Title
The Phonetics and Phonology of So-Called Vowel Devoicing in Malagasy

Author
Aziz, Jake

Publication Date
2024

Peer reviewed|Thesis/dissertation

UNIVERISTY OF CALIFORNIA

Los Angeles

The Phonetics and Phonology

of So-Called Vowel Devoicing in Malagasy

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Jake Aziz

2024

ABSTRACT OF THE DISSERTATION

The Phonetics and Phonology

of So-Called Vowel Devoicing in Malagasy

by

Jake Aziz

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2024

Professor Sun-Ah Jun, Co-Chair

Professor Kie Ross Zuraw, Co-Chair

Previous descriptions of Malagasy (Austronesian, Madagascar) describe a vowel devoicing process affecting unstressed high vowels, similar to other languages said to have devoiced vowels (e.g., Japanese, Korean, Turkish). However, a closer inspection of the data reveals that most Malagasy vowels are not truly devoiced; rather, they appear to be nearly deleted, only measurable as coarticulatory residue on adjacent segments. In this dissertation, I present a phonetic and phonological study of this "devoicing" process in Malagasy, arguing that what is heard as vowel devoicing is actually complete gestural overlap of the vowel by the consonants that surround it.

The dissertation includes two phonetic studies of vowels said to be devoiced. In the first, I show that Malagasy "vowel devoicing" is, in fact, two distinct processes: in Intonational Phrase-

final positions, vowels may be truly devoiced; in non-final positions, however, unstressed vowels seem to disappear altogether, with no segmental interval clearly identified as the vowel. In the second study, I look closely at the acoustics of these non-final "devoiced" vowels, finding that traces of the underlying vowel are present in the acoustic output, even when these vowels appear to be deleted. For example, measures of spectral peak (i.e., the most amplified frequency) of /s/ reveal a lower spectral peak before underlying /u/ compared to /i/, consistent with the effects of typical consonant-vowel coarticulation. This finding suggests that non-final vowels are neither truly devoiced nor deleted in Malagasy; rather, they appear to still be present as an articulatory gesture that is overlapped and obscured by adjacent consonantal gestures.

I model this apparent gestural overlap using a theory of Articulatory Phonology, in which speech sounds are represented as "constellations" of gestures that may overlap each other in time: when a vowel's gestures overlap with adjacent consonants sufficiently, the vowel may be completely obscured, heard only as coarticulation on those consonants. A complete gestural analysis of Malagasy vowel devoicing is presented, accounting not only for the phonetic realization described but also the environment where "devoicing" is most likely to occur: high vowels devoice more than low vowels, unstressed vowels more than stressed vowels and so on. A surprising finding, confirmed quantitatively, is that the likelihood of a vowel undergoing "devoicing" is directly tied to the sonority of the consonants that surround it: vowels are more likely to "devoice" when they *follow* a sonorant but *precede* an obstruent. For this, I invoke a theory of Sonority-Driven Gestural Timing (Gu & Durvasula, 2024), arguing that the degree of overlap between two gestures is modulated by the difference in sonority between them. This work has implications for theories of phonological representation and gestural timing.

The dissertation of Jake Aziz is approved.

Nancy Hall

Bruce P. Hayes

Sun-Ah Jun, Committee Co-Chair

Kie Ross Zuraw, Committee Co-Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1

1.1. MERINA MALAGASY 3
1.1.1. Prosody 4
1.1.1.1. Word prosody 4
1.1.1.2. Phrasal prosody 6
1.1.2. Segmental Phonology 8
1.1.2.1. Consonants 8
1.1.2.2. Vowels 10
1.1.2.3. Allophony involving vowels 10
1.1.2.4. Orthography 11
1.1.3. Morphology and syntax 11
1.2. SO-CALLED "VOWEL DEVOICING" 12
1.2.1. Phonetics of devoicing 12
1.2.2. Analyzing devoicing 14
1.2.2.1. Final devoicing 14
1.2.2.2. Non-final devoicing 15
1.2.3. Vowel devoicing in Malagasy 17
1.3. EVIDENCE OF AN UNDERLYING VOWEL 19
1.3.1. Slow speech 20
1.3.2. Morpho-phonological alternations 21
1.4. OTHER PHONOLOGICAL PROPERTIES OF DEVOICING 25
1.4.1. Palatalization 25
1.4.2. Consecutive devoiced vowels 26
1.5. WHAT WILL NOT BE DISCUSSED IN THIS DISSERTATION 27

CHAPTER 2 TWO TYPES OF VOWEL DEVOICING 28

2.1. INTRODUCTION 28
2.2. FINAL VS. NON-FINAL DEVOICING: THE TYPOLOGY OF VOWEL DEVOICING 29
2.3. THE PHONETICS OF VOWEL DEVOICING 30
2.3.1. Final devoicing produces voiceless vowels 30
2.3.2. Non-final devoicing looks more like deletion 34
2.4. VOWELS NOT UNDERGOING DEVOICING 38
2.5. FINAL DEVOICING REMAINS IN SLOW SPEECH 40
2.6. INTERIM CONCLUSION: TWO TYPES OF VOWEL DEVOICING 42

CHAPTER 3 THE ENVIRONMENT OF DEVOICING: A PROBABILISTIC MODEL 44

3.1. INTRODUCTION 44
3.2. DATA COLLECTION 45
3.2.1. Modelling 47
3.3. FACTORS AFFECTING THE PROBABILITY OF DEVOICING 48
3.3.1. Categorical voicing and devoicing 48
3.3.2. Summary of the model 50
3.3.3. Vowel 51

3.3.4. Stress 51
3.3.5. Phrasal Prosody 53
3.3.6. Segmental Context 57
3.3.7. Factors not found to improve model fit 61
3.4. MODEL FIT 62
CHAPTER 4 VOWEL DEVOICING IS COARTICULATION: AN ACOUSTIC ANALYSIS 64
4.1. INTRODUCTION 64
4.2. METHODS 65
4.2.1. Task One: Real words 65
4.2.2. Task Two: Nonce words 66
4.3. SPECTRAL TRACES OF "DEVOICED" VOWELS 67
4.3.1. /s/ 68
4.3.1.1. Spectral peak 71
4.3.1.2. Spectral moments 74
4.3.1.2.1. Center of gravity 74
4.3.1.2.2. Spectral standard deviation 76
4.3.1.2.3. Skewness 79
4.3.1.2.4. Kurtosis 79
4.3.1.3. Interim summary: vowels following /s/ 81
4.3.2. /t/ 81
4.3.2.1. Spectral peak 82
4.3.2.2. Spectral moments 84
4.3.2.2.1. Center of gravity 84
4.3.2.2.2. Spectral standard deviation 84
4.3.2.2.3. Skewness 85
4.3.2.2.4. Kurtosis 85
4.3.2.3. Interim summary: vowels following /t/ 86
4.3.3. /n/ 87
4.4. DURATION 89
4.4.1. /s/ duration 89
4.4.2. /t/ aspiration duration 92
4.5. SUMMARY OF ACOUSTIC RESULTS: DEVOICING RESEMBLES COARTICULATION 97
CHAPTER 5 A GESTURAL ACCOUNT OF VOWEL DEVOICING 100
5.1. INTRODUCTION 100
5.2. BACKGROUND: ARTICULATORY PHONOLOGY 101
5.2.1. Browman & Goldstein's AP 101
5.2.2. Gafos (2002) and gestural coordination 106
5.2.3. Gestural accounts for vowel devoicing and deletion 109
5.3. A GESTURAL ANALYSIS OF MALAGASY VOWEL DEVOICING 112
5.3.1. Recall: The environment of devoicing 112
5.3.2. Basic gestural timing 113
5.3.3. Vowel height and intrinsic duration 114
5.3.4. Preceding consonants and Sonority-Driven Gestural Timing 118
5.3.4.1. Some acoustic evidence for sonority-driven timing 121
5.3.5. Following consonants and more evidence for sonority-driven timing 124
5.3.6. Prosody and prosodic gestures 129

5.4. DISCUSSION ... 133
5.4.1. Malagasy vowel devoicing: a phonetic and phonological conspiracy ... 133
5.4.2. Variation ... 134
CHAPTER 6 DISCUSSION AND CONCLUSION ... 137
6.1. SUMMARY OF FINDINGS: SO-CALLED VOWEL DEVOICING AS GESTURAL OVERLAP ... 137
6.2. AGAINST A FEATURE-BASED DELETION ANALYSIS OF DEVOICING ... 139
6.2.1. Feature-based theories struggle with opacity ... 140
6.2.2. Features do not capture sonority effects ... 143
6.2.3. A featural analysis falls short ... 146
6.3. ON THE UNIVERSALITY OF SONORITY-DRIVEN TIMING ... 146
6.4. CONFLICTING PRIORITIES IN THE GRAMMAR ... 148
6.5. PHONETIC DETAIL IN PHONOLOGICAL ANALYSES ... 150
6.6. FUTURE DIRECTIONS ... 152
6.6.1. Instrumental articulatory study ... 152
6.6.2. Syllable structure ... 153
6.6.3. Tonogenesis ... 153
6.6.4. Sound change ... 154
APPENDIX A LIST OF STIMULI USED IN CHAPTER 3 DATASET ... 155
APPENDIX B LIST OF STIMULI USED IN CHAPTER 4 TASK ONE ... 160
APPENDIX C LIST OF STIMULI USED IN CHAPTER 4 TASK TWO ... 162
APPENDIX D PRAAT SCRIPT FOR MEASURING SPECTRAL PEAK AND MOMENTS ... 163
REFERENCES ... 170

LIST OF FIGURES

Figure 1.1 Pitch track for the utterance *Tsy ananako ilay basy. Naly no manana azy.* 'I don't have a gun. It's Naly who has it.' 8
Figure 1.2 Spectrogram of *kasoka* /kasuka/ 'rubbing' in slow, careful speech with /u/ clearly voiced. 21
Figure 1.3 Spectrogram of *toky* /tu.ki/, where the stressed syllable contains a voiced /u/. 22
Figure 1.4 Spectrogram of /ahatukiana/, showing that underlying /u/ surfaces as devoiced when unstressed. 23
Figure 1.5 Spectrogram of the indicative verb *manasitrana* /manaˈsiʈʂana/, showing the voiced /i/ associated with the stressed syllable. 24
Figure 1.6 Spectrogram of the imperative verb *manasitràna* /manasiˈʈʂana/. Here, stress shifts rightward, leaving /i/ in the devoicing environment. 24
Figure 1.7 Spectrogram for *tsikaritra* /t͡siˈkariʈ͡ʐa/ 'seen for a moment'. The second tier has the phonetic transcription, showing that both /i/ are devoiced but the /k/ is still palatalized, evident by the high F2 (about 2400 Hz) and low F1 (570 Hz) at the offset of /k/, relative to typical values for /a/. 25
Figure 1.8 Spectrogram for the excerpt ...*(solo)masoko sy fitaratro*... /masuku si fitaraʈ͡ʂu/ 'glasses and mirror' showing four consecutive syllables with devoiced vowels. 26
Figure 2.1 Spectrogram of *tanàna* /taˈnana/ 'town' in the IP-final position with devoiced /a/. Below the spectrogram, the first tier shows the underlying representation of the word in addition to stress, and the second tier shows a phonetic transcription. The same format is used in all figures in this chapter. 31
Figure 2.2 Spectrogram of *zazavavy* /zazaˈvavi/ 'girl' in the IP-final position with devoiced /i/. 32
Figure 2.3 Spectrogram of the name *Ranaivo* /raˈnajvu/ in the IP-final position with devoiced /u/. 32
Figure 2.4 Spectrogram of *tanàna* /taˈnana/ 'town' in the IP-final position with weakly released devoiced /a/. 33
Figure 2.5 Spectrogram of *zazavavy* /zazaˈvavi/ 'girl' in the IP-final position with a fricated devoiced /i/. 34

Figure 2.6 Spectrogram of *dadasika* /daˈdasika/ ‘spacious’ in a non-final position, showing that the vowel /i/ is not easily demarcated after /s/. 35

Figure 2.7 Spectrogram of ...*basy sanduka*... /ˈbasi ˈsa^{n}duka/ ‘fake gun’ 36

Figure 2.8 Spectrogram of *behatoka* /beˈhatuka/ ‘nape’ in a non-final position; while it may look like there is a short voiceless vowel following the release of /t/, Malagasy stops are typically aspirated, and so here there is no additional voiceless vowel. 37

Figure 2.9 Spectrogram of *nanome* /nanuˈme/ ‘give’ in a non-final position where the “devoiced” vowel appears to be deleted. 37

Figure 2.10 Spectrogram of the pronoun *ianao* /iaˈnaw/ in the IP-final position, where the stressed diphthong /aw/ (realized as [o]) is voiced. 38

Figure 2.11 Spectrogram of the demonstrative *itỳ* /iˈti/ showing frication of the final part of /i/ in the IP-final position. 39

Figure 2.12 Spectrogram of *hataka* /ˈhataka/ ‘a request’ showing that non-final /a/ is voiced, even in a devoicing environment (in this case, between two obstruents in an unstressed syllable). The final /a/ is realized as [e] as the following word begins with /i/. 39

Figure 2.13 Spectrogram of *nokekerina* /nukeˈkerina/ ‘to be bitten (past)’ showing voiced /e/ in the non-final position. 40

Figure 2.14 Spectrogram of *zazavavy* /zazaˈvavi/ ‘girl’ in the IP-final position, said at a slower speech rate. 41

Figure 2.15 Spectrogram of *vasoka* /ˈvasuka/ ‘discoloured’ spoken at a slower speech rate, showing that each vowel is clearly voiced. 42

Figure 3.1 Raw counts for devoiced vowels by Stress type. 53

Figure 3.2 Raw counts for devoiced vowels by position of the syllable within the Intonational Phrase. 56

Figure 3.3 Raw counts for devoiced vowels for position of the syllable within the Phonological Phrase. 56

Figure 3.4 Raw counts for devoiced vowels by the sonority of the preceding consonant. 58

Figure 3.5 Coefficients for the manner of articulation of the following consonant as a predictor of vowel devoicing, with “nasal” as the reference level. 60

Figure 3.6 Raw counts for devoiced vowels by the manner of articulation of the following consonant. 60

Figure 3.7 Predicted vs. observed rates of vowel voicing based on the model in (3.2).............. 63
Figure 4.1 Example segmentation of /s/ for measurement of spectral properties. The short duration of low-amplitude noise between frication and the onset of the vowel was not included for these measurements.. 70
Figure 4.2 Spectral peak (F_M) measurements across /s/ by vowel context in the voicing environment. 73
Figure 4.3 Spectral peak (F_M) measurements across /s/ by vowel context in the devoicing environment. 73
Figure 4.4 Center of gravity measurements across /s/ by vowel context in the voicing environment. 75
Figure 4.5 Center of gravity measurements across /s/ by vowel context in the devoicing environment. 75
Figure 4.6 Spectral standard deviation measurements across /s/ by vowel context in the voicing environment. 77
Figure 4.7 Spectral standard deviation measurements across /s/ by vowel context in the devoicing environment. 77
Figure 4.8 Skewness measurements across /s/ by vowel context in the voicing environment. ... 78
Figure 4.9 Skewness measurements across /s/ by vowel context in the devoicing environment. 78
Figure 4.10 Kurtosis measurements across /s/ by vowel context in the voicing environment. ... 80
Figure 4.11 Kurtosis measurements across /s/ by vowel context in the devoicing environment. 80
Figure 4.12 Beanplot of /t/ spectral peak measurements by vowel context and voicing environment. 83
Figure 4.13 Beanplot of /t/ CoG measurements by vowel context and voicing environment. 83
Figure 4.14 Beanplot of /t/ spectral standard deviation measurements by vowel context and voicing environment. 84
Figure 4.15 Beanplot of /t/ skewness measurements by vowel context and voicing environment. 85
Figure 4.16 Beanplot of /t/ kurtosis measurements by vowel context and voicing environment. 86
Figure 4.17 F2 measurements of /a/ before underlying /ni/ and /nu/ in the devoicing environment. 88

Figure 4.18 Example of /s/ segmentation in the context of a voiced vowel, where /s/ is segmented beginning at frication and ending at the onset of voiced /a/. 90
Figure 4.19 Example of /s/ segmentation in the context of a devoiced vowel, where /s/ is segmented beginning at frication and ending at the silent period associated with the closure of /k/. 90
Figure 4.20 /s/ duration in different vowel contexts in the voicing environment 93
Figure 4.21 /s/ duration in different vowel contexts in the devoicing environment. 93
Figure 4.22 Duration of /t/ aspiration in different vowel contexts in the voicing environment. . 95
Figure 4.23 Normalized duration of /t/ aspiration in different vowel contexts in the devoicing environment. 95
Figure 5.1 Gestural score for the American English word *palm* [pʰɑm], originating in Browman & Goldstein (1989, p. 212) and modified based on Goldstein et al. (2006). 103
Figure 5.2 Hypothetical gestural score for the Malagasy nonce word /suna/. 105
Figure 5.3 Gestural landmarks (Gafos, 2002). 107
Figure 5.4 (a) satisfies ALIGN(C_1, RELEASE, C_2, TARGET) because the target of the second consonant is aligned to the release of the first. 107
Figure 5.5 Proposed basic alignment for a Malagasy CV syllable, where the onset of the vowel gesture (dotted red line) is in-phase with the onset of the consonant gesture (solid black line). 114
Figure 5.6 CV alignment predicted by Gu and Durvasula's principle of sonority-driven gestural timing. Because /ni/ has a lower sonority difference, the vowel gesture is timed earlier than for /si/. 120
Figure 5.7 F2 values for /a/ before underlying /ni/ compared to /nu/, showing that the difference between the two widens toward the end of /a/. 123
Figure 5.8 F2 values for /a/ before underlying /si/ compared to /su/. 123
Figure 5.9 Example overlap between two gestures showing the phase window within which the onset of the second gesture may fall, represented as a bell curve... 135
Figure 5.10 Gestural representation of Malagasy …/sin/… showing the phase windows for both gestural timing relations.. 136
Figure 6.1 Gestural schematic for vowel intrusion in a VC_1C_2 sequence, where V is represented as a dotted red line, C_1 as a dashed grey line and C_2 as a solid black line................................ 147

LIST OF TABLES

Table 1.1 Consonant inventory of Merina Malagasy, adapted from Howe (2019). Note that the voicing distinction has been largely neutralized. 9

Table 1.2 Non-transparent orthography in Malagasy.......... 11

Table 3.1 Factors found to affect the probability of a vowel being devoiced and the environments that show the most and least amount of vowel devoicing. 48

Table 3.2 Factors that have a 95% or higher rate of voicing or devoicing in the dataset. 49

Table 3.3 Summary of the generalized mixed effects model for Malagasy vowel devoicing, excluding factors shown to categorically predict voicing/devoicing.......... 51

Table 3.4 ANOVA results comparing position in the utterance (Utt.), in the IP (IP), and in the phonological phrase (Phon. Phrase) as predictors of vowel devoicing. 55

Table 4.1 Duration (un-normalized) of /s/ in different underlying vowel contexts in both the voicing (stressed) and devoicing (unstressed) environments. Note that /a/ is never devoiced..... 91

Table 4.2 Duration (un-normalized) of /t/ aspiration in different vowel contexts in both the voicing (stressed) and devoicing (unstressed) environments.......... 94

Table 4.3 Summary of phonetic findings, comparing various acoustic measurements in the voicing and devoicing environments. 98

Table 5.1 Factors that have a 95% or higher rate of voicing or devoicing in the dataset. 113

Table 5.2 Sonority hierarchy, adapted from Parker (2008).......... 119

Table 5.3 Rate of devoicing of high vowels in unstressed medial syllables, by segmental environment. 128

Table 5.4 Gestural scores predicted by the Principle of Sonority-Driven Gestural Timing for different segmental environments.......... 129

Table 6.1 Rate of vowel devoicing by consonantal environment. 144

LIST OF TABLEAUX

Tableau 5.1 Gestural alignment candidates for /sak/ showing that the inherently long gestures associated with /a/ evade complete overlap by adjacent consonants. 116
Tableau 5.2 Gestural alignment candidates for underlying /suk/. Here, Candidate (a) has the same violation profile as in Tableau 5.1 (c), but produces devoicing for /u/, unlike /a/. 117
Tableau 5.3 Gestural alignment candidates for /nuk/, showing that the candidate that violates neither SDT constraint produces the desired output with extensive overlap. 125
Tableau 5.4 Gestural alignment candidates for /sun/, showing that only the candidate that violates neither SDT constraint produces an output without complete overlap of the vowel. 127
Tableau 5.5 Gestural alignment candidates for /nit/ at the IP-initial boundary, showing how the π-gesture in Candidate (b) makes gestures longer and further apart. 131
Tableau 6.1 Tableaux for vowel "devoicing" if it is seen as vowel deletion constrained by MAX constraints penalizing vowel deletion following obstruents or before sonorants. These constraints predict the incorrect output for underlying /t_k/ and /n_n/. 145
Tableau 6.2 Tableaux for vowel "devoicing" if it is seen as vowel deletion that does not apply between an obstruent and a sonorant, specifically. 145

ACKNOWLEDGEMENTS

First, I want to acknowledge the Malagasy community in Montreal who made this dissertation possible. This research is truly a collaborative work with the speakers who shared their language and culture with me and the consultants who helped me prepare experimental stimuli and answered a million and one questions about Malagasy for me: Kezia Andrianintsoa, Alan Ho Thi Meng, Tojo Njataorimanga, Johann Rabisoa, Anais Rafolisy, Jessie Rambeloson, Brenda Randrianary, Aina Rajaonisaona, Prisca Rakotomalala, Anaïs Rakotomavo, Mitia Rakotorahalahy, Mia Ramilijaona, Sitraka Randriamparany, Harena Ratsitohaina, Antsa Ratsizafy and Andraina Razakamanantsoa. I want to especially thank Vololona Razafimbelo and Rija Rasolonandrasana for the many hours that they spent consulting on this project. Misaotra!

There are many other people in the UCLA and linguistics communities who have helped this dissertation come together. I am extremely grateful for the mentorship provided by my two co-chairs. I am very proud to have received my training in experimental phonetics from Sun-Ah Jun, even before I came to UCLA–she was the one who prepared me to collect intonation data during first field trip to Madagascar, and since then I have valued her expertise in each of our meetings. Sun-Ah has helped me to understand the value of detailed phonetic analysis even in theoretical phonology work, which has been an important lesson in preparing this dissertation. My other co-chair, Kie Zuraw, has been an invaluable mentor in being a phonologist. She is an encyclopedia and database of all things phonology and always has the exact paper in mind to help me answer the latest question that's had me stumped. In fact, it was Kie who first directed me to Articulatory Phonology and inspired the analysis that emerged in this dissertation. I am also thankful for the mentorship of both Sun-Ah and Kie in teaching me how to be an academic: from them, I've learned how to ask questions, seek answers, hypothesize and re-hypothesize and communicate my findings. Thank for reading the million drafts, chapters, abstracts and PowerPoint slides that eventually became this dissertation.

I am also incredibly thankful for my two other committee members, Nancy Hall and Bruce Hayes. Nancy's own work on vowels was a big inspiration for me, and her comments in many ways challenged me to fine-tune my analysis. My meetings with Bruce were always productive and encouraged me to think more deeply about my work's place in the wider theoretical landscape. Bruce was also the one who first trained me in quantitative approaches to phonology, which form an important part of this dissertation. Thank you both for your guidance and advice!

There were many other people in the linguistics community who I must thank: the Malagasyist linguists who have inspired my work in many ways, even if indirectly: Ed Keenan, Ileana Paul, Lisa Travis and many others; audiences at AMP 2022, AFLA 30, UCLA's phonetics and phonology seminars and UofT's Speech Group; Michael Wagner and the staff at McGill for setting me up with a recording booth in Montreal; my research assistants Pauline Antonio-Nguyen, Dhanya Charan, Megan McLane and Noah Rubio Moreno.

Of course, I must acknowledge the many people at UCLA who have made my time here not only academically enriching but an amazingly fun place to spend my 20s. My cohort-mates Colin Brown, Noah Elkins, Jinyoung Jo and Jennifer Kuo were among my first friends here and I've learned something from each of them. I'm thankful for my other friends in the department who have talked about this work with me in the reading room, or over dinner or at a party: John McGahay, Jahnavi Narkar, Isa Cabrera, Hunter Johnson, ZL Zhou, Jian-Leat Siah and Canaan Breiss.

The people outside of my academic life have also been important players in my completing the PhD. My parents have supported me even when they were not sure exactly what I do (my mom now carries a sticky-note with my dissertation title on it to tell her friends) and my friends outside of academia have reminded me of who I am beyond my work: Nicole Skinner, all of my friends back in London and Toronto, friends from trivia and bowling here in LA, friends in other departments. Thank you all!

This dissertation was supported by UCLA Dissertation Year Fellowship.

VITA

2020	MA in Linguistics
	University of California, Los Angeles
2018	BA in Linguistics and Spanish Language and Linguistics
	University of Western Ontario

PUBLICATIONS

Aziz, J. (forthcoming). Imperatives and the Syntax-Prosody Interface in Malagasy. *Natural Language and Linguistic Theory*.

Aziz, J. & Jun, S.-A. (forthcoming). The intonational phonology of Malagasy. In S.-A. Jun & S. D. Khan (Eds.), *Prosodic Typology III: The Phonology of Intonation and Phrasing.* Oxford University Press.

Aziz J., Machado, V., Valdivia, C., Swiderski, N., Rafat, Y., Stevenson, R. & Rao, R. (2022). The intonation of absolute questions in Argentinean- and Venezuelan- Canadian heritage speakers of Spanish: Investigating parental and English influences. *Spanish as a Heritage Language*, *2*(1), 60-90.

Aziz, J. & Elkins, N. (2022). Malagasy /nr/ strengthening within and across prosodic boundaries. In *Proceedings of the Twenty-Eighth Meeting of the Austronesian Formal Linguistics Association (AFLA)*. University of Western Ontario, 2022.

Aziz, J. (2020). Intonational phonology of Malagasy: Pitch accents demarcate syntactic constituents. In *Proc. 10th International Conference on Speech Prosody 2020* (pp. 201-204). Accessible from https://www.isca-speech.org/archive/SpeechProsody_2020/pdfs/236.pdf

Aziz, J. & Paul, I. (2019). The Intonation of Malagasy: A preliminary look. In S. Calhoun, P. Escuerdo, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of the Phonetic Sciences, Melbourne, Australia 2019* (pp. 3792-3796). Canberra, Australia: Australasian Speech Science and Technology Association Inc. Accessible from https://assta.org/proceedings/ICPhS2019/papers/ICPhS_3841.pdf.

Chapter 1
Introduction

Are Malagasy vowels disappearing? Certain vowels in Malagasy are widely reported to have nearly disappeared from surface forms. Wikipedia user *Jagwar* transcribes the Northern Malagasy city of Antsiranana as [antsʲˈraŋanə̥], the peninitial high front vowel reduced to coarticulation on /s/ and the final vowel marked with a voiceless diacritic (*Antsiranana*, 2010). Malagasy language learning app *Assimil* teaches that “en position inaccentuée, les voyelles *i* et *o* ne se prononcent pas” [in the unaccented position, the vowels <i> and <o> aren’t pronounced] (*Assimil Malgache*, n.d.). In the academic literature, processes variably described as *devoicing*, *deletion*, or otherwise are reported to affect vowels across the Malagasy dialects, particularly those that are in final or unstressed syllables (e.g., Dahl, 1952; Howe, 2019; O’Neill, 2015). This dissertation investigates this process in one variety of Malagasy, Merina, posing three primary research question: (1) what is the phonetic reality of Malagasy “devoiced” vowels (i.e., are Malagasy vowels undergoing devoicing, deletion, or something else)? (2) in what phonological environment does devoicing occur? and (3) what in the grammar motivates vowel devoicing?

As a preview, the dissertation will show that: (1) Malagasy “devoiced” vowels are realized as coarticulation on adjacent consonants—i.e., neither deletion nor true devoicing, (2) non-final vowel devoicing is sensitive to prosodic and segmental factors, being most likely when the vowel is (a) a high vowel, (b) in an unstressed, medial syllable and (c) following a sonorant and preceding an obstruent, (3) under a gestural theory of phonology, Malagasy vowels undergo extreme gestural overlap, which is especially driven by what I have termed Sonority-Driven Gestural Timing. The structure of the dissertation is as follows: in this chapter, I outline the literature on Malagasy and vowel devoicing/deletion processes. In Chapter 2, I introduce the vowel devoicing data for Malagasy and present a descriptive account, concluding that “devoicing” is best characterized as two distinct processes. Chapter 3 presents a detailed acoustic analysis of the non-final “devoicing” process, showing that it is acoustically similar to consonant-vowel coarticulation. Chapter 4 is a probabilistic statistical model of the phonological properties affecting devoicing.

Chapter 5 presents an analysis of the phenomenon using Articulatory Phonology in which non-final “devoicing” arises as a result of extensive overlap of the vowel’s gestures with adjacent consonants; in that chapter, I invoke the idea of Sonority-Driven Gestural Timing, which claims that the timing of two segements’ gestures is modulated by the difference in sonority between them. The chapter puts forward a structure of the grammar where devoicing arises as a conspiracy of three factors: the inherent duration of the segments involved, prosodic gestures that lengthen prominent syllables and sonority-driven gestural timing. In short, the phonological grammar assumed in this dissertation is summarized in three parts, based on the principles of Articulatory Phonology (Browman & Goldstein, 1986):

(i) **The Lexicon**

The lexicon takes articulatory gestures as its primitive organizing unit. Malagasy words are made up of constellations of gestures (i.e., segments) that have inherent durations; high vowels have shorter gestures than low vowels, etc.

(ii) **The Grammar**

The phonological grammar constrains the alignment of gestures to one another, presumably in language-specific ways. For example, a principle of Sonority-Driven Gestural Timing modulates the timing of gestures with respect to one another based on sonority; prosodic gestures lengthen and slow gestures at prominent points, including stressed syllables and those at phrase boundaries.

(iii) Production

A task dynamic model (Saltzman, 1986) dictates the trajectory of the articulators associated with each gesture; overlapping gestures are coarticulated when possible.

1.1. Merina Malagasy

Malagasy is a Western Austronesian language spoken by several million people in Madagascar as well as communities in the nearby French Department of Mayotte and the diaspora. The variety under discussion in this dissertation is Malagasy as it is spoken in and around Madagascar's capital, Antananarivo. Ethnologue (*Malagasy*, n.d.) calls Malagasy a "macrolanguage," constituting a dialect continuum with varying degrees of mutual intelligibility between varieties (Bouwer, 2007). The variety spoken in and around the capital is typically called *Merina*, after the Merina people who are indigenous to the region, and in the literature, it is sometimes referred to as *Plateau Malagasy*, *Central Malagasy*, *Ankova* or *Hova*. Further, Merina forms the basis of the standardized variety of Malagasy, *Malagasy ôfisialy* 'Official Malagasy,' and so Merina is frequently referred to simply as *Malagasy*. While vowel devoicing or deletion processes have documented for several varieties (to be reviewed in Section 1.2.3), the dissertation specifically presents data from speakers of Merina; going forward, mention of the "Malagasy" language will refer to the language spoken by these Merina speakers unless otherwise specified. In this section, I will present an overview of the grammar of Merina; certain other details will be withheld until they become relevant in other chapters.

1.1.1. Prosody

1.1.1.1. Word prosody

Malagasy is said to allow only V and CV syllables (e.g., Rasoloson & Rubino, 2004), with some loan words excepted (e.g., *plastika* 'plastic'). One possible exception to this prohibition against codas is nasals, as Malagasy allows nasal-stop and nasal-affricative sequences (such as *lamba* 'clothing' and *antsa* 'singing'). However, these sequences are often analyzed as pre-nasalized stops (e.g., Howe, 2019) because of their presence in word-initial position, such as *ntaolo* 'ancestors' and *ngidy* 'bitterness'. Additionally, it has been suggested that Malagasy has some underlying word-final consonants, which neutralize to either /n/, /k/, or /t͡ʂ/ (Erwin, 1996). This violation of the constrain on codas is repaired in Merina by epenthesis of the vowel /a/. The word *kasoka* /kasuka/, then, is said to have the underlying form /kasuk/. Words with this epenthetic vowel are called *weak stems* and have a distinct stress pattern.

The prototypical position for primary stress is penultimate: /mu.raˈmu.ra/ 'slow', /maˈfa.na/ 'warm', /miˈte.ni/ 'speak'. Weak stems have stress on the antepenult: /ˈka.su.ka/ 'rubbing', /ma.haˈfa.ⁿta.t͡ʂa/ 'recognize', /ˈma.si.na/ 'holy'. Given the hypothesis that the final vowel in weak stems is epenthetic, we can make the generalization that basic stress is assigned to the penultimate syllable of the underlying form. Underlying forms like /ka.suk/ have stress assigned regularly, /ˈka.suk/, before epenthesis and resyllabification: /ˈka.su.ka/. The result is a small number of minimal pairs in Malagasy, such as /ˈta.na.na/ 'hand' and /taˈna.na/ 'town': the former is a weak stem and so has antepenultimate stress, while the latter is not and has the usual penultimate stress assignment.

Additionally, certain words have final stress. Single-syllable words are often stressed (e.g., /ˈna/ 'or', and compounds that end in a stressed single-syllable root maintain stress on that syllable (e.g., /va.vaˈfu/ ‘solar plexus’, which is made up of /ˈvava/ ‘mouth’ + /ˈfu/ ‘heart’. Certain loan words, especially those from French, have final stress, such as /su.kuˈla/ ‘chocolate’, which Martin (2005) suggests is made possible because those words were borrowed into Malagasy as pseudo-compounds. Diphthongs tend to attract stress, and so words ending in diphthongs have final stress: /mi.laˈlaw/ ‘play’, /iˈⁿɖʐaj/ ‘again’. Words ending in /e/ also have final stress, such as /ma.nuˈme/ ‘give’. For these observations, we might revise our general stress rule to say that primary stress falls on the syllable that contains the penultimate *mora* of the underlying form, as suggested by Erwin (1996); this requires us to posit that /e/ is bimoraic. Finally, a small number of lexical items have final stress not otherwise explained, including various demonstratives and locative adverbs: /iˈti/ ‘this (nearby)’, /aˈri/ ‘there (far away)’.

Malagasy has also been claimed to have secondary stress on every second syllable to the left of main stress (e.g., /ˌma.naˈsi.t͡ʂa.na/; Rasoloson & Rubino, 2004), but the acoustic evidence for this is more scarce in the literature; in an unpublished manuscript, Pearson (1994) analyses Malagasy stress impressionistically, citing duration as a correlate of stress. Pearson also cites high vowel devoicing as a correlate of stress (occurring only in unstressed syllables), but, given that the aim of the dissertation is to examine the effect of phonological factors like stress on devoicing, we can’t make any assumptions about the status of secondary stress based on the presence or absence of vowel devoicing alone. Pearson observes that secondary stress is somewhat-regularly assigned on alternative syllables, starting at the stressed syllable and moving leftward. For *rahampitso* /raha^{m}pitsu/ ‘tomorrow’, for example, primary stress is assigned to the penult /^{m}pi/, and secondary stress two syllables to the left: /ˌra.haˈ^{m}pi.tsu/. He notes that some suffixes seemingly shift stress, including secondary stress, while others don’t. The imperative morpheme /-a/ changes the stress pattern of the entire word: the indicative verb *mahazo* /mahazu/ ‘obtain’ has stress on the penult only: /maˈha.zu/. With the imperative suffix added, *mahazoa* /mahazua/, primary stress now appears on a different syllable, with secondary stressed assigned leftward from that: /ˌma.haˈzu.a/. Other suffixes, however, do not move secondary stress. One such suffix is the second-person plural genitive /-(n)areu/. In the unsuffixed form, the verb *fantatra* /fa^{n}taʈʐa/ ‘known’ has antepenultimate stress: /ˈfa.^{n}ta.ʈʐa/. In the suffixed form, /fa^{n}taʈʐ-areu/, primary stress appears on the penult, and given the usual pattern of Malagasy stress, we would expect secondary stress to appear two syllables to its left. In fact, this is not true, and secondary stress appears on the initial syllable, the same syllable where primary stress appeared in the unsuffixed form: /ˌfa.^{n}ta.ʈʐaˈre.u/. We can take this to mean that the interaction of morphology and phonology varies between morphemes: in the case of the imperative, all stress is assigned following suffixation; for the second person plural genitive, secondary stress is seemingly assigned before suffixation, but primary stress after.

The result is that secondary stress is, according to Pearson, predictable: for most words, secondary stress appears leftward from primary stress on alternating syllables; with certain suffixes, however secondary stress is faithful to the stress pattern of the root, even when primary stress has shifted. The suffixes with the second pattern include the second person singular /-(n)ao/ and plural /-(n)areu/ genitives as well as the first person plural exclusive genitive /-(n)aj/.

1.1.1.2. Phrasal prosody

Malagasy has two levels of prosody above the word: the phonological phrase (sometimes called the intermediate phrase) and the Intonational Phrase. Phonological phrasing corresponds to syntactic phrasing, with the syntactic predicate and subject each forming a phonological phrase, as well as predicate-external adjuncts like adverbials; this has been shown for declaratives and questions (Aziz, 2020) and argued for imperatives (Aziz, forthcoming). The phonological phrase is marked with a rising pitch accent on its rightmost stressed syllable (Barjam, 2003; Dahl, 1952; Raoniarisoa, 1990) followed by an optional falling boundary tone.

The Malagasy Intonational Phrase (IP) generally corresponds to the entire clause and is marked with a falling boundary tone and optional pause at the right edge. The IP-final boundary tone is frequently truncated when there is not enough voiced phonological material on which to realize the tone; for example, if the final syllable is already occupied by a pitch accent, or in many cases where the final vowel is devoiced (Aziz, 2020). In these cases, the IP ends on a high tone (marked with % to represent the underlying IP boundary).

Longer utterances and complex sentence types may be divided into multiple IPs along clausal lines. Figure 1.1, for example, depicts the spectrogram and pitch track of an utterance containing two separate clauses: *Tsy ananako ilay basy* ‘I don’t have a gun.’ and *Naly no manana azy* ‘It’s Naly who has it.’ In the figure below, the bold line represents the fundamental frequency (associated with pitch). The second tier shows a broad phonetic transcription for the Malagasy words and the fourth tier shows a phonetic transcription of the intonational tones. The first clause, a passive, has two syntactic phrases below the clause: the predicate *tsy ananako*, which we might translate into English as ‘had by me’ and the subject *ilay basy* ‘the gun’. Each of these syntactic phrases has a steep rising pitch accent, labelled L+<H* on the final stressed syllable of the phrase. Additionally, the predicate is marked with a low tone (annotated L-) at its right edge. The second clause, a focus construction, also consists of two phrases directly below the clause: the focussed subject noun phrase *Naly* (a proper name) and the following predicate *no manana azy*.[1] As with the initial clause, the syntactic structure of this second clause are marked with intonation: rising

[1] Paul (2001) argues that Malagasy focus constructions are, in fact, pseudo-clefts, where *Naly* is a copular predicate and *no manana azy* is a headless relative subject. In any case, the two phrases are distinct syntactic phrases marked with prosody.

pitch accents on the rightmost stressed syllable of each and a (shallow) low boundary tone at the right edge of the IP. Later in the dissertation (in Chapter 3), I will show that Malagasy vowel devoicing processes are sensitive to phrasal prosody; assumptions about prosodic structure are based on the syntax-prosody correspondence described here.

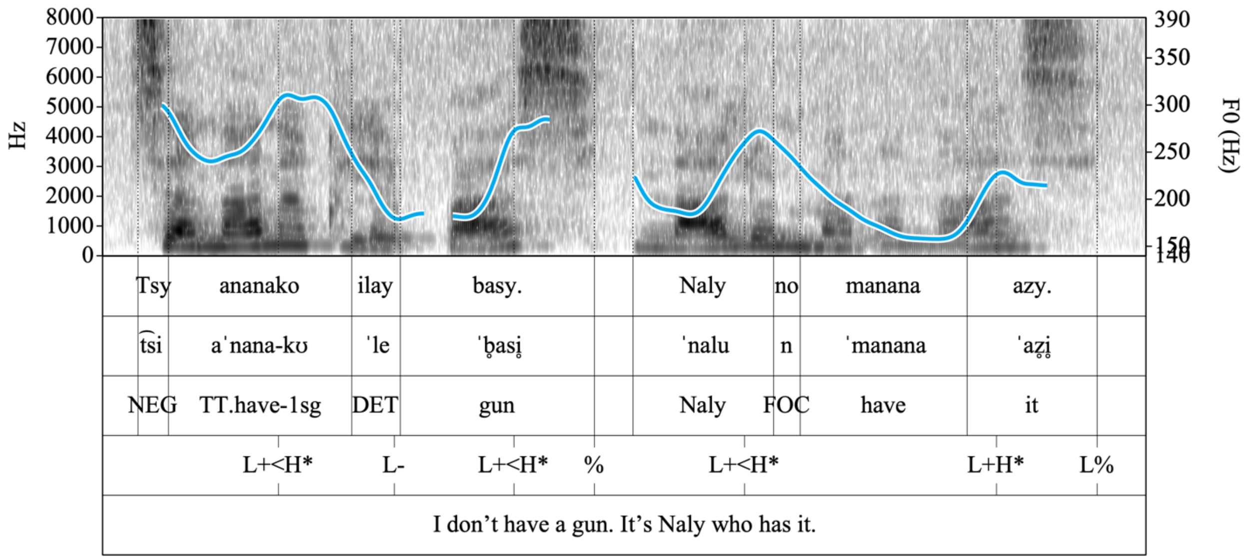


Figure 1.1 Pitch track for the utterance *Tsy ananako ilay basy. Naly no manana azy.* 'I don't have a gun. It's Naly who has it.'

1.1.2. Segmental Phonology

1.1.2.1. Consonants

Merina has 29 consonants, summarized in Table 1.1 below, adapted from Howe (2019). This includes alveolar and retroflex affricates, pre-nasalized stops and affricates and a voicing contrast for most obstruents. However, as Howe (2017) identifies, the voicing contrast has been largely neutralized in Merina and replaced with a tonal distinction. While both voiced and voiceless obstruents are voiceless in most prosodic positions, syllables with historically-voiced obstruents are now associated with a low lexical tone, and voiceless obstruents with a high tone.

Howe reports that the voicing contrast has been neutralized for both oral obstruents and the oral portion of the pre-nasalized obstruents.

	Bilabial	Labio-dental	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Plosive	p b		t d				k g	
Pre-nasalized plosive	^{m}p ^{m}b		^{n}t ^{n}d				^{ŋ}k ^{ŋ}g	
Affricate			t͡s	d͡z	ʈ͡ʂ ɖ͡ʐ			
Pre-nasalized affricate			^{n}t͡s	^{n}d͡z	^{ɳ}ʈ͡ʂ ^{ɳ}ɖ͡ʐ			
Nasal	m		n	r/ɾ				
Trill/Flap								
Fricative		f v		s z				h
Approximant								
Lateral approximant			l					

Table 1.1 Consonant inventory of Merina Malagasy, adapted from Howe (2019). Note that the voicing distinction has been largely neutralized.

Two phonemes are worth discussing in further detail. Howe (2019) states that /h/ is frequently elided in Merina, and I find this to be true in my own field research as well. The result is cases of vowel hiatus that would otherwise be resolved (see below for more discussion on Malagasy vowels). For example, the verb *mahamasaka* /ma.haˈma.sa.ka/ 'cook until well done' is frequently produced as [ma.aˈma.sa.ka]. The rhotic (transcribed as /r/ in the chart above) is highly variable between and within speakers. While the exact distribution (allophonic and sociolinguistic) of these different sounds is not well documented, in my own work I have come across the following productions for the phoneme represented by orthographic <r>: alveolar trill [r], alveolar tap [ɾ], voiced uvular fricative [ʁ] or approximant [ʁ̞], voiced retroflex [ʐ] or postalveolar [ʒ] fricative and something resembling a voiced alveolar tapped fricative [ɾ̝]. Howe (2019) also reports the use of the voiceless uvular fricative [χ] in Merina; given the immense amount of variability in just the present study's datasets, it is possible that even more variation is present in the dialect.

1.1.2.2. Vowels

Merina has four canonical monophthongs: /a, e, i, u/. /o/ also appears in some loan words and the vocative particle *ô*. In unstressed syllables, the monophthongs (when voiced) tend to centralize: /i/ to [ɨ] or [ɪ], /u/ to [ʊ], /e/ to [ɛ] and /a/ to [ə] (Howe, 2019). Additionally, certain authors (Howe, 2019, who also cites Rakotofiringa, 1982) argue that [o] may have phonemic status in native Malagasy roots written with orthographic <oa> (usually /ua/) and <ao> (usually /aw/). [o] and the diphthongs are in near-complementary distribution, with the diphthongs in the word-final position and [o] elsewhere; however, Howe shows that certain high-frequency lexical items are exceptions, such as *vao* /vo/ 'just'. In Section 1.1.2.1, I discussed how diphthongs attract stress; this is true even if the orthographic diphthongs are produced as monophthongal [o], as in *izao* /iˈzo/ 'now'.

As already mentioned, Merina has at least two rising diphthongs: /aj/ and /aw/. Just as /aw/ and /ua/ may be realized as [o], discussed above, /aj/ and /ia/ may coalesce to be realized as [e]. For other sequences of vowels, it is controversial whether they form diphthongs or are simple cases of hiatus (see Howe (2019) for an overview). When two unstressed vowels of the same quality are adjacent, only one is produced (Howe, 2019). For example, in the sequence …*rivotra avaratra*…, only a single /a/ is produced at the boundary of the two words. The following is a complete list of attested vowel sequences: /ia/, /iu/, /ie/, /eu/, /ei/, /ue/, /ui/, /ua/.

1.1.2.3. Allophony involving vowels

It has long been documented that velar consonants and /h/ are palatalized after /i/ in Merina (Griffiths, 1854; Howe, 2019; G. W. Parker, 1883) and in my own fieldwork I have occasionally observed it after /e/. The name of Madagascar itself exhibits this: /madagasikara/ may be realized as [ma.da.ga.siˈkʲa.ra]. In some cases, palatalization fails to apply, though the motivation is unclear. Howe (2019) suggests that stressed syllables are more resistant to palatalization, giving *mikatona* [miˈka.tu.na] ‘be closed’. However, in my own dataset, the verb *mikarama* ‘work’ also resists palatalization, despite the velar appearing in an unstressed syllable: [mi.kaˈra.ma]. It’s possible that palatalization is more likely to occur within the stem; for both *mikatona* and *mikarama*, the conditioning /i/ is across a morpheme boundary, as part of the verbal prefix /mi-/. However, this is not true of all active verbs: *mikasa*, on the other hand, does frequently apply palatalization: [miˈkʲa.sa] ‘intend’.

1.1.2.4. Orthography

The Malagasy orthography is mostly transparent to the Merina phonology, and most consonants and vowels are pronounced as the corresponding IPA symbol (with some phonetic variation, outlined in Howe, 2019). Orthographic nasal-obstruent sequences will be transcribed as pre-nasalized (homorganic) obstruents. Other exceptions are highlighted in Table 1.2 below. Throughout the dissertation, Malagasy will be occasionally represented in its orthographic form, and so worth emphasizing is the orthography for vowels: orthographic <o> represents the high back vowel /u/, and in final position, /i/ is written as <y>. The word *ony* ‘river’, then, is said /uni/.

Orthography	**IPA**
tr	ʈ͡ʂ
dr	ɖ͡ʐ
j	d͡z
y	i
o	u
ô, ao	o

Table 1.2 Non-transparent orthography in Malagasy.

1.1.3. Morphology and syntax

Much of the attention that Malagasy has gotten in the generative linguistics literature is for its morphosyntax, which has several typologically interesting features. Most Malagasy varieties have VOS word order, described in the syntax literature as being derived by predicate raising (e.g., Pearson, 2018); the simple sentence in (1.1), for example, shows the predicate preceding its subject.

(1.1)	Mana-dio	ilay	trano	i	Sariaka.
	AT-clean	DET	house	DET	Sariaka
	'Sariaka cleans the house.'				

The Malagasy verb is inflected for "voice," which depends on the semantic role of the subject, similar to some other Austronesian languages (Keenan, 1976). In (1.2), the subject is the Agent of the verb, and so the verb is marked with an active (also called Actor Topic; AT) prefix, *mana-*; however, various arguments can appear in that position. If the subject is a Theme or Recipient, the verb is marked with passive (Theme Topic; TT) morphology (usually a suffix *-ina*/*-ana* or the prefix *a-*). Most other thematic roles fall under the Circumstantial Topic (CT) voice, often marked with both a prefix and a suffix. The result is that most roots have multiple inflected verbal forms: compare the active verb in (1) with the passive (2) and circumstantial (3):

(1.2)	Diov-in'	i	Sariaka	ilay	trano.
	clean-TT	DET	Sariaka	DET	house
	'The house is cleaned by Sariaka.'				

(1.3)	Andiovana	an'	ilay	trano	ilay	famafa.
	CT-clean-CT	there	DET	house	DET	broom
	'The broom is used to clean the house'					

1.2. So-called "vowel devoicing"

1.2.1. Phonetics of devoicing

The term *vowel devoicing* is frequently deployed by linguists to describe some process where an underlying vowel neglects to be produced as a full, voiced vowel. This has been reported for Japanese (e.g., Tsuchida, 1997 among many others), Korean (Jun et al., 1997), Brazilian

Portuguese (Meneses & Albano, 2015), Cheyenne (R. Vogel, 2021) and many, many more. Phonetically speaking, we might expect that *vowel devoicing* entails some segment that is articulatorily equivalent to a modal vowel except that it lacks vibration of the vocal folds; in IPA, we transcribe such segments with a diacritic (e.g., [ḁ]). Indeed, truly voiceless vowels have been reported using phonetic evidence including in Mexico City Spanish (Dabkowski, 2018). However, in other cases, the existence of vowel devoicing is either asserted without evidence or a misnomer, in fact referring to some other phonetic or phonological phenomenon.

For Japanese, which is said to have devoiced vowels between voiceless obstruents, both Jun and Beckman (1993) and Tsuchida (1997) show spectrograms of purported devoiced vowels with no trace of a vowel at all; rather, in a Japanese word like /sɯki/, the frication of /s/ seems to transition directly to the stop closure of /k/, acoustically resembling a [sk] cluster. In Turkish, too, Jannedy (1995) shows examples where “the vowel has completely disappeared” between two voiceless consonants.

