi^2 (2) = 2.8381; \ p = 0.2419$). Since Place + Gender were both explanatory factors, their interaction (Place*gender) did account for more variations in the unaspirated data set ($\chi^2 (3) = 143.23; \ p < .001$), and the results of this final model are reported in Table 3-10.

![Table 3-10 Loanword VOT coefficients from town N, according to the last described model (Place + Gender + Place*Gender)]](image)

As was previously seen in the Phonation estimates, a parallel pattern was observed for the Loanword category. Men had lower VOT values than women by about 20 ms. Also, the overall hierarchy for place of articulation followed the usual VOT ordering: Post-Alveolar > Velar > Alveolar > Bilabial. Moreover, the Gender (male) interaction with Place introduced a common expected VOT direction by again having higher VOTs for Post-Alveolar followed by Velar, Alveolar and then Bilabial.

The Loanword category, then, has proven to be of less interest in town N participants’ productions due to its statistical insignificance, which means that the origin of the words was of no importance.
3.14  **F0 in Nairobi**

In this second set of data from town N speakers, F0 was examined to form the basis of a confident report about whether the speakers tended to apply a type of pitch variation to account for any difference regarding Phonation or Loanword.

3.14.1  **F0/Phonation**

The Phonation model that I built to test against the null model demonstrated non-significance ($\chi^2 (1) = 0.0224; p = 0.8809$), but Gender ($\chi^2 (1) = 54.653; p < .001$) and Place ($\chi^2 (3) = 22.182; p < .001$) were significant. The combination of Gender and Place improved upon the Gender model ($\chi^2 (3) = 22.192; p < .001$), but Gender and Phonation did not ($\chi^2 (1) = 0.0214; p = 0.8838$). In order to test for more variations, I then looked into the interaction between Gender and Place, which turned out to be significant and account for more deviance in the data. Adding Phonation to this interaction of Gender and Place did not yield any significance ($\chi^2 (1) = 0.1114; p = 0.7385$), so I went with only Gender and Place as the final model, as shown in Table 3-11.

![Table 3-11: Phonation F0 coefficients from town N, according to the last described model (Gender + Place + Gender*Place)](image)

Table 3-11: Phonation F0 coefficients from town N, according to the last described model (Gender + Place + Gender*Place)

The estimates in Table 3-11 present an average F0 drop among male speakers by about 104 Hz. The overall place of articulation variations were seen to increase when compared to the alveolar intercept: the values were higher for post-alveolars; then a bit lower
for velars; and the least increase in the table, though not statistically significant, was for bilabials which was still higher than alveolars. This ordering is totally similar when only males’ responses are considered. In male responses, all places of articulation tended to decrease from the overall intercept, following this pattern: Post-Alveolar > Velar > Alveolar > Bilabial. It is also confirmed so far that Phonation is not distinguished based on the F0 cue.

3.14.2 F0/Loanword

To evaluate the F0/Loanword category for town N, the usual three predictors were again tested against the null model. Loanword was not significant ($\chi^2 (2) = 4.6562; p = 0.09748$), but Gender and Place, on the other hand, had bigger statistical difference ($\chi^2 (1) = 54.024; p < 0.001$ and $\chi^2 (3) = 9.2139; p = 0.02658$, respectively). The combination of Gender and Place reached the level of significance ($\chi^2 (3) = 9.2371; p = 0.0263$), yet Gender and Loanword failed to show any explanatory power ($\chi^2 (2) = 4.6535; p = 0.09761$). Further testing of the interaction between Gender and Place showed an improvement on the model ($\chi^2 (3) = 13.801; p = 0.003189$), but adding Loanword to this interaction did not reveal any significance ($\chi^2 (2) = 5.779; p = 0.0556$), so only Gender and Place were taken as part of the final model, as shown in Table 3-12.

<table>
<thead>
<tr>
<th>Random effects:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups Name</td>
</tr>
<tr>
<td>Word (Intercept)</td>
</tr>
<tr>
<td>Subject (Intercept)</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Number of obs: 2026, groups: Word, 111; Subject, 24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate Std. Error df t value Pr(&gt;</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>GenderMale</td>
</tr>
<tr>
<td>Placebilabial</td>
</tr>
<tr>
<td>Placepost-alveolar</td>
</tr>
<tr>
<td>Placevelar</td>
</tr>
<tr>
<td>GenderMale:Placebilabial</td>
</tr>
<tr>
<td>GenderMale:Placepost-alveolar</td>
</tr>
<tr>
<td>GenderMale:Placevelar</td>
</tr>
</tbody>
</table>

Table 3-12 Loanword F0 coefficients from town N, according to the last described model (Gender + Place + Gender*Place)
Table 3-12 reveals that the F0 estimates for the Loanword category in town N were not much different from the slopes in Table 3-11. Men still maintained the lowest pitch, which was reduced by 103 Hz. The overall place of articulation pattern was higher for Post-alveolars, less high for velars, less still for alveolars, and the lowest values were for bilabials. The interaction of males’ responses with Place was similar to what we saw in the Phonation category: there, post-alveolar F0s were higher than the velar, alveolar and bilabial ones, and here, post-alveolar is a bit higher than velar. However, while the overall bilabial estimate was insignificant in the Phonation table, the bilabial estimate in the Loanword table has proved to be of some statistically significant difference.

3.15 Nairobi interim summary

Just like the speakers from town M, participants from town N did not use VOT to construct clear-cut phonetic cues in either the Phonation or the Loanword category. These two explanatory factors were not part of the Nairobians’ phonetics; instead, in line with the town M speakers, the Swahili speakers in town N showed strong, significant slopes for Gender and Place. The Gender-VOT variation decreased significantly when a speaker was male. Also, the place of articulation pattern was seen to follow the usual hierarchy whether the overall result was considered, or when the males’ responses were examined.

For F0, whether the data represented aspiration (the Phonation category) or non-aspiration (the Loanword category), they showed similar statistics. As was observed with VOT, Phonation and Loanword as predictors were not distinguishable by participants, who did not use pitch as an explanatory phonetic cue. Gender, as has been the case in many F0 studies, was found to be significant in that the F0s were higher for females than for males. When we examine the overall result (in both categories), the places of articulation follow the hierarchy Post-Alveolar > Velar > Alveolar > Bilabial. Again, males’ utterances came up
with different patterns: Post-Alveolar > Velar > Alveolar > Bilabial (Table 3-11) and Post-Alveolar > Velar > Bilabial > Alveolar (Table 3-12).

3.16 **VOT in Iringa**

As was done for the other towns, two data subsets were used: one for Phonation, which included aspiration; and another one for Loanword, which excluded aspiration.

3.16.1 **VOT/Phonation**

When the three explanatory factors were tested against the null model, Phonation was not a significant predictor ($\chi^2 (1) = 0.2191; p = 0.6398$), but Gender and Place were ($\chi^2 (1) = 5.7089; p = 0.01688$ and $\chi^2 (3) = 302.22; p < .001$, respectively). When I added Phonation to Place, these together did not have any statistical difference ($\chi^2 (1) = 0.608; p = 0.4355$), yet the combination of Place and Gender did turn out to be significant ($\chi^2 (1) = 5.7805; p = 0.0162$). Even better than the combination (Place + Gender) was the interaction (Place*Gender) ($\chi^2 (3) = 79.097; p < .001$), and so it was picked as the final model, as shown in Table 3-13.

<table>
<thead>
<tr>
<th>Random effects:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups Name</td>
<td>Variance Std.Dev.</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>48.7 6.978</td>
</tr>
<tr>
<td>Subject (Intercept)</td>
<td>152.8 11.525</td>
</tr>
<tr>
<td>Residual</td>
<td>209.7 14.482</td>
</tr>
<tr>
<td>Number of obs</td>
<td>3158, groups: Word, 137; Subject, 24</td>
</tr>
</tbody>
</table>

| Fixed effects: | Estimate | Std. Error | df | t value | Pr(>|t|) |
|----------------|----------|------------|----|----------|---------|
| (Intercept)    | 43.525   | 3.593      | 31734 | 12.116 | 1.9e-13 *** |
| Placepost-alveolar | -9.721 | 2.145 | 174.482 | -4.531 | 1.0e-05 *** |
| Placebilabial  | 56.691   | 1.986      | 182.331 | 28.562 | < 2e-16 *** |
| Placevelar     | 20.154   | 1.848      | 173.949 | 10.907 | < 2e-16 *** |
| GenderMale     | -12.470  | 4.087      | 25.533 | -3.025 | 0.005 * |
| Placepost-alveolar:GenderMale | 6.107 | 1.564 | 3000.857 | 3.906 | 9.6e-05 *** |
| Placebilabial:GenderMale     | -7.366   | 1.470      | 3824.855 | -5.018 | 5.7e-07 *** |
| Placevelar:GenderMale         | 2.859    | 1.344      | 2999.619 | 2.127 | 0.0335 * |

| Table 3-13 Phonation VOT coefficients from town I, according to the last described model (Gender + Place + Gender*Place) |

On the one hand, the estimates above offer evidence that Phonation is not an explanatory factor in the town I data. Gender and Place, on the other hand, strongly support
the idea of a VOT mismatch between males and females, as well as the hypothesis that places of articulation tend to yield different VOT durations based on their phonetic mechanisms. In comparison to females, males decreased their VOTs by about 12 ms. Moreover, the overall place of articulation hierarchy follows the usual pattern, in this order: Post-Alveolar > Velar > Alveolar > Bilabial. Among male speakers, also, a usual pattern was again observed in that Post-Alveolars maintained the longest values while Bilabials got the lowest VOTs: Post-Alveolar > Velar > Alveolar > Bilabial.

3.16.2 VOT/Loanword

In the VOT/Loanword category for town I, the Loanword predictor was not an improvement upon the null model ($\chi^2 (2) = 3.8832; p = 0.1435$). Gender and Place, however, proved to be better explanatory factors ($\chi^2 (1) = 5.7999; p = 0.01603$ and $\chi^2 (3) = 229.01; p < .001$, respectively). The combination of Place and Gender was significant ($\chi^2 (1) = 5.9119; p = .01504$), but that of Place and Loanword was not ($\chi^2 (2) = 5.4357; p = 0.06602$). A further examination of the interaction between Place and Gender improved the model ($\chi^2 (3) = 65.749; p < .001$), but adding Loanword did not yield a significant result ($\chi^2 (2) = 5.418; p = 0.0666$), so only the interaction of Place and Gender were used in the final model, as shown in Table 3-14.

Table 3-14 Loanword VOT coefficients from town I, according to the last described model (Gender + Place + Gender*Place)
The VOTs in the Loanword category had all the characteristics of those in town I’s Phonation category. First, men had lower VOT values. Second, the overall place of articulation hierarchy had the longest VOT values for post-alveolars, then velars, then alveolars, and the shortest values were for bilabials. Furthermore, when examining the interaction between Place and Gender, we again notice that male speakers followed the same pattern, in which they gave the highest values for post-alveolars and the lowest ones for bilabials. Finally, the specific origins of words did not seem to play a role in town I phonetics.

3.17  **F0 in Iringa**

The same method that was used for the other towns was performed here. Data points with a zero value were eliminated, and two data subsets were created: one with aspiration for the Phonation category, and one with only non-aspiration for the Loanword category.

3.17.1  **F0/Phonation**

In the Phonation category, neither Phonation ($\chi^2 (1) = 3.0163; p = 0.08243$) nor Place ($\chi^2 (3) = 2.4646; p = 0.4817$) showed any improvement upon the null model; Gender was the only predictor that could account for more variation in the data ($\chi^2 (1) = 39.487; p < .001$). Going a step further and testing the combinations of Gender + Place and Gender + Phonation did not yield models with statistical significance ($\chi^2 (3) = 2.4753; p = 0.4798$ and $\chi^2 (1) = 0.0214; p = 0.081$, respectively). Therefore, a final model was chosen with Gender as the only explanatory predictor, as shown in Table 3-15.
As can be seen in Table 3-15, the Gender factor showed males’ F0s decreasing by 101 Hz from the intercept of females, but other factors (Phonation and Place, in particular) had no statistical significance in the F0 data.

### 3.17.2 F0/Loanword

Before I arrived at the final model, the usual three factors were tested against the null model. Place was not significant ($\chi^2(3) = 1.4339; p = 0.6976$), but both Gender and Loanword proved to be of statistical difference ($\chi^2(1) = 40.144; p < .001$ and $\chi^2(2) = 8.427; p = 0.01479$, respectively). Further testing showed that, unlike the combination of Gender and Place ($\chi^2(3) = 1.4545; p = 0.6928$), Gender + Loanword had stronger significance ($\chi^2(2) = 8.4294; p = 0.01478$). The interaction of Gender*Loanword did account for more variation ($\chi^2(2) = 6.618; p = 0.03655$), but adding Place did not yield any improvement ($\chi^2(3) = 2.251; p = 0.522$), so Gender*Loanword was used in the final model, as shown in Table 3-16.
The estimates in Table 3-16 confirm that there were lower F0 values for males in comparison to females. Although Place was not an explanatory predictor, words were assigned different F0 values based on their origin. The overall result has the highest values for English loanwords, a bit lower for native words, and the lowest F0s for words of Arabic origin. When only male speakers are considered, however, the pattern is quite different in that English and native words have the same statistically decreased height from the intercept of Arabic.

3.18 *Iringa interim summary*

The VOTs in the two categories exhibited a symmetrical pattern. Just as was found for towns M and N, speakers from town I did not produce the tokens in a way that signaled any variation based on Phonation or Loanword. Moreover, unlike town Z but in line with towns M and N, VOT was different based on gender in that male productions had lower durations compared to those of women. The overall VOT was also observed to vary according to place of articulation: longest for post-alveolars and shortest for bilabials. Further, when only men’s data are examined, the usual hierarchy pattern is observed. Post-alveolars are produced with the longest duration and bilabials with the shortest (this holds true in both categories of Phonation and Loanword). So far, then, VOT is not a cue that offers any meaningful interpretation regarding the productions of town I speakers, except for the evidence that males tend to have lower VOT values than females.

For F0, on the other hand, different variations were observed as we move from Phonation to Loanword. Predictors in the Phonation data set failed to account for more deviance in the data, except by supporting the commonly observed fact that women produce higher F0 values than men. Loanword, meanwhile, has also confirmed that there is a mismatch of F0 based on gender. However, although the overall Loanword category
suggested higher F0s for English tokens, followed by an intermediate range for native ones and the lowest pitch for Arabic ones, male speakers gave Arabic words a higher F0 while English and native stimuli were in the same range. In sum, F0 did not seem to be a reliable cue for yielding any strong evidence about town I subjects’ productions.

3.19 Overall F0 of all towns’ words by Phonation

The total number of tokens generated by all the participants was 11,394: 6,118 from females and 5,276 from males. The first stage of my F0/Phonation analysis involved examining the F0s in all of these data to see if the F0 heights would change based on any change in the phonation of initial stops.

A pairwise t-test was used to perform multiple comparisons of different groups within each town, and this test showed that males and females may be similar or different in their F0s according to their location (town). Specifically, the F0 data yielded highly significant $p$-values among female participants in towns Z and M. The male participants in town M showed a similar pattern, but with a smaller significant $p$-value; thus, it is hard at this stage to confirm whether all the people in town M are sensitive towards a laryngeal phonetic contrast. Furthermore, neither the town Z males nor any participants from towns N or I produced a contrast in phonation based on the spectral measure of F0. Figure 3-2 below shows these specific trends in F0 heights based on changes in phonation.

---

67 Since the linear mixed-effects analyses focus on the entire distribution and interaction of models, I test the means (averages) against each other in the subsequent sections. I am aware of the redundancy with linear mixed-effects in what follows. My purpose is to share what the data would look like when familiar t-test and ANOVA are used.
Figure 3-2 The distribution of F0 ranges in aspirated vs. unaspirated environments (pooled/overall)

For females in town Z, I found a statistically significant difference in F0 heights ($p = 0.00017$, in the top left box). However, as shown by the next box over, there was no statistically significant difference in the F0 heights of town Z males ($p = 0.18$). In town M, the results for both women and men were significant ($p < 0.001$ and $p = 0.0023$, respectively). For towns N and I, however, the opposite trend was observed: both genders showed an insignificant difference in their F0s based on phonation type.

Based on this analysis, it is clear that the females of town M showed a significant effect of a laryngeal phonetic contrast on F0 while the other groups (Z females and M males) showed little or no effect. Yet we have to understand that these values represent only the overall result for all obstruents; we still need to find out which place(s) of articulation these levels of significance (for town Z women and both genders in town M) are derived from.
3.19.1 *Overall F0 of all towns’ words by Phonation (Place of articulation)*

In this portion of the study, for *bilabials*, I considered 1,089 tokens from female participants and 934 tokens from male participants. After a t-test was applied, no group except town N females showed statistically significant differences (see Appendix B-1).

The female and male data from towns Z, M, and I all showed insignificant *p*-values. For town N, however, as mentioned above, the data from female participants yielded a significant result (*p* = 0.0011) while tokens generated by males also had an insignificant value (*p* = 0.12). This analysis, therefore, shows a different trend as compared to the analysis performed above and shown in Figure 3-2 for the pooled F0 results from all places of articulation. Bilabial consonants do not exhibit the same pattern of significance as the pooled data do.

In the analysis of *alveolars*, 1,704 female and 1,494 male tokens were considered and a t-test was applied. Here, similar to what we observed for bilabials across the different genders and towns, only one group sample (both genders in town M) was statistically significant (see Appendix B-2).

A comparison between the data pools for town Z showed insignificant *p*-values for both genders. Similarly, in towns N and I, insignificant *p*-values were observed for both the female and the male samples. In town M, the female tokens yielded a significant difference (*p* = 0.033), while the male samples had a slightly bigger significant *p*-value for alveolar stops (*p* = 0.021). So, the alveolar results from town M followed the same overall significance pattern that was reported in Figure 3-2.

1,941 female and 1,688 male tokens were considered in the analysis of *velars*. In town Z, the female participants’ data showed a significant difference (*p* = 0.0027). The male sample from town Z, by contrast, showed an insignificant difference, as the calculated *p*-
In town M, unlike males ($p = 0.08$), females did exhibit a significant difference in their F0 heights ($p = 0.036$).

On the other hand, a common trend was seen in towns N and I: both genders, when producing word-initial velar sounds, showed insignificant differences in their F0 heights after a t-test was applied. Therefore (as is evident in Appendix B-3), only tokens generated by women in towns Z and M exhibited significant differences in the height of F0. Neither the males nor the females in the other towns showed any significant changes in height when pronouncing an aspirated vs. unaspirated velar sound.

Then, my study of 2,544 post-alveolar F0 tokens in the four towns revealed that many groups exhibited significant differences in their F0 heights when pronouncing post-alveolar sounds. In town Z, females showed a significant difference ($p = 0.018$), while males had an insignificant $p$-value. In town M, on the other hand, the observed $p$-values were significant for both female ($p < 0.001$) and male ($p = 0.00013$) tokens. The town N $p$-values, like those for town Z, were significant for one gender only (at 0.016 for women and 0.1 for men). The only town in which insignificant $p$-values were observed for both females and males was town I (see Appendix B-4).

Interestingly, in the data set for this one place of articulation, half of the participant groups maintained significant F0 differences in aspirated vs. unaspirated environments. This observation can be compared with the overall F0 Phonation data across different genders and towns (Figure 3-2), with the exception of town N females. Both genders in town M and the females of town Z are the ones that showed a significant level of consistency in this regard.

3.19.2 F0/Phonation summary

F0 as a phonetic cue for contrasting words, as observed in the four locations, did not seem to be very strong. For /p/ and /pʰ/, only the females of town N used F0 to signal the contrast in a
significant way; this was done by assigning the unaspirated bilabial /p/ to a higher F0 range. The phonation of alveolars, on the other hand, was significant for both genders in town M, as they gave the aspirated /tʰ/ higher F0 values. Furthermore, females’ productions of velars in towns Z and M yielded the only significant velar p-values that I obtained: in both groups, aspirated velars were given a higher F0 range than their unaspirated counterparts. Finally, both genders in town M and females in towns Z and N exhibited higher pitch values for the aspirated /tʃʰ/. So, the common trend for signaling the contrast was that a higher F0 range was used when an obstruent was aspirated, with the exception of the bilabial /p/ as produced by town N females.