In fact, it is difficult to find detailed phonetic accounts of so-called vowel devoicing processes where the data actually show voiceless vowels; instead, the term “vowel devoicing” frequently refers to some other vowel reduction process that occurs next to voiceless consonants. It’s easy to see where the nomenclature comes from: when a vowel seemingly disappears next to a voiceless consonant, it may be perceptually similar to a voiceless vowel, especially following an aspirated stop. For Malagasy, the term “vowel devoicing” immediately rings alarm bells because it freely happens between sonorants (to be discussed in more detail in Chapter 3). As we will see in Chapter 2, in a word like *nanome* /nanume/ ‘give (past)’, the vowel seems to disappear between the nasals, resembling [nanme]. In that word, there is nothing “voiceless” about the disappearing /u/ at all.

There are other languages in addition to Malagasy whose unstressed high vowels seem to undergo a process that is acoustically similar to "vowel devoicing", except that it applies to vowels even next to voiced segments. Uspanteko, for example, has vowel "deletion" between sonorants, among other environments: Bennett et al. (2023) give examples like [músmul] ~ [músml] 'light rain' (#14e), showing how the high vowel /u/ disappears between /m/ and /l/. Bennett et al. (2023) arrive at a gestural-overlap account of Uspanteko that is fundamentally similar to Korean vowel devoicing, for example, but in Korean it is called *devoicing* because it only happens next to voiceless segments, while in Uspanteko, it is called *deletion*, presumably because it can happen between two segments of any voicing or sonority specification.

The phonetic reality of so-called vowel devoicing matters, not just for descriptive accuracy, but also for the implications for a phonological analysis of devoicing. A language whose vowels truly do have voiceless allophones warrants a different approach than if those vowels are simply deleted or otherwise. In the next section, I will outline the different approaches that have been taken to describe vowel devoicing processes.

1.2.2. Analyzing devoicing

1.2.2.1. Final devoicing

Analyses of vowel devoicing typically vary depending on the prosodic position targeted for devoicing. Final devoicing has been proposed to be caused by aerodynamic factors. In order to sustain voicing, subglottal pressure must be higher than oral pressure so that air flows through the glottis (van den Berg, 1958); devoicing, then, could arise due to a rise in oral pressure in the final position (e.g., by increased oral constriction, suggested by Fagyal and Moisset (1999)) or by a drop in subglottal pressure, which is consistent with findings that subglottal pressure decreases over the course of an utterance (Collier, 1975). Gordon (1998) suggests the latter as an explanation for final vowel devoicing, analogous to aerodynamic explanations of final obstruent devoicing (e.g., Smith, 1997; Westbury & Keating, 1986). However, this is not totally explanatory, as “final” devoicing does not necessarily refer to the utterance- or even phrase-final position but may include word-final vowels. However, descriptions of devoicing may conflate the word-final and utterance-final position, depending on the method of data collection (i.e., if words were elicited in isolation). A description of vowel devoicing in Oromo, for example, describes voiceless vowels in the “word final position”, but transcribes the words only in isolation (Dissassa, 1980).

For other languages, like Cheyenne (R. Vogel, 2021), there is evidence that final devoicing has been phonologized, even if motivated by phonetic factors at an earlier stage in the language. Cheyenne has both phrase-final vowel devoicing and “penultimate devoicing”, which affects penultimate vowels when they precede a voiceless consonant and only in certain lexical items. Vogel gives [vóhpomaʔo̥htse] ‘salt’ and [mo̥hoʔohtse] ‘part of Ursa Major’ as a near-minimal pair. Vogel proposes that words with penultimate devoicing are underlyingly consonant final; penultimate devoicing arises at a stage in the derivation where the affected vowel is in the word-final syllable. The final vowel (e.g., the [e] in [vóhpomaʔo̥htse]) is epenthesized only after the vowel is devoiced, giving the appearance of penultimate devoicing. And so, while a drop in sub-glottal pressure is a plausible phonetic explanation for final vowel devoicing, particularly at the utterance edge, it may not be viable for every language.

1.2.2.2. Non-final devoicing

Non-final vowel devoicing is usually explained as either a categorical phonetic process or a consequence of gestural overlap with adjacent glottal gestures. For example, feature-based theories of phonology are sometimes employed to explain voiceless vowels: devoicing is considered a categorical process by which vowels take on one or more features from adjacent consonants, causing the vowel to emerge as a phonetically devoiced allophone. In Japanese, whose vowel devoicing process has received perhaps more attention than any other language, featural analyses have been proposed by some authors. In Japanese, most high vowels are devoiced between voiceless obstruents, as in [su̥kuu] 'save'. McCawley's (1968) analysis has vowels in the devoicing environment (between voiceless obstruents) assimilating to the [-voice] feature of the adjacent consonants, a seemingly straightforward approach, but one which does not account for the fact that Japanese high vowels are voiced between two voiceless fricatives (e.g., in [susumu] 'proceed').

Tsuchida (1997), citing evidence against the existence of [-voice] as a possible feature, instead suggests that the relevant feature is [spread glottis]. For her, a constraint requires that Japanese high vowels be specified for [spread glottis], which is dominated by another constraint prohibiting [spread glottis] high vowels when adjacent to a voiced segment. In contrast to McCawley's analysis, Tsuchida manages to account for the voicing between fricatives as avoiding an Obligatory Contrast Principle violation: fricatives, which are also specified for [spread glottis], cannot appear next to a [spread glottis] vowel, and so the vowel is voiced. While these analyses of Japanese differ in the relevant feature, in all cases, these authors assume that vowel devoicing is derived by some feature being added or its value being changed.

One purported advantage of the featural account is that it is a more purely "phonological" analysis in that it does not result from an acoustic consequence of co-articulation, as with the gestural overlap account. This phonologization of the devoicing process allows it to interact with other features of the phonology; for example, multiple vowels in the devoicing environment tend to not all be devoiced when in consecutive syllables, as in [ki̥ʃitsu] 'temperament', which Tsuchida

(2001) also attributes to the Obligatory Contour Principle: the grammar prohibits sequential syllables with a [+spread glottis] specification.

But feature-based theories are less helpful for cases like Korean, where the process described as vowel devoicing does not actually involve a voiceless vowel segment. Perhaps more commonly invoked for this type of devoicing is a more "phonetic" approach, where vowel devoicing arises not because of some effort of the grammar but because of the natural consequences of articulation. Such an analysis has been summoned for many of the languages reported to have devoicing (e.g., Lezgi, Chitoran & Iskarous, 2008; Turkish, Jannedy, 1995; and Korean, Jun et al., 1997; and others). The assumption made for these languages is that speech sounds are made up of "gestures" in the spirit of Articulatory Phonology (Browman & Goldstein, 1986). The story goes that vowel devoicing arises when the vowel gestures are overlapped by the gestures of adjacent voiceless consonants; since voiceless consonants have a glottal gesture (associated with voicelessness), the overlap of that gesture with the vowel causes a voiceless vowel. This is, in essence, what I will claim for Malagasy in Chapter 5; a more detailed review of the literature on gestural overlap appears in that chapter.

1.2.3. Vowel devoicing in Malagasy

Descriptions of vowel devoicing or deletion in Malagasy have appeared throughout the literature, even in early grammars. Most of these early references describe devoicing in final positions: Griffiths (1854) states that word-final unstressed vowels are "slightly and indistinctly sounded"; Parker (1883) likewise describes final /i/ as having a "lighter sound", sometimes "mute"; and Rahidy (1895) says that word-final unstressed /a/ "se prononce faiblement" [is pronounced weakly]. Several grammars make reference to devoicing specifically in the final syllable of the weak stems (Parker, 1883; Caussèque, 1886; Ferrand, 1909). More recently, authors that reference some sort of weakening or devoicing of final vowels include Howe (2019), Rahajarizafy (1960) and Rasoloson and Rubino (2004).

The devoicing of non-final vowels is apparently a more recent innovation: Griffiths (1854) remarked that Merina was “abounding with vowels” at the time and Parker (1883) asserted then that “in *speaking*, each vowel must be clearly pronounced” (emphasis his). Since then, a process variably described as devoicing or deletion has been described for medial unstressed vowels in various descriptions. Rakotofiringa (1969) notes that the disappearance of unstressed high vowels leaves the impression that Malagasy has consonant clusters, going on to transcribe various Malagasy vowels with the voiceless diacritic; Keenan and Polinsky (1998) state that “unstressed vowels, other than *e*…are often devoiced”; Rasoloson and Rubino (2004) remark that /i/ and /u/ are often devoiced both in “final position” and between consonants; Howe (2019) says that vowels are “often devoiced” word-medially as well as finally, though her transcription for those vowels varies between voiceless vowels (e.g., [ˈv̥àdɪ̥kʲu̥]) and secondary articulations on preceding consonants (e.g., [miˈtáfʲ]). O’Neill (2015), discussing the Betsimisaraka dialect, references vowel devoicing but acknowledges that the process may be analyzed as vowel deletion instead. Aziz and Elkins (2022) likewise refer to the process as vowel deletion.

A more complete description of medial vowel devoicing comes from Dahl (1952), who describes it for the penultimate syllable of weak stems in Merina. In that position, he says, /i/ and /u/ (but never /a/) have a tendency to “fall” after a nasal, affricate or fricative. He also describes vowel devoicing between voiceless stops, such as the second /u/ in *tompoko* /ˈtumpuku/ ‘sir’. This is a tendency, not a strict rule, he emphasizes, but one that is most likely to occur between two /n/. Dahl suggests that these high vowels are not completely deleted, and minimal pairs like *anina* /ˈanina/ ‘stopping’ and *anona* /ˈanuna/ ‘whatchamacallit” are still easily distinguished. While

Dahl's description primarily concerns those post-stress syllables in weak stems, he acknowledges that this "falling" of the vowel may also occur in the pre-stress position.

An unpublished manuscript by Pearson (1994) also details the environment of vowel devoicing. In addition to word-final devoicing, he observes that unstressed /i/ and /u/ apparently disappear between two consonants. An exception is in the word-initial syllable, where the vowel is not deleted, even when unstressed and between consonants.

From these previous studies on Malagasy, we can form some generalizations about vowel devoicing; however, the precise environment of devoicing is unclear. Most authors agree that there is vowel devoicing in the final position of some domain, usually identified as the word. In most of these cases, however, the authors do not share their elicitation method, so it is unclear whether these "word-final" devoiced vowels are also utterance-final. In the medial position, the authors who discuss it agree that vowel devoicing affects unstressed high vowels. Some specify that these medial vowels are prone to devoicing either after (Dahl, 1952) or between (Pearson, 1994) consonants. Additionally, Pearson complicates the prosodic environment for devoicing, suggesting that it doesn't apply to word-initial syllables. In Chapter 4, I will present a detailed quantitative model of the environment of devoicing in Merina Malagasy.

1.3. Evidence of an underlying vowel

As I've foreshadowed already, over the course of the dissertation, I will arrive at the conclusion that Malagasy devoiced vowels are represented in the output as a vowel gesture that is co-articulated and overlapped by the consonants that surround it. Acoustically, this means that the underlying syllable /su/ is realized as something resembling [s^w] ([s] and /u/ are concurrently articulated), /si/ as [s^j], and so on. But all of this presumes that there is an underlying vowel in the input to begin with, something that needs to be justified first in order to rule out the possibility that Malagasy simply has underlying /s^w/ or /s^j/.

1.3.1. Slow speech

Beyond orthographic evidence (acknowledging that the orthography, developed in the 19th century, is not necessarily representative of present-day Malagasy phonology), there is synchronic support for an underlying vowel as well. The first bit of evidence comes from slow, careful speech. In a small experiment done with one speaker, he read the same test items featured in the Task One to be featured in Chapter 4, but was instructed to say them as if he were speaking slowly and carefully to someone who was having trouble hearing him. While this experiment was designed to elicit vowels in a devoicing environment (i.e., unstressed and word-medially between two voiceless obstruents), in nearly every case, the slow speech rate yielded a full, voiced vowel. In Figure 1.2, taken from that same experiment, the spectrogram reveals a portion between /s/ and /k/, approximately 80 ms, where voiced [u] is produced.

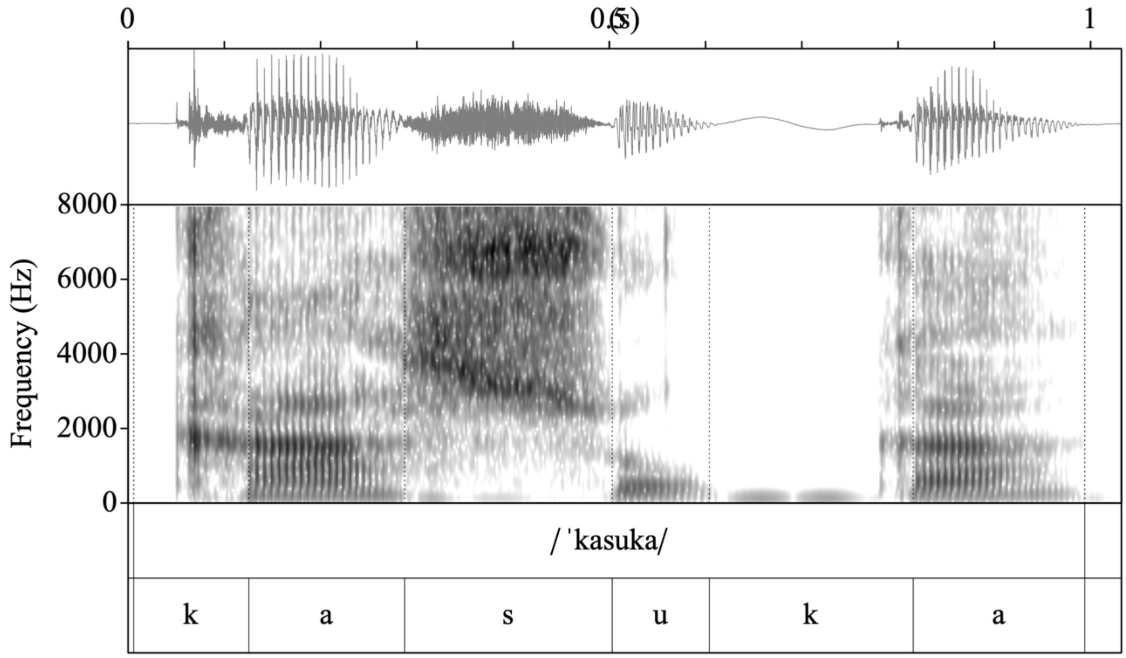


Figure 1.2 Spectrogram of *kasoka* /kasuka/ 'rubbing' in slow, careful speech with /u/ clearly voiced.

1.3.2. Morpho-phonological alternations

Additional evidence comes from alternations involving stress. Malagasy has a rich derivational morphology including affixation, reduplication and compounding (see Keenan and Polinsky (1998) for a detailed overview). Additionally, as we'll see in Chapter 3, vowel devoicing is sensitive to stress, where devoiced vowels typically appear in unstressed syllables, whereas stressed vowels, including secondary stress, are much less likely to be devoiced. As was described in Section 1.1.2.1, Malagasy has phonological (i.e., non-lexicalized) stress, which in most words falls on either the penultimate or antepenultimate syllable. The result of these two facts (morphology and predictable stress) is that there are alternations involving stress where underlying vowels become evident. For example, in the root word *toky* /tuki/ 'confidence', stress falls on the penult: [ˈtu.ki]. In Figure 1.3, we see that the stressed /u/ in *toky* is voiced.

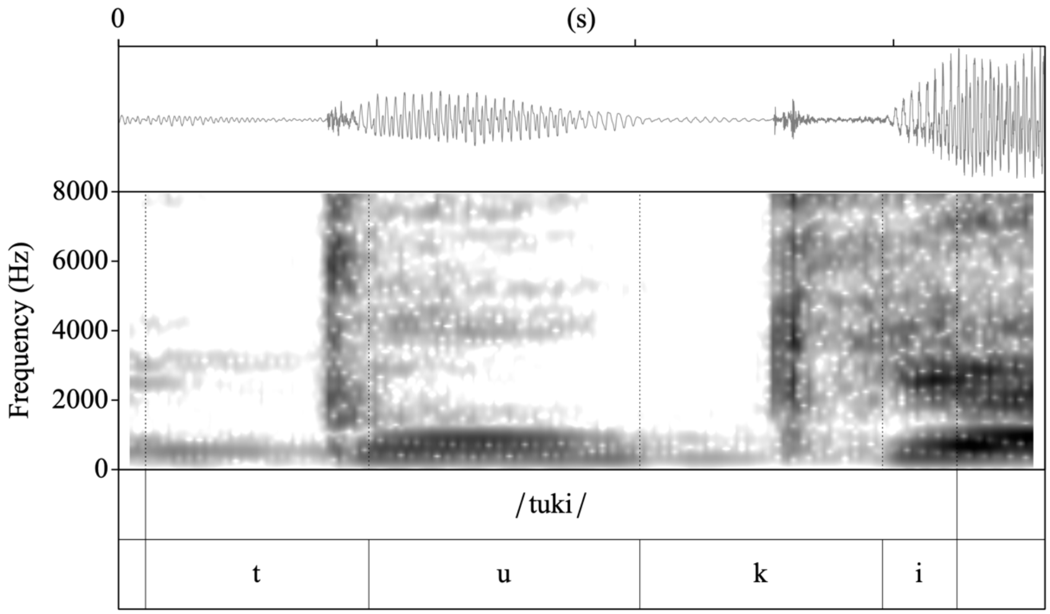


Figure 1.3 Spectrogram of *toky* /tu.ki/, where the stressed syllable contains a voiced /u/.

When verbal morphology is appended, however, the stress pattern changes. Take the Circumstantial Topic (CT; see Section 1.1.4 above) verb *ahatokiana* 'trust (verb)', for example, which has the CT suffix /-*ana*/ (which never attracts stress). The affixed verb form has antepenultimate stress, /a.ha.tuˈki.a.na/, and so the /u/ of the root is no longer stressed and consequently in a devoicing environment. Figure 1.4 confirms that /u/ is not stressed in this suffixed form, showing that Malagasy vowel devoicing is sensitive to the vowel's environment and reveals the underlying vowel in words where that syllable is stressed.

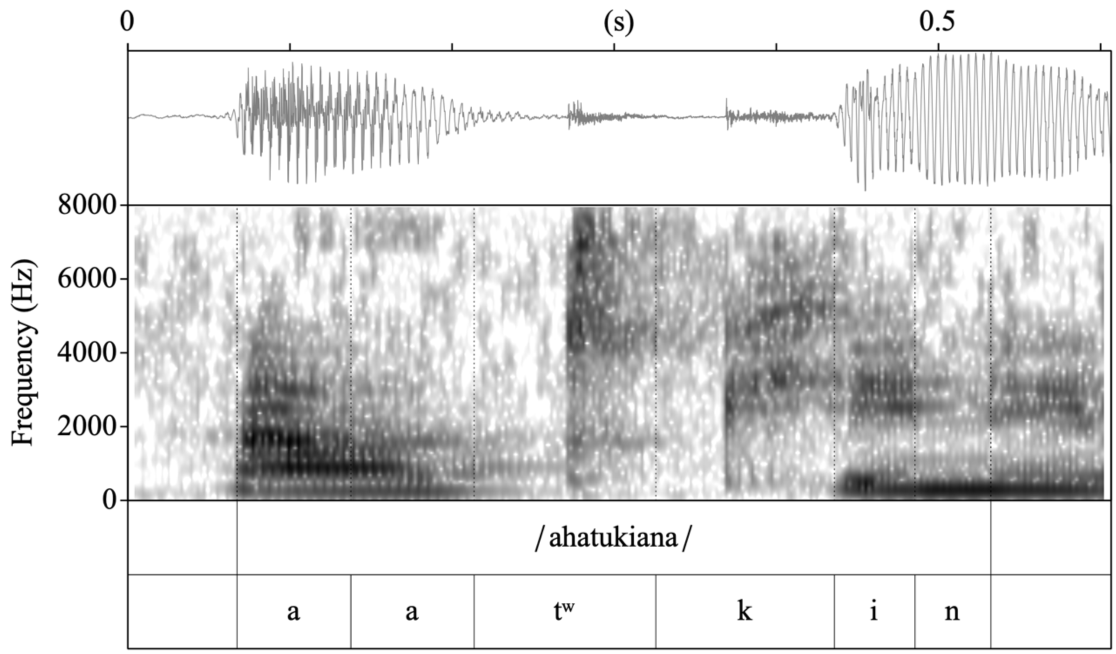


Figure 1.4 Spectrogram of /ahatukiana/, showing that underlying /u/ surfaces as devoiced when unstressed.

An additional, perhaps more transparent, alternation occurs in some imperative verbs. For active verbs in the imperative, the suffix /-a/ is added, or, in cases where the verb already ends in /a/, stress shifts one syllable rightward (Koopman, 2005). The result is minimal stress pairs like *manasitrana* /ma.naˈsi.ʈʐa.na/ 'to heal (indicative)' and *manasitràna* /ma.na.siˈʈʐa.na / 'heal (imperative)'. In fact, the stress alternation is accompanied by a devoicing alternation: in Figure 1.5, the stressed /i/ of the indicative is clearly voiced; when stress shifts rightward in the imperative in Figure 1.6, the vowel disappears.

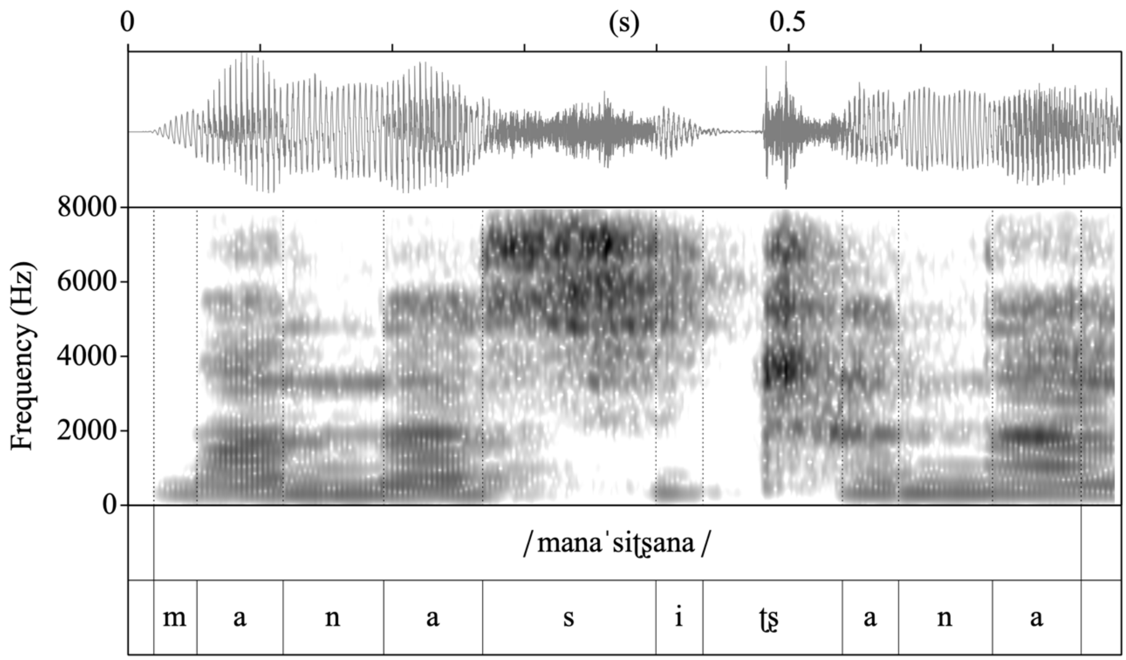


Figure 1.5 Spectrogram of the indicative verb *manasitrana* /manaˈsiʈʂana/, showing the voiced /i/ associated with the stressed syllable.

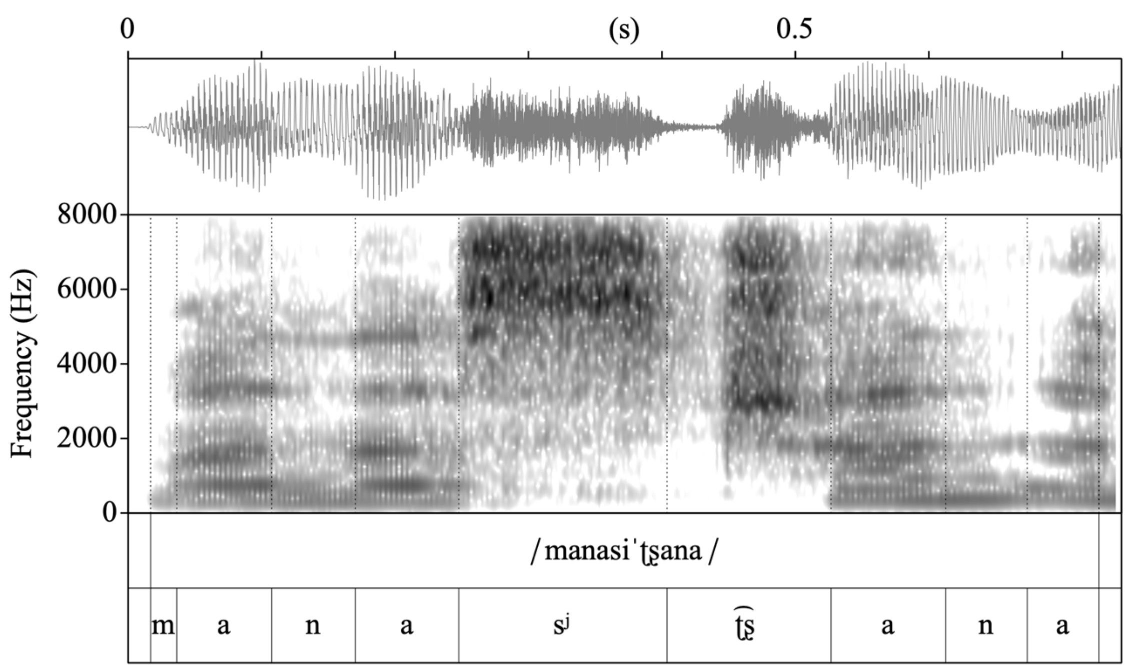


Figure 1.6 Spectrogram of the imperative verb *manasitràna* /manasiˈʈʂana/. Here, stress shifts rightward, leaving /i/ in the devoicing environment.

1.4. Other phonological properties of devoicing

1.4.1. Palatalization

In Section 1.1.3.3, I discussed a process where velar consonants have a secondary palatal coarticulation following /i/; for example, /isika/ is pronounced [iˈsi.kʲa]. Both Howe (2019) and Pearson (1994) suggested that palatalization still happens after a devoiced vowel, and my own data corroborates this claim. Below in Figure 1.7 is the spectrogram for *tsikaritra* /t͡sikariʈ͡ʐa/ 'seen for a moment'. The underlying form places an /i/ before /k/ in the first syllable, predicting palatalization of that /k/, but the vowel is clearly devoiced. Nonetheless, the /k/ is still palatalized; looking at the transition into the following /a/, F2 is quite high and F1 is quite low, typical of the high front glide.

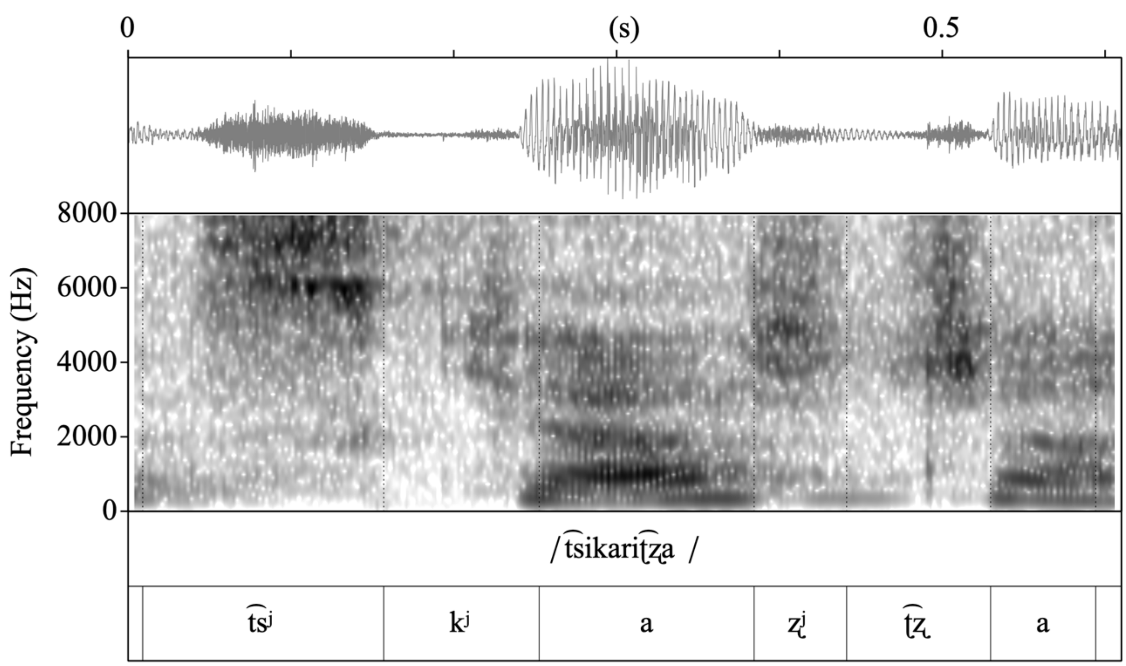


Figure 1.7 Spectrogram for *tsikaritra* /t͡siˈkariʈ͡ʐa/ 'seen for a moment'. The second tier has the phonetic transcription, showing that both /i/ are devoiced but the /k/ is still palatalized, evident by the high F2 (about 2400 Hz) and low F1 (570 Hz) at the offset of /k/, relative to typical values for /a/.

1.4.2. Consecutive devoiced vowels

An additional point worth bringing up is that Malagasy allows multiple consecutive syllables with devoiced vowels. Because Malagasy stress mostly appears on every other syllable (see Section 1.1.2.1), and because vowel devoicing is most likely to occur in unstressed syllables (to be shown in Chapter 3), multiple consecutive devoiced vowels are rare but possible across word boundaries. In Figure 1.8, for example, is an excerpt from the sentence *Vaky ny solomasoko sy fitaratro* ‘My glasses and mirror are broken’, which includes the underlying sequence /sukusifi/, underlined in the example above. The figure shows one speaker’s production of that sentence, including four consecutive syllables with devoiced vowels: the entire sequence, from the /s/ of *masoko* to the /t/ of *fitaratro* contains no voiced vowel (in fact, no voicing at all).

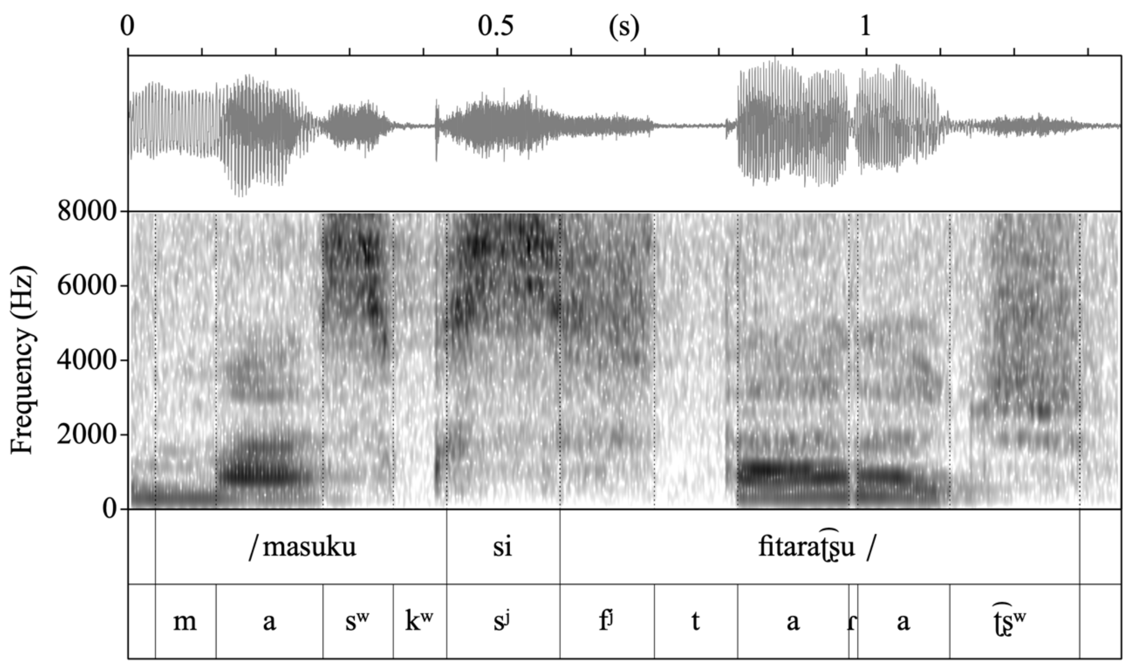


Figure 1.8 Spectrogram for the excerpt …*(solo)masoko sy fitaratro*… /masuku si fitaraʈ͡ʂu/ ‘glasses and mirror’ showing four consecutive syllables with devoiced vowels.

This is relevant because at least one language, Japanese, has been argued to avoid consecutive syllables with vowel devoicing (Tsuchida, 1997); Malagasy speakers seem to have no problem producing multiple devoiced syllables.

1.5. What will not be discussed in this dissertation

The remainder of the dissertation includes an acoustic description of Malagasy vowel devoicing, a statistical model of the environment where devoicing occurs, and an Articulatory Phonology analysis of non-final vowel devoicing. While Chapter 2 does give an overview of both final and non-final devoicing, I will argue that these are two fundamentally different processes in every meaningful way: their phonetic realization, phonological properties and motivation in the grammar. With that being said, the dissertation primarily focusses on what I have been calling *non-final vowel devoicing*; beyond a brief description in Chapter 2, final vowels will not be discussed at length.

Additionally, I have chosen to focus on vowels in the inter-consonantal environment, excluding discussion of underlying vowel hiatus in the present study. As mentioned in Section 1.1.3.2, Malagasy is said to resolve vowel hiatus in various ways, including vowel deletion, diphthongization and coalescence. Thus, I feel that vowel-vowel sequences deserve their own treatment. It is possible that the same mechanisms driving vowel devoicing are also at play in processes like diphthongization or coalescence; I leave these questions to future researchers and now turn back to vowel devoicing. Andao!

Chapter 2
Two Types of Vowel Devoicing

2.1. Introduction

The term *vowel devoicing*, as discussed in Chapter 1, is assigned to various phonetic and phonological processes in the world's languages, many of which do not actually produce truly voiceless vowel allophones. Lacking clear phonetic description, it is not clear by looking at many of these works what, exactly, is meant by devoicing. An accurate depiction of vowel devoicing in Malagasy will serve us in multiple ways: for purely descriptive purposes of concern to future learners or researchers of Malagasy; to accurately place Malagasy within the typology of vowel devoicing; and to provide a phonetic basis for a phonological analysis of devoicing. In this chapter, I will provide a phonetic description of vowel devoicing in environments where it has been reported to occur in Malagasy (both final and non-final devoicing); in Chapter 4, I will present a quantitative acoustic analysis of non-final vowel devoicing.

The description will reveal two types of vowel devoicing found in Malagasy: the first, final devoicing, involves the phonetic devoicing of most vowel segments in Intonational Phrase-final positions; the second, non-final devoicing (the primary subject of the dissertation), affects some unstressed high vowels but looks phonetically quite distinct from final devoicing. I will lay out a qualitative description of some acoustic and phonological properties of these two types of vowel devoicing, ultimately arguing that so-called non-final vowel "devoicing" does not involve a discrete segment clearly discerned as a vowel in the output whatsoever. In other words, these vowels appear to be deleted; as I will argue in Chapter 4, however, there is acoustic evidence of a vowel remnant on the surrounding phonological environment, ultimately leading to the analysis in

Chapter 5 that vowel "devoicing" results from extreme overlap of the adjacent consonants, obscuring the vowel completely.

2.2. Final vs. non-final devoicing: The typology of vowel devoicing

Processes described as "vowel devoicing" have been described in both "final" position (where the domain varies between languages) and medial positions, where vowel devoicing occurs between two consonants. Devoicing of initial vowels is not widely reported (Faust and Pike (1959) claim Cocama as a counterexample). In Chapter 1.2, I outlined the most common phonetic and phonological explanations for vowel devoicing, noting that the analysis proposed often depends on the prosodic position of the vowel: final devoiced vowels are said to arise for aerodynamic reasons, while non-final vowel devoicing is more likely to be described as gestural overlap or using some feature-based analysis.

Additionally, the literature highlights some possible distributional differences between final and non-final devoicing. First, whereas non-final devoicing is often restricted to high vowels (e.g., Turkish, Jannedy, 1995; Korean, Jun et al., 1997; Japanese, Tsuchida, 1997), in some languages with final devoicing, it is reported to also affect low vowels (e.g., French, Smith, 2003), albeit at a lower rate than high vowels. Second, the segmental environment is often different for final vowels. In Gordon's (1998) typology, he notes that final devoicing may occur after voiced and voiceless consonants alike, whereas non-final devoicing typically only occurs next to (and often obligatorily between) voiceless consonants. However, it is possible that the term "vowel devoicing" is biased toward vowel reduction processes that occur next to voiceless consonants, and similar processes that occur next to voiced segments may be called deletion.

In the following sections, I will show that Malagasy final and non-final vowel devoicing show distinct phonetic and phonological properties that warrant their treatment as different processes. First, final devoiced vowels are phonetically distinct: they are more likely to appear as actual voiceless vowel segments, with formant structure resembling their voiced counterparts and occupying a surface interval that is distinct from the preceding consonant. Non-final "devoiced" vowels, in contrast, neither occupy their own interval nor show the formant structure of typical vowels at all. Instead, non-final "devoiced" vowels seem to have disappeared from the output entirely (a hypothesis that will be adjusted in Chapter 4). Additionally, final vowel devoicing affects both high and low vowels, whereas non-final "devoicing" is restricted to high vowels. Finally, preliminary evidence indicates that final devoicing occurs even in slow, careful speech, while non-final "devoicing" does not.

2.3. The phonetics of vowel devoicing

The data presented in this section come from the two experiments performed for this study (see Chapters 3.2.1 and 4.2 for a description of the participants and items that formed the dataset) as well as other data collected from my own fieldwork on Merina Malagasy between 2017 and 2024. As I will describe in more detail in Chapter 3, Malagasy unstressed vowels are nearly all devoiced at the end of the Intonational Phrase; here, *final devoicing* refers to vowels in that IP-final position and *non-final devoicing* refers to so-called "devoicing" in any other position (most commonly, IP-medial).

2.3.1. Final devoicing produces voiceless vowels

A key difference between final and non-final devoicing is that they have different phonetic realizations. Final devoicing produces truly voiceless vowels; that is, there is a segment in the output that is clearly distinguished from its environment and that has the acoustic properties of a vowel, except that it is voiceless. Looking at Figure 2.1, we see that the final /a/ has these features: its boundaries are demarcated by the release of the /n/ that precedes it, after which the voicing bar (in the lowest frequencies) ends. However, formants above the first are still visible, and the value of F2 is around 1600 Hz, consistent with the two previous instances of /a/ in the token. In Figure 2.2, the same pattern arises for final /i/, this time next to a clearly demarcated /v/ and showing an F2 value around 1900 Hz. Finally, Figure 2.3 shows final devoiced /u/. Here, the spectral properties of the vowel are less clear, though we can still distinguish the vowel portion from the more fricated /v/ and a faint F2 at around 1200 Hz. Indeed, in all three cases, the final vowel is clearly heard as a voiceless vowel.

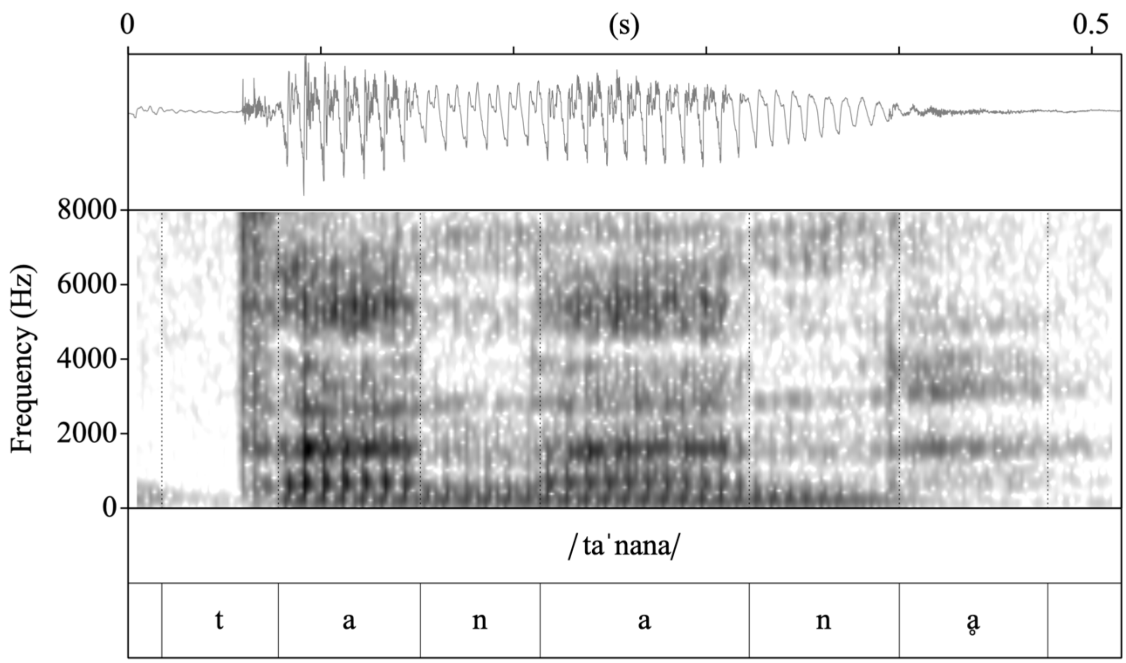


Figure 2.1 Spectrogram of *tanàna* /taˈnana/ ‘town’ in the IP-final position with devoiced /a/. Below the spectrogram, the first tier shows the underlying representation of the word in addition to stress, and the second tier shows a phonetic transcription. The same format is used in all figures in this chapter.

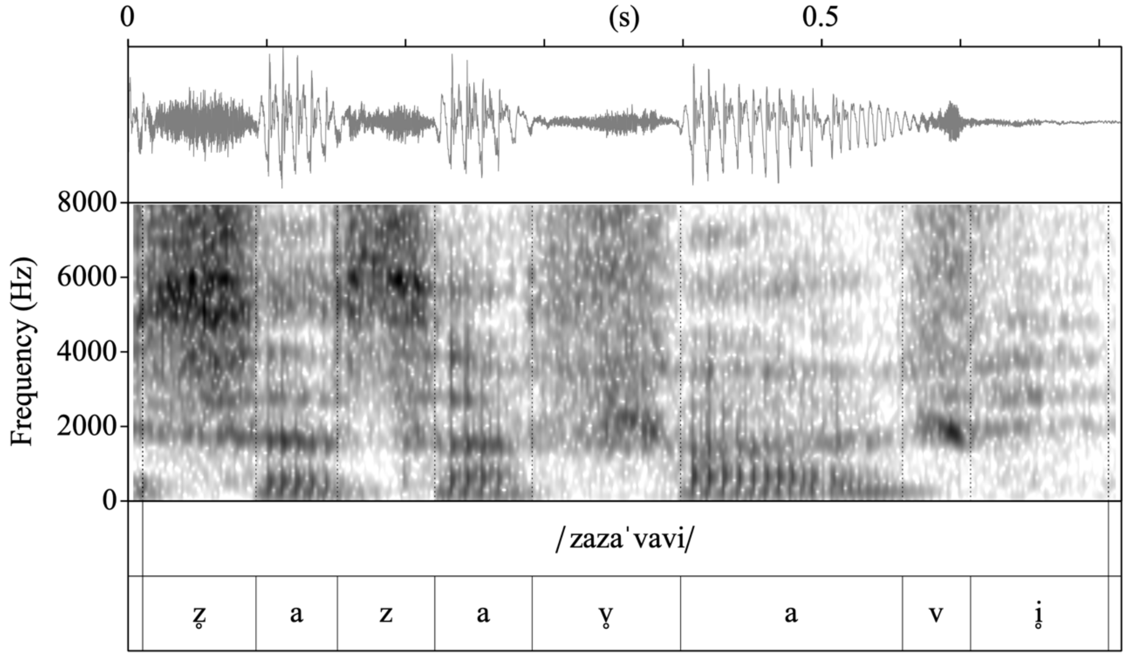


Figure 2.2 Spectrogram of *zazavavy* /zazaˈvavi/ 'girl' in the IP-final position with devoiced /i/.

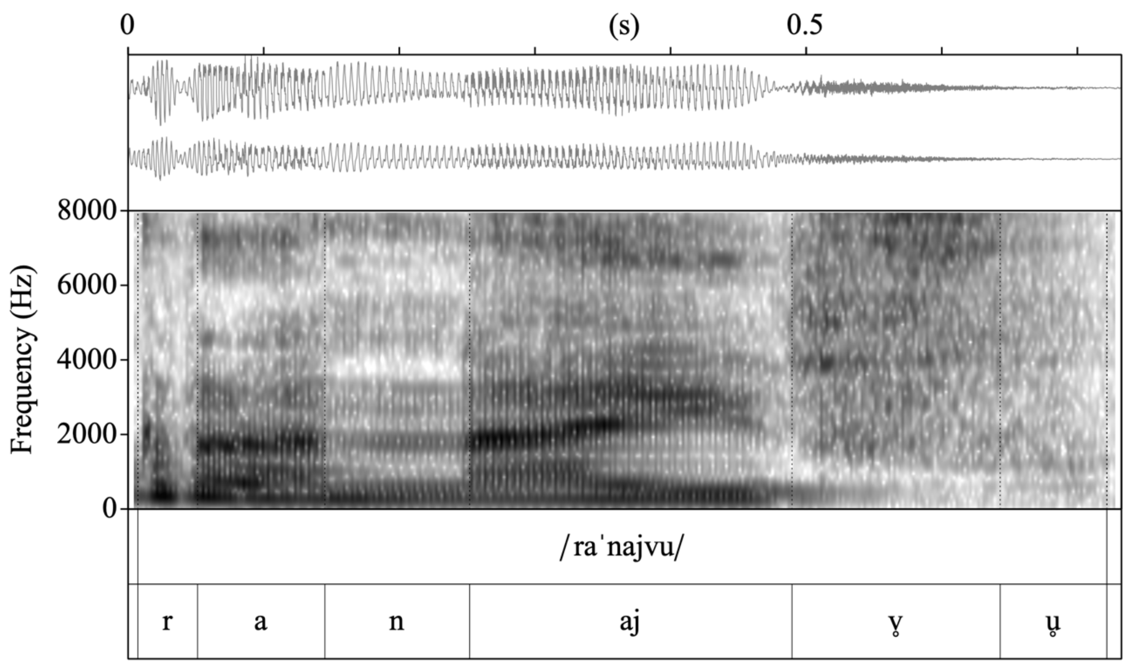


Figure 2.3 Spectrogram of the name *Ranaivo* /raˈnajvu/ in the IP-final position with devoiced /u/.

In some cases, the final devoiced vowel is not as clear and weakly heard. In Figure 2.4, the word *tanàna* is produced by a different speaker. In this case, the devoiced vowel is faintly visible and heard as nothing more than a short, voiceless burst; nonetheless, F2 is still visible at around 1850 Hz, a value that is consistent with the other /a/ vowels in that token.

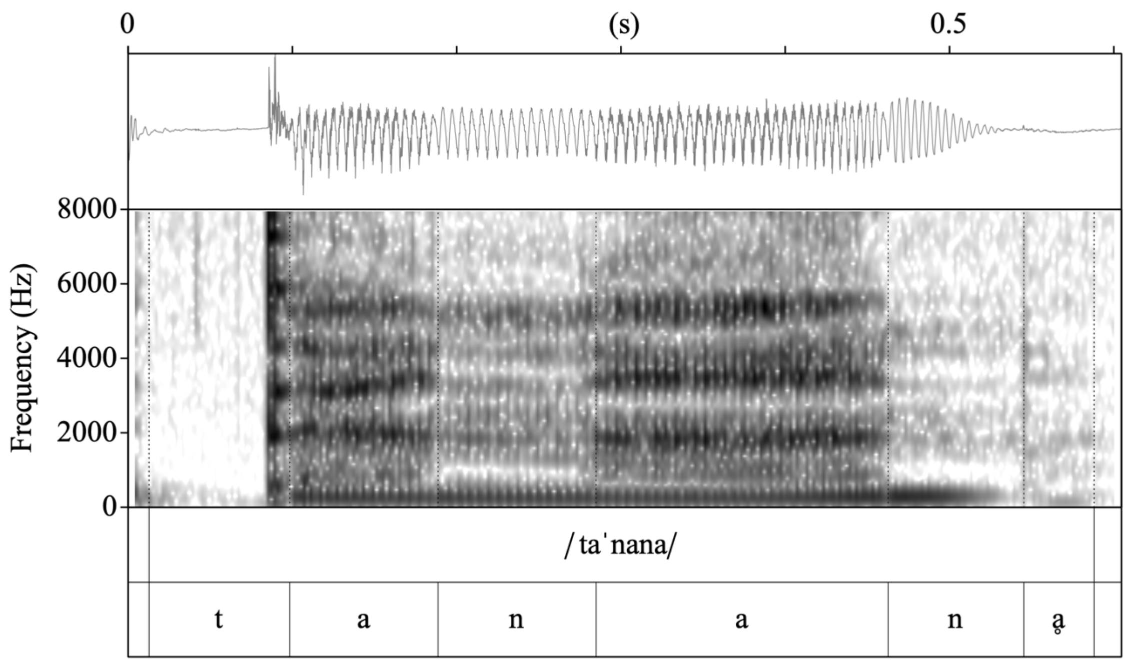


Figure 2.4 Spectrogram of *tanàna* /taˈnana/ ‘town’ in the IP-final position with weakly released devoiced /a/.

Devoiced final /i/ is sometimes heard as a louder, more fricated sound, resembling the palatal fricative [ç]. This is reminiscent of Parisian French, whose IP-final high vowels may be partially devoiced to the point of frication (Fagyal & Moisset, 1999); as most Merina speakers are fluent in French, including the participants in this study, it is not clear whether this feature is due to cross-linguistic influence from French. An example of this appears in Figure 2.5.