3.20 Overall VOT of all towns’ words by Phonation

After looking into the (aspiration/non-aspiration) variations in the production of F0 via t-tests, I did a parallel analysis of the production of VOT in the four towns. Town Z seems to be very sensitive towards the VOT variability of phonation contrasts, as shown earlier in the section on the linear mixed-effects model; this, however (as shown in Figure 3-3), is disappearing from town M and becomes even less prominent as we move from town N to town I. In the analyses below, there are unequal numbers under “Phonation” due to the fact that the aspirated tokens are less common than the unaspirated ones in the language that was under investigation here.
Figure 3-3 The distribution of VOT ranges in aspirated vs. unaspirated environments (pooled/overall)

These VOT analyses were performed on 12,927 tokens, of which 6,455 were produced by females and 6,472 were produced by males. The effect of aspiration was observed to be strongly significant in town Z ($p < 0.001$). For towns N and I, the t-test $p$-values were also significant (0.018 and 0.00015, respectively) when comparing aspiration with non-aspiration. The only town for which an insignificant $p$-value was observed was town M (0.14).

Thus, all the towns except town M exhibited significant differences in their VOT ranges with respect to aspiration and non-aspiration. Interestingly, however, although towns N and I both had significant $p$-values, the two statistically significant areas show that the unaspirated VOTs are very slightly longer. This is an unexpected pattern.

3.20.1 *Overall VOT of all towns’ words by Phonation (Place of articulation)*

This part of the analysis examined around 13,000 tokens. All four places of articulation are discussed below.
3.20.1.1 VOT/Phonation of bilabials

As Figure 3-4 below shows, I performed a statistical analysis on the VOTs that the participants produced for bilabials (/p/, /pʰ/), and I found that the difference between aspiration and non-aspiration was significant across all towns and genders.

In town Z, the $p$-values for both males and females were highly significant ($p < 0.001$). The difference in VOT lengths for the females and males in town M was also found to be significant ($p = 0.000164$ and $p = 0.0255$, respectively). For town N, the calculated $p$-value for the entire sample was significant at 0.043, and the female bilabial data were also significant at 0.033. Thus, for both genders in town N, the $p$-values related to the aspiration/non-aspiration contrast were statistically significant. A similar statistical test was applied on the data that had been generated by the males and females of town I, and each group’s $p$-value was observed to be significant: males ($p = 0.013$) and females ($p < 0.001$).

![Figure 3-4 Line plots showing VOT (duration/length) ranges in aspirated vs. non-aspirated environments across all genders and towns (bilabials in A / alveolars in B)](image)

3.20.1.2 VOT/Phonation of alveolars

A t-test was then performed on the VOT data for the alveolars (/t/, /tʰ/) as produced by both genders (Figure 3-4B above). Unlike what was found for the bilabials, however, significant $p$-values were observed for towns Z and I only.
In the town Z data, the average values for these two groups were lower for the unaspirated sound and higher for the aspirated one, and the difference between them was significant for both males and females \((p < 0.001)\). In town M, however, neither gender showed any significant difference in VOT length; a t-test was used to calculate \(p\)-values, but it revealed no significance for either men \((p = 0.091)\) or women \((p = 0.29)\). The same test was applied on the town N data, but here, too, the \(p\)-value was insignificant \((0.12)\), indicating that there is no significant difference between the mean VOT values of town N females \((p = 0.12)\) and males \((p = 0.063)\). A similar analysis of the town I data revealed that both genders also had significant \(p\)-values; however, a noticeable trend was observed for this town, in which both genders’ average values for the aspirated samples were lower than those for the unaspirated samples. The calculated \(p\)-value for men was \((p = 0.03705)\), and that for women was of slightly stronger significance \((p = 0.017)\).

3.20.1.3 \textit{VOT/Phonation of velars}

Figure 3-4 A illustrates the statistical analysis that was done on the VOT durations of the velars \((/k/, /k^h/)\) across all towns and genders. In town Z, significant \(p\)-values were observed for both the male and the female samples \((p < 0.001)\). However, the durational differences in town M were found to be insignificant \((p = 0.86\) and \(p = 0.055)\) for males and females, respectively).

In town N, the calculated \(p\)-value for the male samples was insignificant \((p = 0.21)\); however, the one that was obtained for the female velar samples was statistically significant \((p = 0.0025)\). A similar statistical test was applied to investigate males’ and females’ VOT variation in town I, but the \(p\)-value was observed to be insignificant for both groups \((p = 0.098)\) for men and \(p = 0.91)\) for women).
3.20.1.4 VOT/Phonation of post-alveolars

A t-test analysis of VOT length was also done for the post-alveolar /tʃ/ and /tʃʰ/ (Figure 3-5B). For town Z, the difference between non-aspiration and aspiration in the male and female samples was significant ($p < 0.001$). In town M, however, there was little difference between the mean values of men and women, which was also shown by their insignificant $p$-values ($p = 0.15$ and $p = 0.40$, respectively). In line with the results for town M, the calculated $p$-value for the town N samples was not of any significance for either males ($p = 0.052$) or females ($p = 0.17$).

In contrast to towns M and N, the male and female samples from town I had significant $p$-values ($p = 0.0031$ and $p = 0.0244$, respectively). Yet if we look further to see how town I is different from town Z, we notice that the trend for the post-alveolars is similar to that for the alveolars (lower averages for aspirated obstruents and higher averages for unaspirated ones). Thus, the results are statistically significant; they just do not follow the authentic pattern of VOT and aspiration as was found for town Z.

Figure 3-5 Line plots showing VOT (duration/length) ranges in aspirated vs. non-aspirated environments across all genders and towns (velars in A / post-alveolars in B)
3.20.2 *VOT/Phonation summary*

For bilabials, then, all genders and towns showed statistically significant differences in their average VOT values after a t-test was applied. In all the towns, the average value for non-aspiration was lower than that for the aspirated sample; however, the only town that showed consistent and very strong *p*-values for the productions of males and females was town Z.

For alveolars, on the other hand, three categorizations were created. While town Z showed a coherent pattern for the aspiration contrast, neither town M nor town N revealed any significance. Town I, meanwhile, did reach the level of significance in both genders’ productions, but with higher values for the unaspirated alveolars than the aspirated ones.

Moreover, the study of velar tokens yielded statistically significant results for both genders in town Z, but only for female speakers in town N. The production of velars by males in town N, and by both genders in towns M and I, did not show a significant difference.

Following up on the same consistency that was seen in the phonemic contrast for the other places of articulation, both genders in town Z had clear-cut VOT values for post-alveolars. Town I men and women did exhibit a significant contrast, but by having statistically longer VOTs for the unaspirated tokens, and neither gender in towns M or N showed any distinction between the post-alveolar tokens.

In short, the conclusion is that the VOT aspiration contrast seems to be very strong overall in town Z, but not necessarily in the other towns. This contrast holds true among all the subjects of town Z (Appendix B-12). For other towns’ participants the case is different. In town M, only one participant did show a phonation contrast similar to town Z (Appendix B-13), however, only one participant from town N (Appendix B-14) and three from town I (Appendix B-15) did show a significant contrast but with a higher VOT for the unaspirated tokens.
3.21 **Overall Loanword/VOT differences**

In this category of over 9,500 tokens, origin-wise and gender-wise variations were examined to re-evaluate some results that had been found using the linear mixed-effects models (as reported earlier in this chapter). Figure 3-6 shows the results of a three-way ANOVA, which reveal gender-wise differences in the VOT durations of native words and loanwords (Swahili, English, and Arabic; S, E, and A for short) across towns Z, M, N, and I.

![Figure 3-6 Differences in the VOT durations of native and borrowed words across towns Z, M, N, and I](image)

In town Z, the overall VOT durations among female respondents differed across the word types (*p* < 0.01). The median VOT duration of native S words was significantly lower than that of E loanwords (*p* = 0.0005), but significantly higher than that of A loanwords (*p* < 0.001). Among loanwords, meanwhile, the median VOT duration of E words was significantly higher than that of A words (*p* < 0.001). In the data from town Z females, therefore, E words had the highest VOT duration as compared to those of S words and A words, and the difference was significant between the loanword types as well.
When comparing the VOT durations among male respondents from town Z, I found results that were similar to those found for the female respondents. Overall, the words differed significantly in their VOT duration (\( p < 0.001 \)). The median VOT duration of native S words was significantly lower than that of E loanwords (\( p = 0.00014 \)), but significantly higher than that of A loanwords (\( p < 0.001 \)). Among loanwords, the median VOT duration of E words was significantly higher than that of A words (\( p < 0.001 \)). In the data from the males of town Z, therefore, E words again had the highest VOT duration as compared to those of S words and A words, and differences were also evident between the loanword types.

In town M, the overall VOT durations among female respondents differed significantly across the three word types (at \( p < 0.001 \)). The median VOT duration of native S words was significantly higher than those of both loanword types: E (\( p = 0.00044 \)) and A (\( p < 0.001 \)). On the other hand, no significant difference in VOT duration was found between E and A words (\( p = 0.05 \)). Among the females of town M, therefore, S words had the highest VOT duration as compared to those of E and A words; however, loanwords were only slightly different from one another in their VOT durations.

Similar results were found for the male respondents from town M, in that the overall VOT durations differed significantly across the three word types (\( p < 0.001 \)). The median VOT duration of native S words was again found to be significantly higher than those of E (\( p = 0.0067 \)) and A (\( p < 0.001 \)) loanwords. However, no significant difference was found between the VOT durations of E and A loanwords (\( p = 0.25 \)). Among town M males, then, S words had the highest VOT duration as compared to those of E and A words; however, loanwords were not different from one another in their durations.

In town N, the overall VOT durations among female respondents differed across the three word types (at \( p < 0.001 \)). As in town M, the median VOT duration of native S words...
was significantly higher than those of E ($p < 0.001$) and A ($p < 0.001$) loanwords. In the loanword categories, the median VOT duration of E words was significantly higher than that of A words ($p = 0.0018$). Thus, among town N females, S words had the highest VOT duration as compared to those of E and A words, and differences were also evident between the loanword types.

Among the male respondents from town N, the overall VOT duration differed significantly ($p = 0.00043$). The median VOT duration of native S words was again found to be significantly higher than those of E ($p = 0.0015$) and A ($p = 0.00047$) loanwords. However, no significant durational difference was found between E and A words ($p = 0.99$). Therefore, among the males of town N, S words had the highest VOT duration as compared to those of E and A words; however, loanwords were not different from each other.

In town I, the overall VOT durations among females differed significantly ($p < 0.001$). As in towns M and N, the median VOT duration of native S words was found to be significantly higher than those of E ($p < 0.001$) and A ($p < 0.001$) loanwords. However, no significant difference was found between the VOT durations of E and A loanwords ($p = 0.96$).

Among the males of town I, the overall VOT durations differed significantly ($p < 0.001$). The median VOT duration of native S words was again found to be significantly higher those of E ($p < 0.001$) and A ($p < 0.001$) loanwords. However, no significant durational difference was found between E and A words ($p = 0.77$). Therefore, among the town I males, S words had the highest VOT duration as compared to those of E and A words; however, loanwords were not different from each other.

Overall, the VOT durations of native words were found to be significantly different from those of loanwords in all the towns, irrespective of participants’ gender. However, no
difference in the duration of loanwords was found among either gender from town M or town I. In town N, meanwhile, only females were found to have a significant difference in the VOT durations of loanwords.

3.2.1.1 Overall VOT of all towns’ words by Loanword (Place of articulation)

Here, a further examination will look for effects based on the words’ origin, with reference to each place of articulation. Utterances with unaspirated bilabials, alveolars, velars, and post-alveolars were the only data points used in this section.

3.2.1.1.1 VOT/Loanword of bilabials

In town Z, as can be seen in Appendix B-5, a significant difference in the VOT durations of bilabials was found among females. In addition, the median VOT duration of native S words was significantly lower than that of E loanwords \( (p = 0.014) \). The males of town Z also exhibited the same variation, where the median VOT duration of S words was lower than that of E words \( (p = 0.017) \). In town M, meanwhile, no significant differences were found in VOT durations (between native and loanword bilabials) among females \( (p = 0.7) \) or males \( (p = 0.79) \). In town N, females had a significantly higher median VOT duration for S words as compared to E words \( (p = 0.028) \); however, no such difference was found among males \( (p = 0.23) \). In town I, no significant differences in VOT duration were found among females \( (p = 0.84) \) or males \( (p = 0.46) \).

3.2.1.1.2 VOT/Loanword of alveolars

In town Z (Appendix B-6), the overall VOT durations (for the alveolar /t/) among female respondents differed across the word types \( (p < 0.01) \). The median VOT duration of native S words was significantly lower than that of E loanwords \( (p < 0.001) \), but significantly higher than that of A loanwords \( (p = 0.0042) \). In the loanword categories, the median VOT duration of E words was significantly higher than that of A words \( (p < 0.001) \). Among the females of
town Z, then, E words had the highest VOT duration as compared to those of the S and A categories, and the difference was significant between the loanword types as well.

When comparing the VOT durations (for the alveolar /t/) among the male respondents from the same town, I found results that were similar to those of the female respondents. Overall, the words differed significantly in their VOT durations \((p < 0.001)\). The median VOT duration of native S words was again significantly lower than that of E loanwords \((p < 0.001)\), but significantly higher than that of A loanwords \((p < 0.001)\). In the loanword categories, the median VOT duration of E words was significantly higher than that of A words \((p < 0.001)\). Therefore, among the males of town Z, E words again had the highest VOT duration as compared to those of the S and A categories, and differences were also evident between the loanword types.

In town M, the overall VOT durations (for the alveolar /t/) among females did not differ significantly across the three word types \((at p = 0.09)\). Nor were significant differences found between native S words and E loanwords \((p = 0.05)\), or within the loanword types of E and A \((p = 0.58)\). Yet native words were, in fact, significantly different than A loanwords \((p = 0.044)\). Slightly different results were found for the town M males, in that no significant differences were found across the three word types or between the loanwords \((at p > 0.05)\).

In town N, the overall VOT durations (for the alveolar /t/) did not differ for females across the three word types \((at p = 0.083)\). However, a post-hoc analysis showed that the VOT duration of native S words was significantly lower than that of E loanwords \((p = 0.035)\). No significant differences were found between S words and A words \((p = 0.075)\), or between the loanword categories \((p = 0.62)\).

Among the male respondents from town N, the overall VOT duration of the words did differ significantly \((p = 0.0092)\). This difference was significant due to a statistical difference
in the median VOT durations of native S words and A loanwords \( (p = 0.0043) \), as well as a
difference between E and A loanwords \( (p = 0.036) \).

In town I, the overall VOT duration (for the alveolar /t/) among females differed
significantly \( (p = 0.0031) \) across the three word types. The median VOT duration of A
loanwords was significantly higher than those of native S words \( (p = 0.0016) \) and E
loanwords \( (p = 0.015) \). However, no difference was found between S words and E words.

Similar results were also evident among the male participants from town I, whose
median VOT duration (for the alveolar /t/) of A loanwords was again significantly higher
than those of native S words \( (p = 0.011) \) and E loanwords \( (p = 0.0041) \). Furthermore, no
significant durational difference was found between S words and E words \( (p = 0.87) \).

3.21.1.3 VOT/Loanword of velars
In town Z (Appendix B-7), the overall VOT durations (for the velar /k/) among female
respondents differed significantly across the three word types \( (p < 0.01) \). The median VOT
duration of E loanwords was found to be significantly higher than those of native S words \( (p = 0.0036) \) and A loanwords \( (p < 0.001) \). Furthermore, the VOT duration of S words was
significantly higher than that of A words \( (p < 0.001) \).

Yet, slightly different results were found for the males in town Z. There, males did
not have a higher median VOT duration (for the velar /k/) of E loanwords as compared to
native S words \( (p = 0.062) \); however, a significant difference was seen between S words and
A loanwords \( (p = 0.00086) \). A significant difference in velar VOT durations was also found
between E and A loanwords \( (p < 0.001) \).

In town M, the overall VOT durations (for the velar /k/) among females differed
significantly across the three word types \( (p < 0.001) \). However, this pattern stemmed only
from a significant difference in VOT duration between native S words and A loanwords ($p = 0.00056$), as well as between the two types of loanwords ($p = 0.003$). No significant durational differences were found between S words and E words ($p = 0.3$).

Among male respondents from town M, the words with velar sounds differed significantly in their VOT duration (at $p = 0.037$). However, this pattern stemmed only from a significant difference in VOT duration between native S words and A loanwords ($p = 0.032$). No significant differences were found between S words and E words ($p = 1$), or between the loanword types ($p = 0.052$).

In town N, the overall VOT duration (for the velar /k/) in native words and loanwords differed significantly among both females ($p < 0.001$) and males ($p = 0.0031$). The velar VOT duration for native S words was also significantly higher than that of A loanwords for both females ($p < 0.001$) and males ($p = 0.0016$); however, no such difference was found between S words and E words for either females ($p = 0.57$) or males ($p = 0.53$). On the other hand, a significant durational difference was found between E words and A words for both females ($p < 0.001$) and males ($p = 0.035$).

When I compared the VOT durations for the unaspirated velar /k/ across native words and loanwords among both genders in town I, no significant differences were found (for females $p = 0.81$, for males $p = 0.32$).

### 3.21.1.4 VOT/Loanword of post-alveolars

In town Z (Appendix B-8), both genders exhibited significant differences in their VOT durations for the unaspirated post-alveolar /tʃ/ (females $p < 0.001$, males $p < 0.001$). The median VOT duration of E loanwords was also significantly higher than that of native S words among both genders.
In town M, no significant post-alveolar durational differences were found among participants of either gender (female $p = 0.4$, male $p = 0.96$).

In town N, neither gender had a higher post-alveolar VOT duration for native S words as compared to that of E loanwords (female $p = 0.07$, male $p = 0.34$).

In town I, however, the results were different. Here, females had a higher post-alveolar VOT duration for native S words as compared to that of E loanwords ($p = 0.0018$), and a significant difference was found among the males, too ($p = 0.015$).

### 3.21.2 VOT/Loanword summary

In the overall Loanword category, the VOT duration of native words was found to be significantly different from that of loanwords in all the towns, irrespective of participants’ gender. However, no difference in the VOT durations of loanwords was found among either gender in town M or town I. In town N, meanwhile, only females were found to exhibit a significant difference in the VOT durations of loanwords.

Overall for unaspirated **bilabials**, females had a higher VOT duration for native S words in towns Z and N only. Among males, this difference was evident in town Z only.

Overall for **alveolars**, VOT durations for A loanwords were significantly different than those for native S words among all groups, with the exception of males in town M and females in town N. The durations of A loanwords were also different from those of E loanwords, except among both genders in town M and the females of town N. In addition, native alveolars were different from those of E loanwords only in town Z and among the females of town N.

Overall for **velars**, VOT durations differed within the genders and across most of the towns. Although the durations differed significantly within both genders in town Z, town I
did not show any significance. In towns M and N, meanwhile, differences were found between native S words and A loanwords, as well as between E and A loanwords, but no significance was observed between E words and S words.

Overall for post-alveolars, significant durational differences in native words and loanwords were evident among both genders in towns Z and I, but among neither gender in towns M and N.68

3.22 Evidence of VOT gender differences

It has been noticed so far in these results that the females tended to produce longer VOT durations than the males. Here, I will provide detailed evidence for this pattern from the pooled data for the aspirated and unaspirated obstruents, as well as a confirmation of a gender-based VOT distinction in the Loanword category.

3.22.1 VOT differences in aspirated tokens

In Town Z, a slightly significant, aspiration-based difference in VOT duration was found between females and males ($p = 0.0042$). However, between-gender differences were also evident in the other towns: M ($p < 0.001$), N ($p < 0.001$), and I ($p < 0.001$). In all three of the latter cases (but not the first), the median VOT duration, based on the aspiration contrast, was higher among the females. Overall, females were found to have higher aspiration-based VOT durations in three of the four towns (see Figure 3-7).

A further examination of the aspirated tokens (/pʰ/, /tʰ/, /kʰ/ and /tʃʰ/) showed that the females from towns M, N, and I maintained higher VOT values for all places of articulation. Town Z, on the other hand, had significantly higher VOTs for males in aspirated bilabials and alveolars, but there was no significance at all between town Z men and women in aspirated

---

68 For an overview of Loanword VOT category (based on word origins) as produced by all participants in towns Z, M, N and I, see appendices B-16, B-17, B-18 and B-19, respectively.
velars and post-alveolars.

**Figure 3-7** Between-gender differences in VOT duration, based on aspiration, in towns Z, M, N, and I

### 3.22.2 VOT differences in unaspirated tokens

Overall between-gender differences (see Figure 3-8) in the VOT durations of unaspirated tokens were evident in all four towns. Similar to the observation that was made earlier about the aspirated sounds, town Z showed a small, significant difference in unaspirated VOT durations, which was higher in males than females ($p = 0.0047$). Elsewhere, however, the gender differences in unaspirated VOT durations were highly significant in that the median VOT duration was significantly higher for females as compared to males in towns M ($p < 0.001$), N ($p < 0.001$), and I ($p < 0.001$). So, it can be concluded that females, as compared to males, had higher unaspirated VOT durations in towns M, N, and I, whereas town Z males had durations that were only slightly, but still significantly, higher.