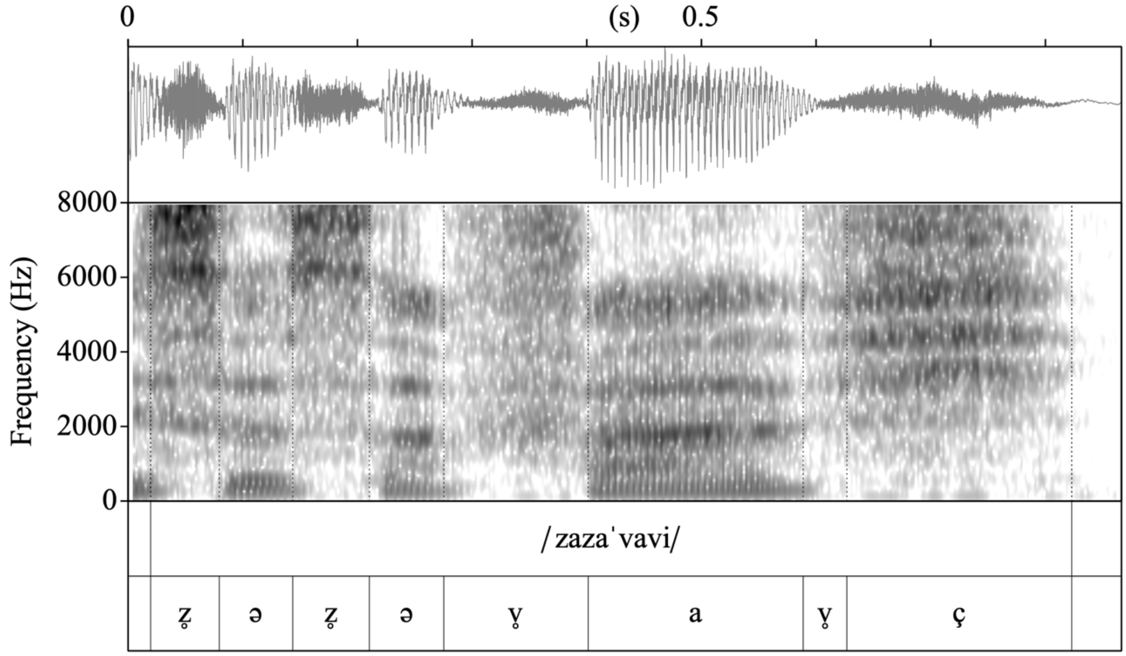


Figure 2.5 Spectrogram of *zazavavy* /zazaˈvavi/ ‘girl’ in the IP-final position with a fricated devoiced /i/.

2.3.2. Non-final devoicing looks more like deletion

In contrast to the IP-final vowels shown above, in non-final positions, vowel “devoicing” does not involve voiceless vowels; instead, these vowels occupy no interval of their own whatsoever. In Figure 2.6, for example, the word *dadasika* /daˈdasika/ ‘spacious’ has the vowel /i/ in the environment /s_k/. Here, there is no portion of the spectrogram that is clearly demarcated as a vowel segment; as the frication of /s/ ends, the stop closure of /k/ begins, with no intervening vowel-like occupant. Indeed, this non-final vowel does not resemble a voiceless vowel at all and instead looks like deletion; though, as I will detail in Chapter 4, deletion is an inadequate description as acoustic traces of the vowel appear on adjacent segments.

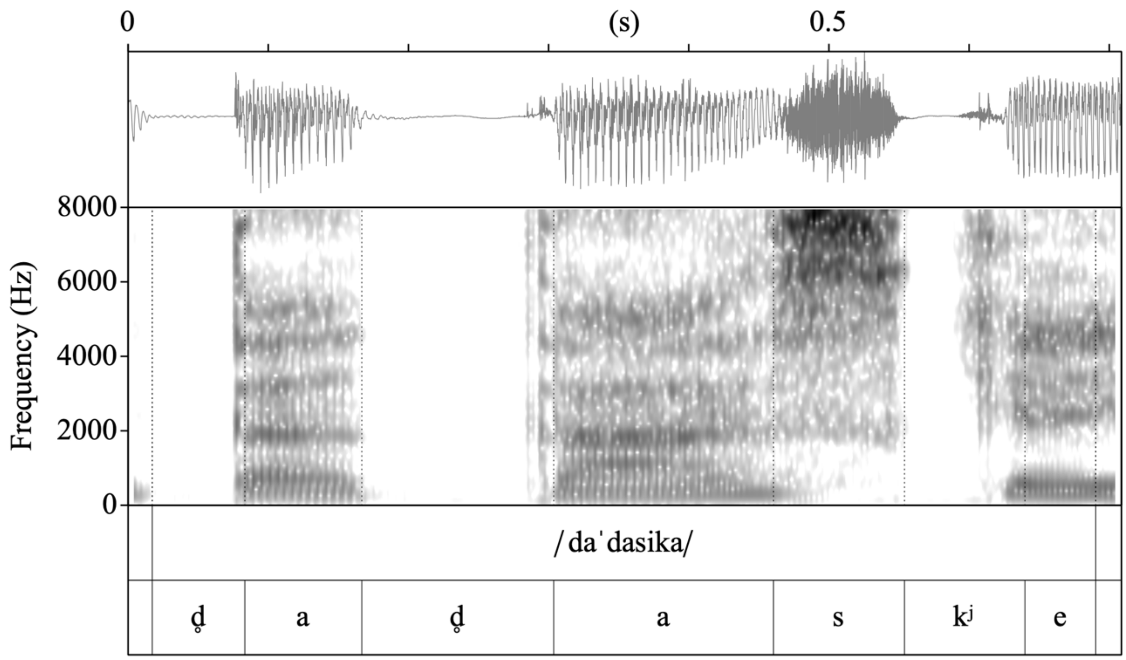


Figure 2.6 Spectrogram of *dadasika* /daˈdasika/ 'spacious' in a non-final position, showing that the vowel /i/ is not easily demarcated after /s/.

The absence of a clear vowel segment is especially evident when the devoiced vowel falls between two identical segments. Figure 2.7 shows an excerpt from the sentence *Tsy manana basy sandoka aho* 'I don't have a fake gun', which includes the non-IP-final vowel /i/ between two /s/. In the spectrogram, the "devoiced" vowel is nowhere to be seen; instead, one uninterrupted period of fricative noise is visible, from the /s/ of *basy* to that of *sandoka*.

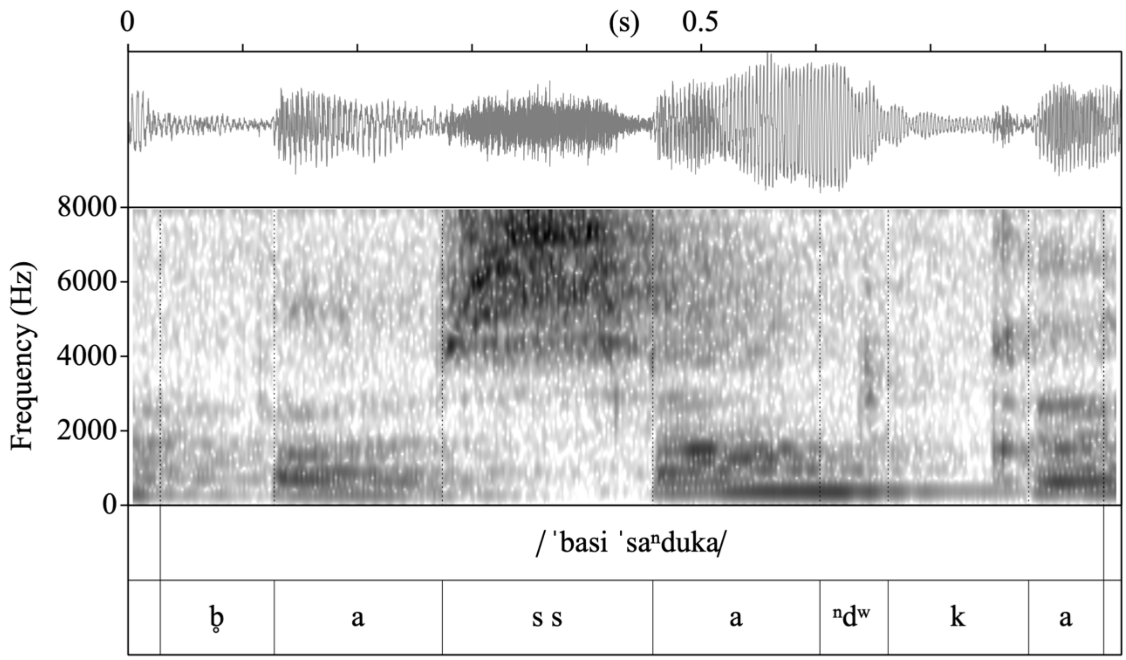


Figure 2.7 Spectrogram of *...basy sanduka...* /ˈbasi ˈsa^{n}duka/ ‘fake gun’

Figure 2.8 shows an example of non-final “devoiced” /u/ following a stop, in the word *behatoka* /behatuka/ ‘nape’. In this case, something resembling a voiceless vowel does appear: following the release of /t/, there is a short duration of aspiration with a formant structure resembling the final devoiced vowels described in Section 2.3.1. However, Malagasy stops are typically aspirated, especially before high vowels (Howe, 2019), even when the vowel is voiced; this will be described in detail in Chapter 4. So, while something resembling a voiceless vowel does appear after stops in non-final positions, this can be explained as aspiration associated with the stop rather than a separate voiceless vowel segment.

Figure 2.9 shows an additional example of non-final “devoicing” when the underlying vowel occurs between two nasals, in the word *nanome* /nanuˈme/ ‘give (past)’. Again, the non-final vowel appears to be absent from the spectrogram; instead, the /n/ transitions abruptly into /m/, noted by the drop in F2 visible at the right edge of /n/ (from about 1500 Hz to 1200 Hz).

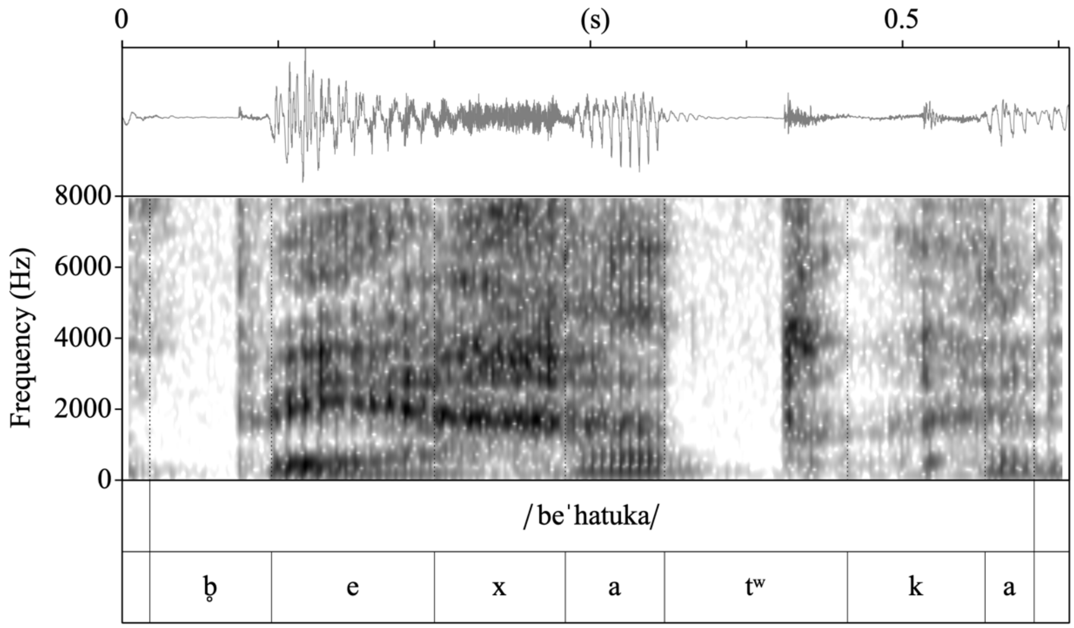


Figure 2.8 Spectrogram of *behatoka* /beˈhatuka/ 'nape' in a non-final position; while it may look like there is a short voiceless vowel following the release of /t/, Malagasy stops are typically aspirated, and so here there is no additional voiceless vowel.

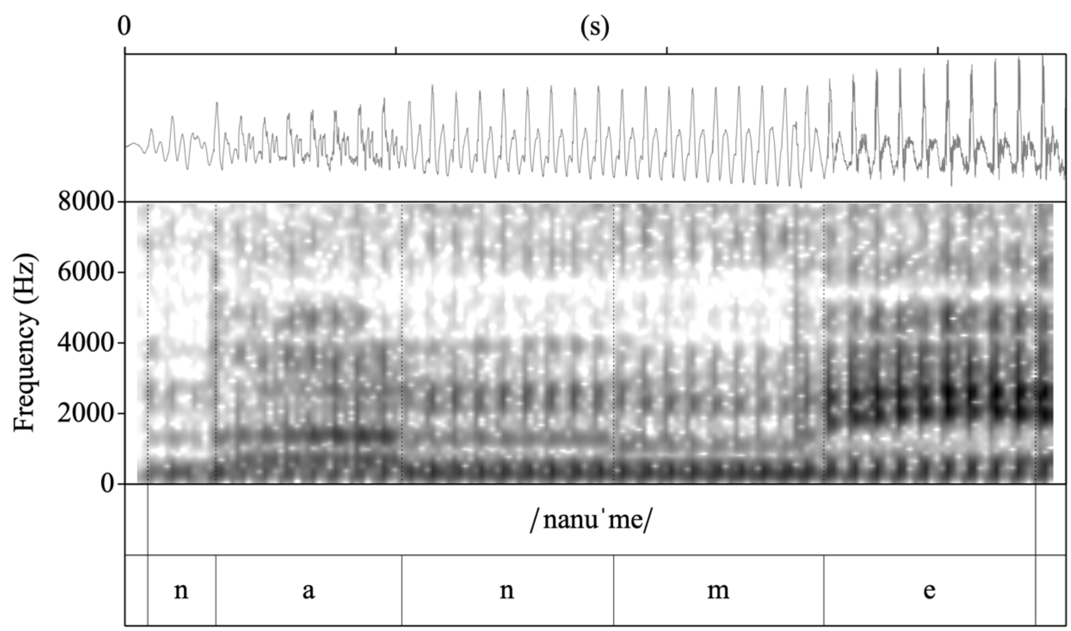


Figure 2.9 Spectrogram of *nanome* /nanuˈme/ 'give' in a non-final position where the "devoiced" vowel appears to be deleted.

2.4. Vowels not undergoing devoicing

In the IP-final position, all unstressed vowels are devoiced regardless of their height. Because the diphthongs /aj/ and /aw/ as well as the mid vowels /e/ and /o/ are always stressed, only /i/, /u/ and /a/ remain to be fully devoiced. Figures 2.1, 2.2 and 2.3 above show final devoiced /a/, /i/, and /u/, respectively. Final stressed vowels, in contrast, are always produced with at least some voicing. Figure 2.10 shows an utterance ending in the second-person singular pronoun *ianao* /iaˈnaw/, where the final diphthong (here realized as [o]) is voiced, indicated by the voicing bar in the lowest frequencies.

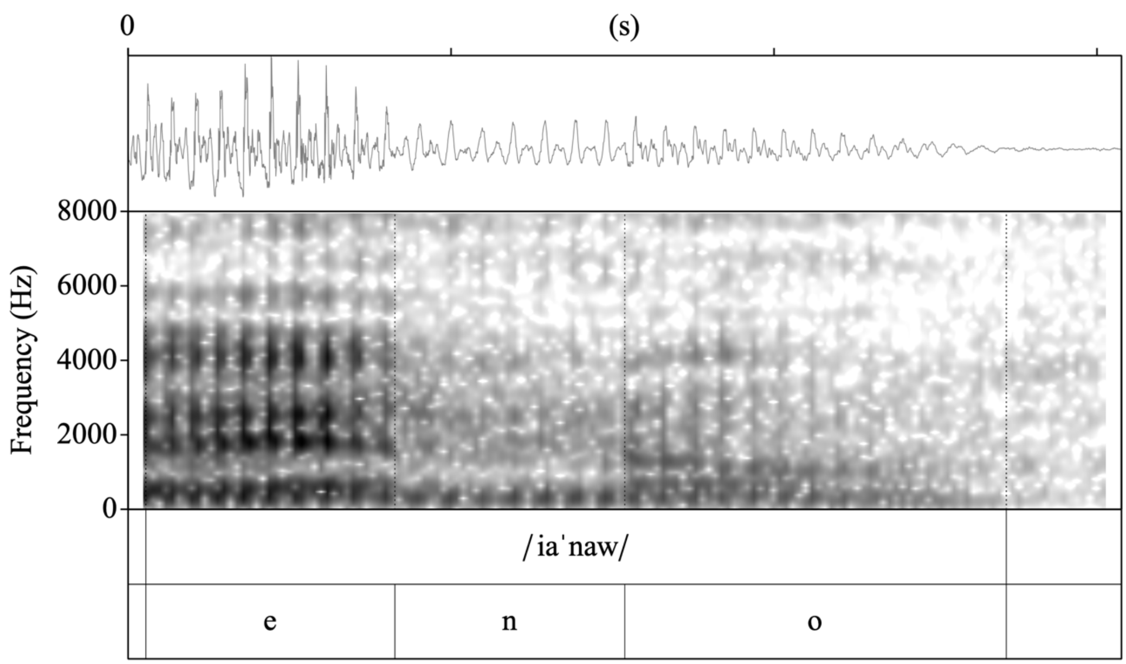


Figure 2.10 Spectrogram of the pronoun *ianao* /iaˈnaw/ in the IP-final position, where the stressed diphthong /aw/ (realized as [o]) is voiced.

However, certain speakers did sometimes partially devoice final stressed vowels. Figure 2.11 shows the demonstrative *itỳ* /iˈti/ with stressed /i/ in the IP-final position. Here, the second half of the vowel is devoiced and fricated, indicated by the abrupt end of the voicing bar and the increased energy above 5000 Hz.

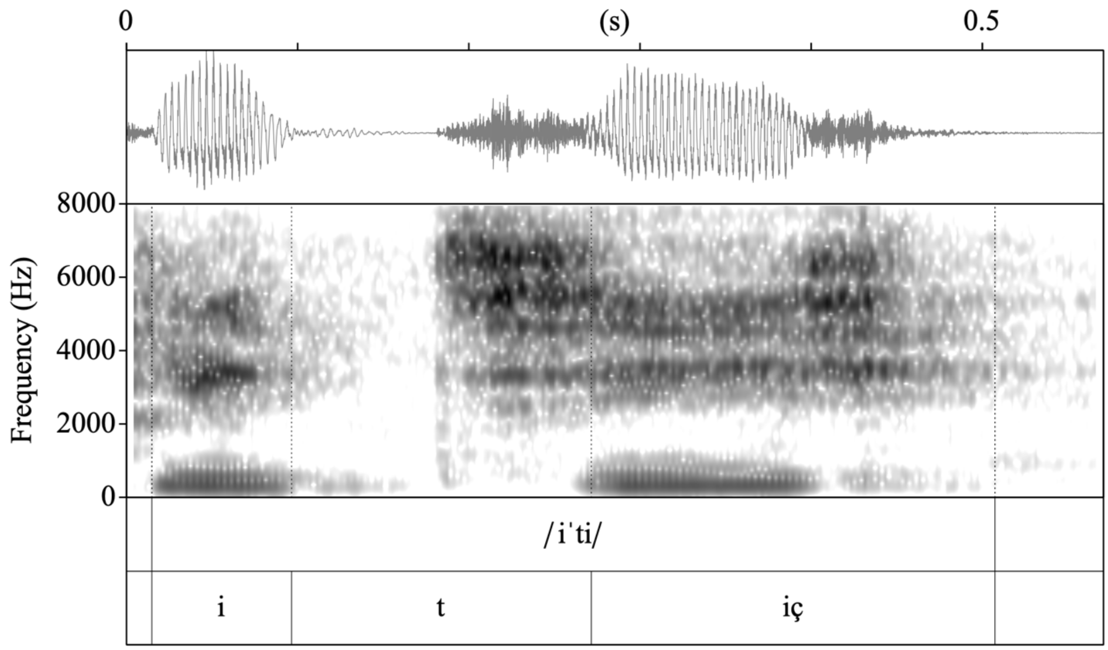


Figure 2.11 Spectrogram of the demonstrative *itỳ* /iˈti/ showing frication of the final part of /i/ in the IP-final position.

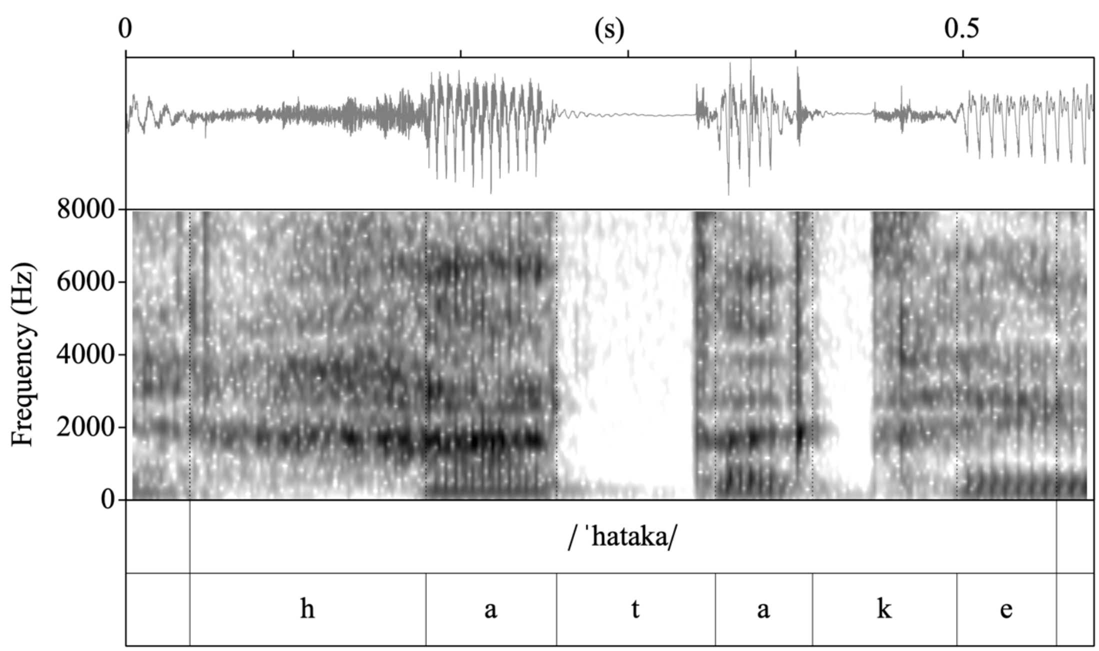


Figure 2.12 Spectrogram of *hataka* /ˈhataka/ 'a request' showing that non-final /a/ is voiced, even in a devoicing environment (in this case, between two obstruents in an unstressed syllable). The final /a/ is realized as [e] as the following word begins with /i/.

In non-final syllables, however, vowel "devoicing" is restricted to the high vowels /i/ and /u/. Unstressed non-final /a/ (e.g., Figure 2.12) and /e/ (Figure 2.13) are always voiced, even in environments that otherwise cause devoicing.

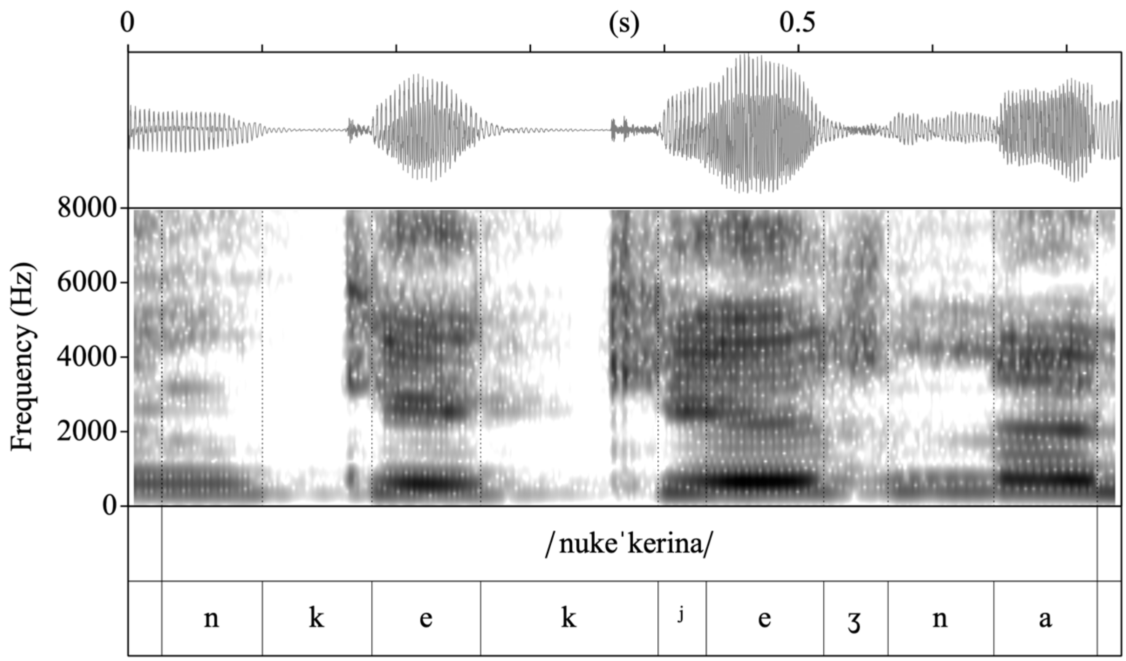


Figure 2.13 Spectrogram of *nokekerina* /nukeˈkerina/ 'to be bitten (past)' showing voiced /e/ in the non-final position.

2.5. Final devoicing remains in slow speech

Preliminary results indicate that an additional difference between final and non-final devoicing is that final devoicing happens even at slower speech rates, whereas non-final devoicing appears to only happen in faster, casual speech rates. In Chapter 1.3.1, I mentioned that Malagasy speakers produce voiced vowels, even in the devoicing environment, in slow, careful speech. This is based on an experiment with one speaker (modelled after Task One to be presented in Chapter 4) where the participant was asked to speak carefully as if to someone who hadn't understood what he had said. In this slower speech rate, that speaker devoiced the final vowel of the Intonational

Phrase (always the /i/ in /zazaˈvavi/, as the experiment used a frame sentence), but voiced other, non-final vowels. This result is preliminary, as it is based on only one speaker (who also speaks Betsileo Malagasy) and only one lexical item (repeated several dozen times). Compare the spectrogram in Figure 2.14, showing a final vowel in slow speech, to Figure 2.2 above, produced by the same speaker at a normal speech rate.

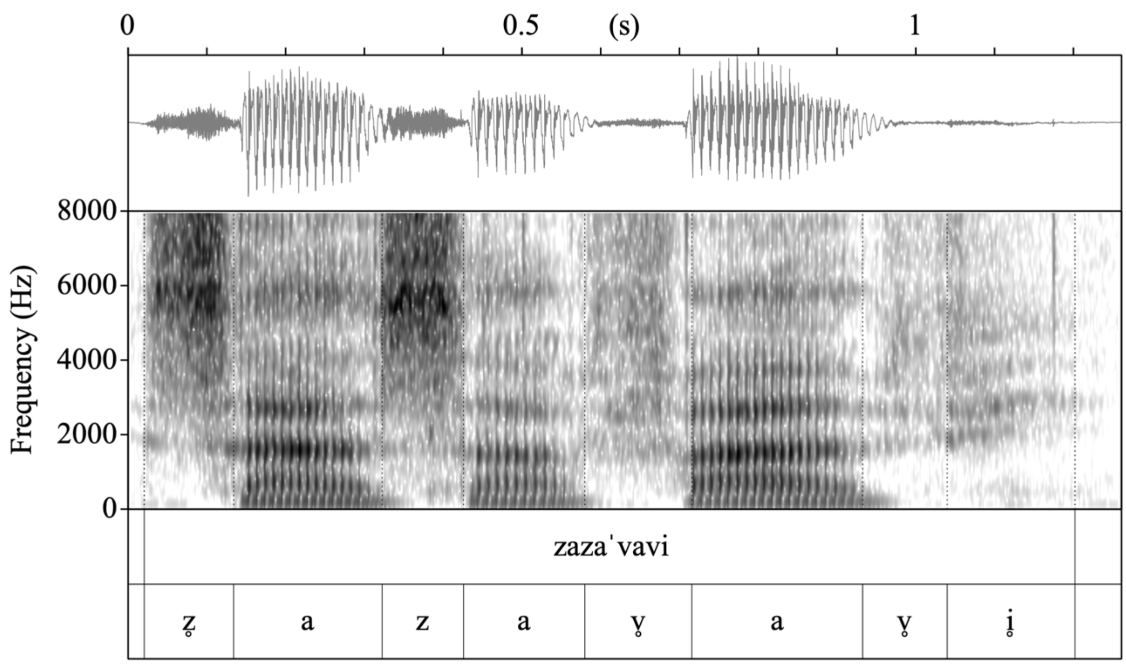


Figure 2.14 Spectrogram of *zazavavy* /zazaˈvavi/ ‘girl’ in the IP-final position, said at a slower speech rate.

This contrasts with Figure 2.15, where the word *vasoka* /ˈvasuka/ ‘discoloured’, is produced with a voiced medial /u/ in slow speech, even though it is in a consonantal environment where vowel “devoicing” is very likely to occur (between obstruents; see Chapter 3).

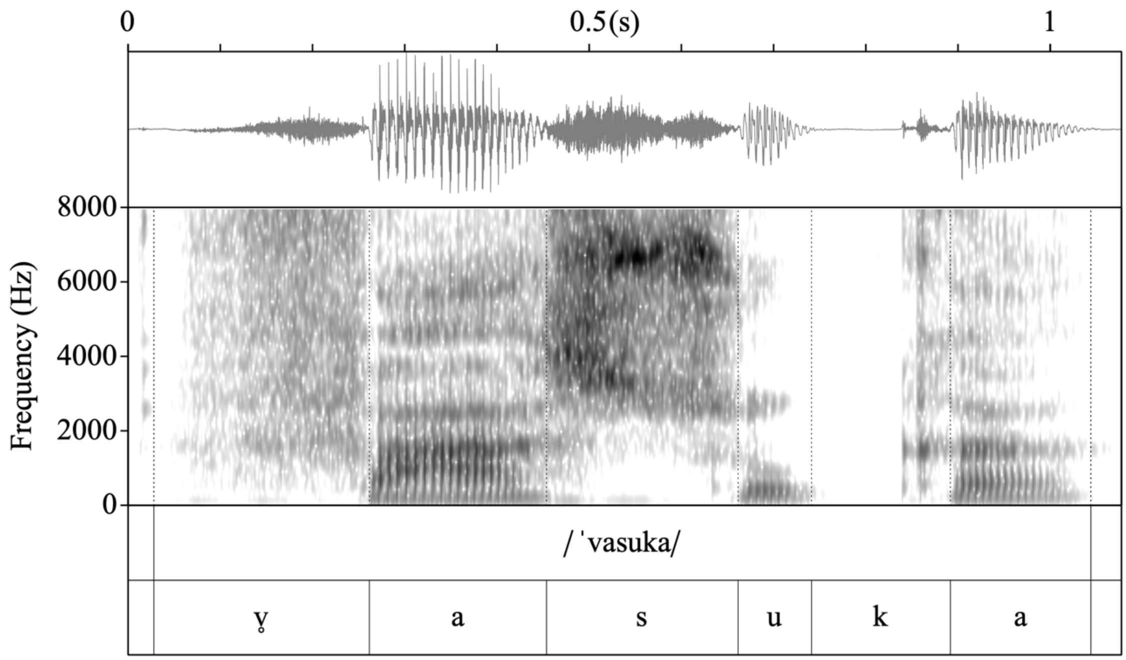


Figure 2.15 Spectrogram of *vasoka* /ˈvasuka/ 'discoloured' spoken at a slower speech rate, showing that each vowel is clearly voiced.

2.6. Interim conclusion: Two types of vowel devoicing

Malagasy has been described as having a "vowel devoicing" process for at least 170 years, though it is only within recent decades that devoicing outside of the final position has been reported. In this chapter, I presented a descriptive phonetic account of so-called devoicing, finding that what has been described for Malagasy is, in fact, two distinct processes. On the one hand, Malagasy does appear to have true vowel devoicing: in the Intonational Phrase-final position, unstressed vowels are frequently realized as voiceless, regardless of their height. In non-final positions, however, "vowel devoicing" is a misnomer; in these cases, the vowel seems to have disappeared from the spectrogram, not occupying any interval of its own and certainly not as a voiceless vowel. However, as we will see in Chapter 4, vowels in this non-final "devoicing" environment have not been completely deleted; rather, remnants of the vowel are acoustically present on adjacent segments. In Chapter 5, I will analyze this so-called vowel “devoicing” as extreme consonant-vowel gestural overlap: in these cases, the vowel is overlapped to the point of total obscuration; it is not heard as a unique segment, but rather as a coarticulatory gesture on its environment.

Chapter 3
The Environment of Devoicing:
A Probabilistic Model

3.1. Introduction

In this chapter, I present a statistical model of the environments where devoiced vowels appear in Malagasy. The purpose of this approach is to identify the factors that contribute to the probability that a given vowel undergoes devoicing, which will inform the theoretical account of the Malagasy phonological grammar to be presented in Chapter 5. Here, I make use of a generalized linear mixed effects model to make predictions about the environment of vowel devoicing in Malagasy. Specifically, the question at hand is what prosodic and segmental factors contribute to the likelihood of devoicing. While this chapter does examine both the final and non-final devoicing processes described in Chapter 2, the main statistical model looks only at non-final "devoicing", the phonetics of which will be described in more detail in Chapter 4.

In this chapter, I show that vowels in certain environments are nearly always voiced or devoiced. Already in Chapter 2, I discussed how vowels are likely to be devoiced in the IP-final position and showed evidence that this final devoicing is a separate phonological process. Additionally, in this chapter, I identify vowels in primary-stressed syllables and /a/ as being near-categorically voiced. Within the remaining environments where vowel "devoicing" is variable (i.e., non-final unstressed or secondary-stressed /i/ and /u/), certain factors will be shown to increase the likelihood of so-called devoicing. Prosodically, vowels are less likely to undergo "devoicing" in secondary stressed syllables and initial syllables. Vowel "devoicing" is also sensitive to the sonority of its segmental environment: on the one hand, "devoicing" is more likely when the vowel *follows* a sonorant compared to an obstruent. But with respect to the following consonant, the likelihood of vowel "devoicing" actually *decreases* as sonority increases. The results presented here will form the basis for the phonological analysis of non-final vowel "devoicing" presented in Chapter 5; each of the factors here shown to affect the likelihood of "devoicing" will be accounted for using a theory of Articulatory Phonology (Browman & Goldstein, 1986) in that chapter.

3.2. Data collection

The data used for this analysis come from an experiment performed with six Merina Malagasy speakers, all of whom also participated in the experiments upcoming in Chapter 4. The experiment was conducted remotely using Riverside.fm (*Riverside*, 2024), a podcasting software that permits uncompressed audio recording up to a sampling rate of 44.1 kHz using the participant's local system. As recording was done remotely, the dataset was not recorded in a sound-attenuated booth; rather, participants were instructed to find a quiet room in their home and mitigate any potential background noise. All six participants were female.

Participants in this experiment were presented with Malagasy sentences on a screen, one at a time. Participants were given the opportunity to read the sentence silently and produce it aloud when they were ready. In cases where a participant was not familiar with one or more words in the sentence, that item was skipped. In some cases, for example if the experimenter detected background noise, the participant was asked to repeat the sentence, and participants were also permitted to repeat any item if they felt they had made a speech error. Items that were produced by less than half of participants were excluded from the analysis. Each sentence was manipulated to contain the four canonical monophthongs of Merina (/a, e, i, u/) among multiple phonetic and phonological factors expected to contribute to devoicing, including stress, segmental environment, and prosodic position. In most cases, each item contained multiple vowels of interest; in (3.1), for example, the sentence was designed with seven vowels in mind (each one underlined). A total of 130 sentences containing 417 target vowels were considered for the analysis; across all six participants, after excluding unproduced items, there was a total of 2313 vowels to be analysed. A complete list of items appears in Appendix A.

(3.1)	Vaky	ny	solomasoko	sy	fitaratro.
	/ˈvaki	ni	ˌsuluˈmasuku	si	fiˈtaraʈʐu/
	broken	DET	glasses.1sg.GEN	CONJ	mirror.1sg.GEN
	‘My glasses and mirror are broken.’				

While the dataset was designed to incorporate as many phonological factors as possible, with an emphasis on the high vowels expected to most undergo devoicing, the dataset is not completely balanced because of constraints within the lexicon. For example, the first vowel of *tilikambo* /ˌtiliˈkambu/ ‘tower’ is a secondary-stressed /i/ between /t/ and /l/, an environment that is not represented for /u/.

Each target vowel was manually labelled as either “voiced” or “devoiced” categorically based on the presence or absence of a clearly voiced vowel in the spectrogram and waveform. When clear formant structure, high-energy in the lowest frequencies (the “voicing bar”) and periodicity in the waveform were present, the token was labelled “voiced”. In all other cases, the vowel was labelled as devoiced; this included the two main types of phonetic devoicing discussed in Chapter 2: true voiceless vowels (in the IP-final position) and apparent vowel deletion (i.e., where the preceding consonant transitions directly into the following consonant). In the remainder of this chapter, the term *devoiced* is used to refer to any underlying vowel that does not surface as voiced (i.e., either or the two processes described in Chapter 2).

When the target vowel appeared next to a sonorant, which typically also have formant structure and a periodic waveform, the presence of a voiced vowel was determined based on a clear delineation in the spectrogram so that the vowel was clearly visible and distinct from the adjacent sonorant, as well as auditory cues. In some cases, the acoustic cues were ambiguous; this was especially the case when /u/ appeared next to a nasal, since both sounds have a low F1 and a relatively low F2. In cases where the ambiguity was unresolvable, that token was excluded from the analysis.

3.2.1. Modelling

A generalized linear mixed effects model (glmer; Bates et al., 2015) was fitted using R. Mixed effects models allow us to group data by factors expected to cause some variability by incorporating random effects. In the present dataset, data is grouped by participant, and it is possible (and, in fact, expected) that the rate of vowel devoicing varies between them. It is reasonable, too, to suspect that the degree to which vowel devoicing is predicted by certain variables may vary between participants. Linear mixed effects models can account for both of these facts, in the form of random intercepts and random slopes, respectively. Factors expected to predict the likelihood of devoicing are included as fixed effects. The model was built stepwise, beginning with a baseline that included no fixed effects, to which individual factors were added incrementally. After each modification to the model, a likelihood ratio test (ANOVA) compared that model to the previous one, and variables significantly improving the model fit ($p < 0.05$) were maintained.

While there are existing computational tools specifically for modelling phonology such as Maximum Entropy Harmonic Grammar (MaxEnt; Goldwater & Johnson, 2003), here, I refrain from making claims directly about the grammar; rather, my goal at this stage is to model the data. A model of the grammar that produces the data discussed here appears in Chapter 5. Additionally, MaxEnt, which uses logistic regression, does not make use of a random effects structure and instead requires the user to collapse data across participants. For this reason, a mixed effects model is superior for the data at hand, which comes from multiple participants.

3.3. Factors affecting the probability of devoicing

The goal of this model is not to make claims about the underlying grammar that causes vowel devoicing; that will be attempted in Chapter 5. In other words, my goal is to model the *data* rather than the *grammar* as agnostically as possible. Of course, it would be impossible to model a linguistic phenomenon without reference to theoretical constructs, and so phonological factors like stress and sonority are invoked here. To summarize what is to come, I find that vowel devoicing is sensitive to sentential prosody, vowel identity, stress, the sonority of the preceding consonant and the manner of articulation of the following consonant. A summary including the levels that are most and least likely to result in a devoiced vowel (collapsing the two types of vowel devoicing described in Chapter 2) appears in Table 3.1.

Factor	**Most Devoicing**	**Least Devoicing**
Sentential prosody	IP-final	IP-initial
Vowel	/i/	/e/
Stress	Post-stress	Primary stress
Preceding consonant	Sonorant	Obstruent
Following consonant	Affricates, stops	Laterals, /r/

Table 3.1 Factors found to affect the probability of a vowel being devoiced and the environments that show the most and least amount of vowel devoicing.

3.3.1. Categorical voicing and devoicing

The dataset reveals several variables that are shown to categorically or near-categorically predict whether or not a vowel is devoiced: vowels are nearly always devoiced in the IP-final position (a phenomenon already described in Chapter 2) and nearly always voiced when the vowel is either primary-stressed, /e/ or non-final /a/. To identify these variables, an iterative process was implemented where the correlation between the rate of vowel devoicing and each level within a testing variable was measured. For example, within the variable "Stress," the percentage of tokens that were devoiced was calculated for each type of syllable in which a vowel appeared (primary stress, secondary stressed, etc.). If a level was shown to have a voicing or devoicing rate of 95% or higher (i.e., <5% or >95% devoiced), all tokens at that level were removed from the dataset and the correlations were calculated for all remaining items. This iterative process allows for us to find both main effects and interactions that (near-)categorically cause voicing/devoicing. Removing these items from the dataset was done to simplify the model and identify those factors that variably affect the likelihood of devoicing. Table 3.2 summarizes all of these findings.

As expected, utterance-final unstressed vowels were nearly all devoiced; as I discussed in Chapter 2, I consider final devoicing a phonetically and phonologically distinct process from medial devoicing, and this result is consistent with my early observation that all unstressed vowels, including /a/, are devoiced in the final position. Additionally, vowels appearing in primary-stressed syllables were nearly categorically voiced, regardless of the vowel. The mid front vowel /e/ was always voiced in the dataset, in both stressed and unstressed syllables. Important to note, though, is that native Merina words do not have unstressed final /e/, as that vowel neutralizes with /i/ in that position (Erwin, 1996; Howe, 2019; e.g., the root of /resena/ 'be defeated' surfaces unsuffixed as [resi] 'defeated'). So, it's not that IP-final unstressed /e/ is not devoiced, but rather that /e/ does not surface in that position. Finally, when stressed and final vowels were removed from the dataset, /a/ was voiced in 100% of tokens, indicating that medial, unstressed /a/ is categorically voiced. The result, then, is a dataset with non-IP-final unstressed and secondary-stressed /i/ and /u/.

Factor/Level	**Result**	**Percent with that result**
IP-final (unstressed) vowels	Devoiced	96.32%
Primary stressed syllables	Voiced	98.91%
/e/	Voiced	100%
Non-final, unstressed /a/	Voiced	100%

Table 3.2 Factors that have a 95% or higher rate of voicing or devoicing in the dataset.

3.3.2. Summary of the model

Here, I present the final state of the statistical model for Malagasy vowel devoicing. The resulting model, then, includes the fixed effects structure that produced the best model fit. The model includes one random slope, IP (Intonational Phrase) Position by Participant. Five fixed effects were found to significantly improve the model fit, added to the model in the following order: Vowel, Stress, IP Position, the Following Segment's Manner of Articulation, and the Preceding Segment's Sonority. The code for the model appears in (3.2). In order to simplify the dataset, the final model does not include those items matching the variables found to categorically predict voicing or devoicing (in Table 3.1). Additionally, certain factors were excluded as they created convergence issues, including the preceding consonant's place of articulation and all interactions.

(3.2)
```
model <- glmer(Voicing ~ Vowel + Stress + IP_position + Following_manner +
Preceding_sonority + (1 + IP_position | Participant),
        data = environment_data, family = binomial,
        control = glmerControl(optimizer = "bobyqa",
                    optCtrl = list(maxfun = 100000)))
```

Table 3.3 shows the summary of the final model. The reference levels are as follows: Vowel = /u/; Stress = pre-stress; IP Position = medial; Following Manner = nasal, Preceding Sonority = obstruent. Because voicing was represented binarily, where 1 = voiced and 0 = devoiced, a positive Estimate corresponds to an increased likelihood of voicing, whereas a negative one points toward more devoicing compared to the reference level.

Fixed Effect	Estimate	Std. Error	z value	p-value
(Intercept)	0.407	0.188	2.16	0.031
Vowel: /i/	-0.329	0.154	-2.14	0.032
Stress: none (single-syllable word)	-0.043	0.332	-0.13	0.898
Stress: post-stress	-0.453	0.178	-2.54	0.011
Stress: secondary	2.829	0.252	11.24	< 0.001
IP Position: initial	1.000	0.545	1.84	0.067
Following Manner: affricate	-2.951	0.382	-7.72	< 0.001
Following Manner: stop	-2.280	0.207	-11.01	< 0.001
Following Manner: fricative	-1.500	0.221	-6.80	< 0.001
Following Manner: pre-nasalized stop	0.167	0.368	0.45	0.650
Following Manner: pre-nasalized affricate	0.290	0.886	0.33	0.743
Following Manner: lateral	1.433	0.367	3.89	< 0.001
Following Manner: /r/	1.560	0.447	3.49	< 0.001
Preceding Sonority: sonorant	-1.803	0.194	-9.31	< 0.001
Preceding Sonority: /r/	0.264	0.247	1.07	0.284

Table 3.3 Summary of the generalized mixed effects model for Malagasy vowel devoicing, excluding factors shown to categorically predict voicing/devoicing.

3.3.3. Vowel

While /a/ was found to be nearly always voiced, even in unstressed syllables, it is the high vowels /i/ and /u/ that are most variable. However, the two vowels do differ in their rate of devoicing; /i/ is significantly more likely to be devoiced than /u/ ($\beta = -0.329$, $p = 0.032$).

3.3.4. Stress

In Section 3.1, we saw that vowels in primary-stressed syllables are nearly categorically voiced, including the high vowels /i/ and /u/. Stress has been suggested as a predictor of vowel devoicing in Malagasy (where only unstressed vowels may undergo devoicing (Pearson, 1994)), which is the typologically usual pattern. Stressed vowels consistently resist devoicing (R. C. Vogel, 2022) and deletion processes (e.g., in Uspanteko, Bennett et al., 2023; Latvian, Kariņš, 1995; and Faialense Portuguese, Silva, 1997; among others). In the present dataset, syllables were annotated as having primary stress based on dictionary entries (de La Beaujardière, 2024) and secondary stress was assumed to be regularly alternating. Whereas I described a series of stress-shifting suffixes in Chapter 1.1.2.1, no words with these suffixes appeared in the dataset. Unstressed syllables were categorized as either pre-primary stress or post-primary stress, with respect to the primary stressed syllable of the same word. The justification for distinguishing pre- from post- stressed syllables comes from certain phonological processes that have been shown to be sensitive to this position, such as the neutralization of /e/ and /i/ that occurs post-stress, described in Section 3.3.1, as well as data from other languages that indicate that the post-stress position may be a target for vowel devoicing (e.g., Brazilian Portuguese; Meneses & Albano, 2015). Additionally, a small number of single syllable words not reported to have any stressed syllable were categorized separately: the conjunction *sy* /si/ and the negation particle *tsy* /t͡si/.

There was a main effect of stress, with secondary stressed vowels significantly more likely to be voiced compared to all unstressed vowels, including post-stress ($\beta = 3.282$, $p < 0.001$), pre-stress ($\beta = 2.829$, $p < 0.001$), and single-syllable unstressed words ($\beta = 2.871$, $p < 0.001$). Among unstressed syllables, vowels were significantly more likely to devoice in the post-stress position compared to pre-stress ($\beta = -0.453$, $p = 0.011$), similar to the Brazilian Portuguese pattern. The raw counts for Stress appear in Figure 3.1, representing all data including those items excluded by the statistical model in Section 3.3.1. In this plot, the y-axis represents the proportion of vowels that are devoiced (shaded dark) and voiced (shaded light grey) in each stress condition. The width of each bar reflects the sample size, where wider bars represent conditions with more tokens.

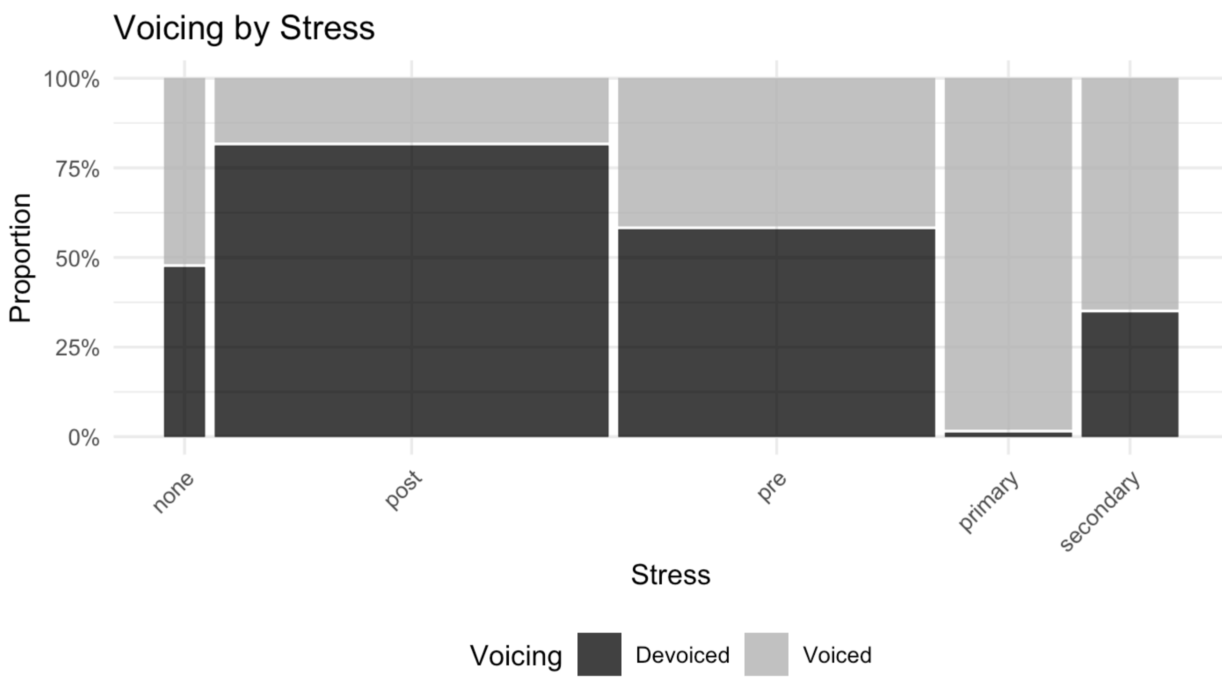


Figure 3.1 Raw counts for devoiced vowels by Stress type.