Going a step further and looking into the four places of articulation, we can confirm that the distinction in VOT length between town Z males and females holds true in only two
places (/t/ and /k/). The other two environments (/p/ and /tʃ/), by contrast, were seen to show no significance in town Z. Towns, M, N, and I, meanwhile, had significant gender differences in all four places of articulation (/p/, /t/, /k/, and /tʃ/).

![Figure 3-8 Between-gender differences in VOT duration, based on non-aspiration, in towns Z, M, N, and I](image)

### 3.22.3 Gender differences in VOT (based on a word’s origin)

As can be seen in Appendix B-9, the application of a t-test showed between-gender durational differences in three of the four towns based on word origin (native words only). In town Z, no significant durational difference was found between females and males \((p = 0.16)\). However, in the other towns, the gender differences were significant: the median VOT duration was significantly higher for females as compared to males in towns M \((p < 0.001)\), N \((p < 0.001)\), and I \((p = 0.00084)\).

Further testing also showed between-gender differences in English loanwords’ VOT durations in three of the four towns (Appendix B-10). Among town Z participants, no significant difference was found between females and males \((p = 0.08)\). Yet in the other towns, the between-gender VOT differences were significant: the median VOT duration in
English loanwords was significantly higher for females relative to males in towns M ($p < 0.001$), N ($p < 0.001$), and I ($p < 0.001$).

Loanwords of Arabic origin, meanwhile, reached the significance level with respect to between-gender differences in all towns (Appendix B-11). However, unlike the significantly higher VOT values ($p < 0.001$) for females in towns M, N, and I, town Z males had higher VOT values than those that were observed for females ($p = 0.014$).

### 3.22.4 VOT gender differences summary

Whether a given word had a native Swahili, English, or Arabic origin, the men and women who participated in the study tended to produce different VOT values. In towns M, N, and I, the female participants had significantly higher values than the males. Town Z, on the other hand, yielded higher VOT values for men as compared to women in all word-origin categories, but this difference was statistical in only one word-type: namely, that of words borrowed from Arabic. Thus, town Z is different than the other locations with respect to the gender differences.
Chapter 4: Imitation Experiment

4.1 Participants

The same participants from the production experiment (except one, n = 97) participated in the imitation task.

4.2 Materials

Four disyllabic nonce words—\textit{pope} for /p/, \textit{tote} for /t/, \textit{koke} for /k/, and \textit{choche} for /tʃ/—were modified to represent different VOT lengths and F0 heights, then randomized and presented to the participants.

4.3 Design

After tokens were divided into VOT and F0 groups, the final stimuli were constructed in two stages.

First, in the VOT group, each initial obstruent was modified in Praat (Boersma & Weenink, 2018) to form three different durations along the VOT continuum: a very short duration for Level 1 and a 20- and 40-millisecond increase for Levels 2 and 3, respectively (see Table 4-1).

Table 4-1 The levels of VOT and F0 for the target obstruents, as presented in the stimuli

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>VOT</th>
<th>F0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>StimLevel</td>
<td>StimLevel</td>
</tr>
<tr>
<td>/p/</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>/t/</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>/k/</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>/tʃ/</td>
<td>50</td>
<td>70</td>
</tr>
</tbody>
</table>
During the construction process, we kept in mind the common-sense understanding that bilabials should have the shortest duration among the sounds we were investigating. So, in the final stimuli, the initial VOT at Level 1 was shortest for bilabials, slightly longer for alveolars, longer still for velars, and longest of all for the post-alveolar affricate.

The original tokens, which constituted the base for our VOT and F0 manipulations, were produced by a female model speaker of Swahili from Dar es Salaam, Tanzania, who was studying at the University of Calgary. To create the modified stimuli, we cut and spliced these tokens (see Figure 4-1 for the word *tote* in the three VOT conditions).

![Figure 4-1 VOT stimuli construction for [t] in the nonce word *tote* at the three levels, from 15 to 35 to 55 ms](image)

Second, F0 stimuli were constructed in Praat via pitch manipulation, with the minimum pitch set at 75 Hz and the maximum pitch set at 600. The first three time points (see Figure 4-2) were always adjusted to be compatible with the pitch contour of the rest of the vowel; this technique was meant to keep the manipulation consistent, with no sharp drop or extreme rise in the F0. Once we had prepared the tokens, and listened to them and checked them for errors, the audio files were transferred to SuperCard for easy running of the experiment.

69 The stimuli range is not based on the talker’s base production as she failed to show any aspiration contrast in Swahili. It is not known if the model speaker speaks a contrasting dialect of another Bantu language.
4.4 Perception procedure

All listeners-turned-speakers produced utterances in both the VOT and F0 conditions. At the time of data collection in East Africa, each participant heard the tokens in a different randomization and was asked to repeat each one twice right after hearing the token (24 stimuli for a total of 48 repetitions) except for Iringa subjects, who produced only one repetition per token. It is worth mentioning that SuperCard software controlled the randomization and while Praat recorded each informant’s responses, SuperCard recorded reaction time. We collected the data using two MacBook Airs, and although a mouse was provided at all times, the touch pad posed a difficulty for some participants in their effort to perform the task efficiently. Reaction time was recorded as a dependent measure, but ended up being excluded from the analysis due to some technical difficulties.

The data reported here comprise 6,151 tokens; these include VOT tokens from VOT modifications, F0 tokens from F0 modifications, and VOT tokens from F0 modifications. That last category was looked at to see if a change in F0 would lead to a change in the produced VOT.

Figure 4-2 F0 stimuli construction for [t] in the nonce word tote at the three levels, from 200 to 220 to 240 Hz
4.5 *Perception annotation method*

The annotation method for the perception (imitation) data was quite similar to the procedure that was performed on the production data (Section 3.5). As in the production experiment, three tiers were created to measure the following: first, the closure before the release of the stop; second, the length of the vowel; and third, the start and end of the imitated token (Figure 4-3 below). What is different about the imitation experiment is that each informant was asked to mimic each nonce word twice, according to a certain randomization that was arranged in a different order for every subject. Also, in the *word* tier, the target non-word was not written because the focus here was on the imitation of the word’s initial VOT. To avoid any bias, the words in the imitation audio files were annotated according to the informant’s listening sequence by assigning each token a number, rather than a word; thus, the first mimicked word got the number “1”, as did the first repetition (e.g. 1-1). In the same vein, the second repetition of the first mimicked word would be referred to as 1-2, and so on. When the annotations were done for all the tokens, a Praat script was run to extract the measurements and associate the imitated VOTs, in relevance, to the original stimulus. Statistical analysis would then be done to see if each novel length and height was close to, or different from, the stimulus presented in the experiment. The manual annotations in this section yielded a total of 4,100 tokens.

![Figure 4-3 The VOT of the non-word koke at 30 ms, the first repetition of word number 6 as produced by NS8MI (a male informant from Town N)](image-url)
4.6 **Results of the imitation task**

To achieve one of the objectives of the study, I analyzed the imitation results using a similar procedure to the one I used for the production data. A linear mixed-effects model from the lme4 package (Bates et al., 2015) in R software (R Core Team, 2012) was used to examine the roles of VOT durations and F0 heights in the imitation of nonce words. In this specific task of imitation, every participant was requested to imitate 24 phonetic modifications divided equally between VOT and F0, the two dependent measures, in two repetitions. The tokens were four non-Swahili words beginning with the obstruents /p/, /t/, /k/ and /tʃ/, and were designed to test the imitation of different phonetic levels (see Table 4-1).

For this purpose, I conducted a step-by-step analysis in order to create the most significant models. Summaries from the lmerTest package (Kuznetsova et al., 2017) of the final mixed-effects models (for both VOT and F0) were then reported. Later, the ggpubr package (Kassambara, 2020) was used for generating the statistics and box-plotting the results of the t-test or the ANOVA – and, on some occasions, both. Each town was then assigned a special color for easy visualization: green for Zanzibar (town Z), blue for Mombasa (town M), grey for Nairobi (town N), and brown for Iringa (town I). The purpose of this experiment was to see whether the participants perceive the changes in the modified levels of VOT and F0 (three each). Figuring out whether the locality/variant of Swahili can have an effect on the dependent measures will shed light on the status of these working phonetic cues in Swahili.
4.7  **Linear mixed-effects analysis**

4.7.1  **VOT variations in all towns**

First, I started by creating a null model just as I did for the production data; this factors out words and subjects. Words and subjects were treated as random effects in the model. After that, I ran tests to determine whether the Level factor (which accounts for the fact that the VOTs were presented at three levels in three different durations) could explain more of the VOT (dependent measure) variation than the null model could. The $p$-value I got, for the Level model, was very small ($\chi^2 (1) = 55.741; p < 0.001$), which means that the speakers did change their VOTs when hearing different durations of VOT in the stimuli. I then added in the factor of Repetition\textsuperscript{70} to see if it would further improve the model, but its $p$-value did not prove to be significant ($\chi^2 (1) = 0.0297; p = 0.8631$), which meant that I was better off with just Level in my model. After excluding Repetition, I added Town as a factor to see if it would improve the model’s explanatory power, and it yielded a significant result ($\chi^2 (3) = 12.971; p = 0.004699$). I then looked at the interaction between Level and Town, which did prove significant ($\chi^2 (3) = 11.313; p = 0.01015$). Neither adding Gender nor adding Repetition yielded any improvement on this interaction of Level and Town ($\chi^2 (1) = 0.7378; p = 0.3904$ and $\chi^2 (1) = 0.172; p = 0.6783$, respectively). Thus, the final model, in the linear mixed-effects model of Level*Town has shown that the VOTs are different across levels and towns, as shown in Table 4-2.

\textsuperscript{70} Repetition = the subjects had to repeat the stimuli more than once.
Table 4-2 Imitation VOT coefficients from all towns, according to the last described model (Level*Town)

By having town I (Iringa) as the intercept at 44 ms, we conclude that the participants in towns Z and M did decrease their VOT values: by 13 and 14 ms, respectively. So, here we have differences based on the explanatory factor Town. The VOTs from town N, on the other hand, increased by 1 ms, but this number is not statistically different from that of town I.

When we then look at the interaction of Town and Level, we see that the results for towns M and N do not have any statistical difference – unlike that which was observed for town Z. Level, as an explanatory predictor estimate, proved that there were more variations in town Z based on Level ($p = 0.027659$).

The findings of the study tell that the overall production of the modified VOT tokens was longer in two towns (I and N; L2 language) and significantly shorter in two towns (Z and M; L1 language). The interaction of Town and Level, however, indicated that the only town with varying significant VOT variation is town Z. In general, localities had different overall perception mechanisms, surviving among L1 speakers and disappearing in L2 varieties. An additional examination of the two main factors (Town and Level) revealed that town Z had stronger imitation than that found in town M.

For further analysis of this interaction in a linear mixed-effects model, the data were then divided up by town and looked at to see how VOT changes by Level in each location.
The explanatory factor Level was then tested against the null model in each town separately. In town Z, the Level model showed very strong improvement upon the null model with a high $\chi^2$ value ($\chi^2 (1) = 52.46; p < 0.001$). Town M also yielded a significant $p$-value for Level; however, its $\chi^2$ value was lower than the one for town Z ($\chi^2 (1) = 29.361; p < 0.001$). Of the other two towns, town N had above-average significance ($\chi^2 (1) = 4.7623; p = 0.02909$), while town I was the only one of the four to show no statistically significant difference for the predictor Level ($\chi^2 (1) = 1.3328; p = 0.2483$).

4.7.2 **F0 variations in all towns**

Several models were tested against the null model. First, a linear mixed-effects model was used to compare the F0 null model with the F0 Gender model, and the results showed that the Gender factor yielded an improvement upon the null model ($\chi^2 (1) = 152.03; p < 0.001$). Second, Level, as an explanatory factor, does improve the model over the null model; that is, it did show some significance ($\chi^2 (1) = 4.4006; p = 0.03593$). After that, a third model was created to test for any variation across towns, but Town as an explanatory factor did not improve upon the null model, either ($\chi^2 (3) = 5.4131; p = 0.1439$). Therefore, unlike Gender and Level, this specific model, Level, implies that the speakers were not changing their own F0 in response to the F0 differences in the stimuli.

Because Gender had the most significance and the highest $\chi^2$ value, the other two predictor slopes (Level and Town) were added to the Gender model independently. I first added Level to the Gender model, these two together did show a significance ($\chi^2 (1) = 4.4093; p = 0.03574$). The combination of Gender + Town was significant ($\chi^2 (3) = 32.634; p < 0.001$). Because the interaction of Gender*Town did not show any improvement ($p = 0.2284$), Gender+Town was then chosen as the final model, as reported in Table 4-3.
The estimates in Table 4-3 provide evidence that the males of all towns had the largest pitch decrease (by 106 Hz). Furthermore, the highest pitch was heard in town M (increased by 40 Hz over the town I intercept). The last two estimates give towns N and Z a significant higher pitch (14 and 17 Hz, respectively) relative to town I intercept. Although the last two values do not appear to be significantly different, the overall F0 in the four localities exhibits this hierarchy: Town M > Town Z > Town N > Town I.

4.7.3 **Summary of variations in imitated VOTs**

The VOT results from these linear mixed-effects models reveal that the town which accounts for the most deviance in the VOT data is town Z because it has the highest \( \chi^2 \) value and the strongest \( p \)-value. Towns M and N, although they yielded some significance when their Level factor was examined independently, they did not differ significantly in VOT from the null model when the data from all of the towns was analyzed together (Table 4-2). Town I, however, was consistent in its statistics in that neither the overall model nor the specific model proved to have any statistical significance.

Town N was insignificantly shorter than the overall intercept of town I (44 ms) by only 1 ms. On the other hand, towns Z and M (L1 speakers) showed lower overall significant VOTs. The difference between the latter two towns did not seem to be significant, too,
because town Z is longer than town M by just 1 ms. So far, and based on VOT, we have seen that speakers of Swahili as a second language (with longer VOTs) did differ significantly from speakers of Swahili as a first language (whose VOTs were short). The interaction between Level and Town as found the VOT last model, provides a further evidence that town Z is not only different from towns I and N. Unlike town M, the aforementioned interaction revealed that the factor Level was found to be significantly stronger in town Z. As a result, we expect Level 1, level 2 and Level 3 to have well-separated VOT durations (and for knowing how these VOTs differ from the stimuli of different levels, see Z in Figure 4-5). However, when the data from each town was analyzed separately, Level was a significant predictor in town M, so we expect to see some VOT differences based on Level (also see M in Figure 4-5).

4.7.4 Summary of variations in imitated F0s

The F0 models as constructed and analyzed above show that F0 was not a phonetic cue to signal the different levels of pitch manipulation: that is, F0 failed to account for any changing F0 values across the three levels of imitated stimuli. Yet, Town, as an explanatory factor, was a factor that could explain variances in the data. Furthermore, in support of many phonetic studies, Gender was an explanatory factor: it offered an improvement over the null model and had a strong p-value. The two-factor model that combined Gender and Level did not show any significance, but the combination of Gender and Town did, which suggests that some towns had some pitch variations.

Overall, the male F0 pattern replicates previous work, males had lower F0 relative to females. Moreover, despite having a close pitch range in towns Z and N (3 Hz difference), the F0 heights of the other three towns were distinct from the intercept estimate of town I.
Thus, we see location-related changes of F0 production based on the locality of subjects (increased in town M and decreased in town I, but intermediate for towns Z and N).

Some of the towns, therefore, exhibited significant variations in their imitated phonetic cues. For this reason, an ANOVA and a t-test will be used to look for the source of these variations.

4.8 Ranges of F0 in all four towns

Figure 4-4 shows the fundamental frequency (F0), in Hz, of the four nonce words at the three different levels (Level 1, Level 2, and Level 3) in the four towns (Z, M, N, and I). Here, I have collapsed the F0 data across all of the stops and the one affricate, but divided them by gender.

Among town Z females, there was no difference between Level 1 and Level 2, Level 1 and Level 3, or Level 2 and Level 3 (p = 0.59, p = 0.46, and p = 0.83, respectively).

Similarly, in the female data from town M, there was no statistical difference between the range at Level 1 and that at Level 2 (p = 0.48), or between Level 1 and Level 3 (p = 0.72). Nor was Level 2 significantly different from Level 3 (p = 0.31). In town N, despite the fact that the median F0 range at Level 3 was higher than those at the other two levels, the female data did not yield a significant difference between Level 1 and Level 2 (p = 0.38), Level 1 and Level 3 (p = 0.22), or Level 2 and Level 3 (p = 0.73). Finally, in town I, the median F0 range increases from Level 1 to Level 3, but despite this difference, no statistical difference was found between Level 1 and Level 2 (p = 0.53), Level 1 and Level 3 (p = 0.3), or Level 2 and Level 3 (p = 0.53).
Among the males of town Z, no significant difference was found between Level 1 and Level 2 \((p = 0.77)\), Level 2 and Level 3 \((p = 0.85)\), or Level 1 and Level 3 \((p = 0.62)\). Also, for the men of town M, there was no significant difference between the range at Level 2 and that at Level 3 \((p = 0.92)\), or between Level 1 and Level 3 \((p = 0.93)\). Level 1 and Level 2 were not significantly different, either \((p = 0.98)\). Among the males of town N, the difference was not significant between Level 2 and Level 3 \((p = 0.71)\), Level 1 and Level 3 \((p = 0.31)\), or Level 1 and Level 2 \((p = 0.26)\). Lastly, for the town I males, no significant difference was found between Level 1 and Level 2 \((p = 0.94)\), Level 1 and Level 3 \((p = 0.51)\), or Level 2 and Level 3 \((p = 0.45)\).

Based on the above analysis, which is supported by the insignificant overall ANOVA \(p\)-value for each gender in each town, we can deduce that the F0 range (in Hz) does not differ significantly at any of the three levels (Level 1, Level 2, and Level 3), in any town, among either gender.
4.9 Durations of VOT in all four towns

Figure 4-5 shows the distribution of VOT durations at the three levels (Level 1, Level 2, and Level 3) in the four towns (Z, M, N, and I). Here, too, I have collapsed the data across all the stops and the one affricate. Overall, some significant variation in VOT durations was found for towns Z and M, whereas no such evidence was found for towns N and I.

Within the town Z data, there is a gradual increase in median VOT duration as we go from Level 1 to Level 3. The application of an ANOVA showed that, overall, the VOT durations were significantly different across all three levels ($p < 0.001$), and further examination showed that the VOT durations at all three levels were also different from one another. Specifically, the VOT durations at Level 1 and Level 2 were significantly different ($p = 0.01$), as were those between Level 2 and Level 3 ($p = 0.014$) and Level 1 and Level 3 ($p < 0.001$). In town M as well, the overall difference in VOT duration at the three levels was found to be significant (ANOVA, $p = 0.0018$). The application of a t-test showed that this durational difference occurs between Level 1 and Level 2 ($p = 0.0085$), and also between Level 1 and Level 3 ($p = 0.00049$). However, no significant difference was found between the VOT durations at Level 2 and Level 3 ($p = 0.4$).
Lastly, in towns N and I, no significant durational difference was evident between any pair of levels.

Based on the above analysis, we can see that the VOT duration at Level 1 differs significantly from those at the other two levels only in towns Z and M. Furthermore, in town Z only, the VOT duration at Level 2 was significantly different from that at Level 3.

4.9.1 Durations of VOT by gender

Here, I will take a closer look (see Figure 4-6) in order to account for the significance levels that were found in the overall VOT results above.

The town Z women had the ability to produce VOTs that were distinct from one another. Their range at Level 1 is statistically different from those at Level 2 \((p = 0.045)\) and Level 3 \((p = 0.0057)\); however, there was no significant difference between Level 2 and Level 3 \((p = 0.36)\). For town M females, Level 1 and Level 3 were significantly different \((p = 0.0011)\), but Level 1 and Level 2 were not \((p = 0.085)\), nor were Level 2 and Level 3 \((p = 0.24)\).
0.16). The town N women, meanwhile, yielded no significant differences in either the overall ANOVA or the three levels’ t-tests. Similarly, the data from the females of town I showed no significant differences in any test or at any level.

Among the male groups, the most receptive to a contrast in the VOT manipulations, as proven statistically, were the men of town Z. They had the lowest overall ANOVA value \((p = 0.00021)\), as well as the lowest \(p\)-value between both Level 1 and Level 3 \((p = 0.00011)\) and Level 2 and Level 3 \((p = 0.013)\), but no significant difference was found between Level 1 and Level 2 \((p = 0.099)\). Furthermore, in a result that was not very distinct than that of their female counterparts, the male speakers of town M had a significant difference between one level pair only, in this case Level 1 and Level 2 \((p = 0.029)\). And, finally, in line with the females of towns N and I, the males from those two towns produced no statistically significant differences at all between any pair of levels.

![Figure 4-6 Distribution of VOT durations across the three levels and the four towns, divided by gender](image)
The ANOVA and t-test results above confirm the results from the linear mixed-effects models, which had showed that only the speakers from towns Z and M exhibited some perceptual sensitivity towards the modified VOT durations. However, town Z also has a stronger $p$-value than town M, which is a clue that the imitation of town Z speakers was much clearer than that of town M speakers, as was their sensitivity to the modifications. Even so, town M is still relatively stronger than towns N and I.

4.9.2 Summary of VOT durations

The overall statistical durations in all towns (Figure 4-5) showed that the best imitation of the three VOT levels occurred in town Z, whose participants produced significant differences between each level pair. Town M participants, meanwhile, gave a good overall imitation of only two level pairs and towns N and I failed to achieve any significance in imitation results across the three levels.

When we look at the VOT durations as per gender in each town (Figure 4-6), we notice that each gender in town Z can distinguish two of the three level pairs (Level 1/Level 2 and Level 1/Level 3 among females, Level 1/Level 3 and Level 2/Level 3 among males). Town M participants, on the other hand, were able to differentiate one level pair (Level 1/Level 3 among women, Level 1/Level 2 among men). And, finally, the gender-based results for towns N and I match the overall results for those towns: namely, that the participants there did not show any distinct imitation patterns.

---

71 For a detailed look into each individual participant’s three-level VOT responses in the four towns (Z, M, N, and I), see Appendices C.1, C.2, C.3, and C.4, respectively.
4.10 **Differences in VOT duration for each obstruent**

This section will discuss variations in the VOT durations of the allophonic sounds [p], [t], [k], and [tʃ] across the three levels in the four towns, as produced by the two genders. In doing so, I will examine differences that may occur within the female or male groups of each town, not between different gender groups.

4.10.1 **Differences in town Z by obstruent and gender**

Figure 4-7 shows the within-gender durational distributions for the four words and target sounds (pope for /p/, tote for /t/, koke for /k/, and choche for /tʃ/) at the three different levels in town Z.

![Town Z by Sound and Gender](image)

**Figure 4-7 Within-gender VOT differences for the four words in town Z**

For /p/, the VOT durations across the three levels (Level 1 to Level 3) were not significantly different for either females or males ($p = 0.58$ and $p = 0.062$, respectively). A further between-levels comparison did not show any significant differences among females.
Males, however, had significant differences between both Level 1 and Level 2 \((p = 0.037)\) and Level 1 and Level 3 \((p = 0.03)\).

For /t/, the overall differences in VOT duration across the levels (Level 1 to Level 3) were significant only among males \((p = 0.037)\). Moreover, males had a significant difference in their VOT durations between Level 1 and Level 3 \((p = 0.013)\). Meanwhile, a between-levels comparison in the data for females showed that their VOT durations for /t/ were significantly different between both Level 1 and Level 2 \((p = 0.029)\) and Level 1 and Level 3 \((p = 0.028)\).

For /k/, the overall differences in VOT duration across the levels (Level 1 to Level 3) were significant only among females \((p = 0.0028)\). However, level-wise within-gender comparisons showed that the VOT durations between Level 1 and Level 3 differed significantly among males as well \((p = 0.025)\). Among the females, the differences in VOT duration were significant between both Level 1 and Level 2 \((p = 0.0063)\) and Level 1 and Level 3 \((p = 0.0011)\).

For /tʃ/, the overall differences in VOT duration across the three levels (Level 1 to Level 3) were insignificant for both females \((p = 0.42)\) and males \((p = 0.13)\). Furthermore, between-levels comparisons showed that neither females nor males had any significant differences in their VOT durations between individual pairs of levels.

Thus, we can conclude that both females and males in town Z showed consistently significant variation in their VOT durations for /t/ and /k/, but mostly between Level 1 and Level 3. In the case of /p/, only males showed a significant difference, and neither gender exhibited any significant variation for /tʃ/. 
4.10.2 Differences in town M by obstruent and gender

Figure 4-8 shows the within-gender distribution of VOT durations for the four words with the examined obstruents (pope for /pl/, tote for /t/, koke for /kl/, and choche for /tʃ/) at the three different levels in town M. Overall, only females showed significant differences in their VOT durations for /t/ ($p = 0.032$), and no significant durational differences were found for /pl/, /kl/, or /tʃ/ among either gender.

As we can see in Figure 4-8, females showed an insignificant overall difference in their VOT durations for /p/ ($p = 0.22$), while males also showed no significant difference ($p = 0.93$). When the VOT durations were compared across the levels, no significance was observed for either gender.

For /t/, only females showed significant overall differences in their VOT durations. Between-levels comparisons showed that both females and males had significant differences between Level 1 and Level 2 ($p = 0.036$ for females, $p = 0.039$ for males), but only females exhibited a significant difference between Level 1 and Level 3 ($p = 0.013$). For /k/ and [tʃ], significant within-gender differences (at $p < 0.001$) were not present between or across the three levels.
4.10.3 Differences in town N by obstruent and gender

Figure 4-9 shows the within-gender distribution of VOT durations for the four words (pope for /p/, tote for /t/, koke for /k/, and choche for /tʃ/) at the three different levels in town N.

Overall, females showed a significant durational difference only in the case of /t/ ($p = 0.011$). For the three other obstruents, neither gender showed any significant differences.

Specifically, there were no significant differences in VOT duration for /p/, /k/, or /tʃ/.

I conducted level-wise comparisons using a series of t-tests, but the results showed that there were no significant differences between any two levels (Level 1 and Level 2, Level 1 and Level 3, Level 2 and Level 3) for the aforementioned obstruents among either gender.

In the case of /t/, only females showed a significant overall difference in their VOT durations. Further examination showed that this fact is due to the significant durational differences between Level 1 and Level 2 ($p = 0.031$) and Level 1 and Level 3 ($p = 0.0039$).

Thus, we can say that in town N, only females exhibited variation in the VOT durations of /t/.
4.10.4 Differences in town I by obstruent and gender

Figure 4-10 shows the within-gender distribution of VOT durations for the four words (\textit{pope} for /p/, \textit{tote} for /t/, \textit{koke} for /k/, and \textit{choche} for /tʃ/) at the three different levels in town I. Overall, no significant durational differences were found for any of the four words within either gender group.
Specifically, there were no significant differences in VOT duration for /p/, /t/, /k/, or /tʃ/. I conducted level-wise comparisons using a series of t-tests, but the results showed that there were no significant differences between any two levels (Level 1 and Level 2, Level 1 and Level 3, Level 2 and Level 3) for the aforementioned obstruents among either gender.

4.10.5 Summary of VOT durations for each obstruent

For the imitated bilabial /p/, the only significant difference I found was among the male speakers of town Z. The alveolar /t/, meanwhile, was the most accurately mimicked phonetic cue: namely, among both genders in towns Z and M and the females of town N. When analyzing the velar /k/, I found that only participants from town Z (both males and females) were able to imitate and trace the modified durations, and only between one pair of levels: namely, Level 1 and Level 3. And, finally, the statistics for the post-alveolar /tʃ/ showed no significance at any of the four locations.

These results show that town Z participants, who had the strongest ability to mimic differences in many cases, succeeded in tracing the differences among the stops (/p/, /t/, and /k/), but could not trace the different levels of the affricate /tʃ/.

4.11 Between-gender VOT differences in all towns

Figure 4-11A shows a comparison of VOT durations (in ms) between females and males at the three different levels in town Z. Overall, a gender difference was found at Level 3, where males had longer VOT durations than females.

At Level 1, despite the fact that the males’ median VOT duration was higher than that of the females, the difference was not statistically significant ($p = 0.11$). At Level 2, meanwhile, the median VOT durations for both males and females were the same. A
difference in the *range* for both genders’ VOT durations at Level 2 was found; however, this difference was not statistically significant (*p* = 0.21). Lastly, a significant between-gender difference in VOT duration was found at Level 3. Based on this, we can conclude that the town Z females and males differed significantly in their VOT durations only at Level 3 (*p* = 0.0075).

![Figure 4-11 Differences in VOT durations between the genders at the three levels, divided by town](image)

A further examination of the *town M* data, as shown in Figure 4-11B, revealed an overall between-gender difference at Level 3, where females had longer VOT durations than males.

At Level 1, even though there was a clear difference in the range of VOT durations between females and males, this difference was not statistically significant (*p* = 0.13). At both Level 1 and Level 2, the durational range was larger for females than males; however, no statistical difference was found (*p* = 0.1 at Level 2). Unlike Level 1 and Level 2, however,
there was a significant difference at Level 3 ($p = 0.00017$). Thus, we can deduce that gender differences in VOT durations were salient in town M only at Level 3.

Different from what was found in towns Z and M, town N’s VOT durations (in ms) proved to be significantly different between females and males at the three levels. Overall, a between-gender difference was found at all three levels, where females had longer VOT durations than males.

As the above chart shows (Figure 4-11C), the between-gender difference at all three levels (Level 1, Level 2, and Level 3) was statistically significant (at $p < 0.001$). Thus, we can say that in town N, the females consistently had longer VOT durations than the males.

In contrast to the above towns, town I’s imitations did not reveal any significance at all. Overall, no difference between the genders was found at any of the three levels. Figure 4-11D shows that the females and males did exhibit variation in the ranges of their VOT durations at all three levels. However, despite this variation, the application of a t-test did not highlight any significant differences at Level 1 ($p = 0.49$), Level 2 ($p = 0.25$), or Level 3 ($p = 0.19$). Thus, we can conclude that between-gender differences in VOT durations were not salient in town I.

4.11.1 Summary of between-gender differences in VOT

When looking at between-gender VOT differences in the four towns, we can see that towns Z and M observe similar patterns in that there is a significant difference between the two genders only at Level 3. However, there is no consistent pattern with respect to which Level 3 duration is longer: the data show longer VOTs for the men of town Z, while the longer VOTs in town M were produced by women.
Towns N and I, meanwhile, were quite different from towns Z and M but also quite different from each other. While all three levels revealed significant gender differences in town N, town I exhibited no statistical difference between the two genders at any of the three levels.

4.12 Gender differences in VOT duration when F0 is imitated

Figure 4-12 below presents a comparison of the modified F0 tokens’ VOT durations (for studying VOT duration when F0s are imitated) among females and males. Overall, neither gender showed significant differences in VOT duration at any of the three levels.

In town Z, an application of ANOVA showed no differences in VOT durations among either females or males when three different F0 levels were heard ($p = 0.99$ and $p = 0.58$, respectively). Furthermore, when the three levels were compared to one another, no significant differences were found for any group of respondents (females or males).

Similar observations were made for town M: no significant differences in VOT durations across the three levels were found for females ($p = 0.85$) or males ($p = 0.69$). Furthermore, a series of t-tests showed that between-levels differences were also insignificant for both females and males.
Within-gender differences in the VOT durations of imitated F0 tokens, divided by town

In town N as well, neither females nor males showed significant differences in their VOT durations at the three given levels ($p = 0.89$ and $p = 0.87$, respectively). Then, the application of t-tests did not show any between-levels differences, either.

Lastly, an ANOVA was run to examine within-gender durational differences at the three levels in town I, but no significant differences were found for females ($p = 0.87$) or males ($p = 0.99$). Moreover, t-tests did not highlight any between-level differences, either. Thus, we can conclude that both level differences and within-gender differences were nonexistent across the three levels when participants were imitating the F0 manipulations.

In sum, the F0 manipulations did not play any role in participants’ VOT productions across the three levels. That is to say, when a listener-turned-speaker was exposed to a modified pitch, the produced VOT remained the same, with no effect on its duration.
4.13 *A short summary of the imitation experiment*

In this test of imitation, six different modified VOTs and F0s (three each) at the beginning of nonce words were examined to see how Swahili speakers in four different locations (towns Z, M, N, and I) would mimic the changes. An analysis using linear mixed-effects models provided a strong confirmation that the phonetically most imitated cue was VOT, and among town Z participants in particular. F0 of all towns, on the other hand, was seen to play a minor role, treated as marginal compared to VOT in towns M, N, and I. Although participants in the last three towns were seen not to be perceptive of the spliced VOT durations, all towns – with no exceptions – showed no effective variation in F0, which, in turn, did not influence any change in the four examined obstruents.
Over a half century ago, Polomé (1967) claimed (on the basis of a non-instrumental method) that the laryngeal contrast in Swahili voiceless obstruents had been gradually disappearing for decades from the productions of L1 speakers, such that only a small number of mostly well educated people continued to articulate aspiration, albeit in arbitrary and unsystematic ways. To date, there has been little empirical work on VOT values in Swahili; the only studies we know of with regard to variations in Swahili are those of Engstrand & Lodhi (1985) and Moxley (1992). Based on some recordings that were obtained from native and non-native speakers of Swahili, Engstrand & Lodhi (1985) asserted that there is no strong evidence that the aspiration is in decline or being reduced. The only support in the two studies for this claim that aspiration is an active process comes from an analysis of data from one informant from Zanzibar Southern Island (one participant in each study), which offers very limited quantitative evidence that minimal pairs show a divergence based on either an aspiration contrast (146 ms for paa ‘gazelle’ and 13 ms for paa ‘roof’; Moxley, 1992: 70) or word origin (66 ms for the native karo ‘washing-place’, but 33 ms for the non-native variant [from Arabic] ‘fee’; Engstrand & Lodhi, 1985: 185). Therefore, the present study was performed in order to determine how VOT is categorized and perceived when L1 and L2 speakers of Swahili produce real words or imitate nonsense words. Note, too, that the previous literature generalized VOT facts from a few speakers to the general Swahili-speaking population as a whole, whereas the present study reports on the status of the laryngeal contrast in question in disparate locations, a point I will return to below.

Maddieson (1984) reports Swahili as making contrastive use of aspiration in its voiceless stops and affricate, a claim that he adopted from Polomé (1967), which had been claimed earlier by Tucker & Ashtone (1942). These earlier claims hold true of the phonemic
inventory of Swahili in town Z (the *KiUnguja* dialect), but they were not in line with the Inter-Territorial Language Committee of the 1930s, which treated the aspiration as a linguistically irrelevant characteristic that, as a consequence, was marginalized and ignored in the standardized Swahili orthographic system. The latter treatment matches, to some extent, the phonetic observations made in the present study for towns M, N, and I. This all goes to show that adopting a common phonemic system in the literature, and a common orthographic system beyond, may only work for a specific (small) territory, rather than for the entire region where L1 and L2 speakers reside.

Before focusing on the first research question — whether Swahili obstruents make a two-way or a three-way laryngeal contrast — it should be admitted that no consensus exists about laryngeal features in the larger phonemic inventory of Swahili. The latter has been posited to have a three-way laryngeal contrast among voiced, unaspirated, and aspirated stops (Engstrand & Lodhi, 2004), or else a three-way aspiration contrast with voiced implosives rather than plosives (Tucker & Ashtone, 1942). Polomé (1967), for his part, suggested another view of the inventory by presenting the voiced stops as allophones of their implosive counterparts. Thus, those who believed in a three-way laryngeal contrast (no matter if the voiced sounds were plosives or implosives)\(^2\) did not disagree that the voiceless sounds are distinct. However, the present work has improved our understanding of Swahili’s voiceless obstruents, by proving that there is a contrast between aspirated and non-aspirated segments, but only in one specific location (town Z). The issue of which voiced phonemes are preferred in the language, whether plosives or implosives, requires more exploration, which is beyond the scope of the current discussion and not among the core research questions\(^3\).

\(^2\) In analyzing Sindhi as a five-way laryngeal language, Cho et al. (2019) included the bilabial implosive /ɓ/ in the contrast.

\(^3\) I did gather some data on this question but it is left for future research.
Building on the scant evidence of Swahili phonetic variations that I obtained from earlier linguists (Engstrand & Lodhi, 1985; Moxley, 1992), the second question of this study is centered on the degree to which the aspiration contrast is implemented. The results of the study indicate that not all Swahili speakers’ language has the same phonological features or identical phonetic properties. As mentioned in the background section on Swahili orthography, the major spoken dialects of Swahili are categorized into four groups: *KiUnguja, KiMvita*, *KiAmu*, and *KiNgwana* (Marshall, 2015). However, it was not my initial plan to focus on these varieties of Swahili in particular, as I was concentrating on locations rather than dialects. Moreover, as we know from previous rough categorizations, the island and coastal dialects (the first three of Marshall’s classification) belong to people who mostly claim them as their L1, while the last one maintains a different status as being spoken far away from the east coast of Africa (Eastern Congo) with no available evidence that it is spoken widely as an L1.

In addition, my research goal was to understand how L1 and L2 speakers of Swahili produce and perceive the variations based on aspiration contrast or word type. The four locations of interest were intended to sample from a wider geographical area, rather than just focusing on the area around the coast and surrounding islands. This approach was meant to facilitate a categorical examination of L1 and L2 speakers from various Swahili backgrounds and linguistic heritages. Based on this objective, strong evidence as gained from the results of this study tells us that not all of the L1 spoken dialects represent similar linguistic facts. *KiUnguja*, the dialect of town Z, preserves more conservative properties both phonologically

74 *KiMvita* (spoken in Old Town Mombasa; Hayward et al., 1989: 51) is the L1 for most of my Mombasa speakers. The consent forms signed by participants did not ask them to state their spoken dialect, but Dr. Swaleh, who facilitated the data collection in Mombasa, confirms (personal communication, July 21, 2020) that “most” of my subjects are *KiMvita* speakers, meaning that I may have recruited some *KiAmu* speakers, too.

75 *KiMrima* is also a variety that is spoken along the coast of Tanzania (Samson, 2015; Githinji & Njoroge, 2017). According to Polomé (1967), this dialect shares some features with *KiUnguja*.
and phonetically than *KiMvita*, as spoken in town M, or *KiAmu*, the dialect of Lamu Island and the coast that faces it. Thus, it seems logical that if the L1 dialects are different, then the L2 dialects are different too. Town N, which represents Swahili as the L2 of *Kikuyu* speakers, and town I, which gives a sample of L2 Swahili from various native linguistic backgrounds, may just be speaking a variant that has a different system of phonemic contrasts than that was seen in town Z. L2 speakers are not necessarily exhibiting some sort of grammatical deficit.

With respect to the third research question about which phonetic detail speakers of the language use to signal a contrast between obstruents in Swahili, the results reveal that the two phonetic cues, voice-onset time (VOT) and fundamental frequency (F0), which were the main two properties I observed in the 1973 UCLA Swahili archive, did not have the same linguistic power. Analyzing VOT, which was first cited in Lisker & Abramson (1964), proved to be the realistic approach to revealing the variability in the spoken Swahili dialects: Swahili speakers (of town Z) rely on the temporal phonetic detail (VOT) to express the language’s contrast. By contrast, the role of spectral phonetic cue (F0) is minimal, if not null. By tracking both VOT and F0 in Swahili, the present study heeded Cho, Whalen, and Docherty’s (2019) suggestion that an examination of a laryngeal contrast ought to include more than one phonetic dimension. Since VOT proved to be a stronger cue for the obstruent contrast in Swahili, I will discuss the primary cue (VOT) results first and the secondary cue (F0) results second.

The answer to the fourth question, about gender differences, is that VOT, as a phonetic instrument, can be used as a differentiating factor across the spoken dialects in the

---

76 One can expect the L2 dialects to be influenced by the speakers’ L1s, independently of whatever L1 Swahili speakers are doing.
different locations. The results provided acoustic evidence that sex has an impact on the produced VOT. What is interesting about my findings is the statistical significance of temporal duration when the women have longer (automatic) VOTs (in towns M, N, and I), which is in line with earlier studies on voiceless stops in American English (Ryalls et al., 1997; Robb et al., 2005; Herd, 2020) and British English (Whiteside & Irving, 1998). On the other hand, when the men produce longer (controlled) VOT durations (in town Z), the difference between men and women turns out to be insignificant, which matches Oh’s (2019) findings on the voiceless stops of American English.

The results regarding the phonetics of native words and loanwords — research question 5, which was prompted by Engstrand & Lodhi’s (1985) observation of varying VOT in a minimal pair — provide a solid confirmation that the nature of adaptation can be tracked through VOT measurements. Through the process of picture naming, town Z participants were able to give the native words and loanwords (from Arabic and English) three distinct categorizations: native Swahili words had an intermediate duration between the long English borrowings and the short Arabic ones. Participants in the other locations (towns M, N, and I), on the other hand, produced native words with a longer duration than the loanwords. Across these three towns, the durations of loanwords from English and Arabic were mostly similar. The present production data, based on word origin (native vs. Arabic only), show that at least some speakers can produce native words with higher VOTs than loanwords from Arabic (75% of participants in town Z, 38% in town M, 29% in town N, and 41% in town I). As for the difference between loanwords from Arabic (shorter VOTs) and English (longer VOTs),

---

77 In localities (towns M, N and I) whose VOT gender differences were statistically significant, VOT differentiates the earlier dialects from the dialect of town Z. Due to these VOT gender differences, male and female groups had different lengths for the four obstruent types.