3.3.5. Phrasal Prosody

Malagasy vowel devoicing is sensitive to prosodic phrase boundaries. Malagasy prosody mirrors syntactic structure (Aziz, 2020, forthcoming; Raoniarisoa, 1990). That is, prosodic phrasing is isomorphic with syntactic phrasing and is marked with intonation. The highest level of the Malagasy prosodic hierarchy is the Intonational Phrase (IP), which corresponds to the sentential or clausal level. IPs are demarcated with a low boundary tone at the right edge (L% in Autosegmental-Metrical notation (Pierrehumbert, 1980)), though that boundary tone may be truncated if there is not enough phonological material on which to realize the tone (e.g., if another tone occupies the final syllables of the IP). Additionally, the final pitch accent (described below) of the IP is often weakened so that it has a relatively lower peak and amplitude compared to other pitch accents. In longer utterances, speakers may also pause after an IP.

Each IP consists of at least one phonological phrase, which corresponds to maximal syntactic phrases below the clause level. For example, in a simple VOS declarative sentence, there are two major syntactic phrases, the predicate (VO) and the subject (S), each of which forms a phonological phrase; more complex structures may have additional phonological phrases. Malagasy phonological phrases are marked with a steep rising pitch accent (usually labelled L+<H*) on the rightmost stressed syllable of the phrase and an optional low boundary tone (L-) athe right edge.

Prosodic boundaries are frequently targeted for devoicing processes across the world's languages. The most common prosodic position for devoicing is the "final" position. While languages differ with respect to what "final" refers to, there is an implicational hierarchy where the presence of devoicing at the edge a smaller domain implies devoicing at the edge of larger domains (Gordon, 1998). In Malagasy, we have already seen in Section 3.3.1 that IP-final unstressed vowels, regardless of height, are nearly always devoiced, and in Chapter 2, I showed that there is good reason to believe that this final devoicing is a distinct process from non-final devoicing. In the present statistical model, then, it is a question of whether other prosodic phrase boundaries contribute to the probability of devoicing.

Each item was coded for the target syllable's position in the utterance (i.e., the entire recorded item), the IP and the phonological phrase. Because the IP usually corresponds to the clause, its boundaries were determined by punctuation: syllables before a punctuation mark were coded as IP-final and those appearing at the beginning of the utterance or following punctuation were coded as IP-initial; all other vowels were coded as IP-medial. For short sentences, the phonological phrase boundaries were coded based on syntactic structure (e.g., if the vowel appeared in the predicate-final position) and for more complex sentences, I identified the position of pitch accents in Praat to confirm phrase boundaries. In (3.3) is an example sentence that appeared in the dataset with the Utterance (Utt.), IP, and phonological phrase (φ) boundaries marked.

(3.3)	Madio		ve	ny	trano?		Nahita ny	famafa		aho.
	Clean		Q	DET	house		Pst.see DET	broom		I
	[									] Utt.
	[				]IP	[				]IP
	[ ]φ	[ ]φ	[		]φ	[		]φ	[	]φ

‘Is the house clean? I saw the broom.’

Recall that the dataset used in the final statistical model includes non-IP-final unstressed high vowels. Three models were attempted with different prosodic positions as predictors: the utterance, the IP and the phonological phrase. For Utterance and IP position, the random effects structure in (3.2) was changed to match the prosodic category being incorporated. For Phonological Phrase position, however, the model fit was singular, indicating that the random effects structure was too complex; in this case, the model was run with the random slope for Phonological Phrase by Participant removed. The model fits were compared using ANOVA, the results of which are below in Table 3.4.

	npar	AIC	BIC	logLik	deviance	Chisq	Df
Phon. Phrase	19	1366	1468	-664	1328		
Utt.	20	1355	1462	-657	1315	13.19	1
IP	20	1352	1459	-656	1312	2.75	0

Table 3.4 ANOVA results comparing position in the utterance (Utt.), in the IP (IP), and in the phonological phrase (Phon. Phrase) as predictors of vowel devoicing.

Position in the IP as a predictor was found to provide the best model fit, though it was only marginally better than position in the utterance (AIC: 1352 vs. 1355; BIC: 1459 vs. 1462; logLik: -656 vs. -657). When IP position was included as the predictor, vowels were found to be more likely voiced in the IP-initial position compared to medial syllables; however, this difference was not found to be significant when $\alpha = 0.05$ ($\beta = 1.004$; $p = 0.067$). Raw counts by IP position appear in Figure 3.2 and by Phonological Phrase position in Figure 3.3.

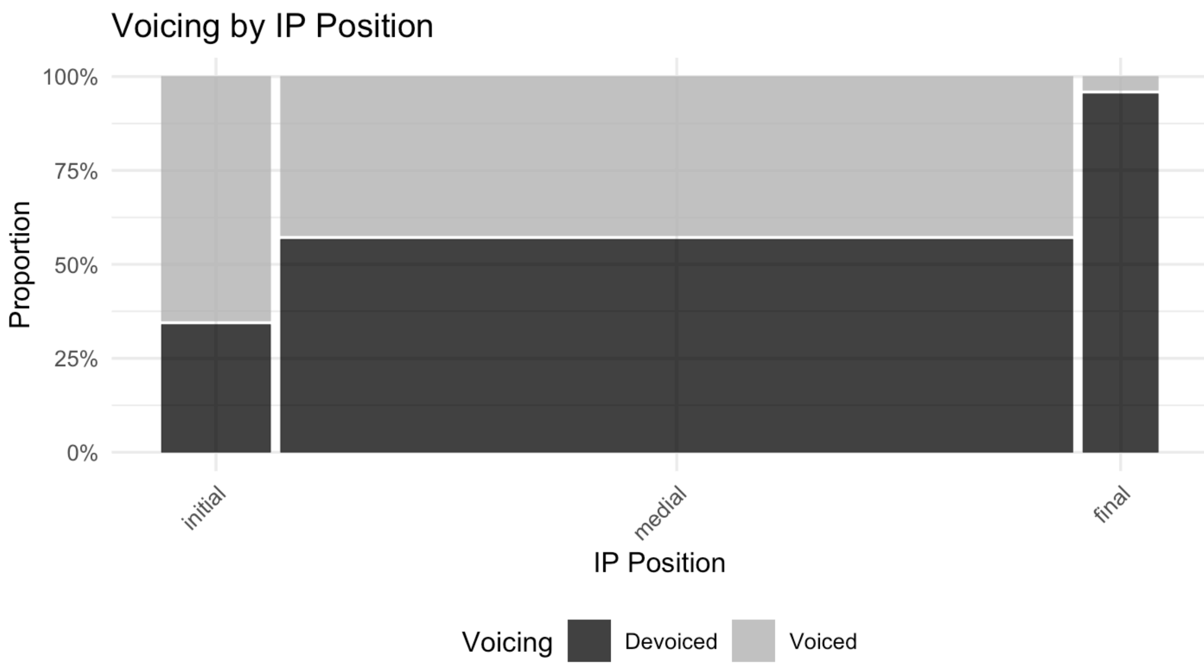


Figure 3.2 Raw counts for devoiced vowels by position of the syllable within the Intonational Phrase.

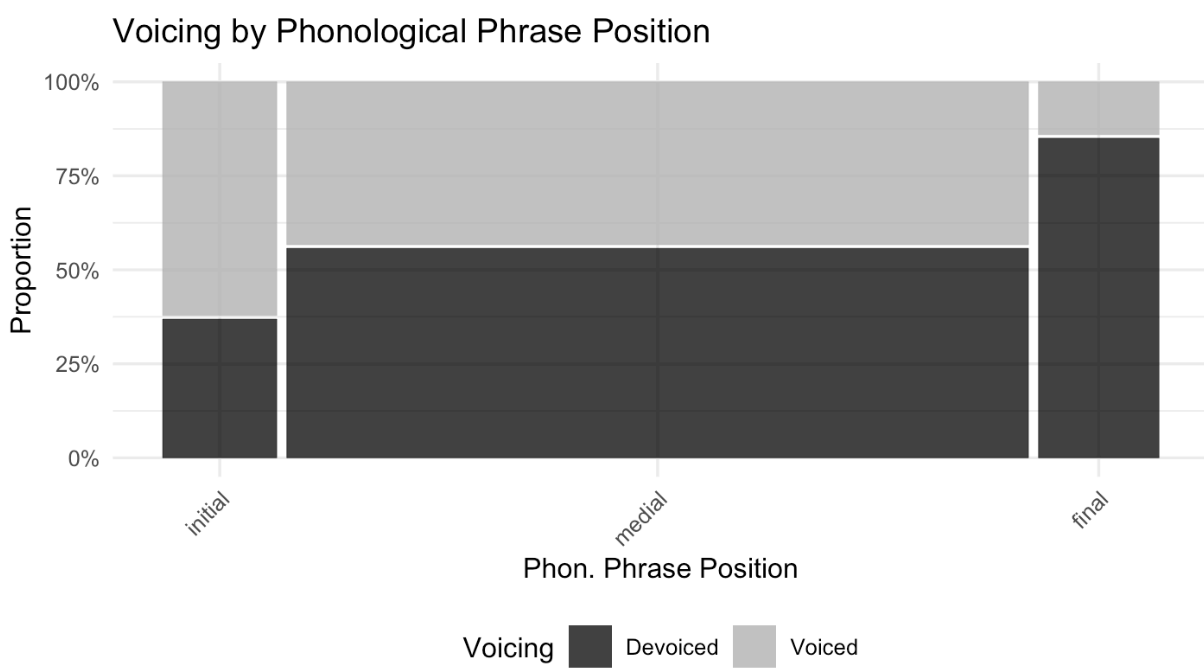


Figure 3.3 Raw counts for devoiced vowels for position of the syllable within the Phonological Phrase.

3.3.6. Segmental Context

The final phonological factor that was found to significantly improve model fit was the segmental environment surrounding the vowel. Languages that do have devoicing and deletion processes vary in the segmental environments where voiced vowels are more or less likely to emerge, but there are some patterns. In Gordon's (1998) and Vogel's (2022) typologies, they identify voiceless consonants as a frequent environment for adjacent vowels to devoice. This is the case for languages like Japanese (Tsuchida, 1997, among others), Andean Spanish (Delforge, 2008), Korean (Jun et al., 1997), Quebec French (Bayles, 2016) and others, where vowels must be flanked on one or both sides by a voiceless consonant. Devoicing is unlikely to occur next to voiced consonants except in domain-final positions where, in fact, vowels are *likely* to devoice: 29 of 36 final-devoicing languages in Gordon's survey show vowel devoicing regardless of the voicing of the preceding consonant.

Other segmental factors have been shown affect the likelihood of devoicing, primarily those factors that relate to the duration of the adjacent consonants. Gordon finds that fricatives and aspirated stops, those consonants that have a relatively longer duration, are the most likely environments next to which vowels are devoiced. Jun et al. (1997) detail the complex ways in which segmental and prosodic factors related to duration conspire to devoice high vowels in Korean. In the Accentual Phrase-initial position, where consonants have a longer closure, following vowels are more likely to be devoiced. Korean's three-way laryngeal contrast further exemplifies the correlation between vowel devoicing and the preceding consonant's duration: vowels are more likely to devoice after an aspirated stop or plain (aspirated) fricative (which have the longest glottal opening) compared to a fortis consonant (an environment that produced almost no vowel devoicing).

For Malagasy, the statistical model indicates that both the preceding and following consonant is an important predictor for non-final vowel devoicing. Target vowels were coded as preceded by either a sonorant (nasal, lateral), obstruent (stop, fricative, affricate, pre-nasalized stop, pre-nasalized affricate), or /r/. /r/ was coded separately because of the variability in its production, discussed in Chapter 1.1.3.1. Pre-nasalized consonants (e.g., /^{n}t/) were coded as obstruents because the component immediately preceding the vowel is oral.

There was a significant main effect of the sonority of the preceding consonant. Vowels were more likely to be devoiced after sonorants compared to obstruents ($\beta = -1.952$, $p < 0.001$) and /r/ ($\beta = -2.23$, $p < 0.001$). The difference between obstruents and /r/ was not significant ($\beta = -0.281$, $p = 0.261$). Figure 3.4 shows the raw counts for the rate of devoicing based on the preceding consonant's sonority.

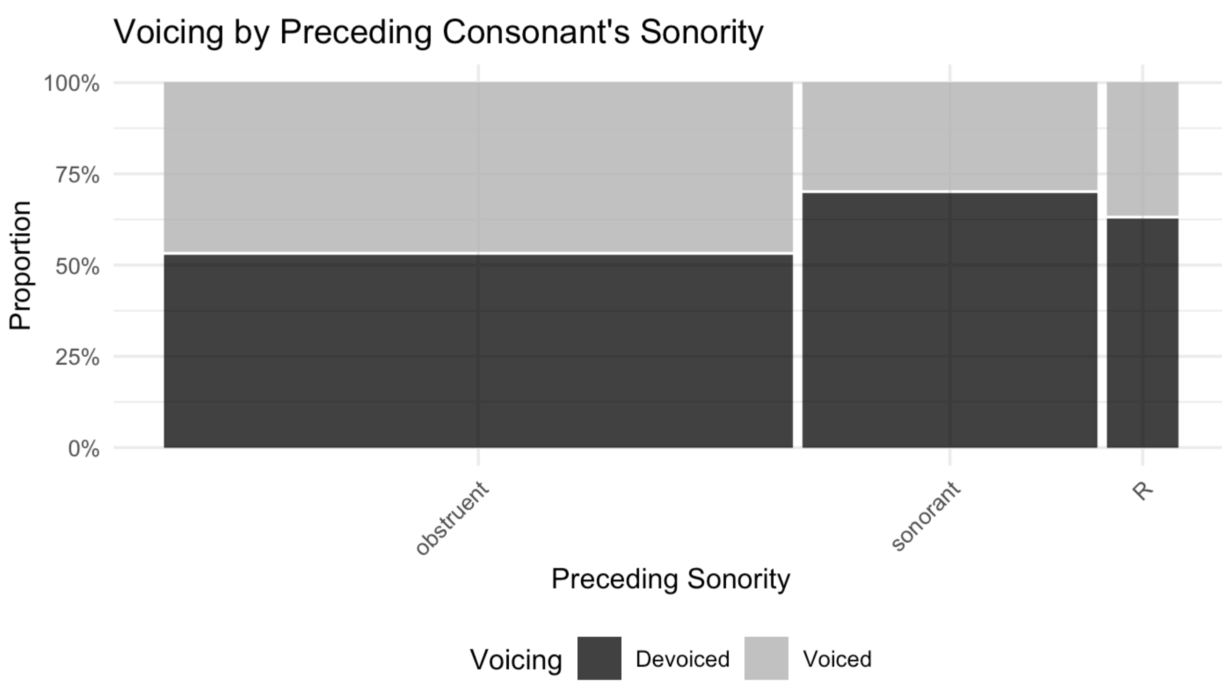


Figure 3.4 Raw counts for devoiced vowels by the sonority of the preceding consonant.

There was also a significant main effect of the consonant following the target vowel. Initially, the model was fit using sonority as a predictor, as with the preceding consonant; however, a likelihood ratio test revealed that specifying the manner of articulation significantly improved the model fit ($p < 0.001$). The model reveals that the likelihood of devoicing is affected by the sonority of the following segment, with a general tendency toward devoicing the less sonorous that segment is. (3.4) shows the likelihood of devoicing based on the coefficients produced by the model.

(3.4) **Likelihood of vowel devoicing by the following consonant**
Affricates ≥ stops > fricatives > nasals ≥ pre-nasalized stops ≥ pre-nasalized affricates ≥ laterals ≥ /r/

In most cases, the comparison between two manners is significant, as summarized in Table 3.5.

Reference	**Predictor**	**Estimate**	**p-value**
Stops	Fricatives	0.779	0.002
Fricatives	Nasals	1.500	< 0.001
Nasals	Laterals	1.433	< 0.001
Pre-nasalized stops	Laterals	1.266	0.009

Table 3.5. Likelihood of vowel devoicing by the following consonant's manner of articulation, showing comparisons that were significant.

However, differences between other categories were not significant; Figure 3.5 shows the coefficients for each manner of articulation with "nasal" as the reference level and a 95% confidence level represented by the error bars, showing the high degree of overlap with the other sonorant categories. Additionally, Figure 3.6 shows the raw counts for each level, with pre-nasalized stops and pre-nasalized affricates collapsed into one category for low counts among the affricates.

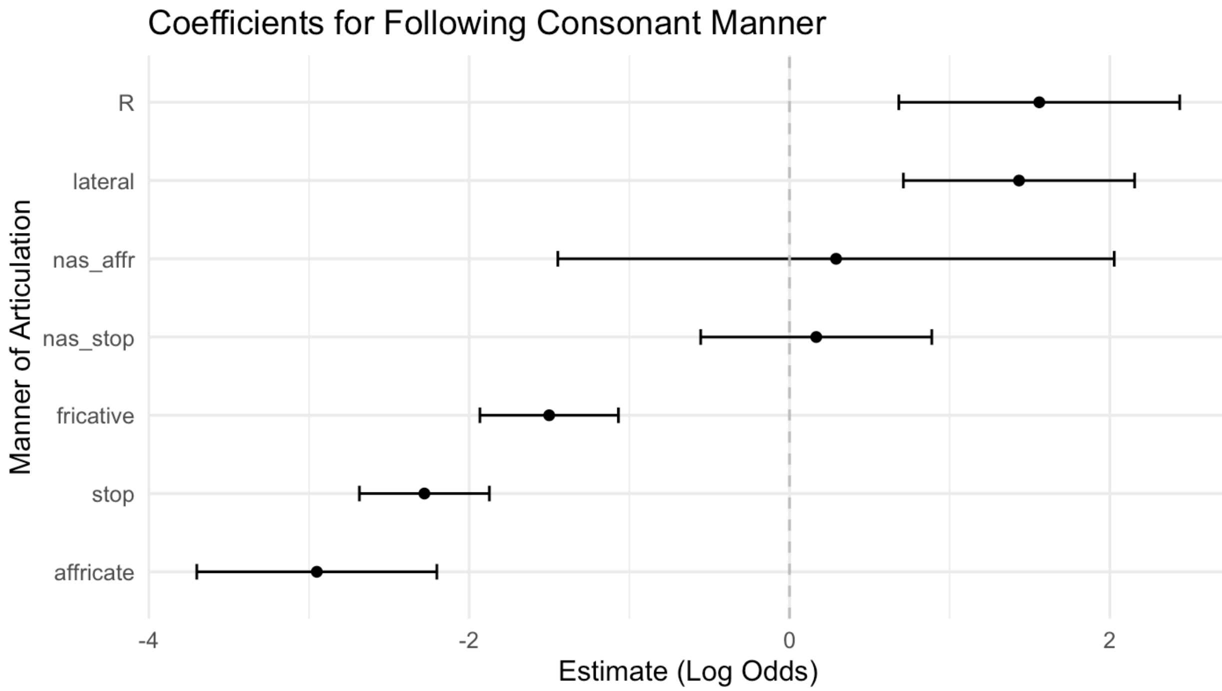


Figure 3.5 Coefficients for the manner of articulation of the following consonant as a predictor of vowel devoicing, with "nasal" as the reference level.

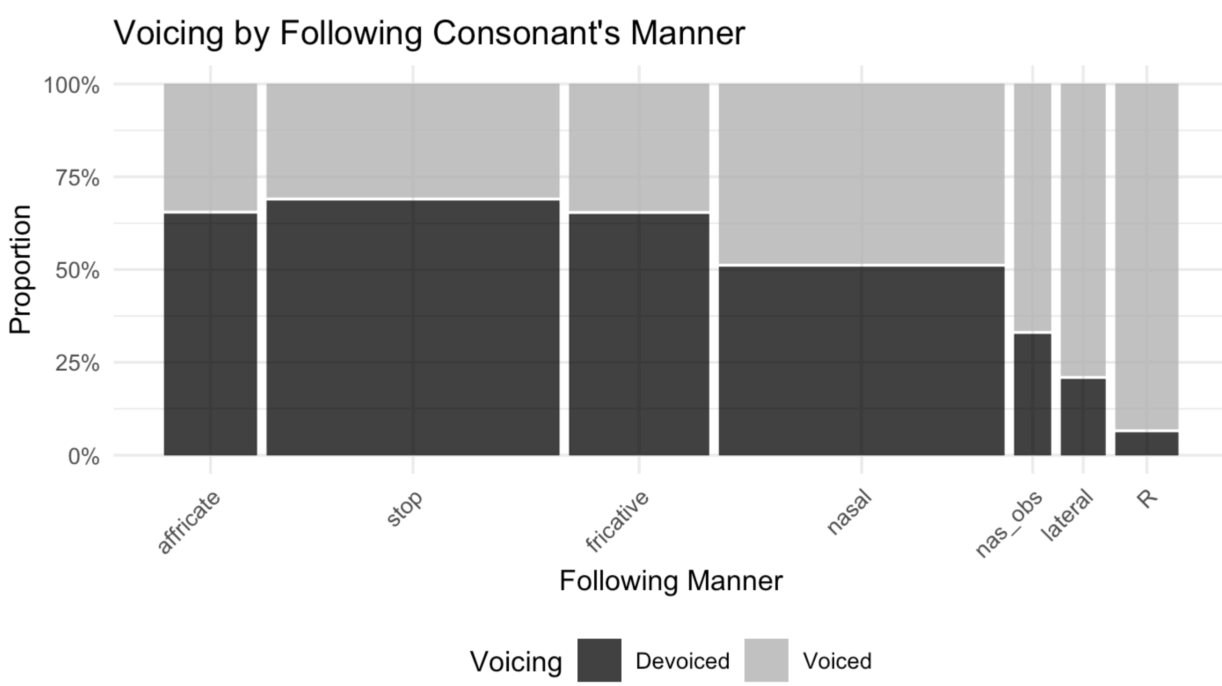


Figure 3.6 Raw counts for devoiced vowels by the manner of articulation of the following consonant.

This result indicates that the likelihood of devoicing of a non-IP-final Malagasy high vowel is conditioned by the following consonant in line with conceptions of the sonority hierarchy. Parker, (2008), for example, proposes the hierarchy in (3.5), from least to most sonorous (excluding categories not relevant here), nearly identical to the predictor of Malagasy vowel devoicing in (3.4). One thing to note is that Malagasy vowels are more likely to devoice before laterals than /r/; however, as mentioned, /r/ is extremely variable and may be realized as a flap, trill or voiced fricative, among other realizations. To conclude, then, Malagasy vowel devoicing is sensitive to the gradient sonority of the following consonant, in line with descriptions of a sonority hierarchy.

(3.5) **Sonority hierarchy proposed by Parker (2008)**
Voiceless stops > voiceless affricates > voiceless fricatives > nasals > flaps > laterals > trills

3.3.7. Factors not found to improve model fit

Certain factors were incorporated into the statistical model but found to not improve the model fit. As described in Section 3.3.5, the Intonational Phrase boundary was found to be the best predictor of vowel devoicing compared to the utterance and phonological phrase. Additionally, manner of articulation of the preceding consonant was incorporated instead of sonority to test whether the gradient effect of sonority found for the following consonant was replicated; however, that model did not improve the log likelihood and so the simpler model was used instead. Finally, morphology was added as a predictor, with vowels categorized as either belonging to the Active verbal prefix *mi-* or not; this was not found to significantly improve the model fit. Other models with additional factors were attempted but would not converge or produced errors; this included the preceding consonant's place of articulation and item production order (based on the assumption that participants' speech patterns change across the course of the experiment). This is not to say that these factors do not affect devoicing, but rather that they were not testable with the present dataset. Finally, it is worth noting that voicing of the adjacent consonants was not included in the model because, as described in Chapter 1.1.3.1, the voicing contrast in obstruents has been largely neutralized so that they are all voiceless.

3.4. Model fit

In this chapter, I have presented a statistical model of Malagasy vowel devoicing trained on the data from a production experiment with six speakers. In order to test the model fit, I used the predict() function in R to predict the probability of voicing of each item type with reference to the fitted training model. "Type" was a combination of vowel-stress-preceding sonority-following sonority-IP position (e.g., post-obstruent, pre-nasal, IP-medial secondary-stressed /i/). The actual observed probability of voicing was calculated as the mean rate of voicing by type (e.g., if 10 of 20 tokens of a type were devoiced, its rate of devoicing was 0.5) and correlation coefficients were calculated to compared predicted vs. observed rates of voicing using the cor() function. The correlation for the dataset was 86%; the absolute error (i.e., the absolute difference between the observed and predicted values) was 13%. Figure 3.7 plots the predicted vs. observed rates of voicing for each type.

The statistical model presented in Section 3.3 is able to explain many of the patterns underlying Malagasy vowel devoicing, pointing to generalizations about the data that will prove helpful in explaining the grammar of devoicing in Chapter 5; however, as we saw in this section, the model is not totally generative. Looking at the absolute errors, the model seems to have an especially hard time predicting devoicing for secondary-stressed vowels; the four types with the greatest errors are all secondary-stressed. As discussed in Chapter 1.1.2.1, secondary stress is somewhat complicated and understudied in Malagasy. Recall that there are certain suffixes in

Malagasy that shift primary stress while secondary stress remains faithful to the stem; other suffixes do cause secondary stress to shift. It is possible that our understanding of Malagasy secondary stress is incomplete; the model's difficulty in predicting devoicing for these syllables emphasizes the need for further research in this area.

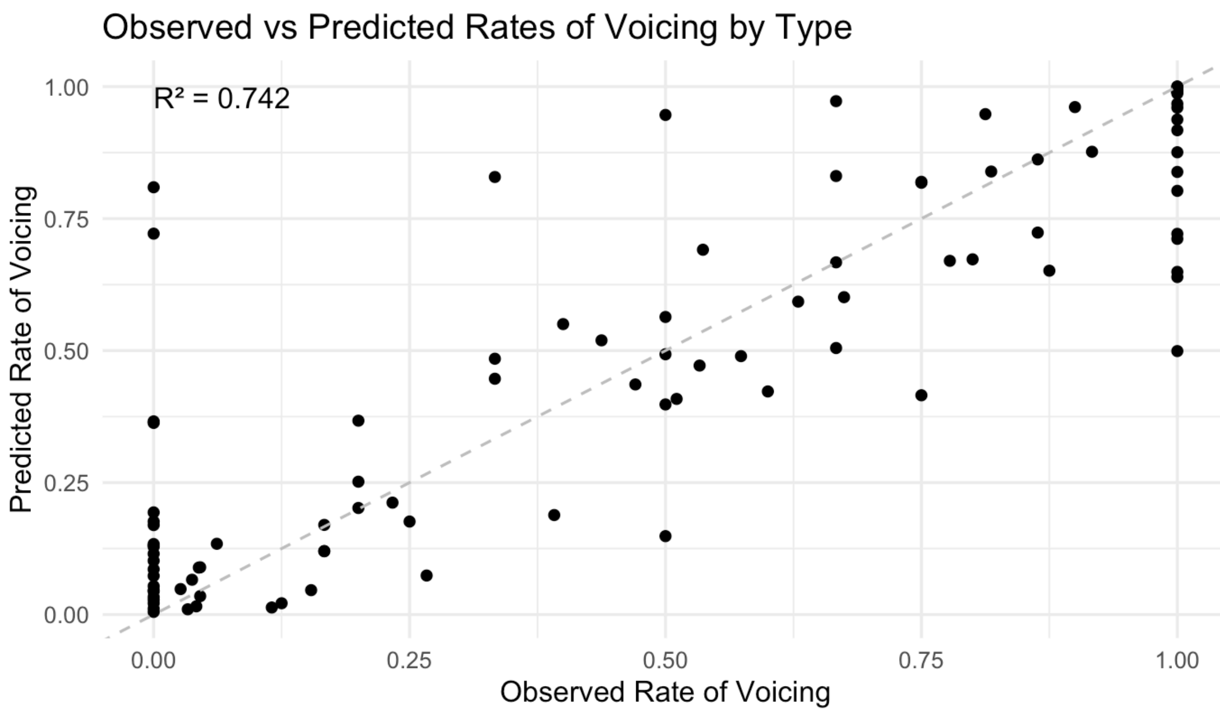


Figure 3.7 Predicted vs. observed rates of vowel voicing based on the model in (3.2).

Chapter 4
Vowel devoicing is coarticulation: An acoustic analysis

4.1. Introduction

In this chapter, I present the results of a quantitative acoustic study of the segments involved in non-final vowel "devoicing" in Malagasy. Following the qualitative sketch of vowel devoicing in Chapter 2, I determined that these so-called "devoiced" vowels in non-final positions do not resemble voiceless vowels at all, unlike devoicing in the phrase-final position. Rather, there appears to be no vowel segment at all, voiced or voiceless, in the devoicing environment. In Chapter 3, I identified the environment of non-final devoicing, showing that vowels are most likely to devoice in unstressed non-final syllables, following a sonorant and preceding an obstruent.

The question remains whether a vowel is *articulatorily* present in the Malagasy at all; that is, are these vowels categorically deleted in the output, or does the Malagasy speaker still attempt a vowel-like articulation, but one that is overlapped by other segments' gestures? The goal of the chapter, then, is to look closely at the acoustic reality of Malagasy vowel devoicing to see if there is evidence of a vowel gesture in the devoicing environment. To do this, I take measurements of the spectral properties and duration of consonants preceding devoiced vowels. These measurements are then compared to those of consonants before clearly voiced vowels, where coarticulatory effects of an underlying vowel should be plainly visible. The conclusion is this: consonants show similar acoustic properties before both voiced and devoiced vowels; i.e., consonants in the devoicing environment still show effects of coarticulation with the following vowel, even if a voiced vowel does not appear as a unique segment in the output. In Chapter 5, I expand on this result and suggest that even in cases of devoicing, the underlying vowel is present as a gesture that is overlapped by its consonantal environment, producing the phonetic result presented here.

The chapter is divided into two principal data sections looking at different acoustic properties: spectral properties and duration. Within each section, I compare the voicing and devoicing environments: in stressed medial syllables, where we expect voiced vowels, we will see the usual consequences of consonant-vowel coarticulation in Malagasy; then, in unstressed medial syllables, an environment that triggers devoicing of /i/ and /u/, the acoustic analysis of the vowel's environment will reveal parallels to the coarticulation shown for voiced vowels, motivating our conclusion that what has been called "vowel devoicing" is acoustically most similar to coarticulation.

4.2. Methods

The acoustic description is based on recordings of 14 speakers of Merina from Antananarivo, recorded in a sound-attenuated booth on McGill's campus in Montreal, Quebec. Each speaker was also fluent in French, and several had knowledge of English and other dialects of Malagasy, including Betsileo, Sihanaka and Sakalava. Participants ranged in age from 19 to 41 and had spent at least their first 14 years living in Madagascar before arriving to Canada, with the average number of years in Madagascar being 24. Two reading tasks were used to elicit the data, which were recorded using an AT2020 microphone at a sampling rate of 44100 Hz; a single participant was recorded with a built-in MacBook microphone.

4.2.1. Task One: Real words

Two tasks were used to elicit the data analysed here. Task 1 was designed to elicit the production of real Malagasy words containing /i/, /u/ and /a/ in unstressed medial syllables, an environment expected to trigger devoicing. Target words manipulated the local segmental environment, intending to elicit vowels next to both sonorants and obstruents; in this chapter, I present data from three environments: as_ka, at_ka, an_ka. The goal here was not to produce an exhaustive set of segmental environments, but rather to examine the acoustic realization of the vowel devoicing process on different phonological natural classes (fricatives, stops and nasals). Each vowel (/a, i, u/) appeared in each environment ten times, including some repetitions of the same word due to low frequency of that environment in the lexicon, allowing a maximum of 90 items per participant. Participants were instructed to exclude any items with words that they didn't recognize, leaving a total of 825 tokens cross all 14 participants. The full list of stimuli appears in Appendix B.

Participants were presented each target word in the frame sentence *Niteny [word] ilay zazavavy*, 'The girl said [word],' each item presented individually on a screen. Participants were instructed to imagine that they are in a café with a Malagasy friend when they think that they overhear a group of teenagers speaking Malagasy. When the friend asks how the participant knows that they were speaking Malagasy, the participant would then respond with the frame sentence.

4.2.2. Task Two: Nonce words

Task 2 elicited additional data using nonce words, the primary goal of which was to elicit minimal pairs manipulating stress. Because the expected environment of vowel devoicing is in unstressed syllables, having stressed/unstressed pairs allows for a clear comparison of the acoustics of voiced and devoiced vowels, but because true minimal stress pairs are rare in Malagasy, nonce words were used. Stress was indicated orthographically using a grave accent (e.g., *Mbaràsaka* vs. *Mbarasàka*), as is often used to mark stress minimal pairs in writing (e.g., *lalana* 'road' and *lalàna* 'law'). Since a small number of Merina Malagasy words are written with <ô> to represent the marginal phoneme /o/ (e.g., the vocative particle *ô*), the experimenter emphasized that the grave accent represented the stressed syllable and gave participants the opportunity to repeat each item until they produced the target stress pattern. Only those nonce words with stressed target vowels are analyzed in this chapter.

This dataset included two environments that appear in the present chapter: as_ka and at_ka. Unfortunately, because this task was not originally designed for direct comparison with Task 1, the environment an_ka is not available in this dataset. Like with Task 1, participants were presented with target words in a frame sentence, this time *Antsoina [word] ilay tanàna*, 'The town is called [word].' Participants were instructed to imagine they are on a road trip through Madagascar with a Malagasy friend, passing by villages with names they've never seen before, at which point they produce the item. The items in this dataset that contained voiced /a, i, u/ are used in this chapter as a baseline for the acoustic properties of consonant-vowel coarticulation in conjunction with existing literature on this topic, and so the number of tokens elicited was smaller, at just three per vowel per environment (two segmental and two prosodic), totalling a maximum of 36 items per participant. The list of stimuli is given in Appendix C.

4.3. Spectral traces of "devoiced" vowels

In order to compare Malagasy "devoiced" vowels to typical vowel coarticulation (and thus test the hypothesis that traces of a vowel-like articulation are present even when the vowel is "devoiced"), this section presents the spectral properties of the environment surrounding these vowels, focussing on three environments: underlying vowels following /s/, /t/ and /n/. For /s/ and /t/, spectral peak and the spectral moments are measured; for /n/, F2 of the preceding /a/ is used as an estimate of vowel-to-vowel coarticulation. In each of these cases, the segmental environment shows evidence of the underlying vowel, even when it is "devoiced".

4.3.1. /s/

The four spectral moments have been used for decades to quantify the acoustic properties of fricatives. The first moment, center of gravity (CoG), is a weighted average of all frequencies, taking into account each frequency's relative intensity. It is correlated with the perception of pitch so that fricatives with higher CoG are perceived to have higher pitch than those with lower CoG. The second moment is the variance, usually measured as (and referred to as) the standard deviation of the spectral frequencies. The third moment, skewness, reflects the symmetry of the spectrum, where a negative value represents more energy in the higher frequencies, and a positive value represents more in the lower frequencies. The fourth moment, kurtosis, reflects the uniformity of the spectrum and is thus sometimes referred to as peakedness; a higher kurtosis reflects a more peaked distribution, and a lower kurtosis a more flat distribution (Forrest et al., 1988; Jongman et al., 2000).

Early research showed potential in using these moments for distinguishing sibilants by place of articulation. Forrest et al. (1988) find that the moments distinguish English /s/ from /ʃ/, with skewness being the most revealing moment. In contrast, Shadle and Mair (1996) identify the other three moments as most helpful. Jongman et al. (2000) find that all four moments are effective in distinguishing /s/ from /ʃ/. Some work has also investigated the effect of consonant-vowel coarticulation on the spectral moments of fricatives; Shadle and Mair (1996) find that kurtosis was higher for /s/ between two /u/ compared to /i/ and /a/ and that standard deviation was lower between /u/. Others acknowledge that lip rounding lowers CoG (e.g., Munson, 2001). Since then, the spectral moments, especially CoG, have been used to study fricatives in languages other than English (e.g., Gordon et al., 2002).

Despite their wide-spread usage, researchers in the signal processing literature have challenged the use of the spectral moments for the study of speech sounds, especially fricatives. Shadle (2023) outlines some of the issues with this approach. Methodologically, measures of the moments can be inconsistent due to differences in sampling rate, the presence of low-frequency energy in voiced segments, and background noise that is captured in the signal, though these issues may be remedied by consistent recording criteria, filters and soundproofing, respectively. Shadle also points to some characteristics of speech that are simply not captured by the moments: whereas the sibilants /s/ and /ʃ/ are reliably distinguished, non-sibilants like /f/ and /θ/ overlap when looking just at the four spectral moments, and factors like overall loudness are not considered due to normalization. Further, she brings up the fact that some of the spectral moments are correlated with each other; for example, a lower CoG is correlated with a higher skewness, rendering these moments redundant.

These issues have led phoneticians to consider alternative acoustic measurements for fricatives. Shadle (2023) argues for the use of several other parameters that, in her view, better reflect the production of fricatives including the filter, source and "underlying source characteristic." These parameters are based on earlier work (Shadle, 1985; 1990), which used mechanical models to manipulate source and filter and straightforwardly measure the acoustic consequences. Of interest to us are measurements of the filter, since coarticulation of a vowel on /s/ may, in principle, involve manipulation of the tongue tip (associated with /s/), tongue dorsum (associated with the vowel) and the lips (associated with rounded /u/). For this, Shadle (2023) proposes measuring spectral peak (F_M), the highest peak within a defined range; for /s/ she proposes 3-7 kHz for men and 3-8 kHz for women. Spectral peak measures have sometimes been used in conjunction with the spectral moments, and indeed it has been shown to vary with vowel context. Jongman et al. (2000), looking at the highest peak within the entire spectrum, find that spectral peak is lower for /s/ and /z/ in the context of the back round vowels [o] and [u]. Soli (1981) finds a spectral peak in the range of 1500-2000 Hz for sibilants, which is higher before the front high vowel [i] compared to [u] and [a], analogous to the effect of backness on the second formant.

In this section I present both F_M and the spectral moments for Malagasy /s/ coarticulation. While Shadle makes a compelling argument against moments for fricatives, I acknowledge the history of those measurements in the phonetics literature and present them here for thoroughness. /s/ was segmented based on the duration of frication identified by high-amplitude aperiodic noise in the spectrogram and waveform. In many cases, /s/ frication is followed by a short period of much lower-amplitude noise, depicted in Figure 4.1. This duration was excluded from spectral measurements.

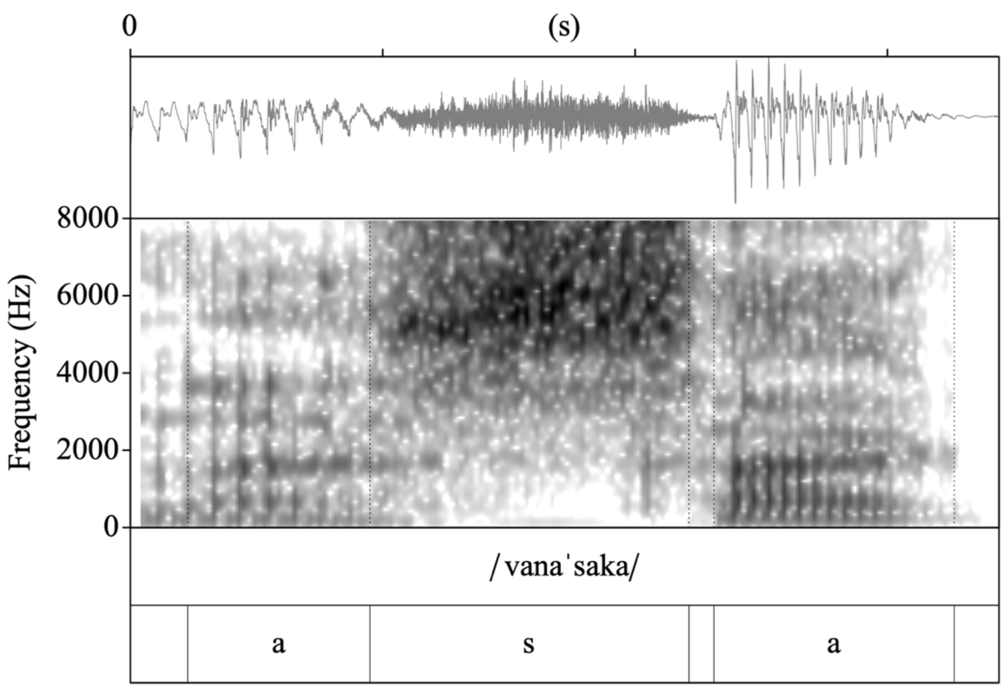


Figure 4.1 Example segmentation of /s/ for measurement of spectral properties. The short duration of low-amplitude noise between frication and the onset of the vowel was not included for these measurements.

4.3.1.1. Spectral peak

Spectral peak measurements for /s/ were extracted in line with Shadle's (2023) definition. A modified Praat script based on DiCanio (2021a; Appendix D) subdivided each /s/ into 3 intervals, each with overlapping windows of 15ms each (five windows per interval for the voicing environment and three windows per interval for the devoicing environments, based on the longer average duration for stressed /s/). For each window, a Fast Fourier Transform (FFT) spectrum is generated using Praat's "To Spectrum…" function and a time-averaged spectrum is created across the windows in each interval. On the time-averaged spectrum, the script loops through the frequency bins in the pre-defined frequency range (3000-7000 Hz for men and 3000-8000 Hz for women, following Shadle) and identifies the frequency bin with the highest magnitude. The use of frequency bins means that the resulting F_M value is rounded to the nearest bin value.

Figure 4.2 shows the average FM values for the voicing environment by interval, across participants. The intervals are linear, where Interval 1 is the beginning third of /s/, 2 is the middle and 3 is the final third. Rather than plotting a single time-averaged spectrum and F_M values, plotting each interval allows us to see how the spectral properties change across the fricative and identify any potential effects of coarticulation with adjacent segments. A linear mixed effect model (lmerTest; Kuznetsova et al., 2017), shown in (4.1), was run for the middle interval only. Spectral peak (F_M) was the dependent variable and vowel context was included as a predictor. Participant was included as a random effect. The middle interval was chosen for statistical investigation as I expect it to be the most stable representation of consonant-vowel coarticulation. In the first third, coarticulation with the preceding /a/ is possible; in the devoicing environment, the final third may show increased coarticulatory effects with the following /k/.

(4.1) fm_model_voicing_s <- lmer(fm ~ vowel + (1 | participant), data = voicing_environment)

In the voicing environment, the model shows an effect of vowel context, with /s/ having a significantly lower spectral peak before /u/ than both /a/ ($\beta = -1542.7$, $p < 0.001$) and /i/ ($\beta = -1488.8$, $p < 0.001$); the difference between /si/ and /sa/ is not significant ($p = 0.79$). Thus, coarticulation with /u/ appears to lower spectral peak of /s/, consistent with the previous studies reported above in Section 4.3.1.

In the devoicing environment, the pattern is similar, shown in Figure 4.3. An analogous linear mixed effects model was run for the devoicing environment. The model reveals a main effect of vowel context: F_M of /s/ is significantly lower when there is an underlying /u/ vowel compared to /i/ ($\beta = -816.7$, $p < 0.001$). While only the high vowels /i/ and /u/ are devoiced in this environment, Figure 4.3 (and the figures that follow) additionally plots the /s/ before (voiced) /a/ to serve as a baseline for the typical spectral properties of /s/ in unstressed medial syllables. The difference between /u/ and /a/ was also significant ($\beta = -879.4$, $p < 0.001$). As in the voicing environment, there was no significant difference between underlying /sa/ and /si/ ($p = 0.07$).

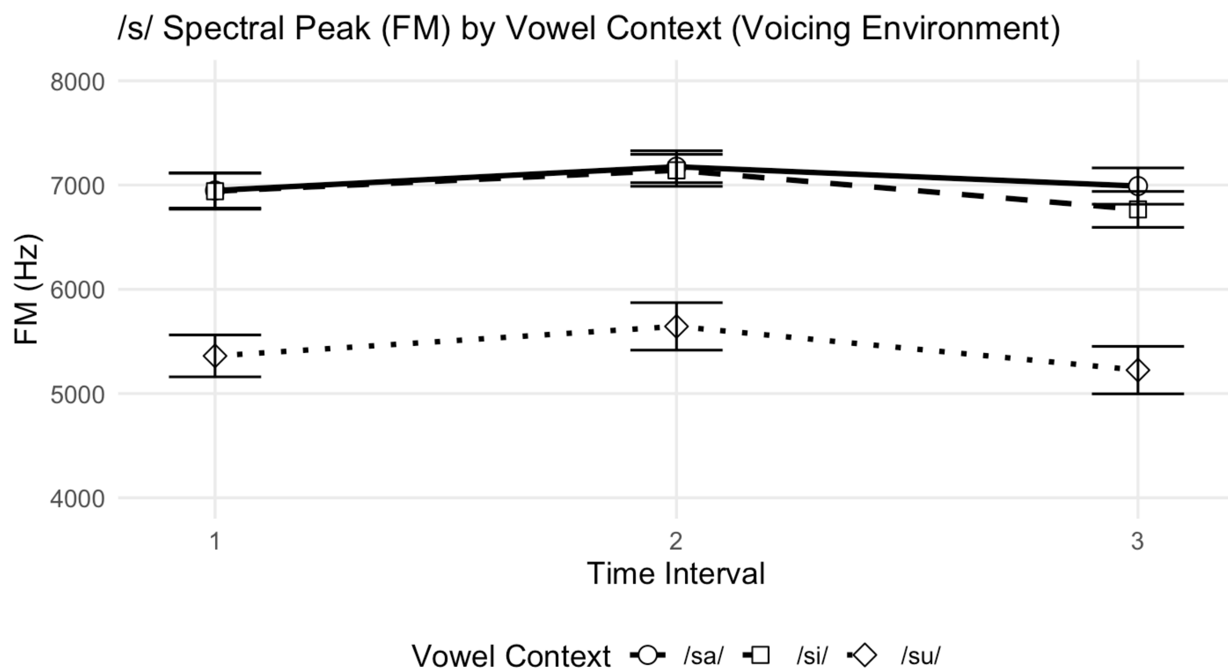


Figure 4.2 Spectral peak (F_M) measurements across /s/ by vowel context in the voicing environment.

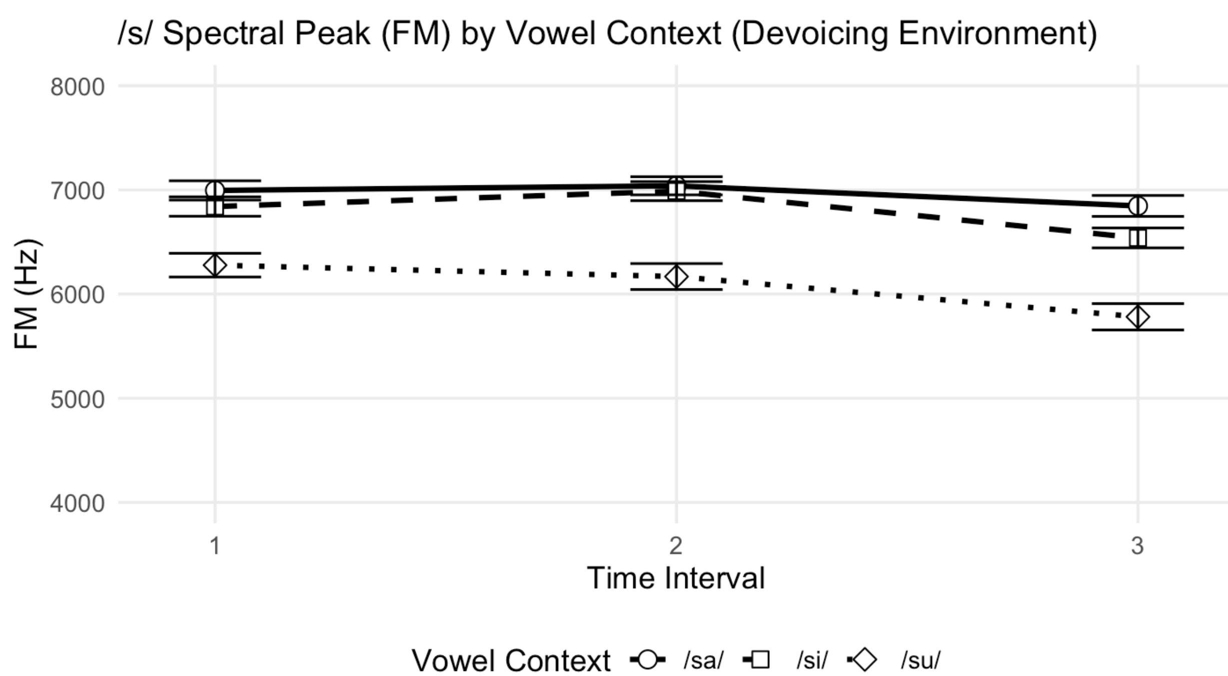


Figure 4.3 Spectral peak (F_M) measurements across /s/ by vowel context in the devoicing environment.

In terms of spectral peak, then, the effect of the underlying vowel can be plainly seen even in the devoicing environment, when there is no voiced vowel segment present: just as F_M was lowered before voiced /u/, this coarticulatory effect is likewise present in cases of vowel devoicing, despite the absence of a vowel segment in the spectrogram.

4.3.1.2. Spectral moments

In additional to spectral peak, the first four spectral moments were measured. As with the spectral peak measurements, a Praat script, based on DiCanio (2021a), generated time-averaged spectra across a specified number of windows within three intervals and calculated the spectral moments based on the method presented by Forrest et al. (1988). Additionally, a high-pass filter at 300 Hz was applied to filter out any low-frequency energy from voicing. For each moment, two linear fixed effects models (for the voicing and devoicing environments) were run for the middle interval of /s/ with the relevant spectral moment as the dependent variable, vowel context as the independent variable and participant as a random effect.

4.3.1.2.1. Center of gravity

In the voicing environment, there was an effect of vowel context on spectral center of gravity (CoG), with /s/ CoG significantly lower before /u/ compared to both /i/ ($\beta = -316.0$, $p = 0.013$) and /a/ ($\beta = -379.7$, $p = 0.004$). The difference between /si/ and /sa/ is not significant ($p = 0.617$). This mirrors the result found for spectral peak above, where coarticulation with /u/ is straightforwardly quantifiable. Figure 4.4 shows the average CoG measurements in the voicing environment.

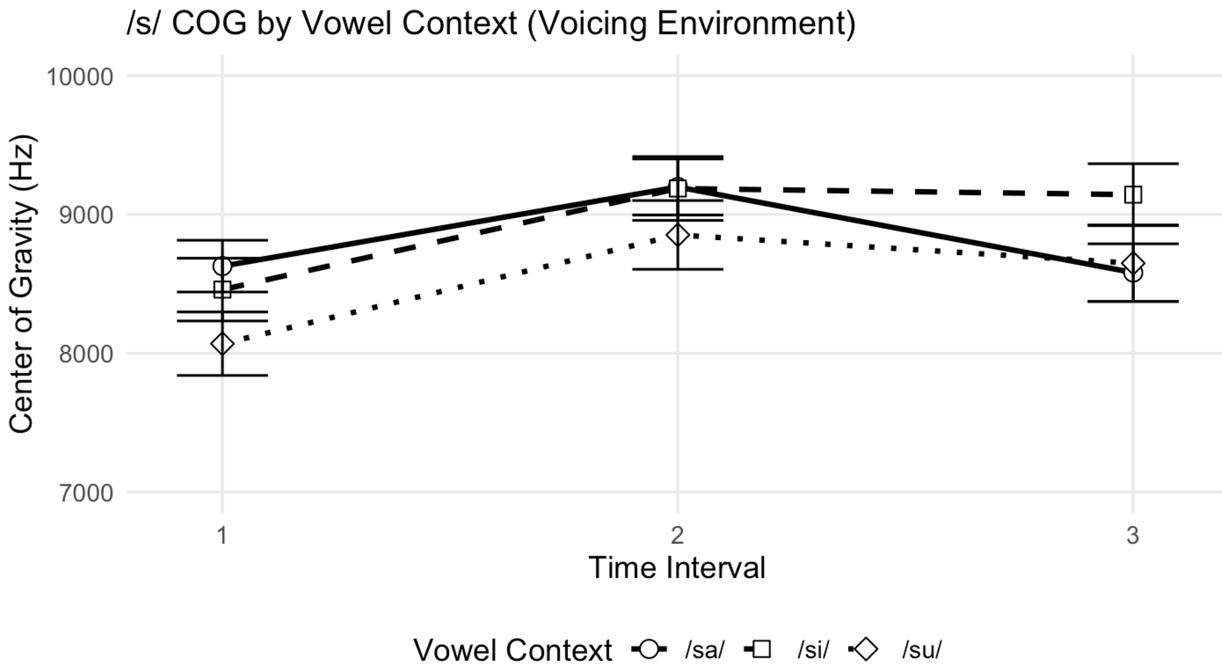


Figure 4.4 Center of gravity measurements across /s/ by vowel context in the voicing environment.

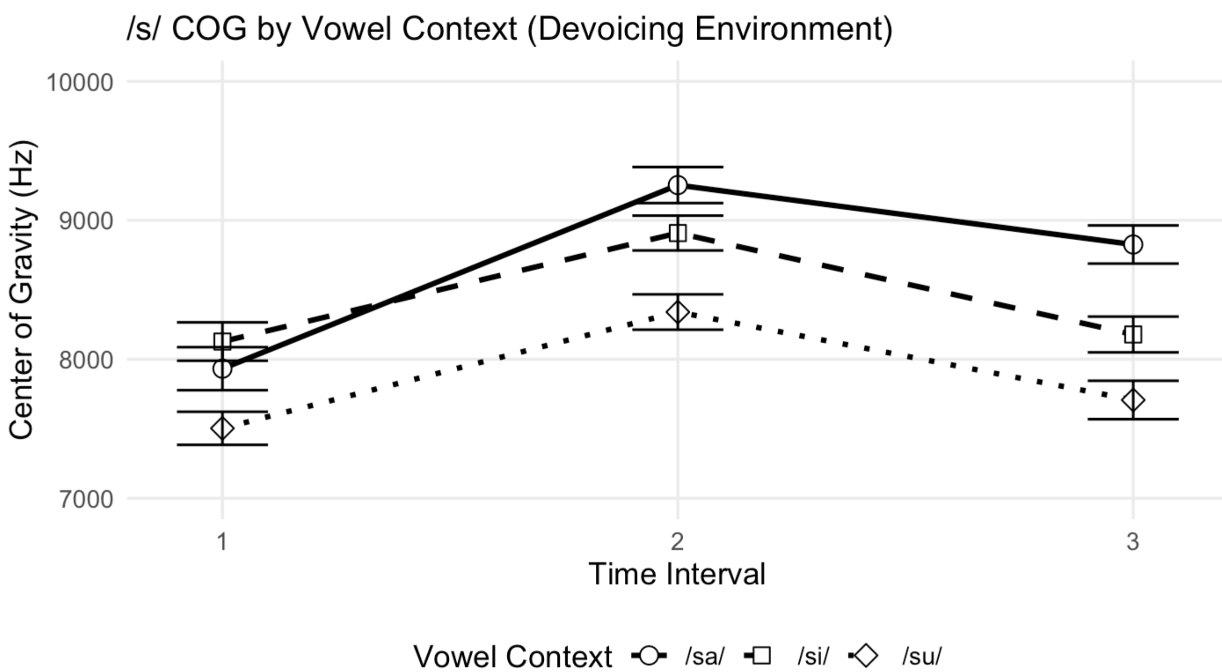


Figure 4.5 Center of gravity measurements across /s/ by vowel context in the devoicing environment.

In the devoicing environment, there is a significant difference in CoG comparing underlying /su/ with both /sa/ ($\beta = -907.3$, $p = 0.001$) and /si/ ($\beta = -546.8$, $p = 0.013$), with /su/ CoG being lower. This is depicted in Figure 4.5. CoG is also significantly lower before /i/ than /a/ ($\beta = -360.5$, $p = 0.027$). This may be explained by the fact that CoG is lower for less-anterior places of articulation (Gordon et al., 2002). For /si/ coarticulation, the tongue blade gesture of /s/ competes with the tongue body gesture of high vowels like /i/, which plausibly results in a retracted tongue tip articulation for underlying /si/ and thus a lowered CoG; this effect has been shown for palatalized and alveopalatal sibilants in other languages (e.g., Kochetov, 2017; Renwick & Cassidy, 2015; Zygis & Hamann, 2003). It is not immediately clear why this effect is not seen in the voicing environment. For /su/, however, the result is consistent with the lowering of CoG seen when /u/ is voiced; as with the FM data above, the CoG data indicates that some vowel-like gesture is being articulated even when the vowel is devoiced.

4.3.1.2.2. Spectral standard deviation

Spectral standard deviation (SD) also shows a significant main effect of vowel context in the voicing environment: /s/ SD is significantly higher before /u/ compared to both /a/ ($\beta = 571.3$, $p < 0.001$) and /i/ ($\beta = 496.5$, $p < 0.001$). There is no significant difference between /sa/ and /si/ ($p = 0.4$). Spectral SD values for each vowel context across all 3 intervals are depicted in Figure 4.6.

Likewise in the devoicing environment (Figure 4.7**)**, spectral SD is higher before /u/ than /i/ ($\beta = 344.6$, $p < 0.001$) and /a/ ($\beta = 309.0$, $p < 0.001$). There was no significant difference between /si/ and /sa/ ($p = 0.48$). The acoustic properties of the devoicing environment, then, mirror the voicing environment with respect to SD, providing further support that a vowel gesture is being coarticulated for devoiced /i/ and /u/.

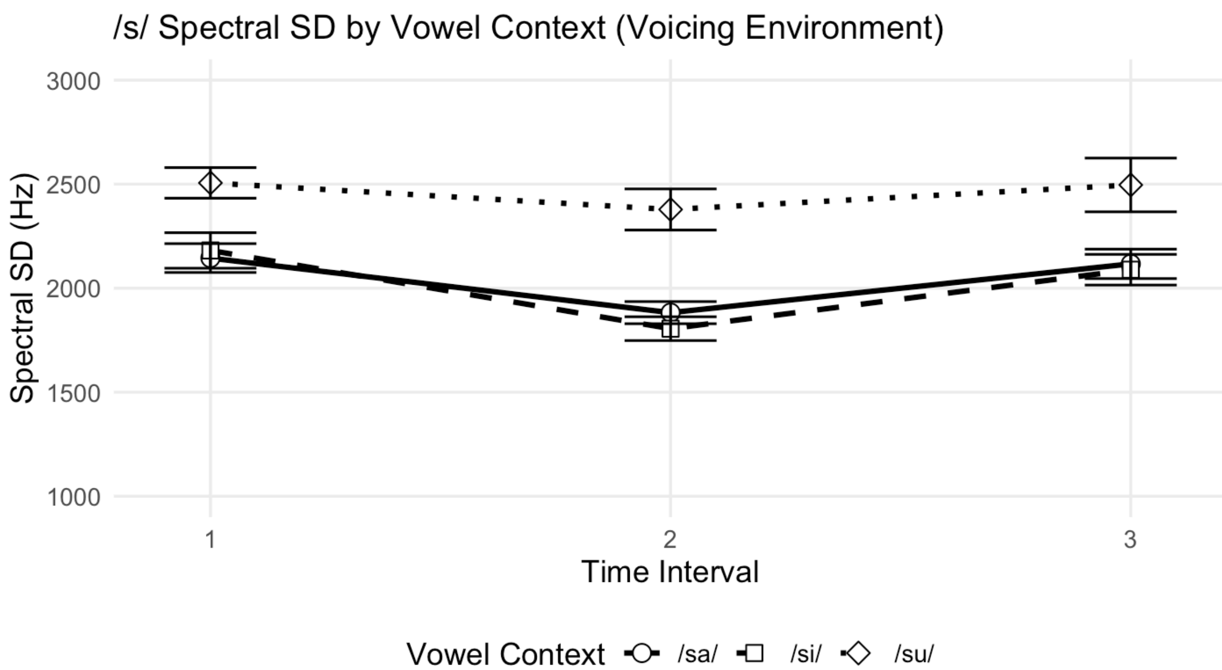


Figure 4.6 Spectral standard deviation measurements across /s/ by vowel context in the voicing environment.

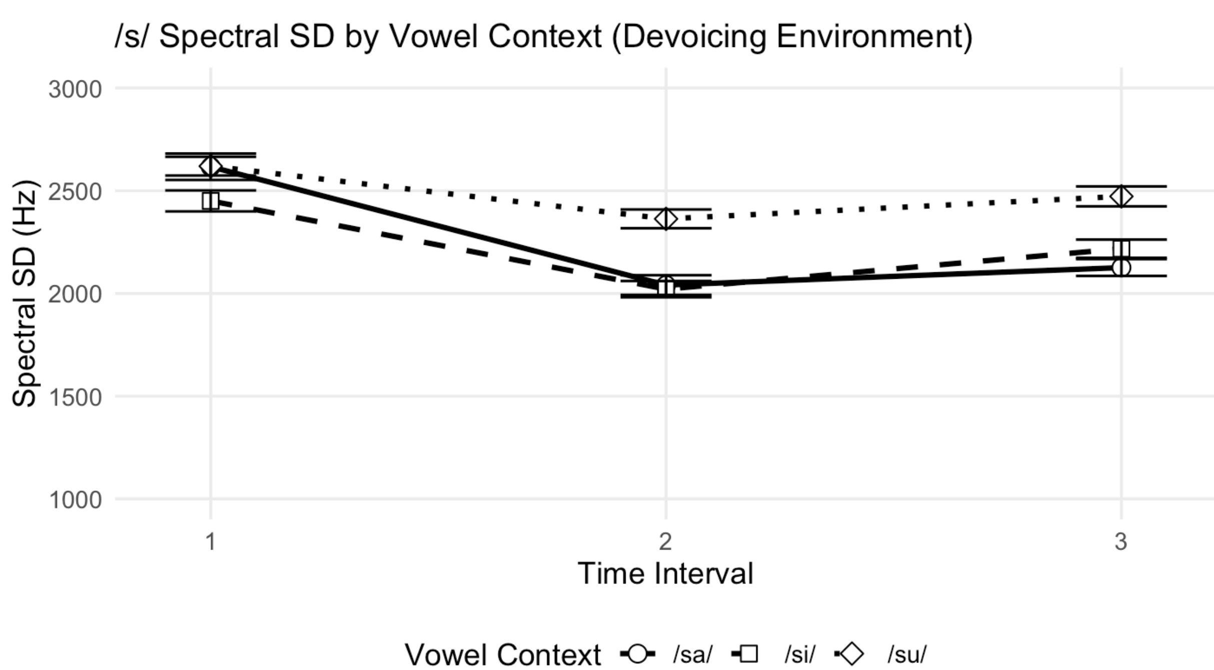


Figure 4.7 Spectral standard deviation measurements across /s/ by vowel context in the devoicing environment.

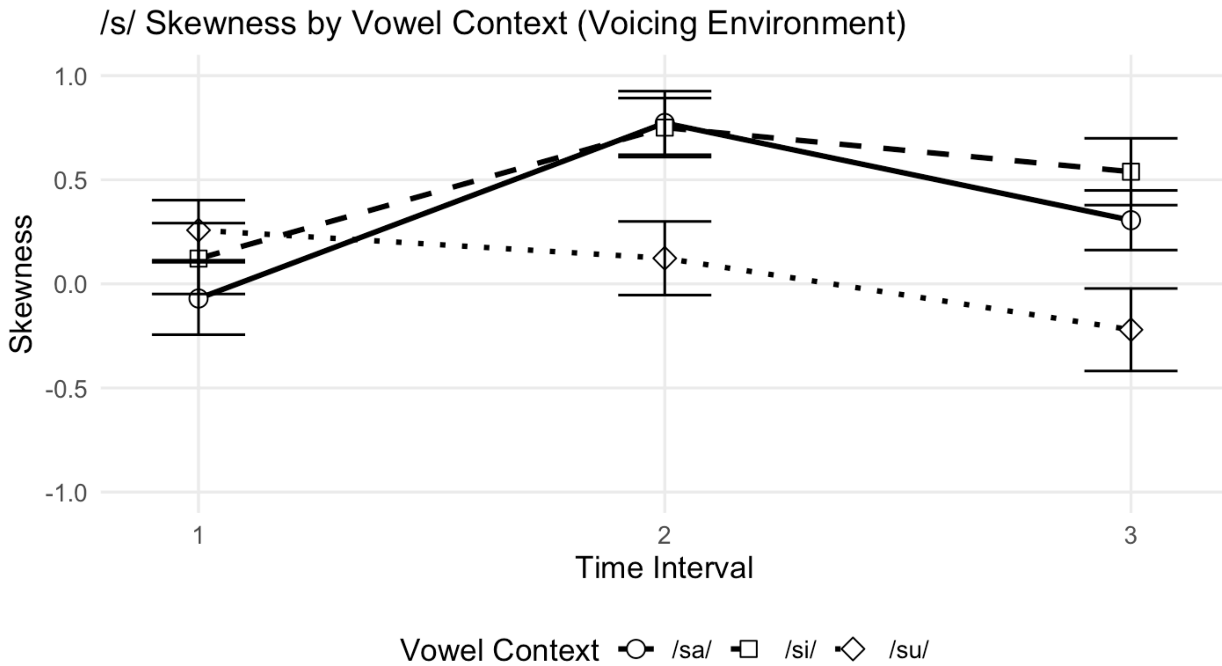


Figure 4.8 Skewness measurements across /s/ by vowel context in the voicing environment.

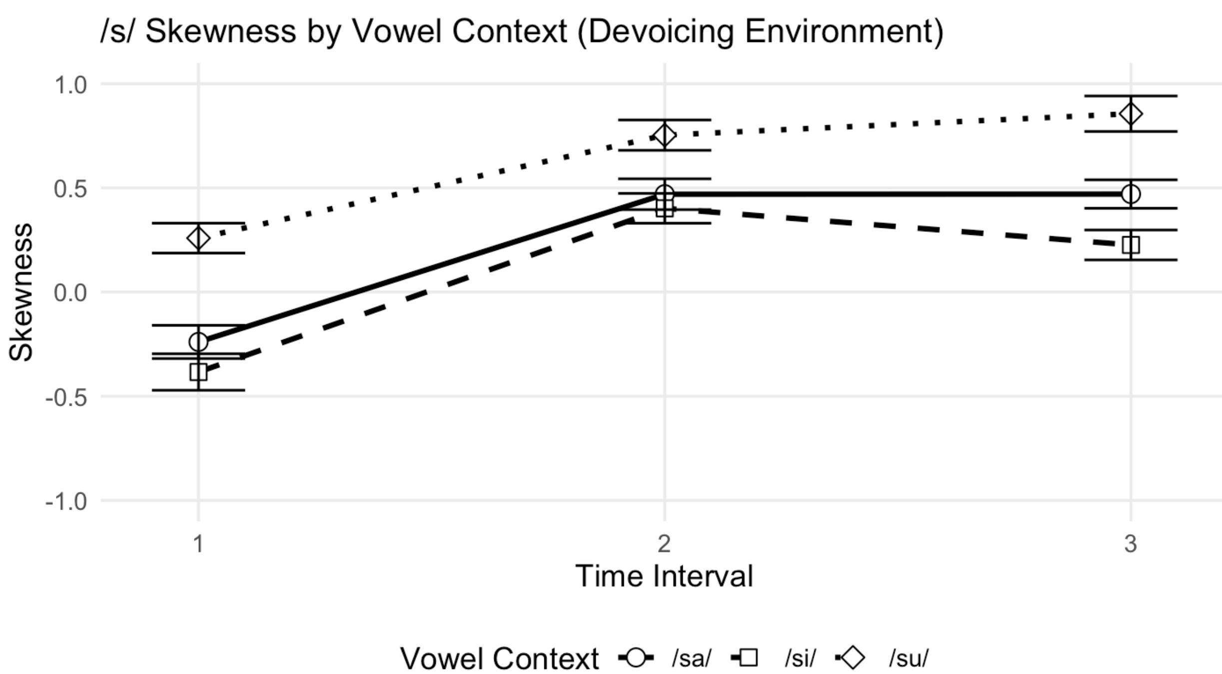


Figure 4.9 Skewness measurements across /s/ by vowel context in the devoicing environment.

4.3.1.2.3. Skewness

In the voicing environment, skewness is significantly lower before /u/ compared to /a/ ($\beta = -0.628$, $p < 0.001$) and /i/ ($\beta = -0.621$, $p < 0.001$). There is no significant difference between /si/ and /sa/ ($p = 0.949$). This is shown in Figure 4.8.

In the devoicing environment, however, the effect was the opposite (Figure 4.9): /s/ was significantly *higher* before underlying /u/ compared to both /i/ ($\beta = 0.295$, $p < 0.001$) and /a/ ($\beta = 0.383$, $p < 0.001$). While this result is unlike that found for the voicing environment, recall that usefulness of skewness as a measure for fricatives has been called into question, even compared to the other spectral moments (Shadle & Mair, 1996), so the discrepancy between this result and those found for the other moments may indicate that skewness is less helpful for capturing consonant-vowel coarticulation.

4.3.1.2.4. Kurtosis

Kurtosis measurements in the voicing environment did show a significant effect of vowel context. /s/ kurtosis was significantly lower before /u/ compared to /a/ ($\beta = -2.216$, $p < 0.001$) and /i/ ($\beta = -2.039$, $p = 0.001$). There was no significant difference between /sa/ and /si/ ($p = 0.773$). This is shown in Figure 4.10.

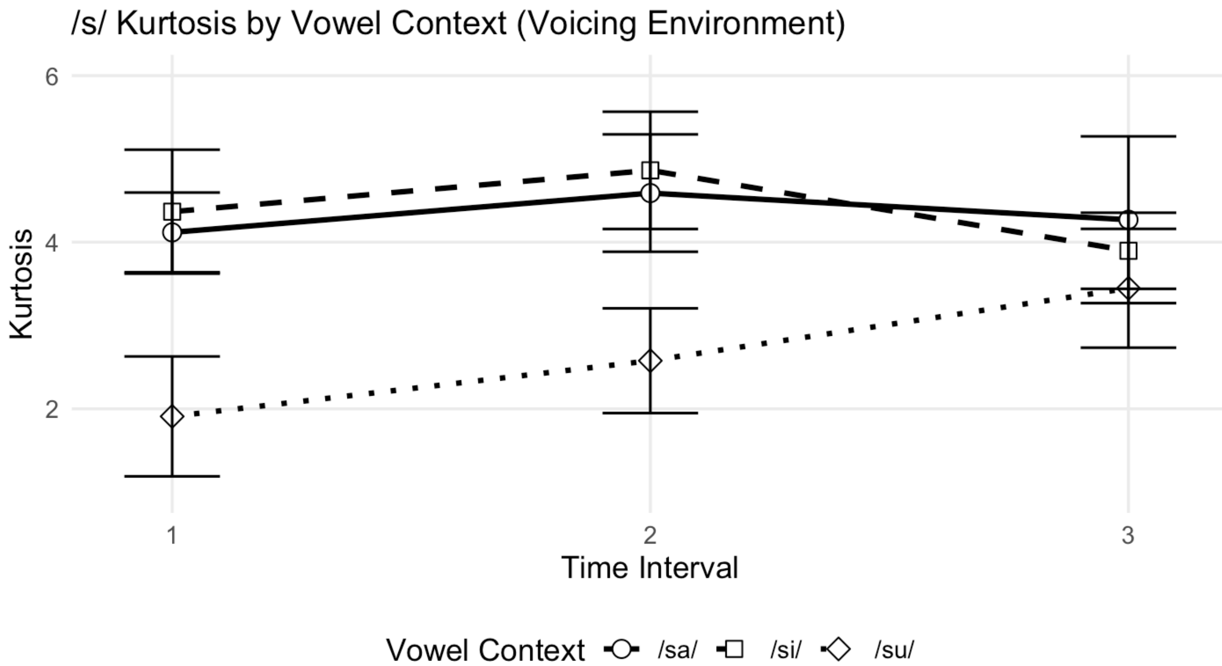


Figure 4.10 Kurtosis measurements across /s/ by vowel context in the voicing environment.

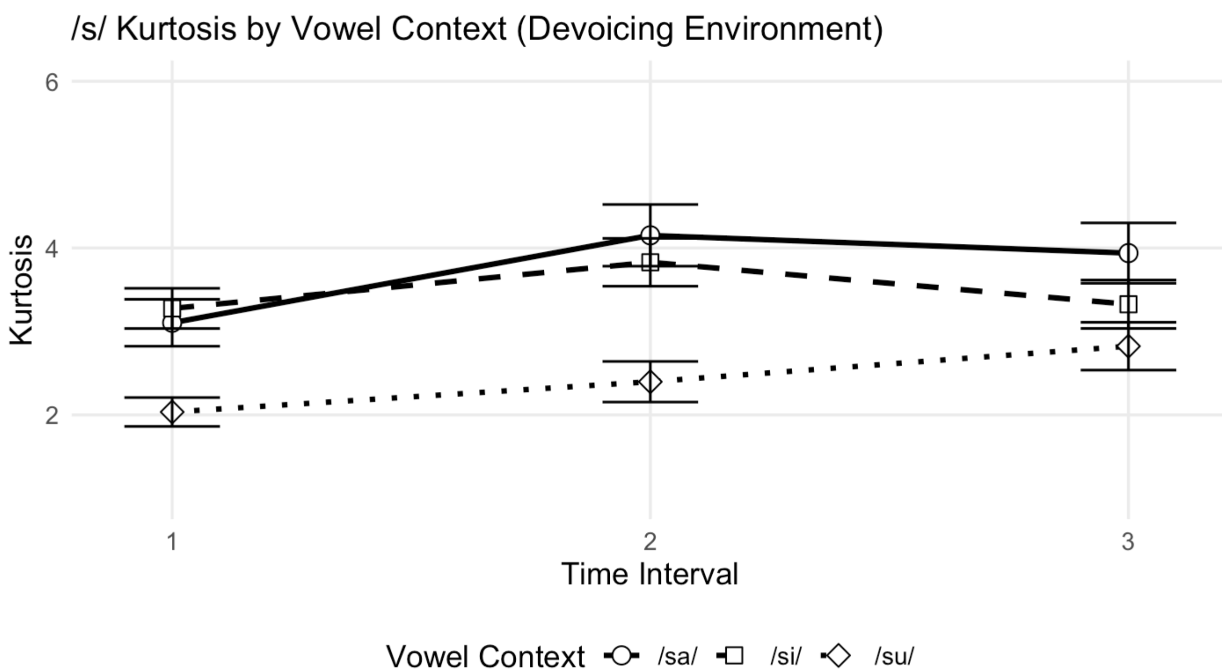


Figure 4.11 Kurtosis measurements across /s/ by vowel context in the devoicing environment.

Likewise, in the devoicing environment (Figure 4.11), there is a significant effect of vowel context analogous to the voicing environment: /s/ is significantly lower before /u/ than /a/

($\beta = -1.774$, $p < 0.001$) and /i/ ($\beta = -1.515$, $p < 0.001$) There is no significant difference between /sa/ and /si/ ($p = 0.48$).

4.3.1.3. *Interim summary: vowels following /s/*

In this section, we saw the spectral consequences of typical consonant-vowel coarticulation following /s/ when the vowel is voiced, as well as when the underlying vowel undergoes non-final "devoicing". Overall, the effect of /u/ was more prominent than /i/; this is unsurprising, as lip rounding is widely reported to effect the spectral properties of sibilants. Spectral peak, center of gravity, spectral standard deviation and kurtosis were all shown to differ between /su/ and /si/ in the voicing environment; analogous effects were found in the devoicing environment, despite there being no voiced vowel segment in this environment. This result lends support to the notion that when a Malagasy vowel is "devoiced", the underlying vowel is present as a coarticulatory gesture on the preceding consonant. However, no measure was found to reliably distinguish underlying /si/ from /sa/.

4.3.2. /t/

Spectral peak and the first four spectral moments were calculated for the stop release of each /t/ (i.e., the burst and aspiration). Here, I look at the spectral properties during the /t/ release before /i/ and /u/; /a/ is excluded because of its short burst duration that makes measurement difficult. The same Praat scripts as above were used to extract spectral peak and the four moments. Based on the duration measurements presented in Section 4.4 below, for /ti/ in the voicing environment, a single time-averaged spectrum with four windows of 15ms was generated for each token; for all other /t/ tokens, three windows of 15ms were used. In order to avoid substantial overlap of the windows, tokens with a duration less than half the total window length (i.e., 30ms for /ti/ in the voicing environment and 22.5ms elsewhere) were excluded. Four participants' data were excluded from the devoicing environment for producing too few /ti/ items (as the words used were relatively low-frequency).

The dataset used here has many fewer data points compared to that used in Section 4.3.1 above, not only because of the participants who were excluded, but also the fewer number of intervals from which the spectral moments were extracted (just one here, compared to three for each /s/). Consequently, some models would not run; in those cases, I present the data qualitatively. Otherwise, a model was created with the spectral moment as the dependent variable, vowel context as the fixed effect, and participant as a random effect; an example involving CoG in the voicing environment is in (4.2).

(4.2) cog_model_voicing_t <- lmer(cog ~ vowel + (1 | participant), data = voicing_environment)

4.3.2.1. Spectral peak

In the voicing environment, spectral peak (F_M) of the stop aspiration is significantly lower before /u/ than /i/ ($\beta = -2811$; $p < 0.001$). In the devoicing environment, F_M is likewise lower before underlying /u/; however the difference is not significant ($p = 0.32$). Figure 4.12 presents the data in a beanplot (Kampstra, 2008), showing the overall distributions of the data and the means for each condition.

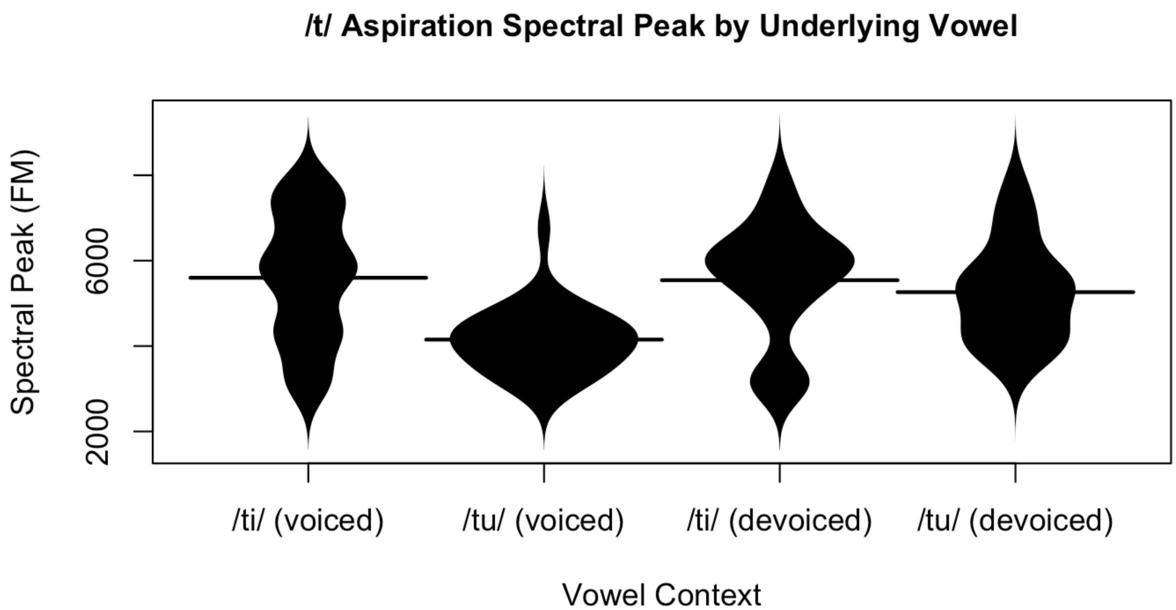


Figure 4.12 Beanplot of /t/ spectral peak measurements by vowel context and voicing environment.

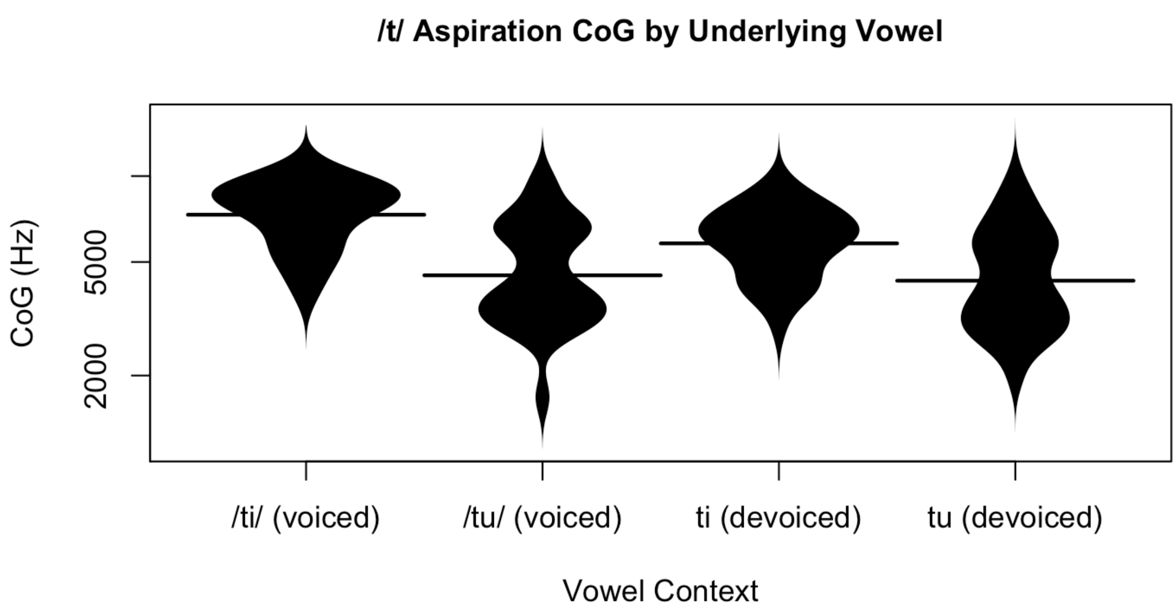


Figure 4.13 Beanplot of /t/ CoG measurements by vowel context and voicing environment.

4.3.2.2. Spectral moments

4.3.2.2.1. Center of gravity

Center of gravity is significantly lower before underlying /u/ in both the voicing ($\beta = 2811$, $p < 0.001$) and devoicing ($\beta = 1365.1$, $p < 0.001$) environments. This mirrors the data for CoG of /s/ and suggests that there is an underlying vowel-like gesture that lowers CoG, such as lip rounding, that is present in the devoicing environment. Figure 4.13 presents beanplots for all of the /t/ data, including /ti/ and /tu/ in the voiced and voiceless environments.

4.3.2.2.2. Spectral standard deviation

The statistical model for spectral standard deviation was not computable in neither the voicing nor devoicing environments; however, qualitatively, we can see in Figure 4.14 that the overall distribution is lower before /i/ compared to /u/. This is in contrast to the devoicing environment, where SD is lower before underlying /u/. This would represent a difference between voiced and devoiced vowel environments, but this has not been confirmed statistically.

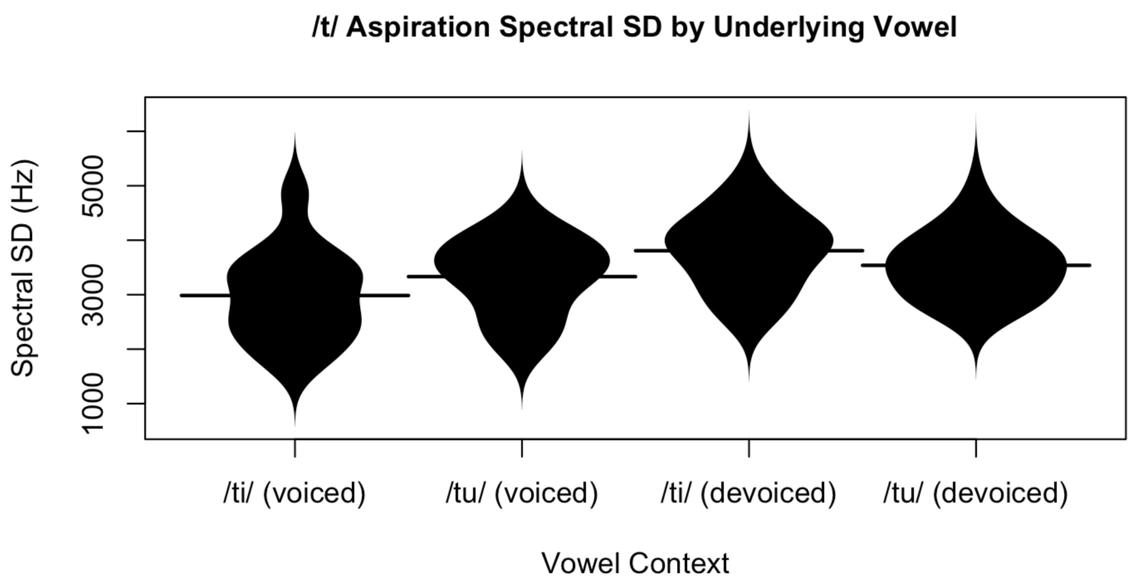


Figure 4.14 Beanplot of /t/ spectral standard deviation measurements by vowel context and voicing environment.

4.3.2.2.3. Skewness

In the voicing environment, skewness of /t/ aspiration is significantly higher before /u/ ($\beta = 0.841$, $p < 0.001$), and this effect is likewise found in the devoicing environment ($\beta = 0.772$, $p < 0.001$). This is shown in Figure 4.15.

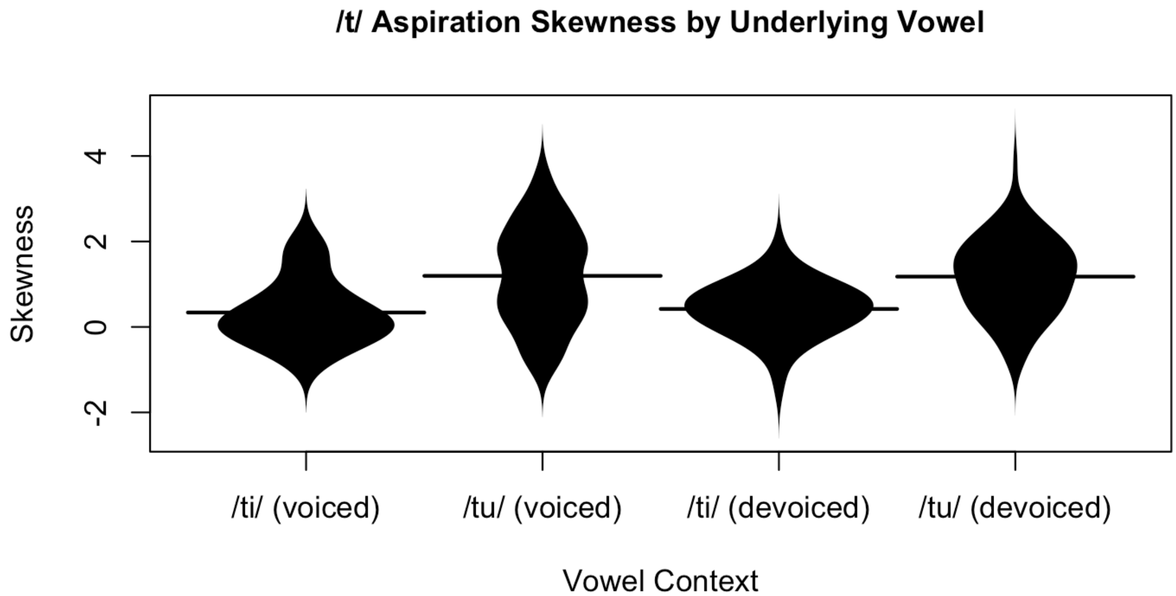


Figure 4.15 Beanplot of /t/ skewness measurements by vowel context and voicing environment.

4.3.2.2.4. Kurtosis

In the voicing environment, the model would not compute, but looking at Figure 4.16 below, we see that the mean kurtosis value is higher before /u/ than /i/. In the devoicing environment, kurtosis is significantly higher before underlying /u/ ($\beta = 1.986$, $p < 0.001$). While we can’t directly compare the devoicing with the voicing environments because of it lacks a statistical model, the result in the devoicing environment is still notable because it shows a clear difference when /t/ precedes underlying /i/ vs. /u/, indicating that there is some representational difference present in the surface forms even when there is no voiced vowel.

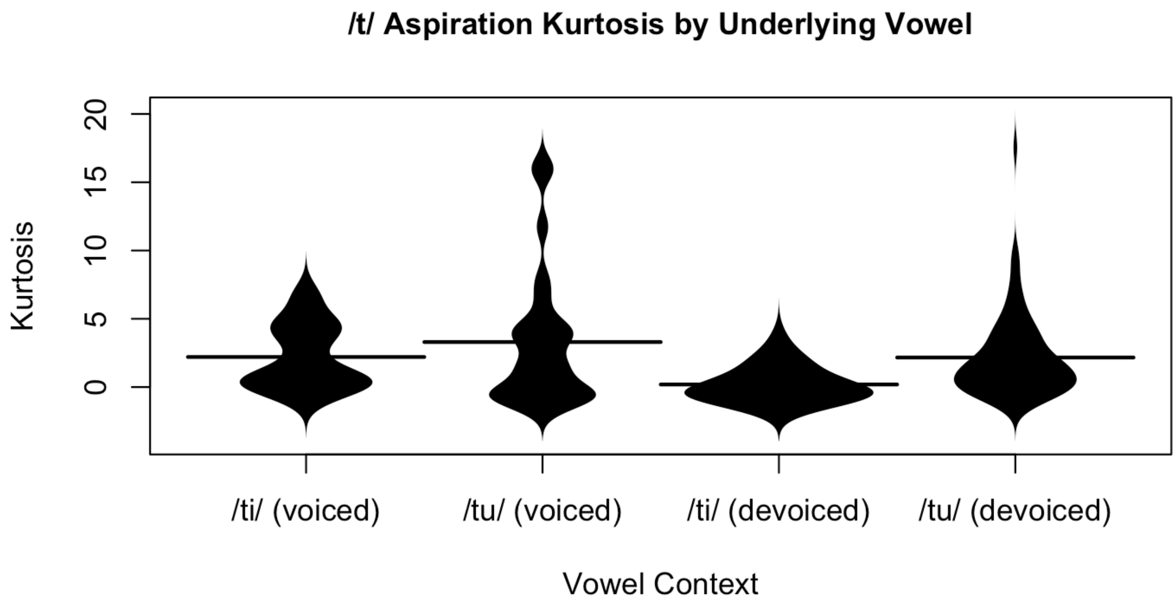


Figure 4.16 Beanplot of /t/ kurtosis measurements by vowel context and voicing environment.

4.3.2.3. Interim summary: vowels following /t/

In this section, I presented the acoustic properties of /t/ aspiration preceding different vowels both when it precedes a voiced vowel and when the vowel is devoiced. Overall, results pointed to similarities between normal consonant-vowel coarticulation and vowel devoicing. That is, certain spectral properties of the underlying vowel were largely present even when the vowel was devoiced; this was shown statistically for measures of center of gravity and skewness. However, the analysis of /t/ was hampered by a limited dataset, especially in the voicing environment, which led to statistically inclusive results for spectral standard deviation and kurtosis. The analysis of spectral peak was found to be different in the voicing and devoicing environments: whereas F_M was significantly higher before voiced /i/, this effect was not found in the devoicing environment. However, while this result is inconsistent with a hypothesis where the vowel gesture is coarticulated with the consonant even when the vowel is devoiced, this discrepancy possibly arose for methodological reasons. F_M was calculated according to Shadle's (2003) definition, which was intended for measurement of fricatives. Importantly, she defined the spectral peak as the frequency with the highest amplitude within a particular range, beginning at 3000 Hz. However, for both /i/ and /u/, any second formant resonances that may be visible in a stop's aspiration noise would be excluded using the same frequency range as fricatives. So, it is possible that a clearer result would arise with a more precise understanding of the spectral peaks of stops.

4.3.3. /n/

Because vowel "devoicing" is possible even in the environment of sonorants in Malagasy, here we look at devoicing after /n/ for evidence of an underlying vowel. As shown in Chapter 2, "vowel devoicing" after /n/ involves no actual vowel segment, and instead, at first glance, looks as if there is no vowel gesture present at all. To investigate whether or not the devoiced vowel is realized in some form in the output, I look at second formant (F2) frequencies. It is well established that back vowels are associated with a lower F2 than front vowels, and that lip rounding lowers F2 (Stevens, 2000). Because antiresonance can make the F2 of nasals very weak and difficult to track, the F2 of the preceding vowel /a/ was measured instead. This type of anticipatory vowel-to-vowel coarticulation has been documented for multiple languages (Manuel & Krakow, 1984; Öhman, 1966, among others) and so, if a vowel gesture is articulated in the devoicing environment, we may expect its effect to be detectable on the preceding vowel.

The data comes from Task 2, and thus contains words with target /i/ and /u/ only in the devoicing environment. Two participants were excluded: one who neglected to produce most words with /n/, and another who did not produce the expected devoicing pattern. Formants were measured using a Praat script that divides each /a/ into eight intervals and extracts formant values using Praat's built in Linear Predictive Coding method (DiCanio, 2021b).

A linear mixed effects model was run with /a/ F2 as the dependent variable and following underlying syllable (/ni/ or /nu/) as the fixed effect. Participant was integrated as a random effect, as well as the segment preceding /a/, as it varied between items and may affect formant values as the consonant transitions into /a/. The model included only data at the final interval of the vowel. The model reveals a main effect of underlying following vowel, with /a/ having a significantly lower F2 when it precedes /nu/ ($\beta = -169.5$, $p < 0.001$). This finding is in line with descriptions of vowel-to-vowel coarticulation, as /u/ has a lower F2 compared to /i/. We can conclude, then, that underlying devoiced vowels following /n/ are articulatorily present, affirming the results found for devoiced vowels following /s/ and /t/. Figure 4.17 depicts the F2 values across all eight timepoints for /a/ preceding both /ni/ and /nu/; the first two timepoints are greyed out because of expected coarticulation with the preceding consonant during those intervals.

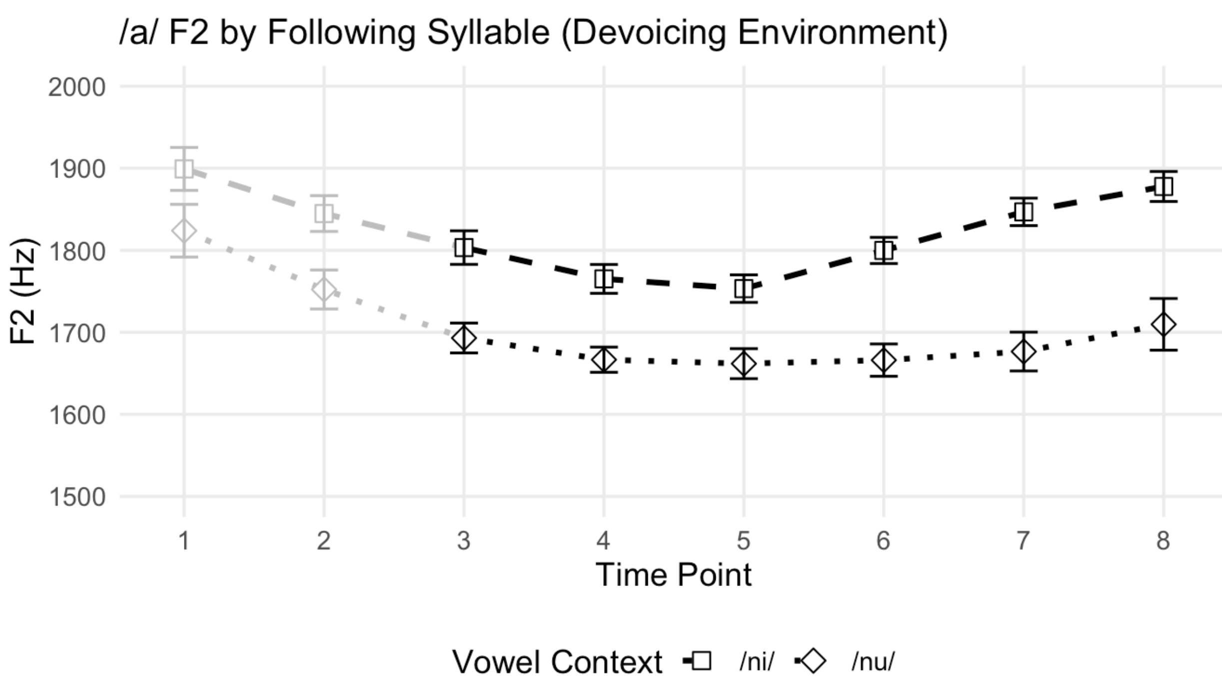


Figure 4.17 F2 measurements of /a/ before underlying /ni/ and /nu/ in the devoicing environment.

4.4. Duration

Durational effects of the vowel environment have been shown for consonants across languages, making duration a natural point of comparison for coarticulation and devoicing. Vowel height has been shown to affect the duration of preceding fricatives in languages like Canadian French (O'Shaughnessy, 1981), English (Berry & Moyle, 2011; Schwartz, 1969), and Greek (Nirgianaki, 2014), where /s/ is longer before high vowels compared to low vowels, though it's not well established whether this finding is cross-linguistically robust. Similarly, voice onset time (VOT) has been shown to be longer before high vowels across languages (e.g., Esposito, 2002; Klatt, 1975; Nearey & Rochet, 1994). The prediction here is this: if an underlying vowel is still present in the output in the devoicing environment, we expect the durational cues related to that vowel environment to still be present.

4.4.1. /s/ duration

Here, I present data that shows that the durational effect of vowel context is present in the devoicing environment: /s/ is longer before underlying /i/ and /u/ compared to /a/, while there is no significant difference between /s/ before /i/ and /u/. Each /s/ in the context /s_k/ was segmented in Praat (Boersma & Weenink, 2017) from the beginning of frication, evaluated as the beginning of aperiodic noise in the waveform. In contexts where the vowel was voiced, the endpoint of /s/ was marked at the onset of voicing of the vowel; in cases where the vowel was devoiced, it was marked at the beginning of the /k/ closure. This differs somewhat from the segmentation of /s/ used in Section 4.3.1 (Figure 4.1); here, the lower-amplitude noise following frication was included in order to measure the complete duration of /s/, not just the fricated portion. Figures 4.18 and 4.19 show examples of /s/ segmentation for voicing and devoicing environments, respectively.

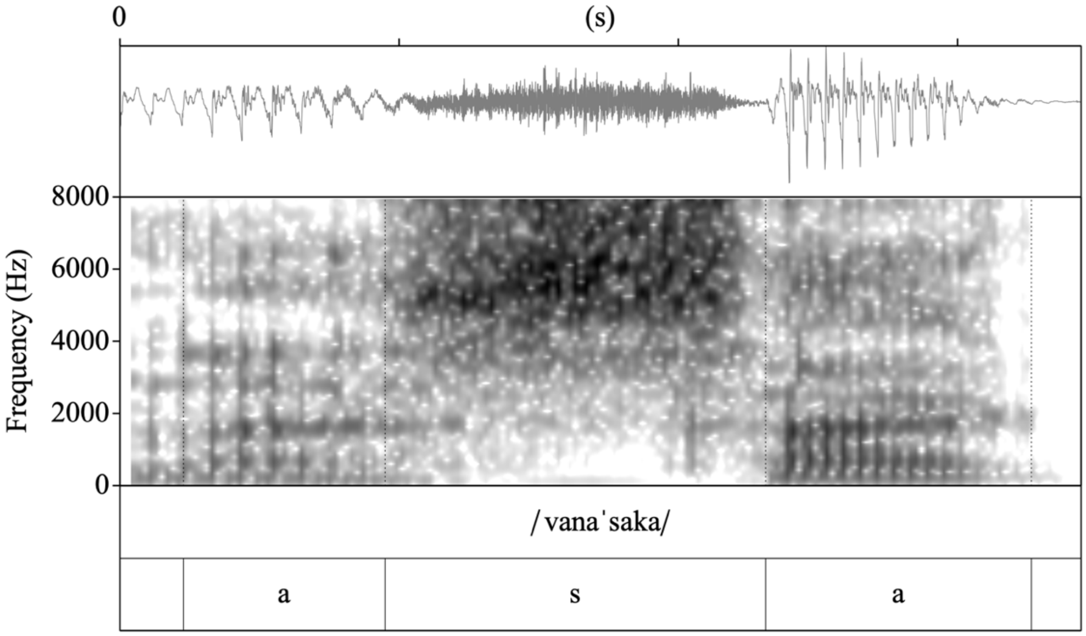


Figure 4.18 Example of /s/ segmentation in the context of a voiced vowel, where /s/ is segmented beginning at frication and ending at the onset of voiced /a/.