78 The only exception to this overall pattern of loanword categorization was the group of town N females, who were able to produce three distinct durations as per word origin: higher for native, intermediate for English, and lower for Arabic.
the percentage is even higher for town Z\textsuperscript{79} but lower for the others (92% in town Z, 8% in town M, 12.5% in town N, and 4% in town I).

Recall that Ito & Kenstowicz (2013) found a shift in the VOTs of Japanese loanwords in Korean in the later period of Japanese as compared to the conservative old period. Likewise, the phonetic implementations of the loanword type and the overall aspiration contrast were quite different in the present study. The aspiration contrast, as a native property, was maintained and preserved maximally in town Z because of the island’s location as being more isolated than the mainland. This laryngeal contrast\textsuperscript{80} was lost further in the inland towns, presumably due to two reasons: the fact that there is minimal contact between the islanders and the coastal inhabitants, and the fact that the latter are in closer contact with other (Bantu) languages that do not use aspiration contrastively.

In the opposite pattern, the loanword types were dealt with in all towns with some preference, but just as I found for the aspiration contrast, the clearest contrasts in loanwords was among town Z participants, who maintained a phonetic contrast for all three loanword types\textsuperscript{81}. Thus, there is no doubt that VOT, as a primary cue, has a high degree of acoustic power in the island dialect, but there is still a puzzling issue in the VOT distinction that participants from the inland towns drew between native words and loanwords. At this time, my only thought on this state of affairs is that the loanwords are assigned to a category,

\textsuperscript{79} However, there is no evidence that the high percentages and longer VOTs for English loanwords (at least in town Z) result from the fact that aspiration plays a major role in English voiceless stops and these speakers of Swahili may be aware of that fact. The town Z participants’ English proficiency is way lower than that of the other locations.

\textsuperscript{80} If we assume that the aspiration contrast was always there to begin with as claimed for Mombasa by Polomé (1967) and Frankl (1991).

\textsuperscript{81} It is unclear, however, why the first male and the first female speaker of town Z (both from Zanzíbar’s northern island, Pemba) failed to show any variabilities based on the category of Loanword. I do not really recall the recording session with the female speaker (ZS1F), but I still remember that the first male speaker (ZS1M) spent quite some time trying to convince me in an articulatory way of how the unaspirated \textit{paa} ‘roof’ is not the same as the aspirated \textit{paa} ‘deer’. Although these two individuals were among the town Z subjects who managed to show that the KiUnguja dialect preserves the aspiration feature phonemically and acoustically, their responses and data were not able to demonstrate that the word-type origins are phonetically different.
separate from the native-word category, and the different VOT associations of these categories represent phonetic arbitrariness\textsuperscript{82} (see Cho, 2015).

Like the production results, which demonstrated frequently that town Z is very different from the other towns, the imitation results showed more clearly delineated contrasts for town Z as compared to the others. This is in line with Caramazza et al. (1973), who found that the productions of French-Canadian English bilinguals surpassed their perception, meaning that although French/English bilinguals were able to separate their French and English systems in terms of VOT contrasts in production, those contrasts seemed to merge in the bilinguals’ perceptual systems\textsuperscript{83}. Similarly, town Z participants did a clearer job of imitating nonsense words’ VOTs with some significance, but the results were not at the level of the production of real words. Again, sensitivity to the acoustic cues of native words/loanwords is crucial to the perception of phonetic differences among fake words, and this further enhanced the differences’ realization during the unfamiliar contrast experiment (the imitation experiment).

Note that it is not the case that participants outside town Z had no sensitivity to VOT contrasts at all; at some point, speakers in towns M and N were able to mimic some of the modified VOTs. At the phoneme level, the alveolar /t/ was easier to imitate relative to the other obstruents. Post-alveolars were the hardest to imitate, and so none of the locations were able to differentiate any of the three VOT levels in post-alveolars. With reference to the levels of VOT, stronger significance is gained the longer the durational difference between levels is: that is, the difference between Level 1 and Level 3 caused a bit less confusion than

\textsuperscript{82} The Swahili speakers are just looking for some way to mark the loanwords as somehow “different” from native Swahili words.

\textsuperscript{83} Caramazza et al. (1973: 427) stress that “the perceptual functions in the two language sets were not very different from each other: both curves had similar shapes, steep yet nonmonotonic, and both showed perceptual crossover points at positions intermediate to the UF and UE functions”.

150
that between Levels 1 and 2, or between 2 and 3. Regarding the overall gender-wise imitation, female speakers demonstrated more perceptual control over the manipulated stops that they heard: they succeeded in mimicking Level 1 with statistical significance, as compared to Levels 2 and 3, for /t/ (in towns Z, M, and N) and /k/ (in town Z).

At the town level, the men of town Z distinguished between the three levels of the primary cue VOT in their imitations for all the stops more consistently, but only between Level 1 and Level 3 (with the exception of /p/, for which they were also able to distinguish between Level 1 and Level 2). One fact that sets town M apart from the other mainland towns (N and I) is the statistical imitation of /t/ between Level 1 and Level 2: town M participants, who are presumably native Swahili speakers, were better at imitating /t/ compared to those from towns N and I. It is evident, therefore, that the mimicking task was more robust and consistent in town Z (island) and showed a bit less consistency in town M, but still clearer by far than what was observed for the other towns on the mainland.

So far, we have seen that VOT is the phonetic cue of the most interest in both production and imitation is VOT; the secondary cue (F0) was not an influential phonetic detail in the present study. It is not clear which location or dialect the 1973 recordings in the UCLA Phonetic Archive (Speaker 2) belong to, but as far as my results go, I can say with confidence that any aspiration contrast based on that speaker’s F0 is a (somehow) phonetically irrelevant spectral cue. Since this speaker’s origins are unspecified, it is impossible to know if she is an L1 or L2 speaker of Swahili and whether the pitch contrast she applied to the Swahili minimal pairs is a lost or reduced L1 acoustic property or a feature inherited from another Bantu mother tongue. However, the other speaker in the UCLA Archive was from Mombasa (Speaker 1), and her productions were based on VOT rather than F0. According to the results here, phonetically contrasted minimal pairs, with reference to
VOT, were nowhere to be found for any of the participants in town M (except one: MS13F)\textsuperscript{84}.

Thus, my VOT results cast doubt on the assumptions of earlier linguists that aspiration has declined (Polomé, 1967) or been lost (Frankl, 1991; Wald, 2009) in the south\textsuperscript{85}. My results also strongly contradict the idea of a living contrast in the north (Frankl, 1991) and further argue against considering L1 speakers from the north and the south as sharing the same linguistic experience (Engstrand & Lodhi, 1985). If one of the two main dialects in the north (KiMvita or KiAmu) maintains any contrast, then we would have seen this in the individual responses (as reported in the appendices for every specific informant, see Appendix B-13), but as was mentioned above, only one town M participant showed any competence in expressing the contrast. Despite the lack of information regarding the participants’ exact L1 dialects, one informant (MS8M) was from Lamu, where KiAmu is spoken. Unfortunately, this speaker gave similar responses for both aspirated and unaspirated tokens with a very small interquartile spread in both datasets. Moreover, it is hardly clear where the difference comes from when we look at his two boxplots; the whiskers are nearly identical, too.

Returning to the F0 issue, specifically how this phonetic cue was used if the tokens have an aspiration contrast or different word origins, the results proved that F0 is an active secondary phonetic detail (that is, when looking at the individual responses, F0 is relevant in some specific parts of the analysis). For the aspiration contrast, F0 was chosen by some individuals at some point to signal a contrast in their production. When we examine the whole data set to see how phonation is different, F0 rates were significantly higher in towns

\textsuperscript{84} It could be the case that the only Mombasa female speaker in my data and the UCLA Mombasa speaker had (a) conservative dialect/s.
\textsuperscript{85} The claim that the aspiration contrast is disappearing in the south was not even there to begin with.
M and Z (by 58% and 42%, respectively). That is to say, 15 out of 26 participants in town M did show, significantly, that aspirated tokens had higher F0s than the unaspirated ones (and 10 out of 24 in town Z). Towns I and N, by contrast, were not in favor of this acoustic feature, and the rates dropped drastically (to 8% in town I, two informants, and 4.2% in town N, only one). The only case in which F0 was statistically higher for unaspirated tokens was recorded for the first male speaker in town Z (ZS1M), and this is not expected to happen (Kirby, 2018; Kwon, 2015). Other than this speaker, no surprising assignment of a higher F0 value to unaspirated sounds was seen anywhere in the data (see Appendices D-1-D-4). That is to say that the unspecified speaker (Speaker 2) from the UCLA Archive, who, as mentioned earlier, made an aspiration contrast based on F0 in the 1970s, implemented a phonetic cue which has been found among some native speakers of the language 45 years later.

F0 has also shown a phonetic effect on loanword types, and this is further evidence that F0 is a secondary acoustic cue in the productions of native Swahili speakers (those from towns Z and M). With reference to the percentages of speakers in my study, the overall ANOVA difference for all word types was significantly higher for towns M and Z (50% and 54%, respectively) than for towns N and I (33% and 25%). At the word-type level, the difference between English words and Arabic words was higher for M and Z (61.5% and 50%) but lower for I and N (29% and 25%). When we compare English words to native words, town Z’s ANOVA difference was higher (33%) than those of towns M, N, and I (15%, 12.5%, and 17%, respectively, see Appendices D-5-D-8). However, town M had significantly more participants who were able to differentiate, in an articulatory way based on F0 alone, the native words from the Arabic ones (42%) than town Z (25%), town I (21%), and town N (12.5%)86. Thus, the production of tokens by relying on F0 for both the aspiration

86 It is a little tricky to say that this cue might be operative in a variety of a language when only a maximum of 42% of speakers (or less) are using it.
contrast and the word-type contrast was best in town M, not quite as good in town Z, but minimal in towns N and I (see Appendices D-1 to D-8).

Therefore, the results so far suggest a relationship between location and the phonetic properties of the language. Choosing one place to examine the phonemes acoustically can yield simpler results than selecting random speakers of the language from different regions. When I first worked on this project, the two pilot-study participants claimed to be native speakers of the language, but no trace at all of the aspiration contrast was seen in their responses. Then I was faced with a literature dilemma: Swahili acoustic studies are scarce and do not give a straightforward and confirmed answer as to where the phonetic variations can be found. From the current study, which examined Swahili speakers in different locations, I have gained supplementary knowledge: not only that being an L1 speaker matters, but also that the level of sensitivity to the perceptual cues is dependent on a person’s place of birth. Aside from the island and the coastal area, the other locations also provided a valuable piece of evidence that men and women do not produce similar VOT values. Recruiting both genders was an advantage, as it enabled me to report the gender differences as found in the data to those who are interested in this specific phonetic property, which, until the present work, was understudied in Bantu languages in general and Swahili in particular.

Unlike the other towns, town Z is on an island and therefore under less linguistic pressure from neighboring dialect/language environments. Town M, on the other hand, as a coastal area, geographically less isolated and culturally less conservative, and its very strong exposure to the outside world, added to contact with many ethnicities, may have forced some linguistic characteristics to disappear gradually. Growing literacy in the simplified orthography, which ignores aspiration, is not causally responsible for the loss of aspiration in speakers, but it is worth noting that it does not help to preserve this laryngeal contrast, either.
In a similar vein, Hill (2007) alleges that a sound change in Standard Tibetan may be due in part to the inconsistency of the writing system, in addition to the competition caused by the spoken dialects of neighboring regions.

Because the two Results chapters did not comment in detail on the mean VOTs of the obstruents’ tokens, this discussion will elaborate more on these mean values and explore what else they can tell us (rather than the mean values of F0, which is the secondary less influential phonetic cue\(^{87}\)). I will look at both production and imitation, describing the “clouds” of mean values for the different places of articulation, in order to see if they abide by the usual (automatic) place of articulation hierarchy and find out how a contrast in either aspiration or the origin of words would lead to any differences. As we move from one discussion to another, we will gain more knowledge that Gender, as a predictor and an explanatory factor, is not marginal when it comes to the VOT durations.

5.1 Production discussion

Here I will discuss the distributions of the obstruents’ VOT values and show how their clouds will look in that acoustic space. When I created the charts below, a clear division was drawn between those for the aspiration contrast and those for the loanword types. The data regarding the aspiration contrast were split into four charts (one chart, with all four obstruents, for each town). After that discussion, word origins will be discussed for each obstruent with one chart for both genders, going from the bilabial /p/ to the alveolar /t/ to the velar /k/ to the post-alveolar /tʃ/.

\(^{87}\) Although F0 did exhibit some significant variations, especially in towns M and Z, these variations did not have the same statistical significance as those of the VOTs produced by town Z participants.
5.1.1 Production discussion (aspiration contrast)

The town Z participants’ unaspirated obstruents (Chart 5-1 below) exhibited a consistent aspiration contrast: all four unaspirated obstruents had significantly shorter VOT means than their aspirated counterparts. From the arrows on the chart, it is very evident that the men and women of town Z used different phonetic cues for signaling the contrast between aspirated and unaspirated voiceless obstruents. The unaspirated (overall) means among town Z participants were as follows: 23 ms for the bilabial /p/, 37 ms for the alveolar /t/, 47 ms for the velar /k/, and 74 ms for the post-alveolar /tʃ/. The aspirated means, on the other hand, were 76, 73, 91, and 94 ms for /pʰ/, /tʰ/, /kʰ/, and /tʃʰ/, respectively.

![Chart 5-1 The overall distribution of the aspiration contrast in town Z’s obstruent VOTs (each arrow targets the other cloud for the same place of articulation/gender)](image)

---

88 Gender-wise comparison of VOT duration is basically the same, although the female speakers generally have lower VOT values for this location.

89 These are the mean values among both men and women when producing the unaspirated obstruents. I am reporting the averages here because town Z was the location with four non-significant differences between male and female VOT values (and three of some significance: namely, 0.01 except for /tʰ/, which was < 0.001). The other towns (M, N, and I) exhibited consistent significant differences between male and female VOT, always way above 0.001.

90 The mean gap between the aspirated and unaspirated obstruents was 53 ms for bilabials, 36 ms for alveolars, 44 ms for velars and 20 ms for post-alveolars.
Unaspirated bilabials, as produced by the two genders, were different by only a few milliseconds: 24 ms for males and 21 ms for females. According to this 3-ms difference, the /p/ productions of both genders were statistically similar. Furthermore, unaspirated alveolars were also longer for men (40 ms) than women (35 ms), with some significance between the two means (0.035). Moreover, unaspirated velars were longer for males (49 ms) than females (45 ms), with higher level of significance (0.012) than alveolars. However, the unaspirated post-alveolars of men and women were statistically similar (74 ms and 73 ms, respectively).

The aspirated clouds for town Z are different from the unaspirated ones in that the former start above 60 ms for both genders (above 70 ms if we consider the overall average for /tʰ/). In line with what has been observed for the unaspirated segments, the general pattern is that men tended to have longer durations than women. The aspirated bilabial VOTs among males (81 ms) were, on average, 9 ms longer than those among females (72 ms), and this difference proved to be significant (0.035). Similarly, the aspirated alveolars showed a significant mean difference (0.001) of 14 ms between men (80 ms) and women (66 ms). Conversely, however, neither the aspirated velars nor the aspirated post-alveolars showed any significant difference between men (94 ms and 92 ms, respectively) and women (87 ms and 97 ms, respectively).

We turn now to the hierarchy patterns across all four obstruents (as seen in Chart 5-1 above). Whether we take only the overall averages for the unaspirated obstruents, as mentioned above, or consider the productions of each gender in isolation, we notice that there is no violation of the usual place-of-articulation-based VOT hierarchy, following this pattern: Post-alveolar > Velar > Alveolar > Bilabial. This result is in line with the conclusions of
many studies (Lisker & Abramson, 1964; Cho & Ladefoged, 1999; Maddieson, 1996; Mikuteit & Reetz, 2007; and Hussain, 2018; among others).91

Unlike the unaspirated obstruents, which exhibited the usual place-of-articulation hierarchy, the aspirated phonemes produced by town Z participants showed no such hierarchy; this is shown by the 3-ms differences (overall average) between /pʰ/ and /tʰ/ and between /kʰ/ and /ʈʰ/. However, looking at the two genders independently revealed different hierarchies because of the durational difference between aspirated and unaspirated post-alveolars (18 ms for males, but 24 ms for females). That is to say that men’s aspirated obstruents had insignificant mean differences between bilabials and alveolars, on the one hand, and between velars and post-alveolars, on the other hand, following this pattern: Post-alveolar/Velar > Alveolar/Bilabial. When the responses of women are considered, the place-of-articulation pattern shows a slight difference between post-alveolars and velars: Post-alveolar > Velar > Alveolar/Bilabial.

Aside from this variability in the place-of-articulation hierarchical pattern, the aspirated VOT lengths showed a salient acoustic contrast, which surpasses the usual, irrelevant (or unimportant) pattern across places of articulation. That is to say, the phonation type of the segment had a stronger effect on the VOT value for the segment than the place of articulation did. For town Z’s people, the phonetic cue of aspiration is a vehicle for representing the phonemic contrast in the dialect. Thus, the VOT for every aspirated segment in the dialect of town Z was significantly greater than each unaspirated segment at the same place of articulation.

Quite different from the speakers from town Z, the participants of town M (in Chart 5-2 below) showed a significant phonation contrast across the two genders between two

---

91 For post-alveolar (affricate), see Hussain, 2018 and languages there.
phonemes only: the bilabials /p/ (22 ms) and /pʰ/ (28 ms)\textsuperscript{92}. The level of significance was seen to be stronger in males (0.001) than in females (0.025)\textsuperscript{93}. Other than this observation, no significance was found between the two phonemic counterparts for the same place of articulation, whether alveolar, velar, or post-alveolar. Thus, all non-bilabial phonemes showed no aspiration contrast, and the phonemes that were assumed to be distinct (unaspirated and aspirated) were seen to merge into one unaspirated phoneme. The difference between the alveolar values was only 2 ms for males and 3 ms for females, and while men’s velars exhibited no difference, women showed an insignificant increase of the (anticipated) aspirated /kʰ/ by 5 ms. Furthermore, unaspirated post-alveolars had a slightly greater VOT than their aspirated counterparts, a pattern that was seen among both men and women. So, earlier claims about awareness (and knowledge) of aspiration in the north—Mombasa in particular—as reported by early works (Polomé, 1967; Frankl, 1991; Wald, 2009), are not supported by the data from the current study. Thus, the generally insignificant, but still higher VOT durations of aspirated obstruents in this variety of the language were not consistent, and there were cases in which unaspirated phonemes yielded a longer VOT, as seen in the production of unaspirated post-alveolars (by both genders) and unaspirated alveolars (by male speakers).

\textsuperscript{92} The significance here is probably the result of the large number of stimuli included in my statistical test, because the effect size itself is quite small.

\textsuperscript{93} This level of significance was higher for males because of the small difference between their means and those of females. Men produced means of 21.56 ms for /p/ and 28.16 ms for /pʰ/, whereas the women’s means were 22.34 ms and 27.68 ms, respectively, with around a 6-ms difference (6.6 ms in men and 5.34 ms in women). Yet, we have to be cautious about this contrast because this overall level of significance is not (strongly) representative, especially among the female participants of town M, and this distinction could disappear when examining individual words.
Another clear pattern in town M’s observed VOT means was that women produced significantly longer VOT durations than men for alveolars, velars, and post-alveolars ($p$-values < 0.001). Hence, town M is similar to town Z in two respects. First, the aspiration distinction that its speakers exhibited in their bilabials resembles the similar pattern in town Z’s data. Second, in line with the unaspirated obstruents in town Z, place-of-articulation-based VOTs were seen to increase in town M as we move from bilabials to alveolars to velars to post-alveolars (the overall VOT averages as produced in town M for /p/, /pʰ/, /t/, /tʰ/, /k/, /kʰ/, /ɾ/, and /ɾʰ/ were 22, 28, 42, 43, 50, 53, 77, and 74 ms, respectively). Town M, however, is different from town Z in its consistently higher VOT durations among female participants, in addition to the documented loss of the aspiration contrast in non-bilabial obstruents.

Now that we have discussed the participants’ productions in towns Z and M, we ought to see if the participants in town N (Chart 5-3 below) have similar phonetic systems for the aspiration contrast, gender differences, and the place-of-articulation VOT pattern. When we look at the productions of town N participants, the results—in most cases—resemble the
responses we got for town M rather than town Z. With reference to aspiration, the two genders’ bilabials were statistically different from each other. However, two more differences in town N, which we did not see in town M, were that only the women’s unaspirated /k/ (62 ms) and aspirated /kh/ (70 ms) were significantly different (p < 0.002) and that all obstruents produced by women, with no exceptions, were significantly longer than the men’s.