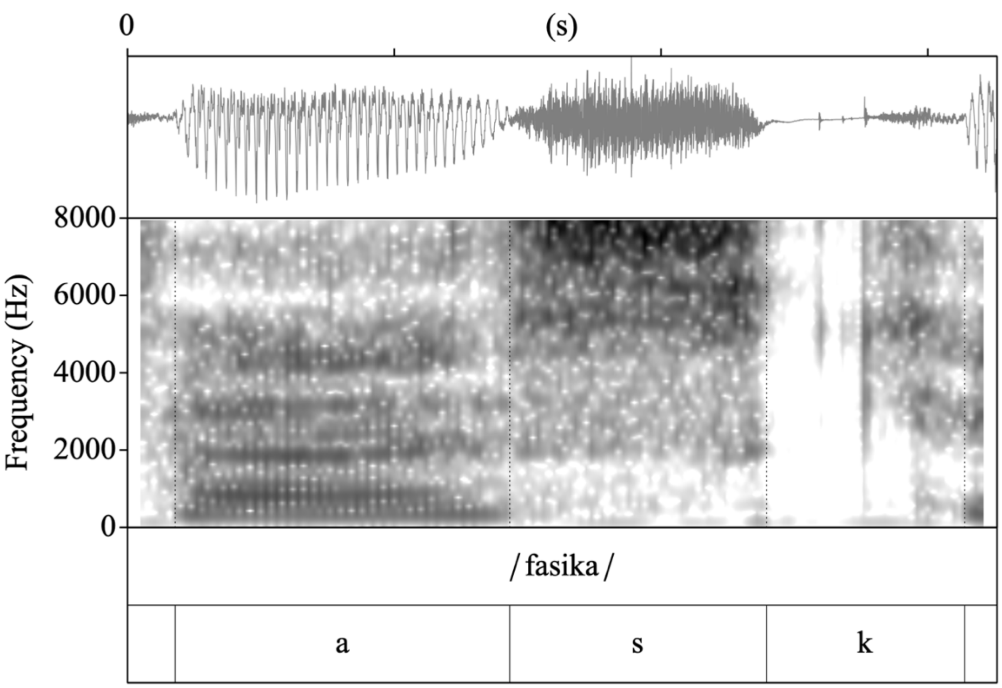


Figure 4.19 Example of /s/ segmentation in the context of a devoiced vowel, where /s/ is segmented beginning at frication and ending at the silent period associated with the closure of /k/.

Duration measurements were automatically extracted using a moderated Praat script (DiCanio, 2021b). For /s/ tokens in the devoicing environment (i.e., collected in Task 1), duration measurements were normalized for statistical testing by dividing the real duration by the duration of part of the frame, *ilay zazavavy*. Because each sentence in this task contained the same frame, its duration could be used as a proxy for speech rate. For tokens in the voicing environment (Task 2), I chose not to normalize durations because of the nature of the task; because speakers were producing nonce words, they tended to slow down when pronouncing the nonce word compared to the frame, which otherwise consisted of high-frequency real Malagasy words.

The relationship between /s/ duration and vowel environment was tested using a linear mixed effects model with duration (normalized in the devoicing environment; un-normalized in the voicing environment) as the dependent variable, underlying vowel environment as a fixed effect and participant as a random effect. In the voicing environment, /s/ was significantly longer in the syllable /si/ ($\beta = 0.024$, $p = 0.001$) and /su/ ($\beta = 0.027$, $p < 0.001$) compared to /sa/, while there was no significant difference between /si/ and /su/ ($\beta = 0.003$, $p > 0.05$). This indicates that vowel height conditions the duration of onset fricatives, where /s/ is longer before voiced high vowels, placing Malagasy among languages like English and Canadian French that show this effect. The mean (un-normalized) duration of /s/ before each vowel, across participants, is summarized in Table 4.1 and visualized as beanplots in Figures 4.20 and 4.21.

	/sa/	**/si/**	**/su/**
Voiced	*144 ms*	*165 ms*	*168 ms*
Devoiced	*87 ms*	*93 ms*	*91 ms*

Table 4.1 Duration (un-normalized) of /s/ in different underlying vowel contexts in both the voicing (stressed) and devoicing (unstressed) environments. Note that /a/ is never devoiced.

Even though /i/ and /u/ are not overt, voiced vowels in the devoicing environment, /s/ still shows a durational effect of the underlying vowel. That is, /s/ is longer before underlying high vowels, regardless of the vowel's realization. Compared to the syllables with /a/ (which, recall, are always voiced, even in the environment where /i/ and /u/ are not), /s/ was significantly longer before underlying /i/ ($\beta = 0.006$, $p < 0.001$) and /u/ ($\beta = 0.005$, $p = 0.002$), analogous to the voicing environment, where /s/ was longer before high vowels. Comparing underlying /si/ and /su/, there was no significant difference in the duration of /s/ ($\beta = -0.002$, $p = 0.158$).

4.4.2. /t/ aspiration duration

Voice onset time (VOT), the duration between a stop's release and the beginning of voicing, has been shown to be sensitive to vowel height, including for Malagasy (Howe, 2019). For voiceless stops, VOT tends to be longer before high vowels, so if the high vowel context exhibits a similar degree of VOT in both the voicing and devoicing environments, it is consistent with an account of vowel devoicing where an underlying vowel is still present in the output in some way. Tokens in the environment t_k were analyzed in order to test this. Because the devoicing environment definitionally contains no voiced segment following /t/, the measurement taken is not VOT in the strictest sense, but rather the duration of aspiration. As with the duration of frication measured in Section 4.4.1 above, aspiration was measured from the beginning of aperiodic noise associated with the aspiration until either (a) the onset of voicing, when the vowel was voiced or (b) the silence associated with the stop closure of /k/.

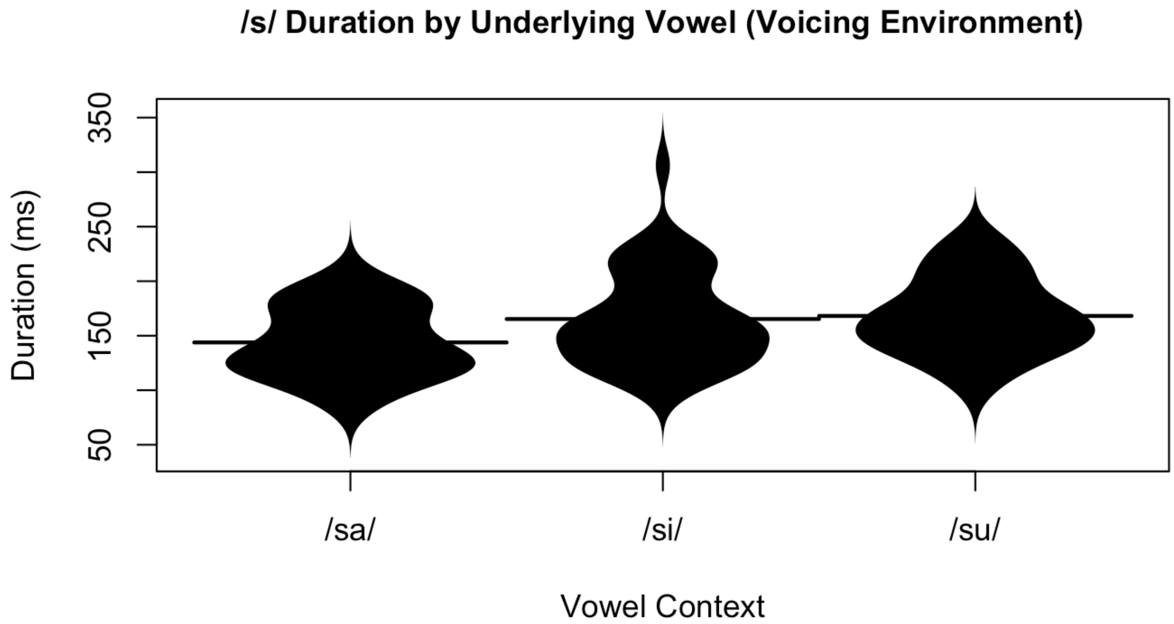


Figure 4.20 /s/ duration in different vowel contexts in the voicing environment .

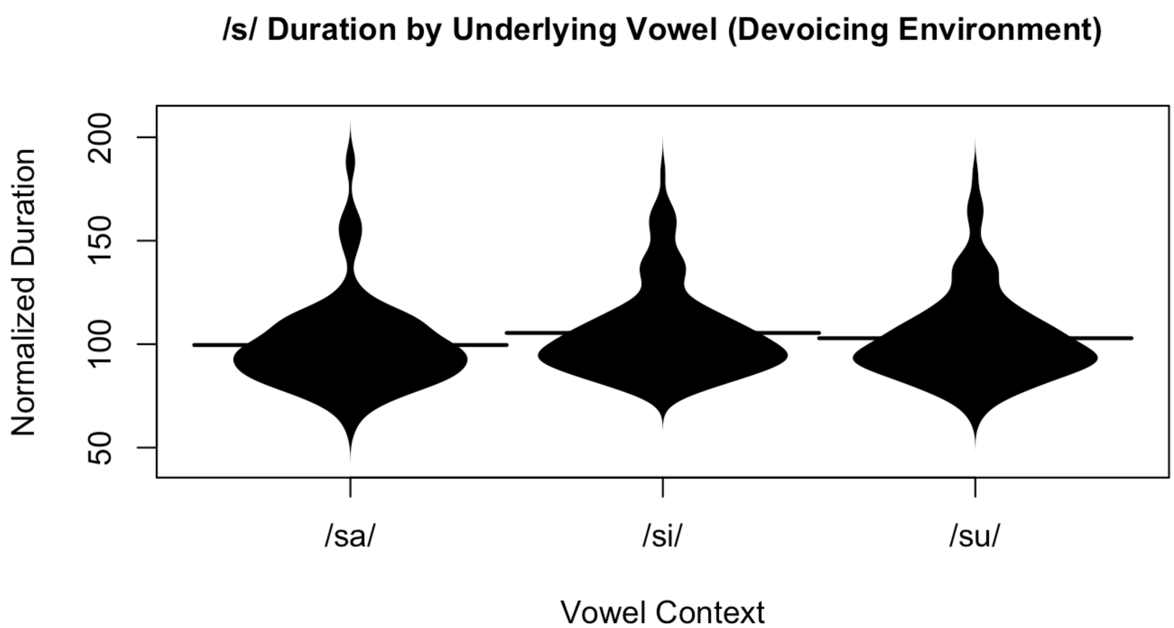


Figure 4.21 /s/ duration in different vowel contexts in the devoicing environment.

	/ta/	/ti/	/tu/
Voiced	*11 ms*	*56 ms*	*33 ms*
Devoiced	*15 ms*	*35 ms*	*37 ms*

Table 4.2 Duration (un-normalized) of /t/ aspiration in different vowel contexts in both the voicing (stressed) and devoicing (unstressed) environments.

Table 4.2 shows real (un-normalized) duration values for /t/ aspiration. Linear mixed effects models were run for the voicing and devoicing environments with duration as the dependent variable, underlying vowel context as the independent variable, and participant as a random effect. As with the /s/ data, duration was normalized for the devoicing environment only. In the voicing environment, there was a three-way difference between the vowel contexts: /t/ aspiration is significantly longer before /i/ than /u/ ($\beta = 0.023$, $p < 0.001$) and before /u/ than /a/ ($\beta = 0.022$, $p < 0.001$). This is depicted in the beanplot in Figure 4.22. For the devoicing model, four participants were excluded for not producing any token with /ti/. In the devoicing environment, /t/ aspiration was significantly longer before both underlying /i/ ($\beta = 0.019$, $p < 0.001$) and /u/ ($\beta = 0.022$, $p < 0.001$) compared to /a/, while there was no significant difference between /ti/ and /tu/ ($p = 0.066$), depicted in Figure 4.23.

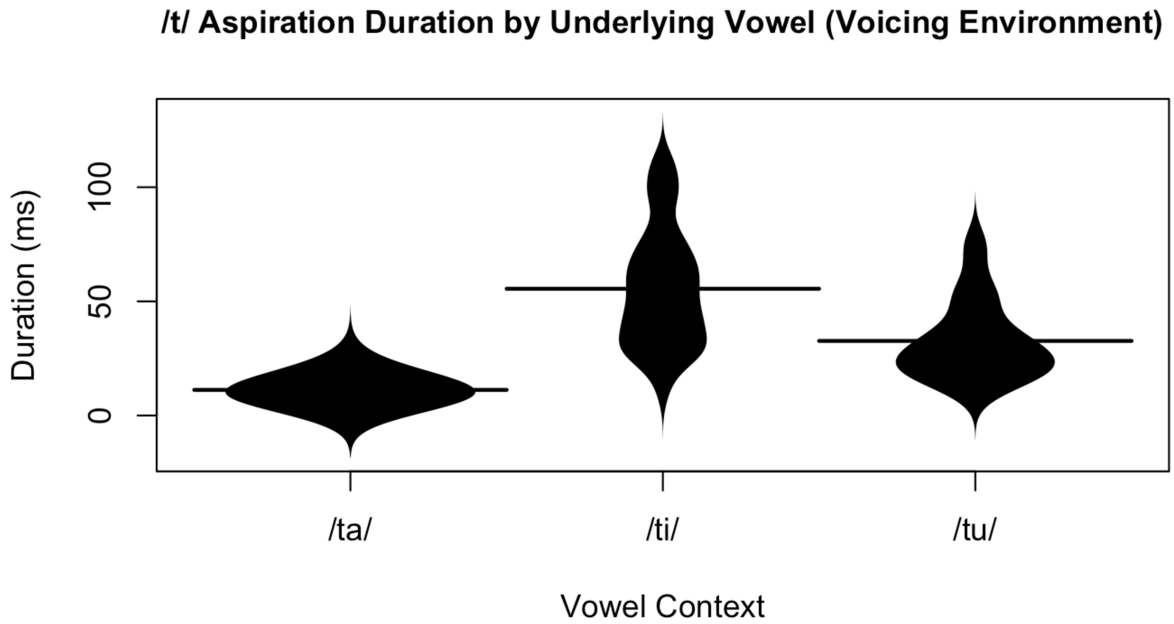


Figure 4.22 Duration of /t/ aspiration in different vowel contexts in the voicing environment.

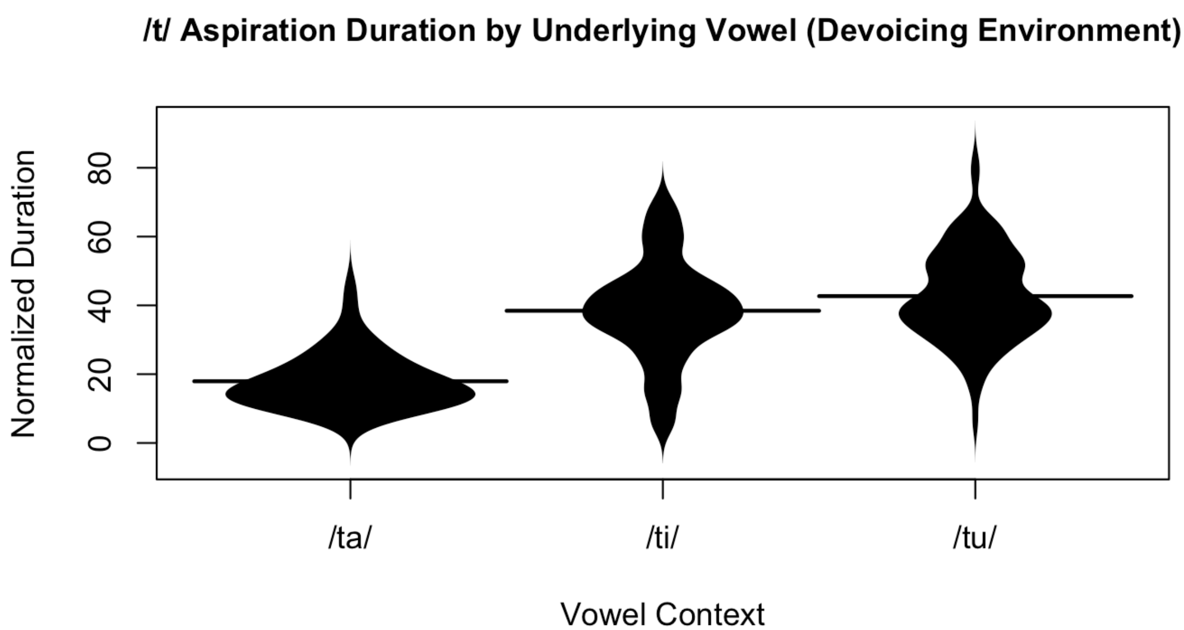


Figure 4.23 Normalized duration of /t/ aspiration in different vowel contexts in the devoicing environment.

The duration results are overall consistent with the presence of a vowel gesture in the devoicing environment: even though the high vowels /i/ and /u/ are not overtly represented as vowel segments, the duration of the consonants that precede them are still longer than consonants that precede voiced /a/, an effect that is also observed in the voicing environment. One inconsistency in the result is with respect to /t/ duration before /i/ vs. /u/: whereas in the voicing environment, /t/ was shown to be longer before /i/, this was not true in the devoicing environment, where /ti/ and /tu/ had equal durations. It is possible that there is no effect of the underlying vowel here; if the high vowel in a word like /fatika/ were deleted, this would place /t/ directly next to /k/. Perhaps, then, the lengthening observed before underlying high vowels is actually an effect of these consonants being either in coda position or before another consonant. However, this would be inconsistent with previous findings that consonants are actually *shorter* in coda position (e.g., Byrd, 1995).

An alternative explanation is that the aspiration of /t/ is interrupted by the gesture associated with the /k/; when the Malagasy vowel is devoiced, it leaves the onset of /k/ adjacent to the aspiration of /t/. The usual explanation for the aspiration of stops before high vowels is that the close constriction of high vowels increases oral pressure, inhibiting voicing for several milliseconds as the tongue makes the transition from the stop to the vowel. In Malagasy, if the vowel gesture is short enough to be completely overlapped by its consonantal environment, it may be the case that the vowel gesture ends and the /k/ gesture begins before the full extent of the aspiration effect can be realized. Gestural overlap as an explanation for Malagasy vowel devoicing will be explored in more detail in Chapter 5.

4.5. Summary of acoustic results: Devoicing resembles coarticulation

In this chapter, I have presented the results of a detailed acoustic study on Malagasy vowel devoicing. While we know from Chapter 2 that vowel devoicing results in no voiced vowel being pronounced, it was unclear whether some underlying vowel gesture was still produced; i.e., it remained a question whether Malagasy devoiced vowels are categorically deleted or present as a co-articulated gesture on an adjacent consonant. The goal of this chapter, then, was to compare the acoustic properties of consonants preceding a voiced vowel, which is an indication of clear consonant-vowel co-articulation (i.e., when the vowel is clearly voiced in a stressed syllable) to those preceding a "devoiced" vowel, to see if the effects of "devoicing" on the acoustic properties of a consonant are analogous to those of CV coarticulation. I looked at vowel devoicing after three consonants, /s/, /t/, and /n/, in order to investigate the effects of devoicing after different types of consonants.

Below in Table 4.3 is a summary of the acoustic study presented here. For each acoustic measurement, the main effects for both the voicing environment (where we see clear consonant-vowel coarticulation) and the devoicing environment (where /i/ and /u/ are devoiced, but /a/ is not) are presented, and then compared as analogous or not, based on whether the result is the same or similar. For /s/ and /t/ measurements, the analogy is based on the relationship between underlying /i/ and /u/ (e.g., since spectral peak of /s/ was found to be higher before /i/ and /u/ in both the voicing and devoicing environments, that measurement was marked as analogous). For /n/ vowel-to-vowel coarticulation F2 measurements, since there is no coarticulation data to compare to, the analogy is based on expected F2 values for the underlying vowels.

Acoustic measurement	Coarticulation data (Voicing environment)	Vowel devoicing environment	Analogous?
/s/ F_M	/i/ > /u/ /a/ > /u/	/i/ > /u/ /a/ > /u/	**Yes**
/s/ CoG	/i/ > /u/ /a/ > /u/	/a/ > /i/ /i/ > /u/ /a/ > /u/	**Yes**
/s/ SD	/u/ > /i/ /u/ > /a/	/u/ > /i/ /u/ > /a/	**Yes**
/s/ skewness	/a/ > /u/ /i/ > /u/	/u/ > /i/ /u/ > /a/	No
/s/ kurtosis	/i/ > /u/ /a/ > /u/	/i/ > /u/ /a/ > /u/	**Yes**
/t/ spectral peak	/i/ > /u/	No significant difference	No
/t/ CoG	/i/ > /u/	/i/ > /u/	**Yes**
/t/ SD	Not computed	Not computed	Inconclusive
/t/ skewness	/u/ > /i/	/u/ > /i/	**Yes**
/t/ kurtosis	Not computed	/u/ > /i/	Inconclusive
/n/ V-to-V coarticulation F2		/i/ > /u/	**Yes**
/s/ duration	/i/ > /a/ /u/ > /a/	/i/ > /a/ /u/ > /a/	**Yes**
/t/ duration	/i/ > /u/ > /a/	/i/ > /a/ /u/ > /a/	Partly

Table 4.3 Summary of phonetic findings, comparing various acoustic measurements in the voicing and devoicing environments.

For eight of the eleven acoustic parameters for which we have quantitative results, there is a clear analogy: the expected acoustic effect of coarticulation is present in the devoicing environment, even though there is no voiced vowel present. Throughout the chapter, I have provided some explanations for the discrepancies seen in the remaining results; for example, the difference in /t/ spectral peak results may be resolved with modified measurement criteria. The findings presented suggest that so-called "vowel devoicing" in Malagasy is neither true vowel devoicing nor deletion, but rather acoustically most similar to extreme coarticulation. For example, we can transcribe the underlying devoiced vowels as secondary glide co-articulations, i.e., /vasuka/ → [vas^{w}ka] and /fatika/ → [fat^{j}ka]. In the following chapter, I will present an analysis of this phenomenon under Articulatory Phonology (Browman & Goldstein, 1986) whereby the underlying vowel gesture *is* present in the output, as the phonetic data presented here suggest. In that chapter, I argue that the acoustic phenomon of "vowel devoicing" arises due to extreme overlap of the vowel's gestures by those of its adjacent consonant.

Chapter 5
A Gestural Account of Vowel Devoicing

5.1. Introduction

In Chapter 2, I showed that medial "devoiced" vowels in Malagasy do not appear in surface forms as discernable segments: unlike both voiced vowels and devoiced vowels in the phrase-final position, medial "devoicing" shows no trace of a timeslot occupied uniquely by a vowel. However, in Chapter 4, I presented acoustic evidence showing remnants of the underlying vowel, as segments adjacent to the "devoiced" vowel show expected coarticulatory effects such as lowered spectral peak of /s/ before underlying /u/. This suggests that, although the vowel is not present as a distinct, voiced unit, it may be present as an articulatory *gesture*. Chapter 3 presented a statistical model of the environment of vowel devoicing, showing that the process is sensitive to both prosodic and segmental factors. The goal of this chapter is to present an analysis of Malagasy non-IP-final vowel "devoicing" that unifies the phonetic realizations found in Chapter 4 with the distributional facts found in Chapter 3; under a theory of gestural representations like Articulatory Phonology, we can do just that. In the analysis that follows, I present a gestural account of "devoicing", arguing that it arises due to extreme overlap of the vowel's gestures with those of the adjacent consonants.

Cases of non-final vowel "devoicing" are explained as follows: in a Malagasy C_1VC_2 sequence, as in other languages, the C_1 and V gestures are aligned to begin roughly simultaneously; C_2 is aligned to begin approximately halfway through the articulation of V. When V is sufficiently short in duration, C_1 and C_2 completely overlap V, resulting in concurrent articulation of the entire vowel with the adjacent consonants; no portion of V remains unoverlapped, resulting in the coarticulatory acoustic effect found in Chapter 4. In cases where the vowel is durationally longer, such as when the vowel is low or stressed, the vowel gestures are long enough to remain partially unoverlapped and therefore heard as a voiced vowel segment. Depending on the sonority of C_1 and C_2, the precise timing of those consonantal gestures with V varies, which may increase overlap of V and thus the likelihood of devoicing. The details of this analysis appear beginning in Section 5.3.

5.2. Background: Articulatory Phonology

Vowel devoicing and deletion have been explained as a consequence of overlapping articulatory gestures before, frequently invoking the principles of Articulatory Phonology (AP; Browman & Goldstein, 1989, 1986, 1992, and subsequent publications). In this chapter, I will propose that this is the case for Malagasy: AP provides a unified explanation of the phonetic and phonological facts of Malagasy vowel devoicing if we think of "vowel devoicing" as complete overlap of vowel gestures by the gestures of adjacent consonants. Variation in the environment where devoicing occurs will be explained by differences in the inherent duration of segments and of gestural timing.

5.2.1. Browman & Goldstein's AP

Articulatory Phonology proposes that the basic units of speech are "gestures," co-ordinated movements of the articulators within the vocal tract to reach some goal. Each gesture is an abstract unit of the phonology that is specified for up to three continuous parameters: Constriction Location (CL), Constriction Degree (CD), and stiffness, which corresponds to the time it takes for the articulators to reach their goal. Because a gesture specifies movement toward some *goal*, rather than movement of the articulators themselves, multiple articulators may be involved. For example, to reach a planned CD of the lips, sometimes called Lip Aperture, a speaker may raise or lower the jaw, the bottom lip, or the top lip (or some combination of the three).

What we think of as segment phonemes, then, are not individual discrete units, but rather a collection (or "constellation") of gestures in time with each other, and phonetic contrasts arise via differences in gestural specification. [k], for example, has two associated gestures: the Tongue Body Gesture is specified for its CL (velar) and CD (closed, referring to the complete constriction of stops). As a voiceless sound, [k] also has a glottal gesture, which can only be specified for CD (in this case, wide, referring to the open vocal folds associated with voicelessness). For [k], all other gestures (e.g., the velum, associated with nasality, or the lip protrusion gesture, associated with rounding) are not under active control; when a gesture is inactive, the articulators involved remain at a defined default resting place (except for any passive movement). [k], then, has a different gestural specification than [t]: while they both have a wide Glottal gesture, for [t], the Tongue Body gesture is inactive, and instead the Tongue Tip gesture is active and specified for its CL (alveolar) and CD (closed). The precise movement trajectory associated with each gesture is explained by Saltzman and Kelso's (1987) model of Task Dynamics, which likens each gesture to a damped mass-spring. Saltzman (1986), modelling speech articulation specifically, provides equations for the trajectory of the different articulators that specify the gesture's dynamic parameters, described in detail in Saltzman (1986) and summarized in Hawkins (1992).

Articulatory gestures are not just represented in the space dimension, as described above, but also in time. Rather than appearing in linear sequences, gestures can overlap each other on the time dimension, whether as part of the same constellation (segment), or when associated with two different segments. Originally, AP represented timing relations with a *gestural score*, a representation of each gesture's articulators (along the Y-axis) and duration (represented as a box across the X-axis). The text inside the box shows the Constriction Degree and Location. Figure

5.1 shows the gestural score for the English word *palm*, adapted based on Browman & Goldstein (1989) and Goldstein et al. (2006).

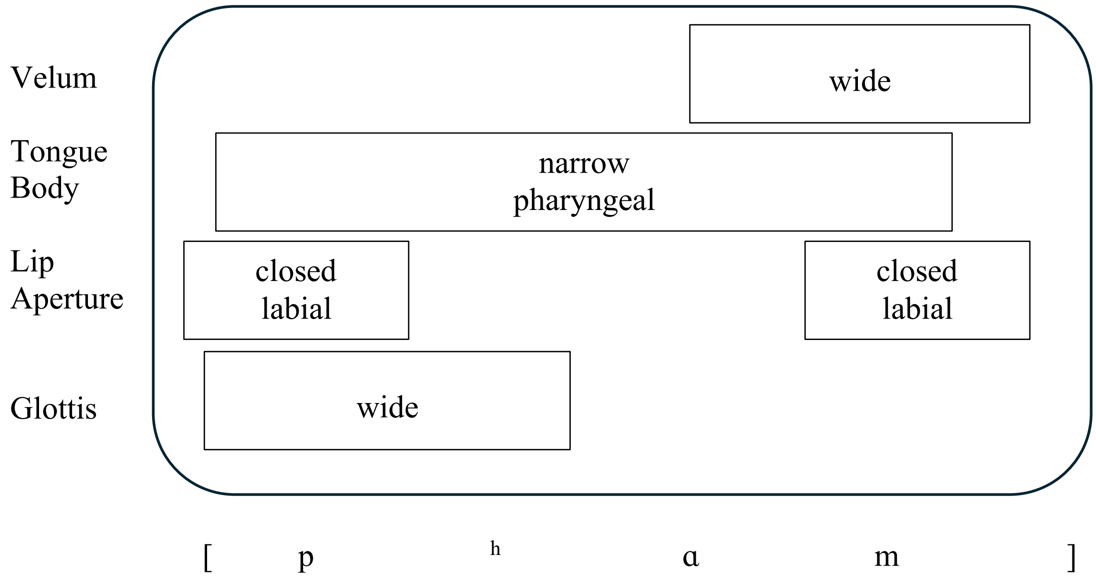


Figure 5.1 Gestural score for the American English word *palm* [pʰɑm], originating in Browman & Goldstein (1989, p. 212) and modified based on Goldstein et al. (2006).

In this gestural score, the lip and glottal gestures associated with [p] overlap, but these gestures also overlap with the tongue body gesture of [ɑ], indicating that speakers begin articulatory movement for vowels approximately at the same time as the onset consonant's gestures begin, consistent with instrumental studies on consonant-vowel coarticulation (Goldstein et al., 2006).

Articulatory Phonology defines two basic timing relations: in-phase and anti-phase. Consonant-vowel sequences are described as being in-phase, meaning that their associated gestures begin at the same time (Browman & Goldstein, 2000; Goldstein et al., 2006). This is in contrast with anti-phase gestural coordination, where one gesture begins at the midway point of the other; this has been proposed for vowel-consonant (coda) sequences (Goldstein et al., 2006). In the gestural score in Figure 5.1, the timing relations between gestures are unproblematic because no two overlapping gestures ever make use of the same articulators. As the tongue body gesture begins its trajectory toward the pharynx (for the vowel [ɑ]), it is articulated concurrently with the glottal gesture. Since these two gestures involve separate variables, they can overlap in time without issue, resulting in what is essentially a voiceless vowel (or aspiration, as it's transcribed in Figure 5.1). When the overlap involves two gestures on the same variable, however, a conflict arises. Since two gestures cannot be produced using the same articulators simultaneously, AP (via the Task Dynamic model) predicts "blending" in these cases, where the articulators move toward some intermediate value between the two gestures (Browman & Goldstein, 1989), which plays out in language-specific ways (Iskarous et al., 2012). For example, in the sequence /ki/, both the consonant /k/ and vowel /i/ are produced with a tongue body gesture, but for /i/, being a front vowel, the target of that gesture is further advanced in the vocal tract. The result is blending: during closure, the tongue body may reach a position between those typical of /k/ and /i/, producing the fronted [k̟] allophone that has been described for many languages (Keating & Lahiri, 1993).

Not only does Articulatory Phonology explain the general temporal relations between speech sounds, it has also been invoked in explaining several phonological phenomena such as allophonic aspiration and consonant assimilation (Browman & Goldstein, 1992), syllable structure (Browman & Goldstein, 1988), harmony (C. M. Smith, 2018), speech disorders (Namasivayam et al., 2020), and many others (see Hall, 2010; 2017 for a review of the literature). Coarticulation, like the consonant-vowel coarticulation described for Malagasy stressed syllables in Chapter 4, was one early locus of inquiry in the Articulatory Phonology literature (e.g., Browman & Goldstein, 1992).

The Malagasy stressed syllable /su/, for example, surfaces as [sʷu]; if we assume the gestural timing relations described above (i.e., that /s/ and /u/ are in-phase), we get coarticulated gestures associated with both sounds in the syllable. In the gestural score in Figure 5.2, the gestures associated with the nonce word /ˈsuna/ are represented. In this score, as in Figure 5.1, each gesture is activated for some period of time, represented along the X-axis. For the first syllable, /su/, two gestures are associated with /s/ (Tongue Tip and Glottis, representing voicelessness), and two with /u/ (Tongue Body and Lips). Since both are activated simultaneously, the lip rounding gesture is produced concurrently with /s/, resulting in the rounded [s^w] allophone that was described in Chapter 4. In the latter half of the syllable, the Tongue Tip and Glottal gestures associated with /s/ are inactive and the articulators return to their resting place, while the vowel gesture is (partially) without overlap, producing the voiced vowel /u/. Note that two gestures overlap on the Tongue Body variable; the duration of overlap is shaded in grey.

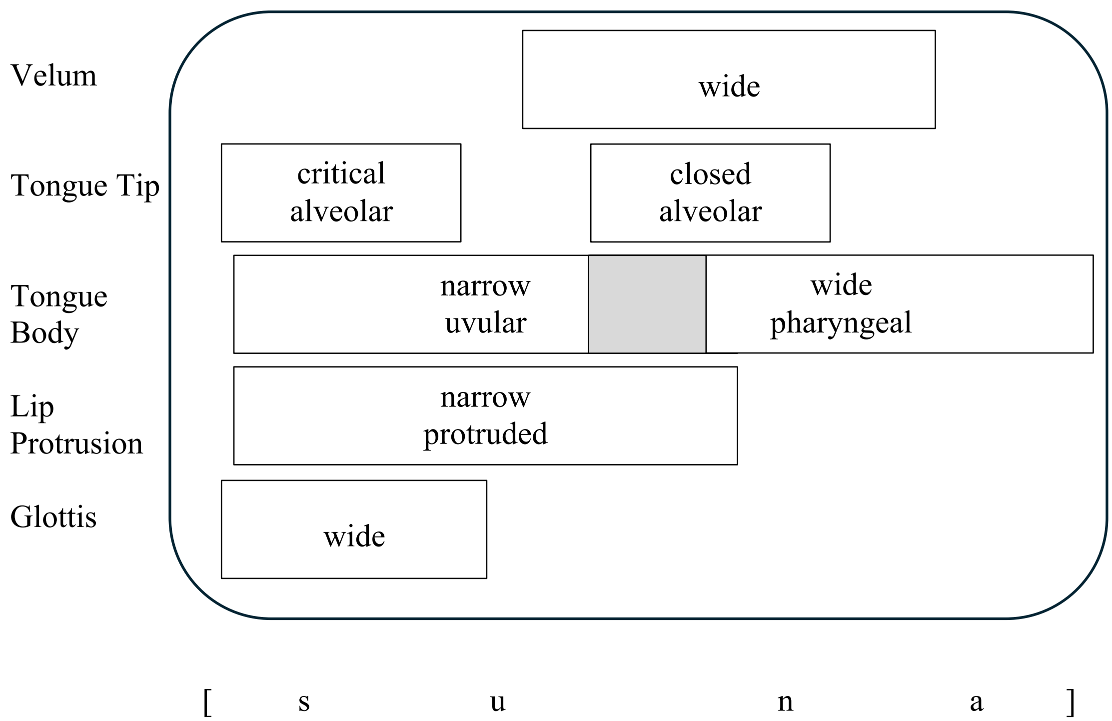


Figure 5.2 Hypothetical gestural score for the Malagasy nonce word /suna/.

5.2.2. Gafos (2002) and gestural coordination

Following Browman and Goldstein's initial body of work, subsequent research in Articulatory Phonology investigated the role of gestural timing and coordination. Gafos (2002), following a similar proposal by Zsiga (2000), presents a theory of the grammar where the simple set of timing relations proposed by Browman and Goldstein, in-phase and anti-phase timing, is inadequate. Looking at cross-linguistic evidence from consonant clusters, Gafos finds language specificity: in English, for example, CC codas occur relatively freely and with substantial overlap, evidenced by the lack of acoustic release in the first consonant in a word like *apt*. In Moroccan Colloquial Arabic (MCA), however, consonants in many clusters exhibit much less overlap, to the point of a short schwa-like release intervening between them, as in [katᵊb] 'write'. To model this precisely, Gafos divides gestures into five "landmarks" along the time dimension.

1. Onset, when movement of the articulators begins
2. Target, when peak constriction is realized
3. Center, between the target and the release
4. Release, when the articulator begins to move away from the point of constriction
5. Offset, when the articulator reaches the default position

Figure 5.3 shows a representation of a gesture under Gafos's view: as with gestural scores, the schematic represents the time dimension on the X-axis, while the Y-axis represents an estimation of the articulator's position with respect to the target Constriction Location. The time between the Target (the point where the articulator reaches constriction) and the Release (when the articulator begins to move away from constriction) is known as the plateau and is estimated to be roughly the same duration as the time between the Onset and Target. Schematics like this will appear throughout the following sections, often simplified to remove the landmark labels.

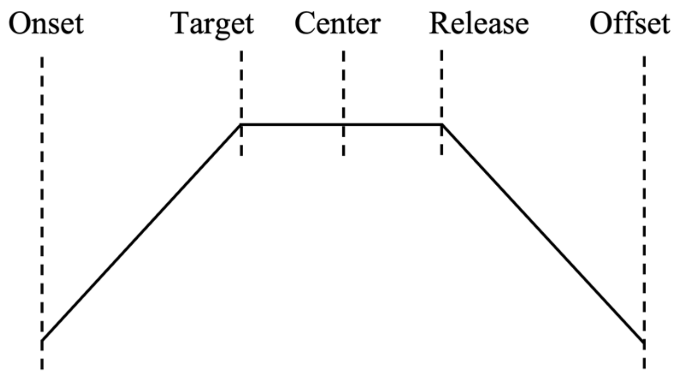


Figure 5.3 Gestural landmarks (Gafos, 2002).

Gafos proposes that Alignment constraints (which are violable, as with all constraints in a framework like Optimality Theory) modulate the coordination of gestures. These constraints take the form ALIGN(G_1, LANDMARK, G_2, LANDMARK), where each G represents a gesture (or multiple gestures, abbreviated). For example, ALIGN(C_1, RELEASE, C_2, TARGET) penalizes any CC sequence for which the release of the oral gestures associated with the first consonant (C_1) do not align with the target of the second (C_2). In Figure 5.4a, we see two gestures that satisfy that constraint, whereas 5.4b violates it because the two gestures are not sufficiently close. In both figures, C_1 is represented with a solid line and C_2 with a dashed line. In this case, (a) corresponds to the CC coordination in a language like English, where there is no audible release between consonants, while (b) could be proposed for a language like Moroccan Colloquial Arabic, where the two consonants' gestures are sufficiently separated and a release is heard.

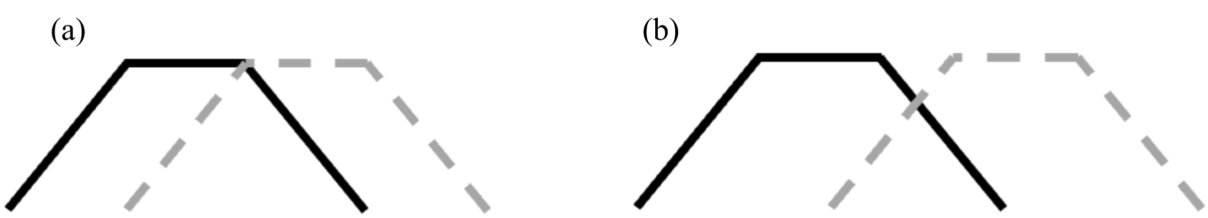


Figure 5.4 (a) satisfies ALIGN(C_1, RELEASE, C_2, TARGET) because the target of the second consonant is aligned to the release of the first.

Gafos defines several Alignment constraints for the MCA data, some of which (e.g., CV coordination (5) and VC coordination (6)) correspond to an assumed universal tendency in gestural coordination for different consonant/vowel sequences. Others, like CC coordination, can vary between languages. The CC coordination constraint defined above for English (Figure 5.4a) is different from that proposed for MCA in (5.3) (also Figure 5.4b). Further, we should note that Gafos's CV coordination constraint assumes a different timing relationship than Goldstein et al. (2006), who proposed that the onset of both C and V are aligned in those syllables.

(5.1) CV coordination: Align(C, CENTER, V, ONSET)
(5.2) VC coordination: Align(V, RELEASE, C, TARGET)
(5.3) CC coordination: Align(C_1, CENTER, C_2, ONSET)

Because these gestural coordination constraints are framed within Optimality Theory, each one is violable and may be outranked by some other constraint, resulting in a gestural coordination that is unlike those presented by the constraints. In MCA, for example, Gafos notes that the CC coordination in (5.3) should not produce a schwa release between homorganic consonants, such as /tt/, according to the Task Dynamics model. Because homorganic sequences have overlapping oral gestures (Tongue Tip, for /tt/), the target of C_2 would be achieved before the release of C_1 could be fully realized, resulting in one continuous Tongue Tip contact. However, as Gafos describes, MCA does exhibit a release between some homorganic clusters. To explain this, Gafos invokes the Obligatory Contour Principle (OCP; Goldsmith, 1976; Leben, 1973) adapted for a gestural phonology in (5.4):

(5.4) OCP: penalizes overlapping identical gestures

For MCA, then, a schwa release emerges between homorganic CC clusters because of the constraint ranking OCP >> Align(C_1, CENTER, C_2, ONSET), causing the gestures of the two Cs to move away from each other and allowing space for the schwa to be released.

This constraint-based approach to gestural phonology can be especially helpful in describing language-specific phenomena that clearly make reference to gestures as the underlying unit of the phonology. In the following sections, I will make use of these Alignment constraints, in combination with other facts about the duration and timing of segments, to explain the specific ways in which Malagasy vowel devoicing plays out.

5.2.3. Gestural accounts for vowel devoicing and deletion

Gestural explanations have been proposed for non-final vowel deletion and devoicing processes in multiple languages, with these studies invoking Articulatory Phonology and its derivatives. In Korean, Jun and Beckman (1994) observe that vowel devoicing is more likely to occur following consonants with larger and longer glottal gestures: aspirated stops, having the longest glottal gesture, caused more devoicing than lenis stops, which have a medium glottal gesture; fortis stops, having the smallest glottal gesture and shortest VOT, triggered almost no devoicing. This acoustic result is in line with the gestural analysis, which predicts that if an onset consonant overlaps the following vowel, longer consonants (i.e., those with longer VOT) should cause more overlap. In the case of Korean, the overlap of the glottal gesture of aspirated stops is sufficiently long to overlap (phonetically short) high vowels, causing devoicing. Further evidence for overlap as the cause of Korean devoicing comes from the fact that the process is not categorical, and in fact vowels even in the devoicing environment may be realized with full or partial voicing. Under feature-based theories, features are categorial (either positive or negative); gestural theories, though, allow for fine phonetic variance like gradience in voicing due to variation in the size and timing of gestures (Browman & Goldstein, 1992).

A gestural explanation for high vowel devoicing is similarly proposed for Turkish (Jannedy, 1995). As with Korean, the context for Turkish vowel devoicing points to gestural overlap. The rate of vowel devoicing increases with speech rate as vowel gestures shorten and thus become more likely to be completely overlapped by adjacent consonants; similarly, vowels in closed syllables, which in Turkish have a longer duration, are less likely to devoice compared to vowels in open syllables (which are durationally shorter). Jannedy points to these facts in analyzing Turkish vowel devoicing as a result of consonant-vowel overlap.

Acoustic evidence has also been used in arguing for gestural overlap as a cause for vowel devoicing, especially when there is debate over whether vowels are devoiced or completely deleted. For Lezgi (Northeast Caucasian; Dagestan and Azerbaijan), unstressed high vowels have been described as deleted, but Chitoran and Iskarous (2008) present spectral evidence that a vowel gesture is still present: in underlying /su/ syllables, spectral energy of /s/ is lower than in underlying /si/ syllables. Since lip rounding lowers spectral energy, Chitoran and Iskarous conclude that an underlying vowel gesture is still present for unstressed high vowels in Lezgi, but it is overlapped by the preceding consonant, resulting in what they call devoicing, rather than deletion.

More recent analyses of vowel devoicing have attempted to model the precise timing of gestures that lead to a vowel's overlap. Building on Gafos's (2002) theory of gestural coordination, Delforge (2008) proposes that CV and VC coordination can be controlled by language-specific Alignment constraints, just as Gafos proposes for CC coordination. In Andean Spanish, Delforge describes, some unstressed vowels undergo devoicing, even in slow speech rates. This is unusual for a vowel devoicing language, where vowels tend to be voiced in slower speech (as in Malagasy). To account for the Andean Spanish pattern, Delforge suggests that there is a constraint driving CV overlap, ALIGN (C, [ONSET-CENTER], V, ONSET), which states that the onset of the vowel must align with any point between the onset and the center of the preceding consonant. This constraint must be highly ranked in the grammar, even in slower speech rates, which accounts for the high degree of consonant-vowel gestural overlap and resulting vowel devoicing. Additionally, her analysis accounts for vowel devoicing as a controlled, language-specific process: languages that have less or no vowel devoicing violate the Andean Spanish Alignment constraint, permitting less-overlapped CV alignment.

The languages described up to this point have used gestural theories of representation to describe "vowel devoicing" processes, but gestural overlap has also been invoked to account for vowel deletion processes as well, such as English schwa deletion (Browman & Goldstein, 1990). More recently, Bennett et al. (2023) have described a vowel deletion process in Uspanteko (Mayan; Quiché, Guatemala). In Uspanteko, some unstressed short vowels optionally undergo deletion. The process is sensitive to the vowel's position relative to stress (i.e., in words with final stress, the pre-stress vowel may undergo deletion; in words with penultimate stress, it is the post-stress vowel) as well as morphological factors (vowels in prefixes do not delete) and other variables. Unlike most other languages described in this section, the low vowel /a/ may undergo deletion, though /a/ is typically centralized to short [ə] or [ɐ]. Bennet et al. propose that Uspanteko vowel deletion is the end point of a gradient vowel reduction process caused by gestural overlap: when (durationally short) vowel gestures are completely overlapped by adjacent consonants, it gives the acoustic impression of vowel deletion, although the vowel *gesture* has not been deleted. Bennett et al. emphasize that Uspanteko vowel deletion is apparently grammatically controlled, evidenced by its sensitivity to the phonological properties like stress described above. Specifically, they argue that the extensive gestural overlap that causes vowel deletion is driven by *prosodic modulation gestures* (Byrd & Krivokapić, 2021; Saltzman et al., 2008; to be described in more detail in Section 5.3.6); unlike articulatory gestures, modulation gestures do not correspond to movement of the articulators, but rather influence the timing and duration of the gestures that they overlap. For Uspanteko, they suggest that a modulation gesture overlapping the unstressed syllable in a foot is responsible for shortening the gestures in that syllable, causing more overlap and thus more deletion; this explains the relationship between deletion and stress described above for Uspanteko.

In sum, then, vowel deletion and devoicing processes are frequently described as a consequence of gestural overlap; language specific differences in the implementation of these processes may result from inherent differences in the phonetics of those languages (e.g., the long durations of some Korean consonants) as well as grammatically-controlled constraints on overlap (e.g., variable Alignment constraints like those proposed for Andean Spanish or modulation gestures described for foot-internal unstressed syllables in Uspanteko).

5.3. A gestural analysis of Malagasy vowel devoicing

In this section, I present an account of the Malagasy vowel "devoicing" process that captures the specific environments in which devoicing arises. First, I show the general principle of gestural coordination under which vowel devoicing arises; then, I introduce a number of gestural principles into the grammar that modulate the duration and timing of the gestures, and account for the specific environments where vowel devoicing is most likely to occur, according to the probabilistic statistical model in Chapter 3.

5.3.1. Recall: The environment of devoicing

In Chapter 3, I presented a statistical model of Malagasy vowel devoicing, identifying the environments where devoicing is or is not likely to occur. Table 5.1 replicates Table 3.1 from that chapter, showing the environments that had a near-categorical preference for either voicing or devoicing. Additionally, among the environments where vowel devoicing is variable (i.e., non-final high vowels in non-primary-stressed syllables), there were factors found to predict the likelihood of devoicing: Malagasy vowels are more likely to devoice in unstressed syllables compared to secondary stressed, in Intonational Phrase-medial syllables compared to initial syllables, and in particular segmental contexts. Specifically, vowels are more likely to devoice after sonorants, but less likely to devoice before more sonorous segments (where the rate of devoicing decreases as sonority increases). Any model of the grammar that underlies devoicing, then, must account for each of these facts.

Factor/Level	Result	Percent with that result
IP-final (unstressed) vowels	Devoiced	96.32%
Primary stressed syllables	Voiced	98.91%
/e/	Voiced	100%
Medial, unstressed /a/	Voiced	100%

Table 5.1 Factors that have a 95% or higher rate of voicing or devoicing in the dataset.

5.3.2. Basic gestural timing

I hypothesize that Malagasy vowel devoicing arises because of nearly complete overlap of the unstressed vowel by adjacent consonants. This hypothesis generates two relevant facts about the phonetics of Malagasy devoiced vowels: first, as we saw in Chapter 2, non-final devoiced vowels do not surface as discernable segments–they are not literally voiceless vowels, as we might expect. However, from Chapter 4, we know that the underlying vowel is represented in the surface form, evident as coarticulation on neighbouring segments even when the vowel is not acoustically present as its own unit. Under the following analysis, overlap of the vowel arises because of a combination of factors that manipulate both the duration of the gestures involved and their alignment to one another. In short, vowels are more likely to be devoiced when vowel gestures are durationally short and aligned more closely to adjacent consonant gestures.

In order to discuss the way that the Malagasy grammar manipulates gestural alignment, we first must establish what the basic timing relation between gestures looks like. Whereas Gafos

(2002) proposes anti-phase timing between gestures in CV sequences, more recent work points to an in-phase alignment (Goldstein et al., 2006). Given the finding in Chapter 4.3 that spectral remnants of the devoiced vowel are visible throughout the preceding consonant (e.g., Figure 4.2), I assume that the basic alignment of CV syllables is in-phase, formalized in the Alignment constraint in (5.5). The result is the schematic in Figure 5.5, showing the relative timing of the consonant's gestures (the black solid line) and the vowels' (red dotted line). This represents a hypothesis for the most basic alignment principle for Malagasy CV syllables; as we will see in Section 5.3.4, the precise coordination of the gestures is modified by the sonority of the segments involved, leading to variation in exact timing.

(5.5) ALIGN(C, ONSET, V, ONSET): In CV sequences, align the onset of the consonant with the onset of the following vowel.