Meanwhile, a phoneme-wise comparison of the aspiration contrast in alveolars and post-alveolars showed that the unaspirated sounds were relatively (but not significantly) longer than the aspirated ones. This pattern of insignificant, but still longer VOTs for non-aspiration was also seen in the unaspirated velar /k/ as produced by male informants.

![Chart 5-3](image)

**Chart 5-3** The overall distribution of the aspiration contrast in town N’s obstruent VOTs (each arrow targets the other cloud for the same place of articulation/gender)

The overall VOT averages as produced in town N for /p/, /ph/, /t/, /th/, /k/, /kh/, /ch/ and /chh/ were 25, 29, 42, 39, 57, 60, 82, and 78 ms, respectively. So, as we saw for the unaspirated obstruents of town Z and all of the obstruent phonemes of town M, the VOTs of town N followed the usual place-of-articulation hierarchy: Post-Alveolar > Velar > Alveolar > Bilabial.
So far, we have seen that town Z participants exhibited an aspiration contrast in all obstruents, with less-significant gender differences. In towns M and N, on the other hand, men and women had an aspiration contrast for bilabials only, and for the rest of the obstruents, males had significantly shorter VOT values than females. In Chart 5-4 below, the town I data show properties that are closer to those of towns N and M; nonetheless, town I is different in that it has a phonation contrast in its bilabials, alveolars, and post-alveolars. Aside from /pʰ/, which had higher VOTs than /p/, the contrasts were quite unexpected due to the fact that the aspirated /kʰ/ and /tʃʰ/ yielded significantly lower VOT values than their unaspirated counterparts. It is interesting to see how the males and females of town I exhibited the same phonation type contrasts, even though the latter (alveolars and post-alveolars) did not follow the same pattern as the former (bilabials). Regarding the place-of-articulation hierarchy, town I was in line with the usual sequence: Post-Alveolar > Velar > Alveolar > Bilabial. That is the case because the mean durations for the unaspirated obstruents /p/, /t/, /k/, and /tʃ/ were 28, 38, 59, and 93 ms, respectively.  

![Chart 5-4](chart.png)

**Chart 5-4** The overall distribution of the aspiration contrast in town I’s obstruent VOTs (each arrow targets the other cloud for the same place of articulation/gender)

---

94 Meanwhile, the averages for the aspirated segments in town I had some overlap between bilabials and alveolars (35 and 34 ms, respectively).
The gender differences in town I were identical to those that were seen in town N (the other town where the participants spoke mainland Swahili). In these two localities, all produced obstruents had statistically longer VOTs for women (and this observation is in the same vein of town M’s results, if both genders’ bilabials are eliminated from the analysis).

In summary, Chart 5-1 above shows that the mean VOT clouds for aspirated and unaspirated sounds proved, with no doubt, that a stable (and more reliable) phonemic aspiration contrast is found in town Z only. So, whenever Swahili’s phonemic inventory is mentioned in the literature (in regard to voiceless obstruents) as having a two-way laryngeal contrast, we know for sure that this is the case for the island (Zanzibar). With reference to the charts for towns M, N, and I (Chart 5-2, Chart 5-3, and Chart 5-4), on the other hand, most durations of the aspirated phonemes were not statistically different from those of the unaspirated ones (disregarding, of course, the inconsistent cases of aspiration in these three towns95). So it is apparent from these charts (for the coast and inland regions) that the voiceless phonemic system resulting from these regions is one in which there is no laryngeal contrast (i.e., there is one series of voiceless obstruents). Also, with the exception of town Z’s unstable gender-difference pattern, the gender differences were very significant, and generally the pattern (outside of Zanzibar) was that female speakers had longer VOTs than the male speakers. That is to say that specific VOT durations vary even among the dialects of one spoken language.

5.1.2 Production discussion (Loanword type)

In this part of the discussion, I will focus on the productions of obstruents in reference to word origin. This portion of the study examined /p/ and /ʃ/ in native Swahili words (S) and

95 That is by either having a significant longer VOT for only aspirated bilabials in towns M, N and I, or by demonstrating statistically higher durations for unaspirated /t/ and /ʃ/ in town I.
English loanwords (E), while the examination of /t/ and /k/ included native Swahili words along with loanwords from both English and Arabic (A). In this section, both males and females are grouped in the same chart so that we can compare and contrast easily. For each obstruent in the four charts below, I will comment on the variants (of males and females) according to the related phonetic clouds of that particular phoneme. More importantly, I will see if the variants from S, E, and A are similar or different.

Chart 5-5 below, for /p/, reveals that the male speakers in all the towns produced a similar S /p/, and all of the towns showed no statistical difference. However, when these speakers produced an E /p/, two set of clouds were created, which are significantly different. The most important observation here is that the S /p/ of town Z males is statistically different from the E /p/ as produced by the same group. The females, on the other hand, had a wider distribution of S /p/s, but this spread gets narrower when they produce E /p/s. The only two towns to show significance in the female data here are the following: town Z with very short VOTs for the S /p/ but longer ones for the E /p/, and – totally in the opposite direction – town N with longer VOTs for the S /p/ and shorter ones for the E /p/. From this overview of the clouds’ spreads, it is evident that the range across locations for /p/ starts at 17.5 ms and ends at 33.5 ms. Within this range, town Z assigns S /p/ words to an unaspirated /p/, but the E tokens’ mean looks as if it should be categorized as slightly aspirated.

96 I am not going to compare the two clouds produced by each gender for one obstruent because, when I tried this technique, the resulting chart was messy and confusing.
As we move from the bilabials to the alveolars, we see stronger significance. In this set of data (Chart 5-6), six variants of /t/ are evident for the males and females (three each). The town with the most significance among the three variants is town Z. In addition, the males and females of town Z reflected the same phonetic pattern by assigning the S, E, and A /t/s to three significantly different VOT durations. Among the males of towns N and I, meanwhile, the S and E /t/s were produced identically; however, these two were significantly different from the A /t/. The males of town M were the only male group to show no significance in this regard. The females, however, presented more variations than the males, and along with the significance of town Z’s /t/s across all variants (S, E, and A), as mentioned earlier in this paragraph, the females of town M produced the S /t/ in a way that was significantly different from the E and A /t/s. The women of town I, meanwhile, were just like their male counterparts in exhibiting no significance between the S and E /t/s, but these two were statistically shorter than the A /t/. The town N females, on the other hand, produced an S /t/ that was shorter than those of E and A. In general, consistency of production for the three variants of the alveolar /t/, for both genders, was found in towns Z and I.
the range of the /t/ clouds starts at 22 ms and ends at 64.5 ms. Since town Z is the one with
the most significant differences between the three word types, we can say that its A /t/ for
both genders was unaspirated, while its S /t/ was slightly aspirated and its E /t/ was aspirated.

Chart 5-6 The overall distribution of /t/ under the Loanword category (each arrow
targets the other cloud with the same color)

For the variants of the velar /k/ in the loanword category (Chart 5-7), all towns
yielded some level of significance except town I, in which both genders produced statistically
similar /k/ for all three types of words. The only difference between the men and women of
town I is that the men had shorter VOTs. It is also surprising that the other locations
conserved the same patterns for both men and women (but not necessarily the same VOT
values). For both genders in towns M and N, the S and E /k/s were statistically similar, but
different from the A /k/. The case of /k/ in town Z, meanwhile, is quite different. Although
the male speakers in town Z yielded three distinct VOT measurements, the S /k/ was
insignificantly different from the E /k/, but these two were very different from the A /k/. The
females of town Z, on the other hand, exhibited more variations, which proved statistically
that the three means VOTs of /k/ for S, E, and A are different. In addition, the overall
aspiration pattern is the same as what we saw for /t/: the men and women of town Z tended to assign the A /k/ to a non-aspirated category, the S /k/ to a slightly aspirated category, and the E /k/ to an aspirated category.

Chart 5-7 The overall distribution of /k/ under the Loanword category (each arrow targets the other cloud with the same color)

Chart 5-8, for the post-alveolar /tʃ/, indicates that participants in towns Z and I produced significantly different /tʃ/ sounds based on word origin. Specifically, town Z’s productions gave the S /tʃ/ shorter VOT durations and the E /tʃ/ longer ones, whether the participants were male or female, while town I did quite the opposite: assigned the S /tʃ/ longer VOTs and the E /tʃ/ shorter VOTs. According to the range of post-alveolar VOTs (from 64 ms to 106.5 ms), town Z’s S /tʃ/ can be treated as slightly aspirated, town Z’s E /tʃ/ and town I’s male E and S /tʃ/ as aspirated, yet the post-alveolars of the females in town I are definitely highly aspirated (also for both types). On the other hand, neither town M nor town N showed any significant differences in VOT. For both male and female participants in these two towns, if we disregard the slight, insignificant durational differences between the S and E word types, we
see that words of these origins carry the same phonetic detail. That is to say, they are not acoustically salient or distinct.

To sum up the loanword category, both genders in town Z had significantly shorter S /p/ VOTs and longer E /p/ VOTs, while the other towns and groups failed to show any significance except for the women of town N, whose S /p/ was longer than their E /p/.

Moreover, the alveolar /t/ for both genders in town Z was categorized into three statistically significant durations (short A /t/s, long S /t/s, and longer E /t/s). No difference was seen among the town M males’ /t/s, and the men of towns N and I had similar VOTs for S and E, but different ones for A. While the females of towns M and N had similar values for E and A, their S /t/ was significantly short. The women of town I, on the other hand, had similar S and E /t/s, but longer A /t/s.

For the velar /k/, among all towns’ participants (except for town I), S /k/s and E /k/s were different from the short A /k/s. The town Z women, however, maintained a further significance between S /k/ and E /k/, which the town Z men did not produce. Post-alveolar
VOTs, meanwhile, were not different for S /tʃ/ and E /tʃ/ in towns M and N, but significantly different in towns Z and I. However, while the S /tʃ/ was shorter among both genders in town Z, it was longer for both men and women in town I. Thus, basically, the consistent patterns we see across the various places of articulation are in town Z. The Z speakers have this hierarchy for VOT: E > S > A.

5.2 Perception discussion (imitation)

The findings from the imitation portion of the study indicate that even speakers of one language can demonstrate widely different linguistic patterns and that the distinct representations even extend to the level of between-gender variabilities. Here, we have set a division between the female and the male F0 ranges because of the physiological differences in producing fundamental frequency (higher for women and lower for men). When we consider the F0 levels (1, 2 and 3) in each particular town, both males and females show no difference, so they do not adjust their F0 to the different levels they heard in the stimuli. However, the production of F0 in general is different between towns. Since we have seen no significant gender differences in imitated F0 levels within each individual town, we will focus on the variabilities that were observed across towns.

The reality of imitated F0 manipulations (in Chart 5-9 and Chart 5-10) visualizes the production mismatch across genders and towns. The imitation of F0 by women, as can be seen in Chart 5-9, showed close patterns (by town) in the three levels of F0, with no level of significance observed. Town I females had the lowest F0 imitations, followed by town Z females and then town N females, while the highest F0s are in the productions of town M females. Moreover, for all female levels in all towns, t-tests were run to evaluate the relation between every level-town pair in the chart, and all yielded significant differences except for Level 3 among females from towns Z and I (p = 0.08507, shown via a red dashed oval). So,
we can confidently conclude that the females from each town have distinct F0s, except for what we have stated in reference to towns Z and I. There is no convincing explanation or story to tell about towns Z and I, but it is worth mentioning that these two are both cities in the south (Tanzania), and the geographical location may have influenced this level of insignificance.

<table>
<thead>
<tr>
<th>Chart 5-9 The overall distribution of F0s (for obstruents), as produced by female subjects</th>
</tr>
</thead>
</table>
| The F0s of the male subjects, meanwhile (Chart 5-10), did not show as many consistently significant differences as what we saw with the females. Similar to the case of Level 3 with town Z and town I women, the men of both those towns showed an insignificant difference at Level 3 ($p = 0.4448$). Level 1, Level 2, and Level 3 in the two inland towns (N and I) are not statistically different, either ($p = 0.8301, 0.5153,$ and $0.7677$, respectively).

These results further support the idea that there are similarities among second-language speakers of Swahili; just as we found for the imitation of VOTs in towns N and I (Chart 5-11 below), these two towns reflect very close prototype patterns. The first-language male speakers of towns Z and M, meanwhile, had statistically significant F0s, as did the female speakers in those towns. To explain the productions of F0 in all four towns, we can conclude
that the F0 is distinct in each female group at almost every level, while the males’ F0 productions are significantly different across towns except for the consistent insignificance between towns N and I.

These F0 results give evidence that, unlike speakers of languages with an F0 contrast (Krishnan et al., 2010; Kwon, 2015 and 2019), Swahili speakers, who show a preference towards the VOT cue, failed to trace the F0 differences (e.g. Level 3 for town M women and Level 3 for town Z men) in that neither men nor women yielded significant results among any of the three levels, just as Kwon (2013) reported for English speakers. The other evidence we can glean from our results is that the fundamental frequency of speakers of a language can vary considerably between locations where the language is spoken, but that gender—as an independent variable—does not follow a similar pattern. As we have seen, the nice-looking distributions of the females’ F0s are contrasted by the males’ slightly different scatterings of the prototype clouds, though the males’ insignificant F0 results for towns N and I are clearer for explaining the F0 similarities when Swahili is spoken as a second language.
If factors related to F0 do not cue the laryngeal contrast in Swahili, then we would expect to see more sensitivity among speakers when it comes to mimicking VOT differences. As we can confirm from the results sections above, and as researchers have found for many languages cross-linguistically (non-tonal languages in particular), Swahili does have a VOT contrast. However, we have to be cautious not to overgeneralize the existence of this property. Realizing that languages are different in this regard, we have to extend this understanding to varieties of the same language as well.

So, what can speakers of Swahili as a first or second language add to the literature on the spread of VOT in East Africa? To get an overall sense of Swahili speakers’ sensitivity to VOT, see Chart 5-11.

**Chart 5-11 The overall distribution of VOTs, as produced by all subjects**

These results confirm that town Z residents were the group of participants whose productions most consistently tracked the different levels of VOT in the stimuli, as the respondents from this town yielded the highest probability of significance at the three imitated levels. Town M, on the other hand, comes second, as its participants were able to distinguish Level 1 from Level 2 and Level 3, but there was no significant difference between
Level 2 and Level 3. However, there was no significance whatsoever in the imitations of town N and town I participants, who represent the two groups of inland second-language speakers. When we look at the clouds that plot the VOT values, the shortest VOTs were produced by town M participants, then those from town Z, followed by those from towns I and N. Although the two inland towns have different, but very close VOT durations, they did not show any statistical significance, and their imitations of Level 3 came out almost identical on the continuum periphery for longer durations. With confidence, therefore, we can report that when imitating modified VOTs of nonce words, town Z residents displayed more consistent contrasts between the three levels than the residents of the mainland towns. Yet we do have to give some credit to town M who produced a consistent difference in response to the level 1 and level 2 stimuli than the respondents in towns N and I did.

In Chart 5-9 and Chart 5-10 above, we did not obtain any level of significance for the F0 imitations of F0 manipulations, but in Chart 5-12 below, we see that the VOT of heard F0s is clearly different from the VOT of heard VOTs (as was illustrated in Chart 5-11). These dissimilarities in the distributions of the clouds give further evidence that Swahili speakers can perceive the changes in the tokens. The modifications, however, were imitated in an unbalanced way. The listeners were not basing their VOT production all that much on the changes in F0 they heard in the stimuli, whether or not they were able to perceive those F0 differences. They were probably just producing a default ("prototype") VOT value across the board. The VOTs that were produced when the speakers were mimicking F0s have given us a rough idea of what the neutral VOT prototype looks like in an environment where the speakers are not able to imitate well enough to produce a non-primary cue such as F0. In Group A (towns Z and M in Chart 5-12), town Z participants, when repeating low-F0 tokens, exhibited a pattern of imitation – though it is not at the level of significance – where the Level 1 VOT values are shorter than those of Level 2 and Level 3.
Town Z is unlike the other towns in both Group A and Group B, which went in quite the opposite direction by shortening their VOTs as the F0 heights increased. Additionally, these imitations of F0 reflect the best example of how first- and second-language speakers of Swahili would react to secondary phonetic details. In Group A, those who speak Swahili as a first language have shown very close VOT values, where the clouds are pretty much tucked together. Group B, on the other hand, shows a similar decrease from Level 1 to Level 2 to Level 3, but with a bit of a gap in between, although those few milliseconds (between the clouds for town N and town I) were not statistically significant. All the clouds in Group B, meanwhile, are statistically different from the clouds in Group A; that is to say, when imitating secondary cues, speakers of Swahili as a first language would create VOT values that are different from those produced by second-language speakers.

![Chart 5-12 The overall distribution of VOT values for F0-based stimuli, as produced by all subjects](image)

To sum up, F0 production was more stable in town M, as the F0s are higher (across both genders) than what was found in the other towns. Town I, on the other hand, tends to have the lowest F0 values. When it comes to VOT, the best imitations were found in town Z with a high level of control over the three synthesized levels, and to a lesser degree in town...
M. When secondary cues (F0) are imitated, two clusters are formed, which separate the first-language speakers from the second-language speakers. However, when we look at the overall result as far as phonetic cues in Swahili are concerned, we conclude that VOT matters but F0 does not.

5.3 Closing remarks

The current sociophonetics of these four Swahili-speaking locations has shown that, despite the early, conflicting conclusions surrounding the status and distribution of laryngeal phonemic variations, speakers of the language still use one phonetic cue when uttering real words based on aspiration or word origin, and also one when mimicking fake words. The primary phonetic parameter is VOT. Speakers on the island (town Z) produce and perceive VOT quite strongly, with little evidence of F0-based productions as well. By contrast, the coast (town M) is observed to use some F0 cues rather than VOT in production, but it relies on VOT rather than F0 in imitation. On the other hand, the inland region (towns N and I) does not use either cue.

In short, there is not one “Swahili,” but rather a collection of different varieties of the language, each of which can express laryngeal features in different phonetic ways, for different historical or linguistic reasons. Moreover, where two groups of native Swahili speakers show significant differences between them, these differences become even larger if the L1 speakers are then compared with L2 speakers.
Chapter 6: CONCLUSION

This dissertation provides a data-focused account of Swahili obstruents, presenting and analyzing cross-dialectal data from four spoken varieties. It fills gaps in the instrumental analysis of Swahili voiceless obstruents and contributes to the existing literature on the quantitative aspect of Swahili variations based on aspiration and word type. This concluding chapter sketches how the relevant phonetic details — VOT and F0 — can add to our understanding of the phonemic representation. It includes a summary of my findings, followed by the limitations of my study and ideas for future research.

6.1 Summary of findings

This dissertation yielded important findings that will be useful for documenting and preserving Swahili as it is spoken in four different locations in East Africa, whether it is spoken as an L1 (towns Z and M) or an L2 (towns N and I). Participants were tested and their production and perception of the language were recorded in all four locations. The production task focused on how the participants articulate VOT and F0 in the obstruents of Swahili (bilabial /p/, alveolar /t/, velar /k/, and post-alveolar /tʃ/). It is acoustically common for speakers to produce bilabials with short VOTs, and this duration keeps increasing as we move to alveolars, then to velars (Maddieson, 1996), and then to post-alveolars (Hussain, 2018). My production study aimed to see whether and how this VOT hierarchy pattern manifests in different Swahili varieties nowadays, and more broadly, to conduct extensive first-hand phonetic observations to better understand the status of aspiration and of word-type distribution during real-word production and a nonce-word imitation.

Many VOT production studies report on multiple languages (Lisker & Abramson, 1964; Cho and Ladefoged, 1999; Hussain, 2018). Other studies focus on a particular

---

97 In some occasions, some of these studies worked on more than one phonetic cue (such as F0 and voice quality).
language (e.g. Turkish: Öğüt et al., 2006; English: Weber, 2007; Hungarian: Gósy & Ringen, 2009; Jordanian Arabic: Mittleb, 2009; Korean: Oh, 2011; Zulu: Midtlyng, 2011; Indonesian: Li et al., 2019), or two languages (e.g. English and Mandarin: Chen et al., 2007 and Chao & Chen, 2008; Hindi and Urdu: Schertz and Khan, 2020), or else dialects of a language (e.g. Korean varieties: Kang et al., in press). Many studies have shown that gender affects VOT measurements in that the VOTs of voiceless stops are longer among women (e.g. Swartz, 1992; Ryalls et al., 1997; Whiteside & Irving, 1998; Koenig, 2000; Robb et al., 2005; Herd, 2020). However, a few studies have reported longer VOTs among men (Gósy & Ringen, 2009; Sokolovic-Perovic, 2012; Mielke & Nielsen, 2018), and others have shown that while the aspirated VOTs are longer for women, the unaspirated VOTs are longer for men (Peng et al., 2009; Ma et al., 2018).

Other studies have focused on the use of both VOT and F0 as phonetic cues to laryngeal contrasts (Lee & Iverson, 2012; Ladd & Schmid, 2018; Coetzee et al., 2018). Although VOT is obviously affected by the voicing/aspiration contrasts, the F0 cue was seen to affect the laryngeal contrasts in these studies.

Finally, imitation has been used to test perception of laryngeal distinctions in other languages (e.g. American English: Nielsen, 2011; English: Babel, 2012; Korean: Kwon, 2015 and 2019; American English: Wade et al., 2020).

In light of these available facts about other languages, African languages remain rather understudied. A thorough examination of Swahili — Africa’s largest Indigenous language — is long overdue to improve our understanding of how its four types of obstruent stops are produced and perceived. Because Swahili is so widely distributed in Africa, this dissertation focused not on a single variety, let alone two or more languages, but rather on four different areas potentially representing different dialects of Swahili.
In reality, in-depth studies are sorely needed on African languages in all fields of linguistics, let alone on the wider phonemic system of Swahili in particular. To ensure that laryngeal contrasts in Swahili warranted a thorough study, I conducted a pilot study on two speakers at the University of Calgary. Both claimed to be speakers of Swahili (from Dar es Salaam, Tanzania), but neither of them was able to exhibit any special phenomenon or provide any promising results. Strikingly, their productions were not in line with the observations of earlier linguists (Engstrand & Lodhi, 1985; Moxley, 1992), who showed instrumentally that the Swahili phonemic system has a three-way contrast. Consequently, I had to prepare myself and make my way to East Africa to personally obtain good-quality recordings from native and non-native speakers of Swahili. I worked directly with all of the participants in three of the four locations (towns Z, N, and M). In the fourth (town I), my research assistant assisted me by running the experiments.

Two experiments were performed in order to answer the research questions. In the production experiment, the participants (who were mostly college students) were presented with a randomized picture-naming task and asked to speak the target Swahili words two times in isolation. The purpose was to come up with basic acoustic measures to determine whether or not Swahili speakers have VOT- or F0-based contrasts for obstruents in the language. The focus of the analysis was twofold: first to see whether and how the informants deploy varying VOT or F0 phonetic ranges based on either the word’s semantic meaning or its origin, and second to find out if the language’s speakers assign any aspiration to non-minimal pairs. In the imitation experiment, the same subjects were asked to mimic manipulated VOTs and F0s in a test of their sensitivity towards the modified phonetic cues.

---

98 In his work on Zulu, Midtlyng (2011) found that words in isolation give rise to intermediate VOT values for plosives between the two speed conditions in a carrier sentence: slow speed (long VOTs) and fast speed (short VOTs).
whether short, intermediate, or long. In both experiments, Praat and R were used for data annotation and analysis, respectively.

6.1.1 Summary of production

In Chapter 3 of this paper, the VOT/Phonation category of the results I obtained from the final models of each town gave different accounts.

Town Z’s final model (Phonation + Place + Phonation*Place) was in favor of VOT as a phonetic cue, which supports the idea that the phonemic representation of the language has a three-way laryngeal contrast, and the four places of articulation have differences in their VOT durations. Every single speaker from town Z exhibited a contrast based on aspiration/non-aspiration, and this distinction was not limited to minimal pairs but also found in singletons, suggesting that aspiration plays a role in non-minimal pairs, too.

Unlike town Z’s results, town M’s final model (Place + Gender + Place*Gender) did not show any effect for VOT/Phonation, suggesting that contrastive aspiration was absent among this location’s speakers. VOT did matter, however, in two predictors: it showed variations based on Place and Gender. Thus, in town M, the production of VOT varies according to place of articulation, and men and women were also quite different in their VOT durations. Towns N and I proved to be just like town M with regard to the production of VOT. That is, in all three of these towns, VOT (as a dependent measure) is effectively a cue for differentiating males’ productions from those of females, and for differentiating places of articulation.

Turning to the VOT/loanword category, as described in the final model (Loanword + Place), town Z speakers’ usage of VOT was seen to differ based on word type. Word origin was a factor that affected the VOT durations — longer for English, intermediate for Swahili, and short for Arabic. Moreover, obstruents had different VOT durations based on their place of articulation.
Town M’s results, on the other hand, indicated that Loanword was not effective as an explanatory factor; the only predictors that affected VOT there were Place and Gender, according to the final model (Place + Gender + Place*Gender). Towns N and I again followed in the footsteps of town M, with only Place and Gender influencing VOT — not Loanword. That is, whether the data set included aspiration (the Phonation data) or only the unaspirated tokens (the Loanword data), towns M, N, and I had approximately similar VOT production mechanisms, which did not match town Z’s.

In the F0/Phonation category, observations from town Z’s final model (Gender + Phonation + Gender*Phonation + Place) provided evidence that pitch was affected by more than one explanatory factor. Gender, as is commonly known, was a factor in that women produced higher F0s than men. The model also included Phonation as a significant factor because of the effect that unaspirated tokens had on F0. Place was also a predictor, but it was not as strong as Gender and Phonation.

Town M was similar yet different from town Z’s final model (Place + Gender + Place*Gender + Phonation): it similarly added Phonation as another explanatory factor, but the difference in F0 with reference to Place and Gender proved to be stronger than the difference with reference to Phonation.

Speakers in towns N and I, on the other hand, had quite different production strategies. Town N’s final model (Gender + Place + Gender*Place) matched those of towns Z and M in that there were F0 differences in terms of Gender and Place, but not Phonation. Finally, town I had only one factor in its F0 final model (Gender), which is also in line with most F0 studies.

In the F0/Loanword category, town Z’s final model (Gender + Loanword) implied that the F0s in this group were, once again, different between the two genders. Loanword (i.e.

---

99 This order is different, but otherwise, this is the same set of factors as the above Z town model.
word origin) was also seen to give rise to different F0 heights. Unlike those of town Z, town M’s explanatory factors showed more variations in that the final model (Place + Gender + Place*Gender + Loanword) had all three factors of Place, Gender, and Loanword, meaning that the F0s are not the same for each of the three-factor estimates. In town N, the final model (Gender + Place + Gender*Place) excluded Loanword as the source of any variation, whereas town I’s final model (Gender + Loanword + Gender*Loanword) excluded Place instead. In general, then, the explanatory factors in the Loanword category are not the same for every town, and each town has its own specific final model.

In sum, the production results indicate that the VOT acoustic measure is a cue to the Phonation and Loanword categories of the different obstruents in town Z. By contrast, Towns M, N, and I had similar models and showed no VOT variations based on Phonation or Loanword. It is worth noting that the place of articulation hierarchy in the two data sets for VOT production (aspiration vs. non-aspiration) were not similar: while the hierarchy for unaspirated tokens among both men and women followed the expected (automatic) pattern Post-alveolar > Alveolar > Velar > Bilabial, the hierarchy for aspiration was somewhat complex and varied. When we took all four towns into consideration, the acoustic space for aspiration had far more overlap between the clouds as compared to that for non-aspiration, which showed bigger significant distances between the clouds’ means.

On the other hand, and quite aside from the usual variation based on Gender, there was a different final model for F0 for each town, and the observed F0 ranges varied based on the reported models. The variabilities in the F0 models give evidence that F0 is secondary and more complicated than the (primary) straightforward dependent measure of VOT as a phonetic cue for Swahili speakers. In the production task of this study, F0 has revealed different acoustic preferences with some significance, but none of these significance levels
were able to compete with the robust VOT/Phonation contrast in town Z, or with the solid overall VOT phonetic change based on Loanword type that was also seen in town Z.

6.1.2 Summary of imitation

In Chapter 4, the statistical tools that I used for the production analysis were implemented to account for the imitation data, which consisted of responses to manipulations of four nonce words starting with /p/, /t/, /k/, and /tʃ/ at three different VOT and F0 levels. First, I looked at the imitation of modified VOT durations. The VOT final model (Level*Town), which included all towns’ participants, indicated that the significance level is reached in two towns only (Z and M) when the explanatory factor is Town (see Table 4-1). Then, a further examination of the interaction between Level and Town revealed that town Z was the only location whose participants more accurately perceived and produced the acoustic VOT variations.

Knowing that the participants in towns Z and M were better at perceiving the VOT differences than those in towns N and I, I created separate models (with VOT Level as a predictor) for each town. The resulting p-values showed that the statistical difference keeps dropping as we move from one town to another: the strongest significance was in town Z (low p-value and higher χ² value), while there was also strong significance in town M (low p-value and high χ² value), little significance in town N (p-value slightly below the significance level and low χ² value), and no significance in town I (high p-value and low χ² value). Therefore, I concluded that the imitation of the unknown VOT phonetic cues was very evident in town Z, where all three levels were imitated with three statistically distinct VOT values. A similar pattern holds, but with weaker statistical significance, in town M, where two levels were produced with statistically different durations. And, finally, none of the three levels were significantly different in towns N and I.
The imitated F0s, on the other hand, were not as significant as the imitated VOTs. Level, as the primary explanatory factor in the imitation data, did not yield any improvement upon the null model. That is, neither the overall F0 model of Level, which included all towns, nor the individual models, which looked into the Level factor of each town independently, were able to account for any variance in the data. The overall Town factor, however, was an explanatory estimate in the three towns M, Z and N. The overall F0 in town M was the highest pitch among all towns. The final explanatory factor (Gender), as reported in the combination of the final model (Gender+Town), showed a significant average difference between men and women, as was the case in the data from the production task of this work.

6.2 **Limitations**

The data collection for this project was challenging in certain respects that are worth reflecting on. First, one of the drawbacks that I had to deal with in all locations was the lack of a sound-attenuated booth, as this was near-impossible to find, much less use. On the other hand, meeting and recording people in their homelands helps the researcher maintain control of any mixed ethnicities and dialects, an opportunity which could not be matched if recruiting the desired sample in a (Western) foreign country (e.g. Swahili speakers in Canada). Still, it can be difficult to make first-hand recordings in Africa. Notably, authorities in certain African territories err on the side of caution and feel suspicious when a foreign national comes to conduct a study on their soil, so one must prepare for the bureaucratic procedures for obtaining final permission from them.\(^{100}\)

---

\(^{100}\) In the case of Kenya, the application for a research permit is available online [https://research-portal.nacosti.go.ke]. Then, the process can be finalized in one day after reporting to the county’s two offices: that of the Commissioner and that of the Director of Education (this was done twice: once in Nairobi and once in Mombasa). However, the data collection permit for Zanzibar required some bureaucracy and back-and-forth among four offices (those of Statistics, Immigration, the Vice President, and the Ministry of Education), and it took around three weeks before I was able to visit my intended research site (Sumait University) on the island. Now, as the dialect of Zanzibar (KiUnguja) has proven to be categorically and acoustically different than the dialects in the other three locations, I am hoping that the government of Zanzibar will consider implementing an easier research application system to motivate researchers (and linguists in particular) to help with documenting and preserving the Swahili language as it is spoken on the two islands: Zanzibar and Pemba.
If the intention is to recruit participants at a college, it is highly recommended to select a time when plenty of students are around, especially given that students may have reservations and hesitate to participate, in spite of being remunerated for their participation. Conducting fieldwork in the summer is convenient for academics, but fewer students are around at that time. Beggars can’t be choosers, so the summer researcher risks recruiting undesired samples (e.g., participants who may be from outside the intended location who may speak a different dialect), which could cause the analysis to show many outliers and create irrelevant data for the project.

In practice, the people at all the locations where I conducted my recordings were in general very responsive and willing to help. The only delay I encountered in this regard was at the University of Nairobi, where I had to wait for over a week due to the fact that the individuals whom I had contacted before coming to Kenya were either on a sabbatical leave or attending a conference overseas. On their return, however, they made every effort to support my research.

In the case of mainland Tanzania, I did not have the required research authorization in hand, and due to the time it takes the authorities to issue a data collection permit, I had to train a research assistant, hand my equipment over to him, and then have him meet and record my Iringa participants on my behalf. Although he did a good job carrying out most of the experiments, some things did not go as expected. In particular, the production experiment went well, but the two perceptual experiments had some problems.

The discrimination experiment was eliminated (not part of the current study) due to the extreme failure of one ultimate condition, namely the “reaction time.” The Iringa participants performed the worst among all the groups and showed a very short reaction time between one stimulus and the following one, which gave me the impression that they did not take the task seriously; basically, the task was performed by chance. For the second
perceptual task (imitation), these participants provided only one repetition per stimulus, which was different from the number of imitations in the other locations. This did not prevent me from analyzing their imitation responses, which were found to have no significance, although they addressed the level of acoustic sensitivity in town I. It is possible that if I’d been able to run the experiment by myself there, I might have gotten stronger significant results.

Another limitation of this study is the choice of minimal pairs as the items to test. Despite Frankl’s (1991: 369) claim to have found around four dozen minimal pairs in BURT (1910), my participants in towns M, N, and I were unfamiliar with the minimal pairs I presented to them. Although I suspected that Swahili speakers were not familiar with certain aspirated tokens, I thought that they would be able to recognize a minimal pair such as paa [paa/pʰaa], meaning either ‘roof’ or ‘deer’. In order to see if town M participants were aware of this contrast, I made a box plot for the two variants of paa above based on their Phonation category (non-aspiration vs aspiration), but it turned out that the contrast left no trace. This suggests a preliminary conclusion that changing the list of minimal pairs to match some of BURT’s word list would not provide much support to the idea of an aspiration contrast (in town M, at least). I was only able to access BURT’s work until after my study, yet I am doubtful that using or adding different minimal pairs would improve speakers’ proficiency

---

101 Only 2 minimal pairs in my list are reported in BURT (1910) pp. 159-160: namely, paa and kaa by having different meanings based on the presence or absence of aspiration. The other minimal pairs in my work are not mentioned in the book, except for two: paka meaning ‘cat’ and ‘to smear’ p. 215, and chungu meaning ‘small cooking-pot’ only p. 84. BURT also mentions the two words ch’uguu ‘ant-hill’ p. 169 and t’ungu ‘small ant’ p. 239, but it is not clear if one of these last two words changed over time to be pronounced with aspiration (ch’ungu), which was then considered a homonym for chungu. tembo ‘palm wine’ is mentioned on p. 233, but the meaning of ‘elephant’ is nowhere to be found. As I went over some pages in BURT’s book, and just like what I found among the speakers of town Z, it was very evident that aspiration is found for the voiceless stops and the voiceless affricate because the apostrophe (’), as an aspiration marker, was applied extensively. In addition, it is worth mentioning that BURT did not include <ch> in the minimal pairs list as the list was entirely for stops (more examples for /l/ and /k/ than /p/) and some post-nasal stops, but she also (1910: 20) states: “beware of uttering ch or k with the English stress, or you may seem to aspirate them, and so possibly say something quite different from what you intended.”
with regard to aspiration, as we have seen that aspiration is also found in non-minimal pairs that town M participants could not produce, even when looking at individual words.

Another limitation is that I do not speak the Swahili language. This did not cause any difficulties during work on this dissertation. Not being able to speak the language could pose a difficulty for certain linguistic aspects such as syntax or morphology, as these might call on the linguist’s intuitions to reach conclusions about the structure of sentences or the formation of lexical items. But my experience doing phonetic work on recorded words suggests that it is possible to examine a big corpus of data from a language without having a deep linguistic knowledge of the language.

Finally, I focused on laryngeal contrasts in voiceless obstruents. My early ambition was to look at all word-initial stops and affricates, whether voiced or voiceless, including the post-nasal stops, but as the work on the dissertation progressed, the scope of the study was narrowed and only 43.5% of my production corpus was used, studied, and analyzed.

6.3 Future directions

As I just mentioned, 56.5% of my production data remain to be analyzed, so there are many possible continuations of this project. The laryngeal contrast has mainly focused on the short-lag/long-lag durations along the VOT continuum, but the negative VOT lead ought to be explored, too.

Of particular interest is whether Polomé (1967) was correct when he assigned the Swahili voiced plosives to a status as allophones of the voiced implosive phonemes (see Appendix-A for the list of words). I have seen some signs of voiced plosives and implosives in the data, but cannot conclusively determine their status in the language until an analysis similar to the account presented in this dissertation is performed. The data for voiced stops

---

102 This 56.5% consists of 83 words in two repetitions, which yield 166 stimuli per subject, and when we multiply the 166 by 98 participants, we get over 16,000 responses. If we further consider that each stimulus has two dependent measures (one for VOT and one for F0), this yields a total of over 30,000 data points.
also contain loanwords (from English and Arabic), so it would be interesting to know if (voiced) borrowed words feature phonetically similar or different acoustic cues compared to native words.

Next, the remainder of the data also contain the voiced post-alveolar affricate /dʒ/, whose phonetic implementation is an issue no less complicated than the status of aspiration in the language (now clarified), or even the ambiguity of how plosives/implosives are distributed. During the course of my annotations, I have encountered many cases in which participants would pronounce the voiced palatal plosive /ɟ/ instead of the voiced post-alveolar affricate /dʒ/; some speakers even pronounce it with very long durations and some presumed sort of aspiration.

Post-nasal stops, whether voiced or voiceless, constitute another interesting category of sounds in the leftover data. As I described in the Background section of this work, some linguists (Wald, 2009; Contini-Morava, 1997) claimed that aspiration in Swahili may have resulted from the loss of nasals in word-initial position. Although I am not in a position to support or refute this possibility, I have heard many words of presumably initial nasality (as they appear in the orthography), but with no produced nasals. Omission of nasality is also observed in both voiced and voiceless post-nasal stops. That is, the phenomenon of dropping nasals and beginning words with stops is still happening among some speakers of the language.

One of the early stages of this dissertation involved looking solely at Swahili loanwords from Arabic and then cross-checking them with the VOTs of Arabic, and that plan drove me to collect data for Arabic stops and affricates (and emphatics) from native speakers of Arabic. For this part of the project, 45 participants from five different nationalities (Egyptian, Omani, Sudanese, Syrian, and Yemeni) were recorded in order to determine which dialect most closely matched the specific VOT values in Swahili words of Arabic origin. It
has been claimed historically that native Swahili speakers along the east coast of Africa (especially in Zanzibar) were in strong contact with speakers of Omani Arabic (as stated in Nurse and Hinnebusch, 1993: 321; also cited in Lodhi, 2000: 97). Thus, there is strong motivation to conduct an acoustic investigation into whether this one Arabic dialect, which is claimed to be the primary donor of Arabic loanwords in Swahili, has any phonetic properties that match any of the production results found in this dissertation.

Future work on the perceptual side could target the discrimination of minimal pairs and determine to what extent speakers of Swahili can differentiate between words with reference to phonation, whether aspirated or unaspirated (long VOT vs. short VOT and high F0 vs. low F0). As I mentioned earlier, the identification of modified VOT/F0 stimuli was planned to be a part of this dissertation, but it was put on hold due to limitations of space and time. Moreover, an acoustic study could analyze the speakers’ productions of the second syllable of the non-words in this imitation task. The study design involved applying similar manipulations on the second syllable as on the first one, but only the mimicking of the first syllable was reported in this dissertation. So the question of what happens to the second syllable when Swahili speakers imitate modified nonsense words still requires an answer.