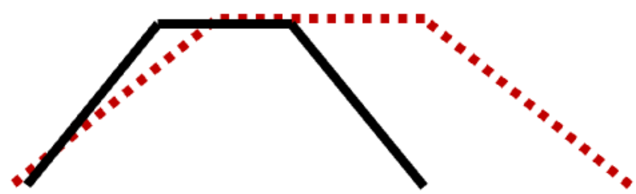


Figure 5.5 Proposed basic alignment for a Malagasy CV syllable, where the onset of the vowel gesture (dotted red line) is in-phase with the onset of the consonant gesture (solid black line).

5.3.3. Vowel height and intrinsic duration

In Table 5.1, we saw the factors that near-categorically predict voicing or devoicing. The first, that IP-final vowels are nearly always devoiced, was explained in Chapter 2 as a distinct devoicing process that is motivated by aerodynamics and results in truly voiceless vowels. We also saw that the non-high vowels /e/ and /a/ are always voiced in non-final positions, whereas high vowels are variably devoiced. This is straightforwardly explained by a theory of intrinsic duration in conspiracy with the basic gestural alignment of Malagasy. It is well documented that cross-linguistically, high vowels have an inherently shorter duration than other vowels (Lehiste, 1970), and this has been cited as a motivator for the tendency of high vowels to undergo deletion and devoicing processes (such as in the languages reviewed in Section 5.2.3). Given the overlap of CV sequences in Figure 5.5, then, it follows that low vowels like /a/, with their intrinsically longer durations, will have longer gestures that escape overlap of the preceding consonant. This is compounded by the fact that Malagasy stops and fricatives are durationally longer before high vowels, as reported in Chapter 4. The union of these facts, then, is that /sa/ sequences consist of relatively short consonant gestures aligned to relatively long vowel gestures, allowing for a greater portion of the vowel to be unobstructed by the consonant and thus heard as a full, voiced vowel segment.

The tableaux that follow show how this comes to be. Each tableau includes an underlying CVC sequence that we can imagine is in a prosodic context where devoicing can occur (i.e., a medial unstressed syllable). Whereas this sequence would appear next to other segments in a longer Malagasy word or sentence (e.g., /fanaka/), here, only the immediate segmental environment and their gestures are represented for brevity. The gestures of each segment in the C_1VC_2 sequence are colour-coded so that C_1 is black and solid, V is red and dotted and C_2 is grey and dashed. The underlying representation is written in IPA but assumed to consist of constellations of gestures corresponding to each segment. Candidates are represented both with gestures and an IPA approximation of the output.

In Tableau 5.1, we see the candidates for a word that includes /sak/, where /a/ is voiced. First, the Alignment constraint in (5.5), reiterated and abbreviated here in (5.6), requires that vowel gestures begin approximately simultaneously with its onset consonant.

(5.6) ALIGN(CV): In CV sequences, align the onset of the vowel with the onset of the preceding consonant.

The second relevant constraint, IDENT [LENGTH] requires that the vowel stay faithful to its intrinsic duration; that is, it penalizes lengthened high vowels and shortened low vowels. Here, I remain agnostic to the exact mechanism that underlies this pattern, as this is debated (see Sole & Ohala, 2010 for a discussion).

(5.7) IDENT [LENGTH]: Incur a violation for any vowel whose gestural duration is not faithful to the intrinsic duration of that vowel.

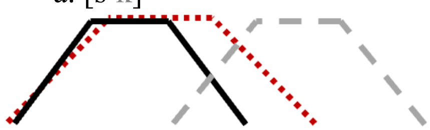
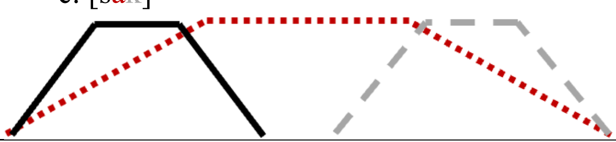
…/sak/…	ALIGN(CV)	IDENT [LENGTH]
a. [s^{a}k] 		*!
b. [saːk] 	*!	
☞ c. [sak] 		

Tableau 5.1 Gestural alignment candidates for /sak/ showing that the inherently long gestures associated with /a/ evade complete overlap by adjacent consonants.

In Tableau 5.1, candidate (a) satisfies ALIGN(CV), as the consonant and vowel are in-phase; however, it violates IDENT [LENGTH] by having a vowel gesture that is too short. Candidate (b), on the other hand, has a sufficiently long vowel gesture, but its alignment is not in-phase as the vowel gesture begins too late, violating ALIGN(CV). For candidate (c) both the gestural alignment and vowel length satisfy the constraints, producing the desired output, [sak], which has the desired vowel length.

For the high vowels /i/ and /u/, their tendency to undergo devoicing is explained by their inherently shorter durations. Tableau 5.2 below, for a word containing /suk/, has candidates analogous to those in Tableau 5.1, except that the /s/ gesture is somewhat longer to account for the lengthening effect seen before high vowels in Chapter 4. For unstressed /su/, IDENT [LENGTH] requires that /u/ be relatively short. Candidate (b) violates ALIGN(CV), and the result is that the vowel is partially unoverlapped and therefore realized as a voiced vowel. Candidate (c) violates IDENT [LENGTH] by having a lengthened vowel gesture, also resulting in voicing of /u/. Candidate (a) satisfies both the alignment constraint and the constraint on high vowel duration, and in fact produces the observed devoicing phenomenon that was described in Chapter 3. While both Candidate (c) in Tableau 5.1 and Candidate (a) in Tableau 5.2 have the same violation profile (violating neither constraint), /suk/ is realized with devoicing because of the intrinsically shorter vowel duration.

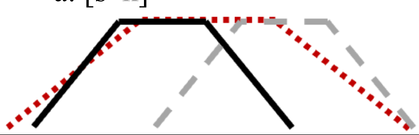
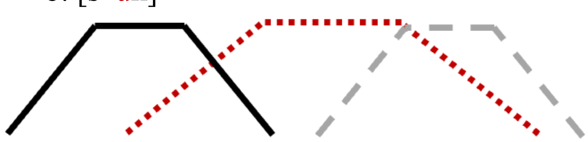
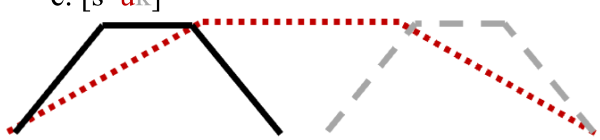
…/suk/…	Align(CV)	IDENT [LENGTH]
☞ a. [s^{w}k] 		
b. [s^{w}uk] 	*!	
c. [s^{w}uk] 		*!

Tableau 5.2 Gestural alignment candidates for underlying /suk/. Here, Candidate (a) has the same violation profile as in Tableau 5.1 (c), but produces devoicing for /u/, unlike /a/.

One additional fact to account for is that /i/ is more likely to be devoiced than /u/, according to the model in Chapter 3. Recall that in Chapter 4 it was found that the duration of aspiration of /t/ (and perhaps of all stops) is longer before voiced /i/ than /u/. Under the gestural view, longer consonants overlap the following vowel more, and thus increase the likelihood of devoicing (similar to what was described for Korean in Section 5.2.3). For Malagasy, then, the lengthening of certain consonants not just before high vowels, but especially before /i/, is straightforwardly linked to the rate of devoicing.

5.3.4. Preceding consonants and Sonority-Driven Gestural Timing

Recall from Chapter 3 that one of the significant factors contributing to the likelihood of vowel devoicing is the sonority of the preceding consonant, with sonorants causing more devoicing than obstruents. Under the gestural overlap account of devoicing, there are three possible explanations for this: either (a) the vowel gesture is longer after obstruents compared to sonorants, (b) the consonant gesture is longer for sonorants than for obstruents, or (c) the alignment of gestures is different for sonorants and obstruents. While I won't completely rule out (a) and (b), recent research on the relationship between sonority and gestural alignment does make (c) a promising hypothesis for the Malagasy data.

In a pre-print (not yet peer-reviewed), Gu and Durvasula (2024) use archival X-ray data to look at gestural timing in English CV syllables. They find that the alignment of C and V gestures can be explained by the difference in sonority between the two (i.e., the sonority of V minus the sonority of C): as the sonority difference (Δs) becomes more negative (i.e., it approaches negative infinity), the relative timing of the vowel gesture is earlier. Not only do they find that gestural timing is sensitive to the sonority of the consonant, but also to that of the vowel: vowel gestures are aligned significantly earlier for high vowels compared to mid and low vowels (though the effect is small). Gestural timing, then, can be predicted based on the difference between the second gesture (in this case, the vowel) and the first (the consonant): a more negative value indicates more overlap, and a more positive value less overlap. The sonority difference is based on sonority values assigned to each class of segments, from Parker (2008).

Natural class	Sonority Index
Low vowels	17
Mid peripheral vowels	16
High peripheral vowels	15
Mid interior vowels (ə)	14
High interior vowels (ɨ)	13
Glides	12
Rhotic approximants	11
Flaps	10
Laterals	9
Trills	8
Nasals	7
Voiced fricatives	6
Voiced affricates	5
Voiced stops	4
Voiceless fricatives	3
Voiceless affricates	2
Voiceless stops	1

Table 5.2 Sonority hierarchy, adapted from Parker (2008).

For example, in the syllable /ma/, the second segment /a/ has a sonority value of 17 and the first, /m/, has a value of 7: $\Delta s = 10$. /ba/, on the other hand has a sonority difference of 13 (17 for /a/ minus 4 for /b/). Since /ma/ has a lower sonority distance than /ba/, Gu and Durvasula's hypothesis predicts more overlap for /ma/, a prediction that is borne out in the X-ray data.

This result mirrors that found for CC onsets in Georgian by Crouch (2022; see also Crouch et al., 2023): CC clusters with rising sonority (e.g., /br/) exhibit less gestural overlap than those with falling sonority (e.g., /rb/). Using the same sonority values in Table 5.2, this gives /br/ a sonority difference of 4 (8-4) and /rb/ a difference of -4 (4-8). As predicted by Gu and Durvasula (2024), the sequences with a more negative sonority difference (such as /rb/) have greater gestural overlap. Thus, as Gu and Durvasula suggest, this relationship between sonority difference and overlap may be a cross-linguistically viable principle, regardless of the segments involved (i.e., CV, CC, VC, or otherwise). For the grammar of Malagasy, I'll call this the Principle of Sonority-Driven (Gestural) Timing:

(5.8) Principle of Sonority-Driven (Gestural) Timing (SDT): As the sonority difference between two segments' gestures ($\Delta s = G_2 - G_1$) decreases (becomes more negative), the timing of G_2 relative to G_1 is earlier.

For Malagasy, the SDT can explain the increased probability of vowel devoicing following a sonorant compared to an obstruent. For example, the CV sequence /ni/ has a sonority difference of 8 (15-7), while for /si/ it is 12 (15-3): because /ni/ has a lower sonority difference, Gu and Durvasula predict an earlier vowel gesture. In Figure 5.6, compare the timing predictions for (a) /ni/ vs. (b) /si/. For /ni/, the SDT pushes the vowel gesture leftward, and the later portion of the vowel is nearly completely overlapped, resulting in devoicing; for /si/, however, the SDT dictates that the vowel gesture be timed relatively later, increasing the probability that the vowel evades overlap and is therefore voiced. Note that this schematic is meant to show the relative difference in timing between /ni/ and /si/ and does not represent actual timing derived from instrumental measurements.

(a) /ni/, $\Delta s = 8$ (b) /si/, $\Delta s = 12$

Figure 5.6 CV alignment predicted by Gu and Durvasula's principle of sonority-driven gestural timing. Because /ni/ has a lower sonority difference, the vowel gesture is timed earlier than for /si/.

In Section 5.3.2, I proposed that the alignment of CV sequences of Malagasy is basically in-phase. Now, we can refine our hypothesis: the precise timing is modulated by the Principle of Sonority-Driven Gestural Timing. We can formalize this prediction in a unified timing constraint for CV syllables, in (5.9).

(5.9) SDT(CV): In CV syllables, the basic timing of gestures is such that the onset of V aligns to the onset of C. As the sonority difference between V and C ($\Delta s = V - C$) decreases, the timing of V relative to C is earlier.

The hypothesis that CV syllables with less sonorant consonants have later vowel gestures successfully explains the relationship between onset sonority and vowel devoicing but has not yet been demonstrated using acoustic or instrumental evidence. While the present study does not include that data, future research could investigate CV timing in Malagasy using methods that can precisely measure timing.

As for the architecture of the grammar, the literature reviewed in Section 5.2 motivates language-specific implementations of gestural timing (which we have accounted for using Gafos's (2002) Alignment constraints). The SDT constraint in (5.9), which collapses the basic gestural alignment for CV syllables described in (5.5) with the principle of Sonority-Driven Timing, is thus violable and assumed to compete with other gestural timing constraints. For this reason, I will leave the original ALIGN(CV) constraint (which does not reference the SDT) in the grammar to account for any language that hypothetically does not obey the SDT.

5.3.4.1. Some acoustic evidence for sonority-driven timing

While the present study does not present articulatory data to support a theory of Sonority-Driven Timing, we can look at the acoustic signal for evidence of differences in timing. Already in Chapter 4, we saw that the acoustic effect of an underlying devoiced vowels was measurable on the vowel of the preceding syllable (i.e., vowel-to-vowel coarticulation). For "devoiced" vowels following /n/, it was clear that the devoiced vowel's gestures began even before the onset of /n/. The prediction of the SDT is that in a CV syllable, the onset of V should be earlier when C is a sonorant (such as /n/) compared to an obstruent; thus, any vowel-to-vowel coarticulation should begin earlier for vowels following /n/.

To test this, we can return to the data from Task One in Chapter 4 (real Malagasy words with devoiced /i/ and /u/). Figure 5.7 shows second formant measurements for the final 70ms of /a/ preceding unstressed /ni/ and /nu/. Using a fixed time interval for F2 measurement anchored to the onset of the consonant (rather than measuring the entire duration of /a/) acts as a normalization measure; any F2 movement associated with the following /i/ or /u/ will be timed with respect to the onset of the consonant; any tokens with /a/ duration less than 70ms were excluded from analysis. In this figure, vowel-to-vowel coarticulation is visible throughout /a/: F2 is measurably lower before /nu/ at all timepoints. Beginning at Time Point 3, however, we see F2 begin to climb steadily before /ni/, and by the end of /a/, the difference in F2 between /ani/ and /anu/ is nearly twice as wide as that at the beginning. From these data, then, we can hypothesize that the vowel-to-vowel coarticulation observed here occurs in two steps: in the first two Time Points, a modest but steady anticipatory coarticulation effect is observed, followed by greater coarticulation with the following vowel beginning at Time Point 3 (an effect that may be caused by a two-part coarticulatory effort reminiscent of Perkell & Chiang's (1986) controversial hybrid model of coarticulation). As a point of reference, note that F2 before /ni/ crosses 1800 Hz around Time Point 3.

Figure 5.8 shows the same vowel-to-vowel coarticulation measurement but before /si/ and /su/. Here, note two differences: first, the difference between /asi/ and /asu/ is considerably smaller in the first half of /a/. Additionally, the point at which /asi/ and /asu/ diverge appears to be later; whereas F2 crossed 1800 Hz at Time Point 3 before /ni/, before /si/ this happens at some point between Time Points 3 and 4.

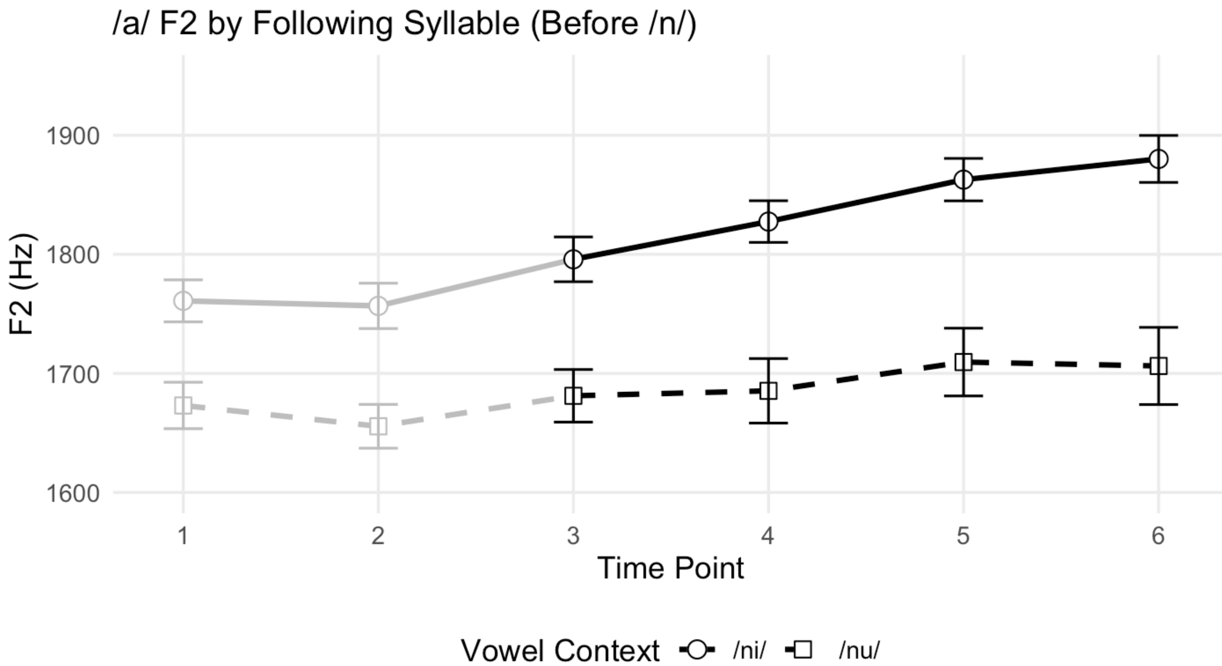


Figure 5.7 F2 values for /a/ before underlying /ni/ compared to /nu/, showing that the difference between the two widens toward the end of /a/.

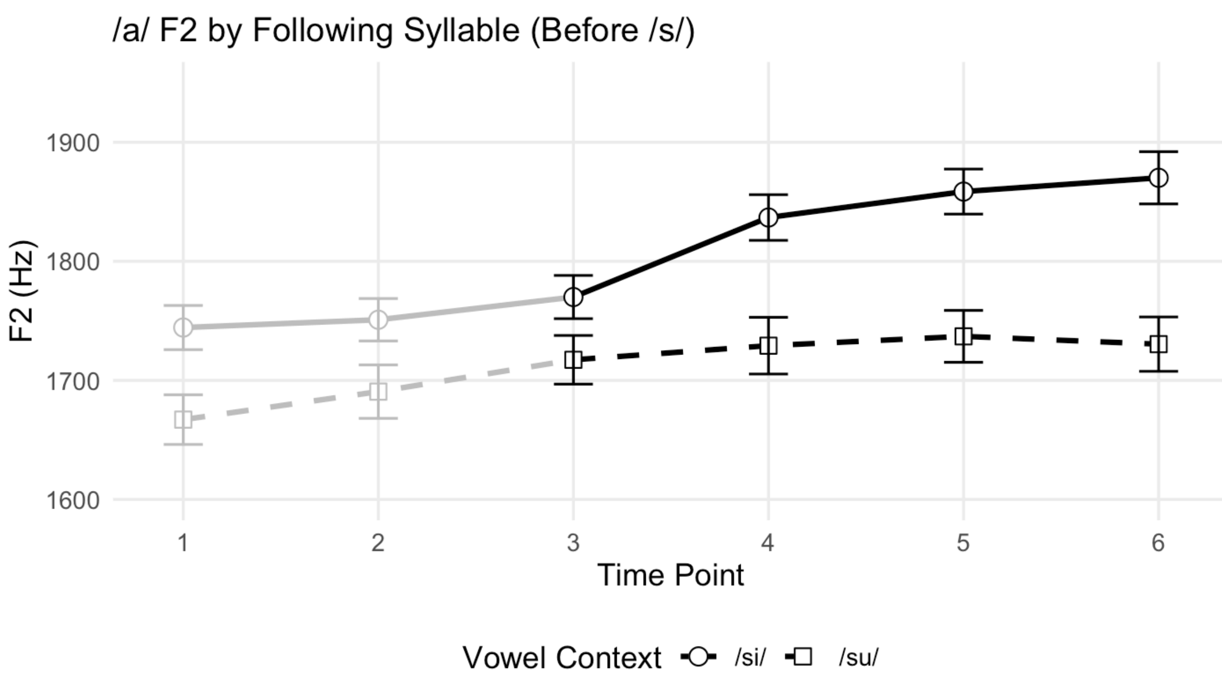


Figure 5.8 F2 values for /a/ before underlying /si/ compared to /su/.

If we take the critical point of diversion as the onset of the devoiced vowel's gestures (again, assuming that some other anticipatory coarticulation mechanism is responsible for the across-the-board difference seen throughout /a/), then it appears that the high vowel gesture begins earlier with respect to /n/ than it does for /s/.

5.3.5. Following consonants and more evidence for sonority-driven timing

In Chapter 4, I also found a relationship between sonority and devoicing in VC sequences: Malagasy vowels are more likely to be devoiced before less sonorous sounds. Whereas above I demonstrated that CV sequences are sensitive to sonority, in that case the statistical model found only two relevant categories worth comparing: sonorants and obstruents. For VC sequences, however, the model found that specifying the manner of articulation better modelled the data in a way that mirrors the sonority hierarchy. (5.10) replicates item (3.4) in Chapter 3.3.6, showing the likelihood of devoicing by manner of articulation, including only those comparisons that were shown to be significant (see that chapter for the full statistical analysis).

(5.10) **Likelihood of vowel devoicing by the following consonant**
Stops > fricatives > nasals > laterals > /r/

This finding can be explained by extending the Principle of Sonority Driven Gestural Timing described above in (5.8): as with CV sequences, the timing between V and C gestures is driven by the relative distance in sonority between the two segments. Since vowels have a high sonority value, the SDT predicts that VC sequences with a relatively shallower fall in sonority will have an earlier C gesture relative to the V. For example, in the sequence /an/, the sonority difference between /n/ and /a/ is -10 (7-17); for /as/, Δs is -14 (3-17). As Δs for /as/ is lower, the SDT predicts closer alignment and therefore more gestural overlap.

For the Malagasy data, this explains the sonority of the following consonant as a predictor for vowel devoicing: as the sonority of C increases, the sonority difference of VC increases and the C gesture is timed relatively later, increasing the likelihood that V escapes overlap of C and is heard as a voiced vowel. I formalize this with a second SDT constraint in (5.11). This constraint takes the basic VC alignment to be anti-phase (following Goldstein et al., 2006), and I assume that SDT (VC) is in competition with a generalized ALIGN(VC) constraint (ALIGN(V, CENTER, C, ONSET)).

(5.11) SDT(VC): In VC sequences, the basic timing of gestures is such that the onset of C aligns to the center of C. As the sonority difference between C and V ($\Delta s = C - V$) decreases, the timing of C relative to V is earlier.

Tableau 5.3 shows the constraint competition for a sample CVC sequence /nuk/. Here, constraint violations for both the SDT constraints and the basic Alignment constraints are assessed categorically; though, with more precise timing measurements, we could assess violations as a measure of the exact distance of a gesture from its ideal timing (e.g., in milliseconds).

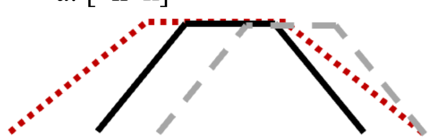
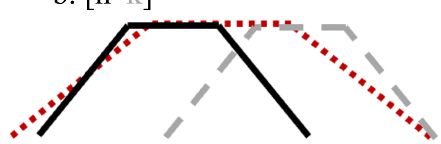
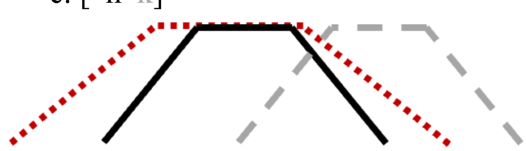
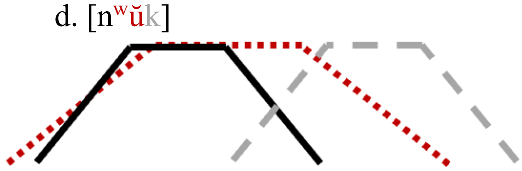
…/nuk/…	SDT(CV)	SDT(VC)	ALIGN(CV)	ALIGN(VC)
☞ a. [^{w}n^{w}k] 			*	*
b. [n^{w}k] 	*!			*
c. [^{w}n^{w}k] 		*!	*	
d. [n^{w}ŭk] 	*	*!		

Tableau 5.3 Gestural alignment candidates for /nuk/, showing that the candidate that violates neither SDT constraint produces the desired output with extensive overlap.

In this tableau, Candidate (b) violates SDT (CV) as the V gesture is timed later than it should; the SDT dictates that vowels be timed earlier with respect to a more sonorous onset consonant. Candidate (c) has the opposite violation profile; whereas the CV timing adheres to the SDT, that candidate violates SDT (VC). Recall that in VC sequences, the SDT predicts that a less-sonorous C gesture (such as the voiceless stop here) will be timed earlier than the basic (anti-phase) alignment. Candidate (d) violates both SDT constraints: like in Candidate (b), the V is timed too late relative to the onset C, and like in Candidate (c), the following C is timed too late relative to V. While Candidate (a) does violate both basic Alignment constraints, as the CV sequence is not strictly in-phase nor is the VC sequence anti-phase, those constraints are outranked by the SDT constraints, neither of which is violated by (a). Candidate (a) emerges as the winner, as it violates neither SDT constraint, and the result is a vowel that is completely overlapped (and therefore "devoiced") by both consonants that surround it: the observed phonetic tendency for unstressed medial /nuk/ in Malagasy.[2]

In Tableau 5.3, the winner (Candidate (a)) is not the only candidate that produces the gestural overlap that could cause "devoicing": Candidates (b) and (c) also show the vowel undergoing near-total overlap, and so at first glance it may appear that violating the SDT can produce the desired outcome. However, in Tableau 5.4 below, we see the tableau for a word with medial unstressed /sun/, where it is evident that SDT is never violated. Recall from Chapter 3 that Malagasy vowels are less likely to be "devoiced" when they precede more sonorous sounds (as in this case, where /u/ precedes /n/). In Tableau 5.4, only one candidate, Candidate (d), produces the desired output, where the vowel is not totally overlapped and thus emerges as a voiced segment. Note that that candidate is the only one that does not violate the SDT in any way; adherence to this

[2] Note again, however, that the precise timing is by hypothesis: here, CV is depicted as being neatly in-phase, for example, but it's possible that the exact timing is somewhat different.

principle of sonority-driven timing successfully predicts the segmental environment where vowel devoicing is least likely to occur.

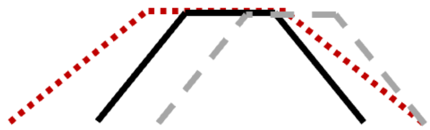
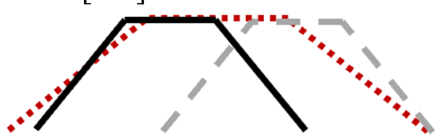
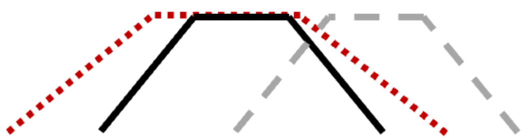
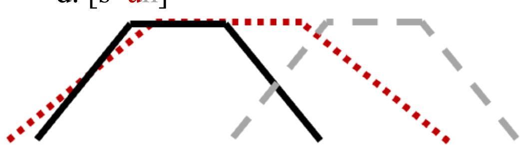
…/sun/…	SDT(CV)	SDT(VC)	ALIGN(CV)	ALIGN(VC)
a. [^{w}s^{w}n] 	*	*	*	*!
b. [s^{w}n] 		*		*!
c. [^{w}s^{w}n] 	*		*!	
☞ d. [s^{w}ŭn] 				

Tableau 5.4 Gestural alignment candidates for /sun/, showing that only the candidate that violates neither SDT constraint produces an output without complete overlap of the vowel.

In fact, we now have the machinery to derive vowel devoicing in each of the desired segmental environments. Returning to the dataset used in Chapter 3, we can look at the raw percentage of tokens that surfaced as voiced in different environments based on sonority. The data comes from high vowels in medial, unstressed syllables (i.e., a prosodic environment where we expect maximal devoicing). Here, all sonorants except /r/ are collapsed into one category, as are all obstruents. Data with /r/ are not included because of its variable phonetic status, discussed in Chapter 1. Table 5.3 includes the rate of vowel voicing in each environment, alongside the predictions of the SDT for alignment of gestures in CV and VC sequences.

Environment	Tokens Voiced (%)	SDT(CV)	SDT(VC)
Obstruent-vowel-sonorant (e.g., s_n)	59.7	Later V	Later C_2
Sonorant-vowel-sonorant (e.g., n_n)	8.97	Early V	Later C_2
Obstruent-vowel-obstruent (e.g., s_k)	5.76	Later V	Early C_2
Sonorant-vowel-obstruent (e.g., n_k)	5.43	Early V	Early C_2

Table 5.3 Rate of devoicing of high vowels in unstressed medial syllables, by segmental environment.

Three of the four environments (where the vowel follows a sonorant or occurs between two obstruents) cause a high rate of devoicing (over 90%), while the rate of devoicing is quite variable between an obstruent and a sonorant. The SDT predicts this. Below in Table 5.4 are the four gestural score estimates, based on the SDT, for each of the segmental environments in Table 5.3. The columns depict the gestural alignment based on the sonority of the consonant preceding the vowel, and the rows the consonant following the vowel. Note that the alignment of the second consonant is gradiently related to its sonority; the SDT predicts more than two categories here (i.e., the alignment is different for fricatives, laterals and so on), but only a stop and a nasal are depicted here for simplicity. You'll notice that each of the possible alignments were already seen as the candidates in Tableaux 5.3 and 5.4; this is because they represent the principle alignment structures for C_1VC_2 based on the SDT (though, given the gradience effect of sonority observed for C_2, SDT(VC) in principle produces an equally gradient range of VC timing relations).

The figures in Table 5.4 demonstrate how the SDT correctly predicts the devoicing environment for Malagasy: (b, c, and d) all show the final part of the vowel gesture completely overlapped by the adjacent consonant gestures; indeed, these are the segmental environments where vowels "devoicing" is most likely to occur. It is only for (a) that the vowel is able to emerge without total overlap, resulting in possible voicing, as was reported in Chapter 3.

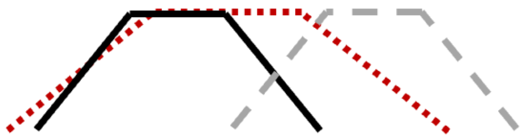
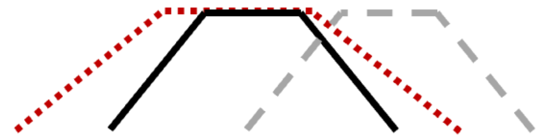
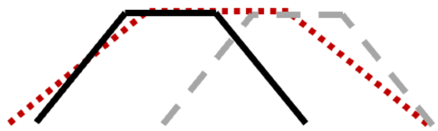
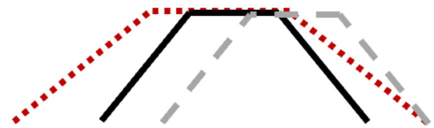
	V is later (C_1 is less sonorous)	V is early (C_1 is more sonorous)
C_2 is later (sonorant)	(a) /sin/ 	(b) /nin/ 
C_2 is early (obstruent)	(c) /sik/ 	(d) /nik/ 

Table 5.4 Gestural scores predicted by the Principle of Sonority-Driven Gestural Timing for different segmental environments.

5.3.6. Prosody and prosodic gestures

In the past two decades, researchers have attempted to incorporate prosody into Articulatory Phonology and Task Dynamics, using these theories to explain things like stress, rhythm, and intonation. In Chapter 3, we saw three ways in which prosody contributed to devoicing in Malagasy. First, although it was just outside the criteria for statistical significance, vowels in IP-initial syllables were more likely than medial vowels to be voiced. Second, stress was a significant factor: primary-stressed syllables were nearly all voiced, while secondary stress significantly increased the likelihood of voicing. Finally, unstressed vowels in the pre-stress position were significantly more likely to be voiced than those in the post-stress position. To account for these facts, we can make use of two types of abstract gestures introduced into AP: prosodic gestures (π-gestures) and modulation gestures (μ_T-gestures).

Byrd and Saltzman (2003) propose π-gestures as an explanation for prosodic effects at phrase boundaries, including lengthening that often occurs both phrase-initially and phrase-finally. π-gestures, like other gestures in AP, are abstract units that overlap other gestures, but unlike those gestures, π-gestures do not correspond to the physical movement and constriction of the articulators. Rather, π-gestures affect the dynamics of the articulatory gestures that they overlap, in a way that "slows the clock that controls the timeflow of an utterance" (Byrd and Saltzman, 2003, p. 160). The effect is that articulatory gestures overlapped by π-gestures are longer and exhibit less overlap with one another. The π-gesture straddles the prosodic phrase boundary, and the slowing effect increases in its magnitude toward the phrase edge. They suggest that the strength of the π-gesture is correlated with the strength of the prosodic boundary.

In Malagasy, the π-gesture straightforwardly explains the decreased probability of voicing in the IP-initial position: in this position, the initial consonant and vowel gestures move apart from each other, resulting in less overlap and therefore a greater likelihood that the vowel is voiced. For CV sequences for which the SDT predicts an earlier vowel (i.e., when C is more sonorous), I assume that the gestures moving "apart" results in the vowel moving rightward, rather than even more extremely leftward. If it were true that the π-gesture pushed the vowel even further left, we would expect to hear a sort of metathesis, where the vowel is audible before the consonant (e.g., [^{i}nit^{j}firana] for /nitifirana/). However, after manually analyzing the spectrogram of every token included in the experimental dataset, this is not something that I observed. As a result, we can conclude that the π-gesture constraint outranks the SDT.

In Tableau 5.5, the π-gesture is represented with a horizontal brace representing the duration over which the π-gesture is active. The gradient beneath the brace represents the strength of the π-gesture, which is darkest at the IP-boundary, where the effects of the π-gesture are strongest. In the grammar, we can propose a simple constraint, π-gesture, which assigns a π-gesture to the IP boundary. While the model in Chapter 3 did not represent the phonological phrase boundary as a predictor of devoicing, I don't rule out the possibility that there is a π-gesture at that boundary. Aziz & Elkins (2022) find that a process of consonant strengthening is more likely to occur within, rather than across, the phonological phrase boundary in Malagasy. The resistance to strengthening at this boundary may be caused by decreased overlap between gestures in that position.

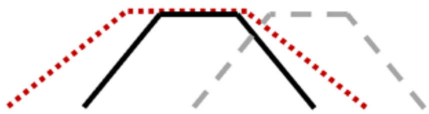
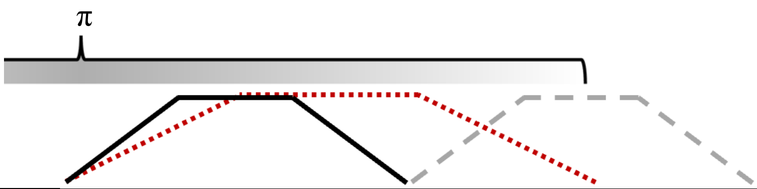
IP[/nit/…	π-gesture	SDT
a. [^{i}n^{j}t] 	*!	
☞ b. [n^{j}it] 		*

Tableau 5.5 Gestural alignment candidates for /nit/ at the IP-initial boundary, showing how the π-gesture in Candidate (b) makes gestures longer and further apart.

In Tableau 5.5, Candidate (a) is eliminated because it violates π-gesture by having no π-gesture, and therefore typical gestural durations and timing. Candidate (b) satisfies π-gesture as the gestures have been lengthened and separated from one another, including moving the vowel rightward. Candidate (b) violates the SDT because the vowel gesture is timed too late (see Figure (d) in Table 5.4 for the anticipated timing); however, because π-gesture outranks it, the candidate is chosen as the winner. In fact, that candidate, having a voiced vowel, is the one that may be observed in Malagasy in the initial position.

Regarding stress, Howe (2017, p. 202) reports that Malagasy vowels are shorter in unstressed syllables, and, as we've seen, primary stressed syllables are near-categorically voiced. As with vowels in IP-initial syllables, we can explain the relationship between stress and duration/devoicing using a function modulating gestural duration and timing. Saltzman et al. (2008) extend the idea behind the π-gesture to that of the temporal modulation gesture (μ_T-gesture), of which the π-gesture, they say, is just one type. The general idea is the same: μ_T-gestures are abstract gestures that don't manipulate the articulators themselves, but rather have an effect on the articulatory gestures that they overlap. For Malagasy stress, then, we can assume that stress syllables are associated with μ_T-gestures that result in longer, less overlapped gestures in a way that is analogous to the IP-final case in Tableau 5.5. The slowing of gestural timing caused by the μ_T-gesture can also explain why vowels are more likely to be devoiced after a stressed syllable than before: in pre-stress syllables, there is a timing relation between the vowel and the following stressed consonant (as with all VC sequences). A μ_T-gesture overlapping the stressed syllable will thus partially overlap that pre-stress vowel, slowing the timeclock and causing less overlap and therefore less devoicing. Vowels in the post-stress position, however, have no inherent timing relation to the stressed syllable; they are only timed to the preceding and following consonants. Thus, if a μ_T-gesture overlaps a stressed syllable, the timing of the following (post-stress) vowel will be unaffected.

For secondary stress, recall from Chapter 3 that, unlike primary stress, vowels in these syllables are not categorically voiced; however, they are more likely to be voiced than unstressed syllables. For this, I can offer two possible explanations: on the one hand, it's possible that our understanding of secondary stress in Malagasy is incomplete, and that some of those vowels coded as secondary stressed in the dataset are actually unstressed. Alternatively, it could be that secondary stressed syllables are associated with an intermediate duration or overlap specification, compared to primary and unstressed syllables. In that case, stressed syllables may be marginal in their likelihood of total overlap, resulting in more variation in whether the vowel is devoiced. While I don't offer a concrete explanation for secondary stress here, in any case, this highlights the need for a more comprehensive understanding of Malagasy stress.

5.4. Discussion

5.4.1. Malagasy vowel devoicing: a phonetic and phonological conspiracy

In this chapter, I have argued that Malagasy vowel devoicing arises due to gestural overlap of the vowel by adjacent consonants. As presented in Chapter 4, Malagasy devoiced vowels appear to be deleted, except that there is acoustic evidence of an underlying vowel gesture on the preceding consonant and even as early as the previous vowel. Additionally, Chapter 3 presented a probabilistic model of the phonological environments where vowels are most likely to devoice; in this chapter, I have shown that the devoicing environment, while quite complex at first glance, is explained neatly using a theory of Articulatory Phonology where gestural duration and timing may vary. Vowel devoicing, then, can arise under different circumstances, as multiple different factors contributing to the likelihood of devoicing may conspire to cause devoicing.

Three main phonetic or phonological factors were found to contribute to the likelihood of devoicing: the inherent duration of segments, prosodic gestures that cause gestures to vary in their overlap and sonority-driven gestural timing. In Section 5.3.3, I argued that high vowels have shorter gestures (consistent with the literature on intrinsic vowel duration); short vowel gestures are more likely to be overlapped and therefore devoiced, which explains why mid and low vowels are resistant to devoicing. Additionally, in Chapter 4 I showed that obstruent consonants are durationally longer before high vowels, increasing the duration of overlap. In Sections 5.3.4 and 5.4.5, I showed how Malagasy vowel devoicing is sensitive to the sonority of the adjacent consonants. Adopting a recent proposal by Gu and Durvasula (2024), I showed how a theory of sonority-driven gestural timing can explain the specific segmental environments where vowel devoicing is most likely to occur. Finally, I have shown that vowels in phonologically-prominent syllables are less likely to undergo devoicing, including stressed syllables and IP-initial syllables.

In Section 5.3.6, I suggested that this is due to prosodic gestures that overlap prominent syllables, decreasing gestural overlap.

Because multiple factors are at play, this means that the likelihood of devoicing increases with the addition of factors that promote devoicing. For example, inherent vowel duration alone will not explain the pattern of vowel devoicing, as the relatively short vowels /i/ and /u/ are voiced in prosodically-prominent syllables, as well as in the obstruent-vowel-sonorant environment. Devoicing, then, emerges as a perfect storm of these phonetic and phonological variables.

5.4.2. Variation

I have described how gestural timing and segmental duration result in complete overlap of the vowel by adjacent consonants, resulting in the non-final vowel "devoicing" process of Malagasy. In certain environments, the rate of devoicing is categorical or near-categorical; primary-stressed syllables, for example, nearly always contain voiced vowels. However, in most other environments, the rate of devoicing is variable, and the probability of devoicing varies between those environments. In Table 5.3, I presented the raw percentages of tokens that surface as voiced segments in unstressed medial syllables, depending on the segmental environment. Between two obstruents, for example, over 96% of vowels are devoiced. For vowels that follow an obstruent but precede a sonorant (collapsing all types of sonorants), that number drops to just 40%. Even looking at individual items, certain participants may devoice the vowel while others do not. Why, then, does this variation arise and how can we account for the fact that certain environments are more prone to variation?

Here, I will not formally model variation in Malagasy vowel devoicing, but we can look to the literature for explanations of variation in gestural theories of phonology. Byrd's (1996) Phase Window framework proposes that the relative timing of a gesture is not precise, but rather constrained to a durational *window* within which the gesture may be timed. The probability that a gesture is timed to any given point in the window is determined by a combination of universal, language-specific and extralinguistic constraints (which she calls *influencers*). I will take Byrd's basic assumption to be true, that there is a window within which gestural timing takes place, but simplify it somewhat in assuming that the probability distribution over the window is normal, leaving aside any influencers (though, I acknowledge that constraints like Sonority-Driven Gestural Timing could be incorporated as influencers in Byrd's model). Figure 5.9 shows two gestures, the second of which (the grey dashed line) is timed to begin somewhere within the phase window (represented with the bell curve). Because the curve is a normal distribution, the gesture is likely to fall near the center of the window but may fall earlier or later. While some authors have incorporated the phase windows directly into the grammar (e.g., Zsiga, 2000), I assume that the target gestural timing is determined by the SDT and Alignment constraints presented in this chapter, and that the phase window here is essentially random noise caused by the natural imprecision of human speech articulation.

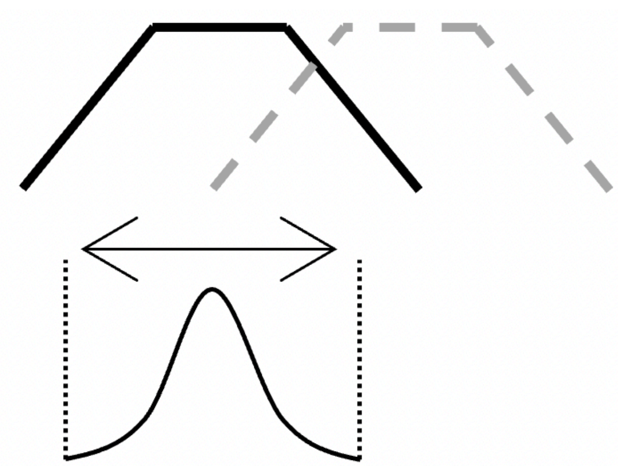


Figure 5.9 Example overlap between two gestures showing the phase window within which the onset of the second gesture may fall, represented as a bell curve.

Because Malagasy vowel "devoicing" is caused by gestural overlap, the variation in gestural timing caused by the phase windows results in variation in overlap and thus in the probability of devoicing. For the environments where vowel devoicing is nearly always realized (e.g., Table 5.4 (b-d)), I assume that the target timing relations cause such extensive overlap that even given the randomness introduced by the phase window, the probability that a vowel is partially unoverlapped remains extremely low. For environments where vowel devoicing is more variable, the timing targeted by the gestures results in a degree of overlap that is marginal. Figure 5.10 replicates Table 5.4 (a), showing the gestures for unstressed medial /sin/ with phase windows. It has two phase windows, one corresponding to C_1V timing and the other to VC_2 timing. If either V or C_2 falls in the left half of the window, it increases the degree of overlap and thus the likelihood of devoicing; if the gestures fall in the right half of the window, the degree of overlap and probability of devoicing decreases. It is in this way that phase windows can account for the variation seen in some phonological environments where devoicing occurs.

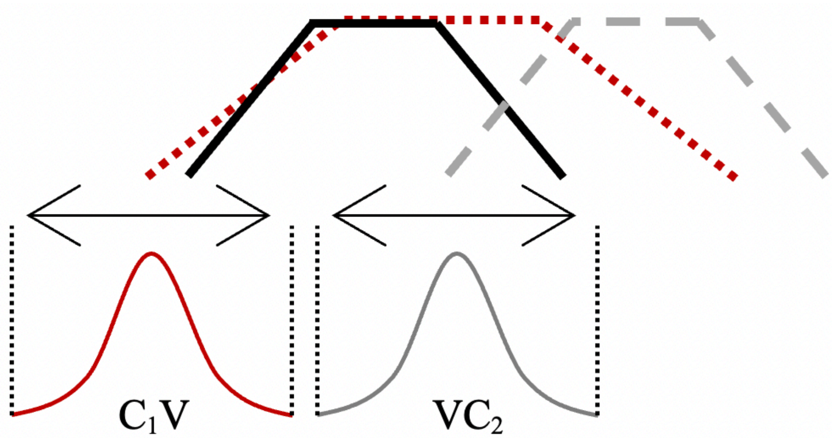


Figure 5.10 Gestural representation of Malagasy …/sin/… showing the phase windows for both gestural timing relations.

Chapter 6
Discussion and Conclusion

6.1. Summary of findings: So-called vowel devoicing as gestural overlap

This dissertation presented a comprehensive look at a phenomenon commonly called "vowel devoicing" in Merina Malagasy. In Chapter 2, I provided a descriptive account of these vowels, finding that what has been called "vowel devoicing" is better characterized as two distinct processes: in the Intonational Phrase-final position, unstressed vowels (/i, u, a/) are commonly realized as voiceless, clearly distinct from the preceding consonant. In non-final positions, however, "devoicing" of the high vowels /i/ and /u/ leaves no evidence of an interval occupied by the underlying vowel. That is, unstressed high vowels undergoing "devoicing" appear be absent from the output as a discrete unit at all. For example, in a word like *pasoka* /ˈpa.su.ka/, the transition from the frication of /s/ to the stop closure of /k/ is abrupt, leaving no evidence of a vowel segment, voiced or devoiced. This is emphasized by the fact that "devoicing" occurs following voiced sonorants. In those cases, as with obstruents, there is no voiceless vowel following the sonorant; rather, there seems to be no interval occupied by the vowel at all.

Chapter 3 presents a probabilistic model of the phonological environment where vowel devoicing occurs. Aside from IP-final devoicing, already discussed in Chapter 2, I find three types of factors that contribute to the likelihood of devoicing: vowel height (high vowels may devoice, non-high vowels do not), prosodic factors (stressed and IP-initial syllables are more likely to contain voiced vowels) and segmental environment (vowels are more likely to devoice *after* sonorants but *before* less sonorous sounds).

In Chapter 4, the acoustics of vowel devoicing is explored in detail. With it already established that these underlying vowels do not appear in a timeslot of their own in the output, the question in this chapter remains whether the vowel is deleted outright, or if some articulation associated with the vowel is still present. The chapter includes acoustic analysis of the segments surrounding "devoiced" medial vowels and compares them to typical consonant-vowel coarticulation (i.e., when the vowel is fully voiced). For each of the three segmental environments analysed (/i/ and /u/ following /s/, /t/ or /n/), acoustic evidence of the underlying vowel was found even in the devoicing environment; for example, /s/ had a lowered spectral peak before underlying /u/ compared to /i/, even though the vowel itself was not spectrographically visible. The finding for the devoicing environment was analogous to consonant-vowel coarticulation: just as underlying /u/ lowered spectral peak of /s/ in the devoicing environment, so did it in the voicing environment. The conclusion, then, is that non-final vowel "devoicing" *is* coarticulation, to the point where the entire vowel is coarticulated with its environment, obscuring it entirely except for its acoustic remnant on adjacent segments.

The phonological patterns identified in Chapter 3 and the phonetic realization of "devoicing" presented in Chapter 4 informed the theoretical account of devoicing presented in Chapter 5. There, I presented an Articulatory Phonology analysis of vowel devoicing where devoicing arises as a result of overlapping gestures: when a short high vowel is overlapped by adjacent consonants, it results in total obscuration of that vowel, producing the phenomenon that has been described as vowel devoicing. A critical component of the theory is the principle of Sonority-Driven Gestural Timing: the precise timing of the vowel gestures to those of its adjacent consonants is modulated by the relative sonority of those consonants. This accounts for the generalization found in Chapter 3: when a vowel *follows* a sonorant, the timing of the vowel is earlier, causing more overlap and therefore a higher likelihood of devoicing; when a vowel *precedes* a sonorant, the timing of the gestures is less close, causing less devoicing.

To conclude the dissertation, the remainder of this chapter addresses some of the major theoretical contributions and methodological considerations of this study. Section 6.2 discusses the theory of Articulatory Phonology and Malagasy vowel devoicing's implication for a theory of phonological representation, arguing against a featural analysis for vowel devoicing. Section 6.3 discusses the viability of sonority-driven timing as a cross-linguistic principle; 6.4 discusses the language-specificity of Malagasy's devoicing process and implications for a theory of perceptual recoverability. Section 6.5 discusses the importance of incorporating phonetic data in answering phonological questions. Finally, Section 6.6 highlights some implications for the present study on the Malagasy grammar and identifies areas for future research.

6.2. Against a feature-based deletion analysis of devoicing

The analysis that I've presented for devoicing assumes gestures as the basic unit of the phonology; devoicing arises when the vowel's gestures undergo complete overlap. The gestural overlap account suits the Malagasy data very nicely, especially once a theory of sonority-driven gestural timing is introduced. However, gestural theories of phonology are far from universally accepted, especially amongst those inclined toward a generative theory of the grammar. Since the early days of Articulatory Phonology, linguists have debated whether phonological features or gestures make up the central representational unit of phonology (e.g., Clements, 1992) and both views have persisted. I won't take a strong stance *against* features as a unit of the grammar; in fact, certain authors have argued for representations that include both features and gestures (e.g., Zsiga, 1997). Readers may wonder, then, why I assume that vowel devoicing arises because of gestural overlap rather than deletion. Not only does a gestural approach neatly capture the Malagasy pattern, it is superior to the featural approach for two reasons: (1) it avoids the complications that come with opacity and (2) it captures the specific timing relations necessary to explain the Malagasy data (i.e., Sonority-Driven Gestural Timing).