In sum, there is huge potential to learn much more about the Swahili sound system by examining produced (true) words for data on voiced plosives vs. implosives; voiced post-alveolar affricates vs. voiced palatal plosives; and post-nasal stops vs. non-post-nasal stops, whether voiced or voiceless. Moreover, perception can be examined via discrimination of spliced real words, and we can further our understanding of the mechanisms which Swahili speakers deploy when imitating the modified second syllable of a non-word.
REFERENCES


Appendix A

Appendix A.1 Pemba Map (Zanzibar Northern Island)

Appendix A.2 Unguja Map (Zanzibar Southern Island)
<table>
<thead>
<tr>
<th>Swahili</th>
<th>Arabic loanwords in Swahili</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;bahari&gt;</td>
<td>‘sea’</td>
<td>/bahr/</td>
</tr>
<tr>
<td>&lt;baridi&gt;</td>
<td>‘cold’</td>
<td>/bārid/</td>
</tr>
<tr>
<td>&lt;bəti&gt;</td>
<td>‘house’</td>
<td>/bəjɪ/</td>
</tr>
<tr>
<td>&lt;bilaʒi&gt;</td>
<td>‘for free’</td>
<td>/balaj/</td>
</tr>
<tr>
<td>&lt;binti&gt;</td>
<td>‘daughter’</td>
<td>/bint/</td>
</tr>
<tr>
<td>&lt;bustani&gt;</td>
<td>‘garden’</td>
<td>/bustān/</td>
</tr>
<tr>
<td>&lt;buni&gt;</td>
<td>‘coffee’</td>
<td>/bunn/</td>
</tr>
<tr>
<td>&lt;tarehe&gt;</td>
<td>‘date’</td>
<td>/tārīx/</td>
</tr>
<tr>
<td>&lt;tafsiri&gt;</td>
<td>‘interpretation’</td>
<td>/tafsīr/</td>
</tr>
<tr>
<td>&lt;tini&gt;</td>
<td>‘fig’</td>
<td>/tīn/</td>
</tr>
<tr>
<td>&lt;tobə&gt;</td>
<td>‘repentance’</td>
<td>/taubah/</td>
</tr>
<tr>
<td>&lt;tabibu&gt;</td>
<td>‘doctor’</td>
<td>/tˤabīb/</td>
</tr>
<tr>
<td>&lt;tibu&gt;</td>
<td>‘scented oil’</td>
<td>/tˤib/</td>
</tr>
<tr>
<td>&lt;tufani&gt;</td>
<td>‘flood’</td>
<td>/tˤufān/</td>
</tr>
<tr>
<td>&lt;dawa&gt;</td>
<td>‘medicine’</td>
<td>/dawaʔ/</td>
</tr>
<tr>
<td>&lt;dalali&gt;</td>
<td>‘broker’</td>
<td>/dallāl/</td>
</tr>
<tr>
<td>&lt;deni&gt;</td>
<td>‘debt’</td>
<td>/dajn/</td>
</tr>
<tr>
<td>&lt;dini&gt;</td>
<td>‘religion’</td>
<td>/dīn/</td>
</tr>
<tr>
<td>&lt;dola&gt;</td>
<td>‘state’</td>
<td>/dawlah/</td>
</tr>
<tr>
<td>&lt;dohani&gt;</td>
<td>‘chimney’</td>
<td>/duxxān/</td>
</tr>
<tr>
<td>&lt;duka&gt;</td>
<td>‘store’</td>
<td>/dukkān/</td>
</tr>
<tr>
<td>&lt;kamili&gt;</td>
<td>‘complete’</td>
<td>/kāmil/</td>
</tr>
<tr>
<td>&lt;kanzu&gt;</td>
<td>‘treasure’</td>
<td>/kanz/</td>
</tr>
<tr>
<td>&lt;kitabu&gt;</td>
<td>‘book’</td>
<td>/kitāb/</td>
</tr>
<tr>
<td>&lt;kiburi&gt;</td>
<td>‘arrogance’</td>
<td>/kibr/</td>
</tr>
<tr>
<td>&lt;kofia&gt;</td>
<td>‘hat’</td>
<td>/kuftah/</td>
</tr>
<tr>
<td>&lt;kurasa&gt;</td>
<td>‘notebook’</td>
<td>/kurṛāṣah/</td>
</tr>
<tr>
<td>&lt;kulabu&gt;</td>
<td>‘hook’</td>
<td>/kullāb/</td>
</tr>
<tr>
<td>&lt;kaburi&gt;</td>
<td>‘grave’</td>
<td>/qabr/</td>
</tr>
<tr>
<td>&lt;kalamu&gt;</td>
<td>‘pen’</td>
<td>/qalam/</td>
</tr>
<tr>
<td>&lt;kasidi&gt;</td>
<td>‘intention’</td>
<td>/qasˤd/</td>
</tr>
<tr>
<td>&lt;kisa&gt;</td>
<td>‘story’</td>
<td>/qisˤsˤah/</td>
</tr>
<tr>
<td>&lt;kiriba&gt;</td>
<td>‘waterskin’</td>
<td>/qirbah/</td>
</tr>
<tr>
<td>&lt;kufruli&gt;</td>
<td>‘lock’</td>
<td>/qufl/</td>
</tr>
<tr>
<td>&lt;kura&gt;</td>
<td>‘ballot’</td>
<td>/qufrˤah/</td>
</tr>
</tbody>
</table>
Appendix A.7  Swahili native words and loanwords from English and Arabic

---------------------------------------------
Laryngeal Contrast---------------------------------------------

<table>
<thead>
<tr>
<th>Minimal Pairs</th>
<th>1- /p/ paa [pa:] ‘roof’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2- /pʰ/ paa [pʰa:] ‘gazelle’</td>
</tr>
<tr>
<td></td>
<td>3- /p/ paka [paka] ‘smear/apply’ (V)</td>
</tr>
<tr>
<td></td>
<td>4- /pʰ/ paka [pʰaka] ‘cat’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aspirated vs. Unaspirated</th>
<th>5- /p/ pesa [pesa] ‘money’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6- /pʰ/ punje [pʰundʒe] ‘seed’</td>
</tr>
<tr>
<td></td>
<td>7- /pʰ/ punda [pʰunda] ‘donkey’</td>
</tr>
<tr>
<td></td>
<td>8- /pʰ/ pemba [pʰemba] ‘the sister Island of Unguja’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minimal Pairs</th>
<th>1- /t/ tembo [tembo] ‘alcohol’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2- /tʰ/ tembo [tʰembo] ‘elephant’</td>
</tr>
<tr>
<td></td>
<td>3- /t/ tasa [tasa] ‘basket made up of bamboo sticks’</td>
</tr>
<tr>
<td></td>
<td>4- /tʰ/ tasa [tʰasa] ‘a woman who doesn’t give birth’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aspirated vs. Unaspirated</th>
<th>5- /t/ tofali [tʰofali] ‘brick’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6- /tʰ/ tope [tʰope] ‘mud’</td>
</tr>
<tr>
<td></td>
<td>7- /t/ tai [tai] ‘eagle’</td>
</tr>
<tr>
<td></td>
<td>8- /tʰ/ taka [tʰaka] ‘garbage’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minimal Pairs</th>
<th>1- /k/ kaa [ka:] ‘charcoal’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2- /kʰ/ kaa [kʰa:] ‘land crab’</td>
</tr>
<tr>
<td></td>
<td>3- /k/ koa [koa] ‘a dry bark of banana plant’</td>
</tr>
<tr>
<td></td>
<td>4- /kʰ/ koa [kʰoa] ‘snail’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aspirated vs. Unaspirated</th>
<th>5- /k/ kiti [kʰiti] ‘chair’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6- /kʰ/ kucha [kʰutʃa] ‘fingernail’</td>
</tr>
<tr>
<td></td>
<td>7- /k/ kuku [kuku] ‘chicken’</td>
</tr>
<tr>
<td></td>
<td>8- /kʰ/ kasa [kʰasa] ‘sea turtle’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minimal Pairs</th>
<th>1- /tʃ/ chana [tʃanə] ‘tear’ (v)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2- /tʃʰ/ chana [tʃʰanə] ‘a bunch of banana’</td>
</tr>
<tr>
<td></td>
<td>3- /tʃ/ chungu [tʃungu] ‘earthen cooking pot’</td>
</tr>
<tr>
<td></td>
<td>4- /tʃʰ/ chungu [tʃʰungu] ‘black ant’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aspirated vs. Unaspirated</th>
<th>5- /tʃʰ/ cheche [tʃʰetʃʰe] ‘fire sparks’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6- /tʃʰ/ chenza [tʃʰen.za] ‘tangerine’</td>
</tr>
<tr>
<td></td>
<td>7- /tʃ/ chiriku [tʃiriku] ‘parrot’</td>
</tr>
</tbody>
</table>

---------------------------------------------
English Loanwords---------------------------------------------

<table>
<thead>
<tr>
<th>Borrowed From English</th>
<th>1- /p/ papai [papai] ‘papaya’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2- /p/ penseli [penseli] ‘pencil’</td>
</tr>
<tr>
<td></td>
<td>3- /p/ picha [pitʃa] ‘picture’</td>
</tr>
<tr>
<td></td>
<td>4- /p/ posta [posta] ‘post office’</td>
</tr>
</tbody>
</table>
### Swahili

<table>
<thead>
<tr>
<th>Swahili</th>
<th>Arabic</th>
<th>Arabic Loanwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. tarehe</td>
<td>/tārī/t/</td>
<td>‘date’</td>
</tr>
<tr>
<td>2. tini</td>
<td>/tīn/</td>
<td>‘fig’</td>
</tr>
<tr>
<td>3. toba</td>
<td>/taubah/</td>
<td>‘penance/repentance’</td>
</tr>
<tr>
<td>4. tabibu</td>
<td>/tˤabīb/</td>
<td>‘doctor/clinic officer’</td>
</tr>
<tr>
<td>5. tibu</td>
<td>/tˤib/</td>
<td>‘scented oil’</td>
</tr>
<tr>
<td>6. tufani</td>
<td>/tˤufān/</td>
<td>‘hurricane/storm’</td>
</tr>
<tr>
<td>7. kanisa</td>
<td>/kanisah/</td>
<td>‘church’</td>
</tr>
<tr>
<td>8. kitabu</td>
<td>/kitāb/</td>
<td>‘book’</td>
</tr>
<tr>
<td>9. kefia</td>
<td>/kufijah/</td>
<td>‘hat’</td>
</tr>
<tr>
<td>10. kurasa</td>
<td>/kurrāsah/</td>
<td>‘book page’</td>
</tr>
<tr>
<td>11. kalamu</td>
<td>/qalam/</td>
<td>‘pen’</td>
</tr>
<tr>
<td>12. kisa</td>
<td>/qisˤsˤah/</td>
<td>‘story’</td>
</tr>
<tr>
<td>13. kiriba</td>
<td>/qirbah/</td>
<td>‘water bottle(waterskin)’</td>
</tr>
<tr>
<td>14. kufuli</td>
<td>/qufl/</td>
<td>‘lock’</td>
</tr>
<tr>
<td>15. kura</td>
<td>/qurʕah/</td>
<td>‘ballot’</td>
</tr>
</tbody>
</table>

### Arabic

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Arabic Loanwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>/k/</td>
<td>‘camp’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘cake’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘kilo’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘coach’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘chalk’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘chimney’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘chips’</td>
</tr>
<tr>
<td>/k/</td>
<td>‘chocolate’</td>
</tr>
</tbody>
</table>

---

*for all locations but chengeu ‘lampshade’ for some Iringa participants (males 1 to 6 and 8)*
Appendix B

Appendix B.1  F0 of All Towns (Bilabial*Gender)

Appendix B.2  F0 of All Towns (Alveolar*Gender)
Appendix.B.3  F0 of All Towns (Velar*Gender)

Appendix.B.4  F0 of All Towns (Post-Alveolar*Gender)
Appendix B.5  VOT of All Towns by Loanword/p/

Appendix B.6  VOT of All Towns by Loanword/t/
Appendix.B.7  VOT of All Towns by Loanword/k/

Appendix.B.8  VOT of All Towns by Loanword/ch/
Gender Differences in Native words

T-test, p = 0.16

T-test, p < 2.2e-16

T-test, p < 2.2e-16

T-test, p = 1.6e-08

Gender Differences in English Loanwords

T-test, p = 0.014

T-test, p < 2.2e-16

T-test, p = 5.9e-16

T-test, p = 2.2e-16

Gender Differences in Arabic Loanwords

T-test, p = 0.0316

T-test, p < 2.2e-16

T-test, p = 5.3e-16

T-test, p = 2.3e-16

Gender

VOT duration (ms)
## Appendix B.12  Aspiration Contrast in Town Z as per Subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>T-test, p = 3.9e-05</th>
<th>T-test, p = 1.5e-11</th>
<th>T-test, p = 2.8e-07</th>
<th>T-test, p = 5.4e-06</th>
<th>T-test, p = 1.5e-09</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unasp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Appendix B.13  Aspiration Contrast in Town M as per Subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>T-test, p = 0.42</th>
<th>T-test, p = 0.14</th>
<th>T-test, p = 0.25</th>
<th>T-test, p = 0.68</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018M3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018F7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unasp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VOT duration (ms)

Aspiration contrast in town Z as per subject.
Appendix.B.14  Aspiration Contrast in Town N as per Subject

Appendix.B.15  Aspiration Contrast in Town I as per Subject
Appendix.B.16  Loanword Type in Town Z as per Subject

Appendix.B.17  Loanword Type in Town M as per Subject
Appendix.B.18  Loanword Type in Town N as per Subject

Appendix.B.19  Loanword Type in Town I as per Subject
Appendix C

Appendix C.1 Level in Town Z as per Subject

Appendix C.2 Level in Town M as per Subject
Appendix.C.3  Level in Town N as per Subject

Appendix.C.4  Level in Town I as per Subject
Appendix D

Appendix D.1  Aspiration Contrast in Town Z as per Subject (F0)

Appendix D.2  Aspiration Contrast in Town M as per Subject (F0)
Appendix.D.3  Aspiration Contrast in Town N as per Subject (F0)

Appendix.D.4  Aspiration Contrast in Town I as per Subject (F0)
Appendix.D.5  
Loanword Type in Town Z as per Subject (F0)

Appendix.D.6  
Loanword Type in Town M as per Subject (F0)
### Appendix E: Wordlist for other words in the data

<table>
<thead>
<tr>
<th>English Loanwords</th>
<th>Arabic Loanwords</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td><strong>Arabic</strong></td>
</tr>
<tr>
<td><strong>Borrowed</strong></td>
<td><strong>Borrowed</strong></td>
</tr>
<tr>
<td>From</td>
<td>From</td>
</tr>
<tr>
<td>English</td>
<td>English</td>
</tr>
<tr>
<td>Native</td>
<td>Native</td>
</tr>
</tbody>
</table>

#### English Loanwords

<table>
<thead>
<tr>
<th>Number</th>
<th>English</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/b/ balbu</td>
<td>/balbu/</td>
<td>‘bulb’</td>
</tr>
<tr>
<td>2</td>
<td>/b/ betri</td>
<td>/bêtri/</td>
<td>‘battery’</td>
</tr>
<tr>
<td>3</td>
<td>/b/ biskuti</td>
<td>/bîskûti/</td>
<td>‘biscuit’</td>
</tr>
<tr>
<td>4</td>
<td>/b/ burashi</td>
<td>/burâʃî/</td>
<td>‘brush’</td>
</tr>
<tr>
<td>1</td>
<td>/b/ bafu</td>
<td>/bafû/</td>
<td>‘bathroom’</td>
</tr>
<tr>
<td>2</td>
<td>/b/ bendera</td>
<td>/bendêrâ/</td>
<td>‘flag’</td>
</tr>
<tr>
<td>3</td>
<td>/b/ bisibisi</td>
<td>/bisîbîsî/</td>
<td>‘screwdriver’</td>
</tr>
<tr>
<td>4</td>
<td>/b/ bundi</td>
<td>/bunđî/</td>
<td>‘owl’</td>
</tr>
</tbody>
</table>

#### Arabic Loanwords

<table>
<thead>
<tr>
<th>Number</th>
<th>English</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/d/ daktari</td>
<td>/daktârî/</td>
<td>‘doctor’</td>
</tr>
<tr>
<td>2</td>
<td>/d/ dereva</td>
<td>/derêvâ/</td>
<td>‘driver’</td>
</tr>
<tr>
<td>3</td>
<td>/d/ dishi</td>
<td>/dîʃî/</td>
<td>‘dish’</td>
</tr>
<tr>
<td>4</td>
<td>/d/ dola</td>
<td>/dola/</td>
<td>‘dollar’</td>
</tr>
<tr>
<td>1</td>
<td>/d/ dada</td>
<td>/dâda/</td>
<td>‘sister’</td>
</tr>
<tr>
<td>2</td>
<td>/d/ dengu</td>
<td>/dengu/</td>
<td>‘lentils’</td>
</tr>
<tr>
<td>3</td>
<td>/d/ dimbwi</td>
<td>/dimbwią/</td>
<td>‘pool/puddle’</td>
</tr>
<tr>
<td>4</td>
<td>/d/ duma</td>
<td>/duma/</td>
<td>‘cheetah’</td>
</tr>
</tbody>
</table>

#### Swahili

<table>
<thead>
<tr>
<th>Number</th>
<th>English</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/ɡ/ gauni</td>
<td>/ɡauni/</td>
<td>‘gown/dress’</td>
</tr>
<tr>
<td>2</td>
<td>/ɡ/ gereji</td>
<td>/geredʒî/</td>
<td>‘garage/repair shop’</td>
</tr>
<tr>
<td>3</td>
<td>/ɡ/ gitaa</td>
<td>/ɡitâ/</td>
<td>‘guitar’</td>
</tr>
<tr>
<td>4</td>
<td>/ɡ/ goli</td>
<td>/ɡoli/</td>
<td>‘goal’</td>
</tr>
<tr>
<td>1</td>
<td>/ɡ/ ganda</td>
<td>/ɡanda/</td>
<td>‘peel’</td>
</tr>
<tr>
<td>2</td>
<td>/ɡ/ gereza</td>
<td>/ɡereza/</td>
<td>‘prison’</td>
</tr>
<tr>
<td>3</td>
<td>/ɡ/ gogo</td>
<td>/ɡogo/</td>
<td>‘log’</td>
</tr>
<tr>
<td>4</td>
<td>/ɡ/ gumba</td>
<td>/ɡumba/</td>
<td>‘thumb’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number</th>
<th>English</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/dʒ/ jaketi</td>
<td>/dʒakêtî/</td>
<td>‘jacket’</td>
</tr>
<tr>
<td>2</td>
<td>/dʒ/ jela</td>
<td>/dʒela/</td>
<td>‘jail’</td>
</tr>
<tr>
<td>3</td>
<td>/dʒ/ jinzi</td>
<td>/dʒinzi/</td>
<td>‘jeans’</td>
</tr>
<tr>
<td>4</td>
<td>/dʒ/ juisi</td>
<td>/dʒuisi/</td>
<td>‘juice’</td>
</tr>
<tr>
<td>1</td>
<td>/dʒ/ joto</td>
<td>/dʒoto/</td>
<td>‘hot/heat’</td>
</tr>
<tr>
<td>2</td>
<td>/dʒ/ jengo</td>
<td>/dʒengo/</td>
<td>‘building’</td>
</tr>
<tr>
<td>3</td>
<td>/dʒ/ jiji</td>
<td>/dʒijî/</td>
<td>‘city’</td>
</tr>
<tr>
<td>4</td>
<td>/dʒ/ jua</td>
<td>/dʒua/</td>
<td>‘sun’</td>
</tr>
</tbody>
</table>

---

**Appendix D**
4. dohani /dużăn/ ‘smoke’
5. duka /dukkăn/ ‘store’

POSTNASAL VOICED STOPS:

/mb/
1. mbaazi /mbaazi/ ‘pigeon pea’
2. mbegu /mbego/ ‘grain/seed’
3. mbingu /mbingu/ ‘sky’
4. mboga /mboga/ ‘vegetable’
5. mbuzi /mbuzi/ ‘goat’

/md/
1. mdalasini /mdalasini/ ‘cinnamon’
2. mdela /mdela/ ‘coffee pot (brass)’
3. mdimu /mdimu/ ‘lime tree’
4. mdomo /mdomo/ ‘lip’
5. mdudu /mdudu/ ‘insect/bug’

/mg/
1. mgahawa /mgahawa/ ‘restaurant’
2. mgeni /mgeni/ ‘guest’
3. mgiligilani /mgiligilani/ ‘coriander plant’
4. mgomba /mgomba/ ‘banana tree’
5. mguu /mguu/ ‘foot/leg’

/nd/
1. ndama /ndama/ ‘calf’
2. ndege /ndege/ ‘airplane/bird’
3. ndimu /ndimu/ ‘lime’
4. ndoto /ndoto/ ‘dream’
5. nduaro /nduaro/ ‘striped marlin fish’

/ng/
1. ngamia /ngamia/ ‘camel’
2. nge /nge/ ‘scorpion’
3. ngiri /ngiri/ ‘worthog’
4. ngoma /ngoma/ ‘drum’
5. nguruwe /nguruwe/ ‘pig’

POSTNASAL VOICELESS STOPS:

/mp/
1- mpanzi /mpanzi/ ‘sower’
2- mpira /mpira/ ‘ball/football’
3- mpokezi /mpokezi/ ‘recipient/receptionist’
4- mpunga /mpunga/ ‘rice plant’

/mt/
1- mto /mto/ ‘river’
2- mtende /mtende/ ‘date/date palm’
3- mtindi /mtindi/ ‘yogurt’
4- mtumbwi /mtumbwi/ ‘canoe’

/mk/
1- mkanda /mkanda/ ‘belt’
2- mkeka /mkeka/ ‘mat’
<table>
<thead>
<tr>
<th></th>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-</td>
<td>nta</td>
<td>/nta/</td>
<td>‘wax’</td>
</tr>
<tr>
<td>2-</td>
<td>nti</td>
<td>/nti/</td>
<td>‘ear ornament’</td>
</tr>
<tr>
<td>3-</td>
<td>mkia</td>
<td>/mkia/</td>
<td>‘tail’</td>
</tr>
<tr>
<td>4-</td>
<td>mkulima</td>
<td>/mkulima/</td>
<td>‘farmer’</td>
</tr>
</tbody>
</table>