6.2.1. Feature-based theories struggle with opacity

Indeed, Malagasy surface forms, when simplified to their IPA transcriptions, look like feature assimilation followed by deletion: underlying /masuka/ is realized as [masʷka]; /lasiʈʐ̞a/ as [lasʲʈʐ̞a]. Looking strictly at the underlying/surface correspondence, it is a classic case of counterbleeding opacity: a phonological process is conditioned by some underlying environment, but that environment is not visible in the surface form. In the Malagasy case, rounding and palatalization are conditioned by underlying /u/ and /i/, respectively, but those vowels are not produced as surface segments. Under rule-based models of the phonology (SPE; Chomsky & Halle, 1968), this type of process is, at first glance, unproblematic if we assume that rules are serially realized: first, the grammar would implement a rule of rounding or palatalization, followed by a rule of unstressed medial vowel deletion (simplified for now), as in (6.1).

(6.1)	/kasuka/	/lasiʈʐ̞a/
Rule 1a: Rounding C → [+round] / _u	kasʷuka	
Rule 1b: Palatalization C → [+high] / _i		lasʲiʈʐ̞a
Rule 2: Vowel deletion (simplified) V[+high] → Ø / C_C	kasʷka	lasʲʈʐ̞a
SR	[kasʷka]	[lasʲtʐ̞a]

However, the analysis in (6.1) is problematic for multiple reasons. First, ordered rules in the style of SPE have generally become dispreferred in favour of constraint-based theories like Optimality Theory (OT; Smolensky & Prince, 1993). OT has proved popular for its ability to capture typological generalizations (e.g., Gordon, 2007), phonological conspiracies (Pater, 1999) and the interaction of markedness and faithfulness constraints (Lombardi, 1999). Because of this, many phonologists operate under constraint-based theories. However, in these constraint-based theories, opaque processes like Malagasy vowel devoicing become problematic. Optimality Theory posits that surface representations correspond to the "optimal" form based on language-specific rankings of universal constraints. To describe the Malagasy devoicing data, then, we could describe two crucial constraints: on the one hand, Malagasy penalizes consonants whose secondary articulations do not agree with the following segment in lip rounding [+/- round] and dorsal features (e.g., [+/- high]), represented with the constraints AGREE[ROUND] and AGREE[HIGH] (Lombardi, 1999). This constraint obligatorily outranks the correspondence constraints IDENT[ROUND] and IDENT[HIGH] (McCarthy & Prince, 1995), which penalize surface segments that do not have the same [round] and [high] features as the input form. The ranking of the AGREE >> IDENT constraints in Malagasy means that speakers will always choose surface forms with coarticulation over forms that are faithful to the underlying form. In (6.2), for example, the input form /masu/ has two possible candidates: (6.2a), which violates the AGREE constraint and (6.2b), which violates the IDENT constraint. Because AGREE >> IDENT, form (6.2b) is the optimal candidate and therefore the form that we observe in Malagasy.

(6.2) /masu/	AGREE[ROUND]	IDENT[ROUND]
a. masu	*!	
☞ b. mas^{w}u		*

On the other hand, Malagasy also shows the apparent deletion of certain unstressed medial high vowels (which, as in (6.1), we can simplify to say that Malagasy deletes unstressed high vowels between consonants). For this, we can say that Malagasy has a constraint NOMEDIALHIGH, which penalizes unstressed high vowels that appear between consonants and outranks the competing correspondence constraint MAX, which penalizes surface forms that contain fewer segments than the input form (i.e., it penalizes deletion). If the Malagasy grammar only had these constraints, and coarticulation was not observed, we would have the constraint violations in (6.3). Here, because the constraint that penalizes medial vowels outranks that which penalizes deletion, the optimal candidate is one that undergoes deletion.

(6.3) /kasuka/	NoMedialHigh	Max
a. kasuka	*!	
☞ b. kaska		*

The problem arises when we attempt to model a grammar where both coarticulation and vowel deletion emerge. In Classical OT, constraint violations are assessed simultaneously, meaning that it is impossible for both the vowel environment that triggers coarticulation to be present in order to satisfy Agree[round] and that same vowel to delete in order to satisfy NoMedialHigh. In the tableau in (6.4), this is demonstrated. The grammar erroneously selects (6.4c), [kaska] as the output of underlying /kasuka/. This is because [kaska] violates neither Agree[round], as both [s] and [k] are [-round], nor NoMedialHigh. However, the actual observed surface form [kas^{w}ka], violates both Ident[round] and Agree[round]; this violation profile arises because the [s^w] in [kas^{w}ka] is neither faithful to the (unrounded) input form /s/, nor does it agree in rounding with the following unrounded [k]. No matter the constraint ranking, the opaque candidate will not be optimal.

(6.4) /kasuka/	Agree[round]	NoMedialHigh	Ident[round]	Max
a. kasuka	*	*		
b. kas^{w}uka		*	*	
c. ☹ kaska				*
d. kas^{w}ka	*		*	*

Early on, Classical OT was criticized for its apparent failure to capture opaque processes like this one (Idsardi, 1998), and since then, variants of OT have been proposed to deal with such cases including Harmonic Serialism (McCarthy, 2000, following Smolensky & Prince, 1993), Stratal OT (e.g., Cohn & McCarthy, 1998) and Phonetic Faithfulness (Kim, 2023).

But what if Malagasy vowel devoicing is not opaque at all? The analysis that I have presented in this dissertation evades discussion of opacity outright since it assumes that the vowel *is* present in the output of words like /kasuka/. Although it is not visible as a separate segment, I have argued that the vowel *gesture* is still articulated, albeit overlapped by adjacent consonant gestures. Multiple stages of derivation are not necessary since a gestural theory can account for both coarticulation and complete overlap simultaneously, as in Tableau 5.3 in Chapter 5. By analyzing Malagasy with gestures, rather than features, we can thus remove Malagasy vowel devoicing from the set of apparent counterbleeding phenomena that pose a challenge to OT.

6.2.2. Features do not capture sonority effects

Perhaps the most novel contribution of this dissertation concerns what I called Sonority-Driven (Gestural) Timing (SDT), adopting a proposal by Gu and Durvasula (2024). In their paper, they put forward the idea that the relative timing between two segments and their associated gestures is determined by the difference in sonority between them. For example, in CC clusters with rising sonority (e.g., /br/), their gestures are further apart than in clusters with falling sonority (e.g., /rb/). Gu and Durvasula showed evidence from CV sequences, following Crouch (2022), who showed a similar effect for CC clusters in onsets. Gu and Durvasula suggest that this principle is generalizable to all segments within a syllable, and possibly across syllable boundaries.

In Chapter 4, I showed that Malagasy vowel devoicing is sensitive to the sonority of adjacent consonants: vowels are *more* likely to devoice when they *follow* a sonorant but *less* likely to devoice when they *precede* more sonorous sounds. In Chapter 5, I showed how the SDT can explain this: in CV sequences, more sonorous consonants drive the C and V gestures closer, producing more overlap and thus a higher likelihood of “devoicing”. In VC sequences, the SDT predicts the opposite effect: a more sonorous C drives the V and C *apart* from each other, decreasing overlap and thus the rate of "devoicing". The account of the SDT is a major asset to the analysis of Malagasy vowel devoicing, and one that is not easily replicated in feature-based theories since it relies on phonological representations that incorporate timing.

To illustrate this, let's imagine a simplified system where sonority is a binary feature [+/-sonorant]. Table 6.1 shows the percentage of high vowels that underwent devoicing in unstressed, IP-medial syllables, replicating the data Table 5.3. /r/ is excluded for the variation in its realization.

Preceding	**Following**	**% Devoiced**
Obstruent	Sonorant	40.30
Sonorant	Sonorant	91.03
Obstruent	Obstruent	94.24
Sonorant	Obstruent	96.57

Table 6.1 Rate of vowel devoicing by consonantal environment.

Going back to our simplified (non-opaque) grammar in (6.3), we could posit two additional constraints in the grammar, reflecting the general tendencies for vowels to not delete (a) after obstruents and (b) before sonorants.

(a) MAX AFTER OBSTRUENT: do not delete vowels after obstruents
(b) MAX BEFORE SONORANT: do not delete vowels before sonorants

These constraints, which mirror the statistical finding in Chapter 4, cannot capture the full range of environments where deletion does (not) occur in multiple ways: it incorrectly predicts that vowels are voiced between two obstruents (since the grammar penalizes deletion after obstruents) and also between sonorants (since it penalizes deletion before sonorants). This is shown in Tableau 6.1.

UR		MAX AFTER OBSTRUENT	MAX BEFORE SONORANT	NOMEDIALHIGH
/t_n/	☞ **Voiced**			*
	Deleted	*	*!	
/n_t/	Voiced			*!
	☞ **Deleted**			
/t_k/	☹ **Voiced**			*
	Deleted	*!		
/n_n/	☹ **Voiced**			*
	Deleted		*!	

Tableau 6.1 Tableaux for vowel "devoicing" if it is seen as vowel deletion constrained by MAX constraints penalizing vowel deletion following obstruents or before sonorants. These constraints predict the incorrect output for underlying /t_k/ and /n_n/.

Instead, we would need an ad hoc constraint (6.5) to specifically avoid deletion in the environment /t_n/, which would correctly predict the deletion pattern, shown in Tableau 6.2.

(6.5) MAX T_N: do not delete vowels when they appear both after an obstruent and before a sonorant.

UR		MAX T_N	NOMEDIALHIGH
/t_n/	☞ **Voiced**		*
	Deleted	*!	
/n_t/	Voiced		*!
	☞ **Deleted**		
/t_k/	Voiced		*!
	☞ **Deleted**		
/n_n/	Voiced		*!
	☞ **Deleted**		

Tableau 6.2 Tableaux for vowel "devoicing" if it is seen as vowel deletion that does not apply between an obstruent and a sonorant, specifically.

This type of constraint is highly specific and so it has no phonetic motivation in and of itself, as opposed to the gestural account of the SDT which has been shown to be true of gestural timing in at least two other languages (English CV sequences and Georgian CC clusters). Not only that, but the grammar presented in Tableau 6.2 is a simplified version of the Malagasy pattern, as it collapses all manners of articulation into a binary sonority feature. In order to capture the observed effect that the rate of vowel devoicing is scaled to the sonority scale (e.g., where more deletion occurs before stops than fricatives), the grammar would require individual constraints for each manner of articulation. Additionally, the grammar would need to account for the varying probability of deletion, since the probability of deletion is never categorical in unstressed syllables (and, in the case of the T_N environment, nearly a tossup).

6.2.3. A featural analysis falls short

In sum, the theoretical machinery required to account for the patterns of Malagasy vowel devoicing is not easily found under a featural theory of phonology. I've highlighted here that such an analysis would require at least two parts:

1. A serial theory of phonology that can capture opacity
2. Ad hoc constraints to replicate the effect of sonority-driven timing

In contrast, the gestural analysis presented in the dissertation neatly accounts for the Malagasy data:

1. Opacity is avoided as the vowel is present (as a gesture) in the output
2. The principle of Sonority-Driven Gestural Timing accounts for the rate of devoicing based on a sonority scale

6.3. On the universality of Sonority-Driven Timing

As I have already highlighted in this chapter, an important part of the story of Malagasy vowel "devoicing" is that gestural timing is affected by the sonority of the segments involved. Gu and Durvasula (2024), in their original proposal of Sonority-Driven Gestural Timing, suggest that the SDT may be universally applicable to all segmental sequence types within the syllable (e.g., CV, VC, etc.) and perhaps across syllable boundaries (as VC sequences are in Malagasy). Gu and Durvasula describe the SDT as a single "force" that, all else being equal, should apply to all types of segmental sequences unless it is in conflict with some other force driving gestural timing. In the present work, this is analogous to the formalization of the SDT as constraints (in this case, distinct constraints governing CV and VC timing) that may be outranked by other gestural-timing constraints, resulting in gestures being timed otherwise.

It is important that sonority-driven timing be implemented as a violable constraint to account for language-specific (non-)implementations of the principle. Indeed, certain phonological phenomena are apparently in conflict with the predictions of the SDT. *Vowel intrusion* is a phenomenon where a vowel appears between two adjacent consonants whose gestures are sufficiently far apart, leaving a gap that is filled by the vowel (Hall, 2006; the Moroccan Colloquial Arabic case described in Section 5.2.2 is one such case). One specific type of vowel intrusion involves a "copy vowel"; i.e., the identity of the intrusive vowel is the same as the vowel in an adjacent syllable. The gestural explanation for copy vowels is that a consonant gesture may be timed so that the adjacent vowel's gesture extends *beyond* the boundaries of the consonant, being heard on the other side. Figure 6.1 shows the gestural schematic for this phenomenon for an underlying VC_1C_2 sequence, which is heard as $[VC^VC]$. In this figure, we can see that the relatively early timing of C_1 and the distance between C_1 and C_2 results in V partially intervening between the consonants.

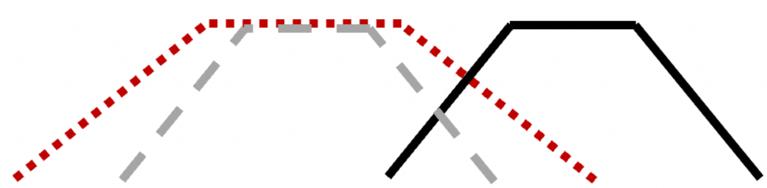


Figure 6.1 Gestural schematic for vowel intrusion in a VC_1C_2 sequence, where V is represented as a dotted red line, C_1 as a dashed grey line and C_2 as a solid black line.

However, Hall (2006) notes an interesting typological generalization with respect to copy vowels: they only appear next to sonorants and gutturals. That is, in VC_1C_2 sequences, C_1 must be a sonorant or guttural for the copy vowel to appear; in C_1C_2V sequences, C_2 must be a sonorant or guttural. Hall gives the example of Finnish /kalvo/ 'transparency', which surfaces as [kalavo]; the /a/ is copied across the sonorant /l/. For C_1C_2V sequences, the copy vowel phenomenon is compatible with that of sonority-driven timing: if C_2 is sonorous, the SDT predicts that V will be timed relatively early, increasing the likelihood of the vowel gesture intervening between the two consonants. For VC_1C_2 sequences, however, the prediction is the opposite of what is observed: the SDT predicts that a more sonorous C_1 will drive that consonant *away* from the vowel, and yet it must be substantially overlapped by it for the copy vowel to appear, as in Figure 6.1.

Whereas the copy vowel phenomenon is consistent with the predictions of the SDT for CCV sequences, it is not immediately clear why this is not the case for VCC sequences. As the present study is the first to look at sonority-related gestural timing for VC sequences, it is impossible to estimate the degree to which the SDT is cross-linguistically viable; it may be the case that CC clusters behave differently, or that Malagasy is unique in applying the SDT to VC sequences. In any case, the fact that sonority-driven timing may not be applicable to some segmental sequences in some languages justifies the implementation of the SDT not only as ranked, violable constraints within a larger gestural grammar, but also as distinct constraints for different sequences.

6.4. Conflicting priorities in the grammar

Conceptualizing gestural timing as driven by violable constraints can also explain the language-specific ways in which Malagasy vowel "devoicing" plays out, compared to languages also analyzed to have vowel devoicing caused by gestural overlap. Additionally, for many other languages, vowel devoicing or deletion is not reported at all, and even unstressed high vowels emerge as full, voiced segments. On the one hand, this is likely partially attributed to the inherent phonetic differences between languages. Malagasy frequently aspirates its stops, especially before high vowels (Howe, 2019; also discussed in Chapter 4), which increases the degree of overlap with the following vowel. This is similar to Korean, where consonants with longer voice onset times (VOT) cause higher rates of devoicing (Jun et al., 1997). But in most dialects of Spanish, whose stops have short-lag VOT (Dmitrieva et al., 2015), non-final devoicing is not reported, perhaps because those stops are not long enough to completely overlap adjacent vowels.

In other cases, though, the presence or absence of devoicing may be under the control of the grammar. For example, it is not clear why devoicing happens after sonorants in Malagasy but not in Japanese or Korean. If, as detailed above in Section 6.3, we take gestural timing to be modulated by constraints like Sonority-Driven Gestural Timing, we can posit that those constraints are in competition with other constraints in the grammar; in Malagasy, the SDT is highly ranked, resulting in extensive overlap between, for example, a sonorant consonant and the vowel that follows it. In other languages, the SDT constraints may be outranked by other constraints, particularly ones that prioritize recoverability of gestures, that result in less overlap than dictated by the SDT.

For example, a language may prohibit devoicing after nasals because of perceptual constraints: for nasals, speakers need more than just the spectral characteristics of the nasal alone (i.e., murmur) to perceive its place of articulation, relying also on transitions in and out of adjacent vowels (Kurowski & Blumstein, 1984). For coarticulated nasals (e.g., a labialized alveolar nasal), the spectral effects may be even more minimal; in fact, labialized nasals are rare amongst the world's languages,[3] indicating that coarticulated nasals may not be perceptually stable. In cases of complete CV overlap involving a nasal, then, the vowel is not recoverable just as a coarticulated

[3] There are apparent exceptions, including Degema (Kari, 2004), which has phonemic [ŋʷ].

gesture on the nasal, and so languages like Korean may avoid devoicing in that environment (presumably either by adjusting gestural timing or duration).

This highlights a balancing act that languages make: that between articulatory efficiency and perceptual recoverability. Gestural studies of phonology have long integrated the idea of recoverability, the idea that speakers produce utterances in a way that allows all relevant phonological information to be perceived (Chitoran et al., 2002). Under Articulatory Phonology, increasing overlap between gestures is more efficient, as information can be transmitted more quickly; decreasing overlap, on the other hand, increases recoverability, and speakers can make better use of acoustic cues.

In essence, the Malagasy speaker prioritizes efficiency and so constraints like SDT(CV) (from Chapter 5) are highly ranked; speakers of a language like Korean prioritize recoverability, and so some other constraint (let's say RECOVER-VOWEL) outranks the SDT, driving the vowel gesture apart from a nasal onset so that the vowel can be perceived. For languages that disallow any vowel devoicing whatsoever, there may be an even stronger constraint requiring that each vowel be heard as a full, distinct vowel segment. The Malagasy speaker, however, prioritizes the articulation of a *gesture* over the perceptual target. However, as noted in Chapter 1, non-final vowels are produced in slow, careful speech, indicating that the relative constraint ranking efficiency and recoverability may be flipped in different speech rates or styles.

6.5. Phonetic detail in phonological analyses

One of the empirical goals of this dissertation was to determine the precise phonetic realization of a process variably described as *vowel devoicing* or *deletion* in the literature. While grammars and descriptions of Malagasy frequently make reference to some process affecting certain vowels (e.g., Dahl, 1952; Howe, 2019), the descriptions lack an instrumental phonetic basis, instead relying on the ears of the authors themselves. This is an adequate and suitable methodology for field linguistics: not only is documenting the phonetic and phonological system of a language laborious work (for which detailed instrumental study of each phenomenon would only delay publication into perpetuity), it serves its purpose for language learners as well as linguists like myself who use these grammars as a point of departure. In a thorough investigation of a phenomenon like vowel devoicing, though, detailed phonetic detail proved helpful in refining the analysis from a phonological perspective.

In recent decades, the field of Laboratory Phonology has aimed to bridge the gap between phonetic methodologies and phonological theory (as the name implies), and a number of influential papers have been successful in doing so. Work in laboratory phonology has led to discoveries in incomplete neutralization (e.g., Port & O'Dell, 1985), language-specific phonetics (Cho & Ladefoged, 1999), intonational phonology (Jun, 2005; Ladd, 2008), among many others. The motivation is straightforward: if phonology is the study of speech sounds as abstract units of the grammar, we should be able to make inferences about that grammar based on real speech data (i.e., experimental phonetic data).

In my dissertation, then, I used experimental phonetic methods to uncover the phonological process underlying this process that has been called vowel "devoicing". If these disappearing vowels are *deleted*, *devoiced* or overlapped, this has straightforward implications for the phonetic data. In Chapter 2, I showed that "devoicing" was not an adequate term for what happens to Malagasy vowels in non-IP-final positions, as there is no vowel occupying a segmental interval in the "devoicing" environment, much less a voiceless vowel. However, simple "deletion" also fails to account for the acoustic situation: in Chapter 4, I showed that there are acoustic remnants of the underlying vowel even in cases of "devoicing". For example, for underlying /su/, the /s/ has a lowered spectral peak, consistent with typical coarticulation with /u/. Indeed, the phonetic output of "vowel devoicing" mirrors consonant-vowel coarticulation, indicating that this is an extreme case of coarticulation, to the point of total overlap of the vowel by adjacent consonants. This experimental evidence, combined with the distributional properties described in Chapter 4, led to the gestural overlap account of vowel "devoicing" presented in Chapter 5. This research, then, highlights the value in the laboratory phonology approach: experimental acoustic data can provide important insights into the data that inform our theory.

6.6. Future directions

In this dissertation, I have presented a detailed phonetic account of the phenomenon of vowel "devoicing" in Merina Malagasy alongside statistical and phonological modelling. The analysis in Chapter 5 hinges on the idea of *gestures* as a phonological primitive that are manipulated by the grammar; in particular, non-final vowel "devoicing" is generated by extreme overlap of the vowel by its consonantal environment to the point of total obscuration of that vowel. The precise environment where vowel devoicing emerges was said to be caused by multiple factors within the grammar, including the idea of Sonority-Driven Gestural Timing, where the timing of two gestures depends on the difference in sonority between them.

6.6.1. Instrumental articulatory study

The most natural next step for future research should be to demonstrate the predictions of the phonological analysis using instrumental articulatory measurements. For example, I argue that the propensity for vowel devoicing following sonorants is explained by greater gestural overlap of a vowel following a sonorant compared to an obstruent. This is empirically testable using technology such as real-time MRI where the articulators can be traced to model the precise timing of gestures. This sort of research will be especially valuable with respect to VC timing; as described in Section 6.3, the application of sonority-driven timing to VC syllables is novel and potentially in conflict with the results of other phonological studies; it would thus be extremely valuable to see further research into VC gestural timing, even for languages other than Malagasy.

6.6.2. Syllable structure

As described in Chapter 1, Malagasy is traditionally analyzed as having a strict (C)V syllable structure; no codas or consonant clusters (with the pre-nasalized obstruents and affricates) are permitted in native Malagasy words. The acoustic reality of vowel devoicing means that consonants regularly appear next to each other in fluent speech, giving the appearance of codas and clusters; for example, *rakotra* 'cover' is pronounced [ˈrakʷʈʂa] and . For *rakotra* to be syllabified as [ˈrakʷ.ʈʂa] assumes that the vowel is deleted entirely from the output; however, as argued for in Chapter 5, a vowel *gesture* is present in the output, even if not realized as its own segment. The result is that there is no necessary requirement for resyllabification of these Malagasy words, as the theory dictates that the Malagasy speaker is still aiming toward the articulatory goal of CV syllables. This ties into earlier discussion in Section 6.4 that Malagasy speakers have articulatory targets rather than recoverable acoustic targets. However, future work on vowel devoicing in Malagasy should confirm whether syllabification is affected by vowel devoicing.

6.6.3. Tonogenesis

Central dialects of Malagasy, including Merina, are undergoing tonogenesis, where the voicing contrast in obstruents has largely been replaced by a contrast in pitch (Howe, 2017). This is surprising, given that vowel devoicing is so prevalent in these dialects; without a voiced vowel following a voiceless consonant, the tonal contrast cannot be realized and pairs like *manavika* /maˈnavika/ 'cling to a tree' and *manafika* /maˈnafika/ 'invade' approach neutralization (to [manafʲkʲa]). In the typology of devoicing, vowels are less likely to be devoiced when they are in a tone-bearing syllable (Gordon, 1998), and so it remains to be seen how Malagasy speakers accommodate this. On the one hand, Malagasy speakers may completely neutralize the contrast in these unstressed syllables where vowels undergo “devoicing”, the tonal contrast only remaining in syllables with voiced vowels. Alternatively, it is possible that tones associated with devoiced vowels are produced on an adjacent syllable. Future research should look at the relationship between vowel devoicing and tonogenesis to investigate how Malagasy speakers maintain contrasts.

6.6.4. Sound change

Malagasy non-final vowel devoicing is apparently a recent innovation, only appearing in grammars beginning in the 1950s (Dahl, 1952). Already, I have discussed how the Malagasy pattern is not phonologically stable, as certain contrasts are minimally recoverable (see Section 6.4 above). This poses a problem for the Malagasy-learning child, who must learn to produce articulatory units even when they cannot hear them. It is logical, then, to assume that some future generation of Malagasy speakers may acquire a rule of vowel deletion, rather than overlap or devoicing. Ohala (1981) suggests that phonetic ambiguity can lead to sound change as listeners must hypothesise on their own the origin of the ambiguity. The Malagasy learner simultaneously hears the underlying vowels in the alternations described in Chapter 1.3.2, but their near absence in some syllables, like the /u/ in *vanona* /ˈvanuna/. Then, the learner must decide whether to attempt an articulatory gesture anyway or to delete the vowel altogether. Meneses and Albano (2015) suggest that vowel deletion is a natural next step for “devoicing” processes like in Malagasy, and I agree. While the speakers in this study have not yet arrived at that stage, it is something that future researchers of Malagasy vowel devoicing should look out for.

Appendix A
List of stimuli used in Chapter 3 dataset

Afaka gasina ve ny fanontaniana?
Afaka kelikely, Hanodidina ny atidoha ny ranoka.
Ahatokiana ilay lehilahy.
Anananareo ilay maso sahala.
Apetraho ao anaty alokaloka ny sarona metaly.
Aza akombona ny molotrao!
Aza atsontsorika ny lambanao fa manintona ny lalitra.
Aza fokarina ilay miaramila!
Aza mifoka ny sisa fa tsy mahasalama ny fototra.
Aza mikarama anio tontolo andro!
Fantatro fa nividy basy tan'ny fivarotana i Rakoto.
Farany dia sitrana ihany ny lohaliny.
Fokarin'ny dokotera ilay miaramila.
Fokarina ilay miaramila.
Hijerevako ny kintana ny masolavitro.
Ianareo dia monina ao amin'ny trano sahala.
Ikaky dia mitomany fa midongy ny reniko.
Ilay miaramila no fokarin'ny dokotera.
Ilay miaramila no fokarina.
Ilay mpiasa no mikarama anio tontolo andro.
Ilay mpiasa no mikarama.
Inay! Toa silahina ny sima-kazo.
Irahina ny irakiraka fa finaritra izy.
Isaina ny sarony.
Iza no mikarama? Itỳ mpiasa itỳ no mikarama.
Madio ilay ati-trano lehibe.
Madio ilay ati-trano.
Madio ny trano andavanandro.
Madio ve ny trano? Nahita ny famafa aho.
Madio ve ny trano? Omaly ve no nodiovina?

Mamatonalina dia matetika kokoa ny fandosirana.
Maminania ny taonako!
Maminany aho fa maloka ny andro.
Mamory vato aho.
Manadio ny trano andavanandro izy.
Manadio ny trano izay aho.
Manadio ny trano ny mpiasa.
Manadio trano Sariaka.
Manafina hanina sy ronono matimaty anatin'ny trano ilay lehilahy.
Manakotsà ny lamba ianao!
Mananika havoana ny osy mba ahazoany aloka.
Manasitrana ilay lohalika ny vato masina.
Manasitràna ilay maso ianao!
Mantsina sy vasoka ny ranovola.
Miatoka sy mikimokimoka ilay lehilahy fa sina izy.
Mifantina satroboninahitra sy sikim-bavy ilay malagasy.
Mijery ny fiaramanidina amin'ny masolavitro aho.
Mikarama anio tontolo andro ilay mpiasa.
Mikarama ilay mpiasa.
Mikaroka basy ao trano ilay ampakarina.
Mikatona ny fonenana mandritra ny herinandro.
Mikorokoro ilay efi-trano.
Milaza ilay miaramila fa sinambotra ny tilikambo.
Milaza ny sombinaiko fa soratana sy soniavina ilay taratasy.
Mino aho fa fanina tokoa ilay lehilahy satria mamo izy.
Mino aho fa fokarina ilay miaramila.
Mino aho fa misongadina ilay vehivavy sosotra fa tiany ny fitokanana.
Mino aho fa somary tomady ilay varika.
Mino aho fa sosotra ilay vehivavy.
Mino aho fa sosotra tokoa ilay vehivavy.
Mino aho fa vorina ny vato.
Mino ilay mpiambina vanona fa misy kanonkanona analinalina miafina ao anaty fasika lalina.
Mino izy ireo fa mora sosorina ny tatsinanana.

Misy maso anankiray sitrana.
Namidinao ilay basy.
Nanadio ny trano omaly aho.
Nanasitrana ny maso ilay lehilahy.
Nanasitrana ny maso ny dokotera.
Nanasitrana ny maso Sariaka.
Nanome toky ahy ilay mpampianatra.
Nanonofy nitoletsika ao anaty ony aho.
Ndao isika hidinidinika ny asa miandry antsika rahampitso.
Nianatra ny fanekena tany an-tsekoly izahay.
Niditra an-tsokosoko tao an-trano izy.
Nihantona ny foko noho ilay tafika.
Nikarama ilay mpiasa omaly.
Nikarama itỳ mpiasa ity omaly.
Nikarama omaly ilay mpiasa.
Nilaza i Mamy fa nosainina ilay fampakaram-bady.
Nilaza i Mamy fa nosaininao ilay fampakaram-bady.
Nilaza i Mamy fa notolorana ny taratasy.
Nilaza i Rakoto fa tsikaritra ny sokina.
Nitifirako ny vorona ilay basy. Omaly no novinidiko izy.
Nitifirana ny vorona ilay basy omaly.
Nitifirana ny vorona ilay basy. Oviana?
Nitoreo ny saka.
Nividy basy omaly aho.
Nodiovin'i Rakoto ny trano. Izy no nanadio ny trano.
Nodiovina ve ny trano? Satria mbola maloto izy.
Nofenoina ary nesorina ilay vera.
Nosinitrana ny maso afakomaly.
Nosinitrana ny maso omaly.
Notresahina sy nokekerina ilay voasary.
Ny fanatitra dia ho an'ny fokonolona.
Ny fantisny no isandratana ny tolàna.
Omaly alina dia niasa aho.

Orohana sy tonena ny vinantovaviko.
Oviana no sitranina ny tsinay?
Raha mandentika ny fantsika aho, tsy mitsinkafona izy ireo.
Raha miantsinanana ianareo, mazàna miroky ny olona.
Raketina sira betsaka ka siraina ilay hena.
Sosotra ilay vehivavy.
Sosotra tokoa ilay vehivavy.
Tolorana ny vakoka sy ny vatosoa.
Tovanana ny tsatsika fa lehibe ilay jomòka.
Tsy ananako ilay basy. Avia mikaroka azy.
Tsy ananako ilay basy. Ilay lehilahy no manana azy.
Tsy ananako ilay basy. Naly no manana azy.
Tsy ananako ilay basy. Sariaka no manana azy.
Tsy fokarin'ny fokonoloko ny rikarikan-taolana.
Tsy madio izany trano izany.
Tsy madio na ny trano na ny zaridaina.
Tsy manana basy any izy.
Tsy manana basy ilay lehilahy.
Tsy manana basy Naly.
Tsy manana basy sandoka aho.
Tsy manana basy Sariaka.
Tsy manana izany basy izany aho.
Tsy manana na basy na bala aho.
Tsy mikasa handamina ny afokasoka ilay vehivavy.
Tsy mitomany ilay lehilahy ory.
Tsy nikasa mangala-tsonia ilay iraka.
Tsy sitrana itỳ maso itỳ.
Tsy sitrana na ilay maso na ilay tanana.
Tsy sitrana ny masoko.
Tsy sitrantsitrana ny maso.
Tsy tiako ny fikambanana fa naleoko nipetraka tany ambanivohitra.
Vaky ny solomasoko sy fitaratro.
Vetivety foana dia fokarina ilay miaramila.

Vorina ny vato.

Appendix B
List of stimuli used in Chapter 4 Task One

The following is the list of Malagasy words used in Task One in Chapter 4. Most words were repeated more than once by each participant.

Target vowel	Environment	Malagasy	English
i	s_k	afokasika	match
a	t_k	ahataka	TT.separate
a	n_k	anaka	child
u	n_k	anoka	contentment in eating/drinking
u	t_k	atoka	nod
u	t_k	behatoka	type of rice
i	s_k	dadasika	wide and level
i	t_k	dipatika	type of shrub
a	n_k	fanaka	furniture
i	s_k	fasika	sand
i	t_k	fatika	thorns
i	n_k	fidanika	manner of heating
a	t_k	hataka	request
u	t_k	hatoka	nape of neck
i	s_k	kasika	action of touching
u	s_k	kasoka	act of rubbing
u	s_k	kasokasoka	rustling sound
u	n_k	lanoka	tired
a	s_k	mahamasaka	AT.cook until well done
u	s_k	manakasoka	AT.rub
a	s_k	manasaka	AT.divide into halves
a	t_k	manataka	AT.split open
u	n_k	mandranoka	AT.sick
a	t_k	mangataka	AT.ask for something
a	s_k	masaka	cooked
a	n_k	mianaka	AT.be together as parents
i	n_k	mianika	AT.climb
u	n_k	mianoka	be slow
i	t_k	miatikatika	pose
u	t_k	miatoka	AT.to nod

i	s_k	midadasika	spacious
i	t_k	mifanatikatika	AT.tickle
a	t_k	mihataka	AT.separate onself
i	s_k	mikasika	AT.touch
u	s_k	mikasoka	rub against
u	s_k	mipasoka	AT.iron clothing
a	s_k	misasaka	AT. be divided into halves
i	n_k	mitanika	AT.cook
a	t_k	mitataka	AT.be split open
a	t_k	mpangataka	beggar
u	s_k	nipasoka	TT.iron
u	s_k	pasoka	smoothness
u	n_k	ranoka	liquid
i	n_k	tanika	cooking
a	n_k	taranaka	posterity
a	t_k	tataka	a slit
i	s_k	torapasika	beach
u	t_k	tsatoka	stability
u	s_k	vasoka	discoloured
i	s_k	voakasika	touched
a	n_k	zanaka	offspring

Appendix C
List of stimuli used in Chapter 4 Task Two

Target vowel	**Environment**	**Word**
a	s_k	Harozasàka
a	s_k	Mbarasàka
a	s_k	Vanasàka
i	s_k	Harozasìka
i	s_k	Mbarasìka
i	s_k	Vanasìka
u	s_k	Harozasòka
u	s_k	Mbarasòka
u	s_k	Vanasòka
a	t_k	Harozatàka
a	t_k	Mbaratàka
a	t_k	Vanatàka
i	t_k	Harozatìka
i	t_k	Mbaratìka
i	t_k	Vanatìka
u	t_k	Harozatòka
u	t_k	Mbaratòka
u	t_k	Vanatòka

Appendix D
Praat script for measuring spectral peak and moments

```
# This script is a modification of Christian DiCanio's spectral moments
script (copyright below).
#It calculates the spectral moments and additionally extracts the spectral
peak (FM) within a
#specified frequency range.
#It does this by iterating through frequency bins within the defined range
and identifying the bin
#with the greatest amplitude.

# The script also differs from DiCanio's original script in a few ways:
#     The script allows you to identify a number of subintervals into which
the fricative is
# divided and the number and size of the windows for each interval.
#     The script is set to measure the entire duration of each interval, with
no 10% margins

# Dicanio's original comments:

#Praat script which produces the first four spectral moments from fricative
spectra. The DFTs are
#averaged using time-averaging (Shadle 2012). The fricative signals should
not be upsampled, so if
#the signal is sampled at under 44.1 kHz, please adjust the Resampling rate
to match that of the
#original recording. Within time-averaging, a number of DFTs are taken from
across the duration of
#the fricative. These DFTs are averaged for each token and then the moments
are calculated. The
#analyzed duration of the fricative is always equivalent to the center 80% of
the total duration,
#cutting off the transitions.

#Note that the duration of these DFTs (window size) multiplied by the number
of DFTs (window number)
#should equal a value no greater than 1.6 times the duration of the sound
file to be analyzed.
#In other words, if you set the window size to 15 ms and the window number to
6, the sum of all
#window durations is equal to 90 ms. This number of windows works fine for a
total fricative duration
#down to 56 ms, but no lower. The reason for this is that the windows should
overlap only up to 50%
#over the duration of the fricative. If they overlap more than this, certain
parts of the duration
#(in the center) get more heavily weighted than others. The value of 1.6
derives from the fact that
#the analyzed duration is only 80% of the duration of the total. Thus, the
sum of the windows may
#only be twice of this 80% value, or 160% of the total duration.
```

```
#Copyright 2013, Christian DiCanio, Haskins Laboratories & SUNY Buffalo.
Please cite this script
#if it is used in a publication or presentation. Special thanks to Christine
Shadle for suggestions
#and troubleshooting.

#Version 2.0 revised 2017 to fix an issue with intensity averaging. Thanks to
Ting Huang at
#Taiwan Tsing Hua University for pointing this error out to me.

#Version 3.0 revised in 2021 to fix some minor issues with sampling rate.

#Version 4.0 revised in 2021 with extensive help by Wei Rong Chen and
Christine
#Shadle at Haskins Labs. The newest version applies an improved function
#in calculating the amplitude of frequency bins and does not utilize Praat's
built-in
#functions for spectral moments, but calculates them using methods discussed
in
#Forrest, K., Weismer, G., Milenkovic, P. & Dougall, R. N. (1988) Statistical
analysis of word-initial
#voiceless obstruents: preliminary data, Journal of the Acoustical Society of
America, 84(1), 115â€“123.

orm Time averaging for fricatives
   sentence Directory_name:
   sentence Interval_labels
   sentence Log_file
   positive Labeled_tier_number 4
   positive Lexical_tier_number 1
   positive Resampling_rate 44100
   positive Window_number 3
   positive Window_size 0.015
   positive High_pass_cutoff 300
   positive Number_of_subintervals 3
   positive Fm_low_freq 3000
   positive Fm_high_freq 7000
endform



# Convert interval_labels to a string array and count labels
interval_labels$ = interval_labels$
number_of_labels = 0
label_length = length(interval_labels$)
word_start = 1
for i to label_length
    if mid$(interval_labels$, i, 1) = " " or i = label_length
        if i = label_length and mid$(interval_labels$, i, 1) <> " "
            i = i + 1
        endif
        number_of_labels = number_of_labels + 1
        word_end = i - 1
        label'number_of_labels'$ = mid$(interval_labels$, word_start,
word_end - word_start + 1)
        word_start = i + 1
    endif
endfor
```

```
Create Strings as file list... list 'directory_name$'/*.wav
num = Get number of strings

# Write header to include interval content, lexical information, FM, etc.
fileappend 'directory_name$''log_file$'.txt
file'tab$'interval_content'tab$'lexical_content'tab$'
for sub_int from 1 to number_of_subintervals
    fileappend 'directory_name$''log_file$'.txt
start'sub_int''tab$'duration'sub_int''tab$'intensity'sub_int''tab$'cog'sub_in
t''tab$'sdev'sub_int''tab$'skew'sub_int''tab$'kurt'sub_int''tab$'fm'sub_int''
tab$'
endfor
fileappend 'directory_name$''log_file$'.txt 'newline$'

for ifile to num
    select Strings list
    fileName$ = Get string... ifile
    Read from file... 'directory_name$'/'fileName$'
    soundID1$ = selected$("Sound")
    Resample... resampling_rate 50
    soundID2 = selected("Sound")
    Read from file... 'directory_name$'/'soundID1$'.TextGrid
    textGridID = selected("TextGrid")
    num_labels = Get number of intervals... labeled_tier_number

    for i to num_labels
        select 'textGridID'
        label$ = Get label of interval... labeled_tier_number i

        # Check if the current label is in our list of labels to analyze
        label_match = 0
        for j to number_of_labels
            if label$ = label'j'$
                label_match = 1
            endif
        endfor

        if label_match
            interval_content$ = label$
            intvl_start = Get starting point... labeled_tier_number i
            intvl_end = Get end point... labeled_tier_number i
            durval = intvl_end - intvl_start

            # Get lexical content
            lexical_interval = Get interval at time... lexical_tier_number
intvl_start
            lexical_content$ = Get label of interval... lexical_tier_number
lexical_interval

            fileappend 'directory_name$''log_file$'.txt
'fileName$''tab$''interval_content$''tab$''lexical_content$''tab$'

            # Calculate sub-interval duration
            sub_interval_duration = durval / number_of_subintervals

            for sub_int from 1 to number_of_subintervals
```

```
                sub_start = intvl_start + (sub_int - 1) *
sub_interval_duration
                sub_end = sub_start + sub_interval_duration

                # Apply the 10% margin to each sub-interval
                #threshold = 0.1 * sub_interval_duration
                #domain_start = sub_start + threshold
                #domain_end = sub_end - threshold

                select 'soundID2'
                Extract part... sub_start sub_end Rectangular 1 no
                intID = selected("Sound")
                select 'intID'
                Filter (stop Hann band)... 0 high_pass_cutoff 1
                intID2 = selected("Sound")
                d1 = Get total duration
                d2 = ((d1-window_size)*window_number)/(window_number-1)
                margin = (window_size - (d2/window_number))/2
                end_d2 = (sub_end-margin)
                start_d2 = (sub_start+margin)

                chunk_length = d2/window_number
                window_end = (chunk_length)+margin
                window_start = window_end-window_size
                bins = round(((resampling_rate/2)*window_size)+1)
                bin_size = (resampling_rate/2)/(bins - 1)
                Create TableOfReal... freqs 1 bins
                freqs = selected("TableOfReal")
                Create TableOfReal... avs 1 bins
                averages = selected("TableOfReal")
                Create TableOfReal... mag window_number bins
                magnitudes = selected("TableOfReal")
                Create TableOfReal... reals window_number bins
                real_table = selected("TableOfReal")
                Create TableOfReal... imags window_number bins
                imag_table = selected("TableOfReal")
                offset = 0.0001

                Create Table with column names: "table", window_number,
"int.val"
                int_table = selected("Table")

                for j to window_number
                    window_end = (chunk_length*j)+margin
                    window_start = window_end-(window_size + offset)
                    select 'intID2'
                    Extract part... window_start window_end Hanning 1 yes
                    chunk_part = selected("Sound")

                    intensity = Get intensity (dB)
                    select 'int_table'
                    Set numeric value: j, "int.val", intensity
                    select 'chunk_part'

                    To Spectrum... no
                    spect = selected("Spectrum")

                    for k to bins
```

```
                    select 'spect'
                    freq = Get frequency from bin number: k
                    select 'freqs'
                    Set value... 1 k freq
                    select 'spect'
                    real = Get real value in bin... k
                    real2 = real^2
                    select 'real_table'
                    Set value... j k real2
                    select 'spect'
                    imaginary = Get imaginary value in bin... k
                    imaginary2 = imaginary^2
                    select 'imag_table'
                    Set value... j k imaginary2
                    select 'magnitudes'
                    Set value... j k real2+imaginary2
                endfor
                Create Table with column names: "table", window_number,
"dsmfc"
                Set numeric value: 1, "dsmfc", 92879
            endfor

            select 'int_table'
            Extract rows where column (text): "int.val", "is not equal
to", "--undefined--"
            int.rev.table = selected("Table")
            int = Get mean: "int.val"

            for q to bins
                select 'magnitudes'
                mag_ave = Get column mean (index)... q
                select 'averages'
                Set value... 1 q mag_ave
            endfor

            start_bin = ceiling(high_pass_cutoff/bin_size)
            select 'averages'
            Extract column ranges... 'start_bin':'bins'
            new_aves = selected("TableOfReal")
            select 'freqs'
            Extract column ranges... 'start_bin':'bins'
            new_freqs = selected("TableOfReal")
            select 'new_aves'
            To Matrix
            ave_mat = selected("Matrix")
            sum_mat = Get sum

            for x to (bins-start_bin+1)
                select 'ave_mat'
                val_x = Get value in cell: 1, x
                Set value: 1, x, val_x/sum_mat
            endfor

            cog = 0
            for b to (bins-start_bin+1)
                select 'new_freqs'
                f = Get value: 1, b
                select 'ave_mat'
```

```
                p = Get value in cell: 1, b
                cog = cog+(f*p)
            endfor

            l2 = 0
            l3 = 0
            l4 = 0
            for c to (bins-start_bin+1)
                select 'new_freqs'
                f = Get value: 1, c
                select 'ave_mat'
                p = Get value in cell: 1, c
                l2 = l2+((f-cog)^2) * p
                l3 = l3+((f-cog)^3) * p
                l4 = l4+((f-cog)^4) * p
            endfor

            sdev = sqrt(l2)
            skew = l3/(l2^(3/2))
            kurt = (l4/(l2^2))-3

            # Calculate FM
                          fm_max_amp = -1000
                          fm = 0
                          for d to (bins-start_bin+1)
                                select 'new_freqs'
                          f = Get value: 1, d
                          if f >= fm_low_freq and f <= fm_high_freq
                                select 'ave_mat'
                                amp = Get value in cell: 1, d
                                if amp > fm_max_amp
                                fm_max_amp = amp
                                      fm = f
                                endif
                          endif
                          endfor

            fileappend 'directory_name$''log_file$'.txt
'sub_start''tab$''sub_interval_duration''tab$''int''tab$''cog''tab$''sdev''ta
b$''skew''tab$''kurt''tab$''fm''tab$'

            # Clean up
            select 'intID'
            plus 'intID2'
            Remove
        endfor

        fileappend 'directory_name$''log_file$'.txt 'newline$'
    endif
  endfor
endfor
select all
Remove      endfor

            # Clean up
            select 'intID'
            plus 'intID2'
            Remove
```

```
            endfor
        endif
    endfor
endfor
select all
Remove
```

References

Antsiranana. (2010). Wikipedia. https://en.wikipedia.org/w/index.php?title=Antsiranana&oldid=397487005

Assimil Malgache. (n.d.). [Computer software].

Aziz, J. (2020). *A Preliminary Model of Malagasy Intonation* [M.A., University of California, Los Angeles]. https://www.proquest.com/docview/2455792634/abstract/A4923FDE1F314B8CPQ/1

Aziz, J. (forthcoming). Imperatives and the Syntax-Prosody Interface in Malagasy. *Natural Language & Linguistic Theory*.

Aziz, J., & Elkins, N. (2022). *02. Malagasy /nr/-strengthening within and across prosodic boundaries*.

Barjam, J. P. (2003). *The Intonation of Malagasy Declarative Sentences*.

Bates, D., Mächler, M., Bolder, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bayles, A. J. (2016). *High-vowel lenition in the French of Quebec and Paris* [Master's thesis, University of Utah].

Bennett, R., Henderson, R., & Harvey, M. (2023). Vowel deletion as grammatically controlled gestural overlap in Uspanteko. *Language*, *99*(3), 399–456.

Berry, J., & Moyle, M. (2011). Covariation among vowel height effects on acoustic measures. *The Journal of the Acoustical Society of America*, *130*, EL365–EL371. https://doi.org/10.1121/1.3651095

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.24) [Computer software].

Bouwer, L. (2007). Intercomprehension and mutual intelligibility among southern Malagasy languages. *Language Matters*, *38*(2), 253–274. https://doi.org/10.1080/10228190701794624

Browman, C. P., & Goldstein, L. (1988). Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica*, *45*(2–4), 140–155. https://doi.org/10.1159/000261823

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201–251. https://doi.org/10.1017/S0952675700001019

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology* (1st ed., pp. 341–376). Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.019

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, *49*(3–4), 155–180. https://doi.org/10.1159/000261913

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'I.C.P.*, *5*, 25–34.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*, 219–252. https://doi.org/10.1017/S0952675700000658

Byrd, D. (1995). Articulatory characteristics of single and blended lingual gestures. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2).

Byrd, D. (1996). A Phase Window Framework for Articulatory Timing. *Phonology*, *13*(2), 139–169. https://doi.org/10.1017/S0952675700002086

Byrd, D., & Krivokapić, J. (2021). Cracking Prosody in Articulatory Phonology. *Annual Review of Linguistics*, *7*(Volume 7, 2021), 31–53. https://doi.org/10.1146/annurev-linguistics-030920-050033

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2

Caussèque, P. (1886). *Grammaire malgache*. Imprimerie catholique.

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 419–448). De Gruyter Mouton. https://doi.org/10.1515/9783110197105.2.419

Chitoran, I., & Iskarous, K. (2008). *Acoustic Evidence for High Vowel Devoicing in Lezgi*.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Harper & Row.

Clements, G. N. (1992). Phonological Primes: Features or Gestures? *Phonetica*, *49*(3–4), 181–193. https://doi.org/10.1159/000261914

Cohn, A., & McCarthy, J. J. (1998). Alignment and parallelism in Indonesian phonology. *Working Papers of the Cornell Phonetics Laboratory*, *12*, 55–137.

Collier, R. (1975). Physiological correlates of intonation patterns. *The Journal of the Acoustical Society of America*, *58*(1), 249–255. https://doi.org/10.1121/1.380654

Crouch, C. (2022). *Postcards from the syllable edge: Sonority and articulatory timing in complex onsets in Georgian* [Doctoral dissertation, University of California, Santa Barbara].

Crouch, C., Katsika, A., & Chitoran, I. (2023). Sonority sequencing and its relationship to articulatory timing in Georgian. *Journal of the International Phonetic Association*, *53*(3), 1049–1072. https://doi.org/10.1017/S0025100323000026

Dabkowski, M. F. (2018). *Variable Vowel Reduction in Mexico City Spanish* [Doctoral dissertation, The Ohio State University]. https://etd.ohiolink.edu/acprod/odb_etd/etd/r/1501/10?clear=10&p10_accession_num=osu1531994893143203

Dahl, O. C. (1952). Étude de phonologie et de phonétique malgaches. *Norsk Tidsskrift for Sprogvidenskap*, *16*, 148–200.

de La Beaujardière, J.-M. (2024). *Malagasy Encyclopedic Dictionary*. Malagasy Encyclopedic Dictionary. malagasyword.org

Delforge, A. M. (2008). Gestural Alignment Constraints and Unstressed Vowel Devoicing in Andean Spanish. *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 147–155.

DiCanio, C. (2012). *Vowel acoustics script for corpus data analysis* (Version 2.0) [Computer software].

DiCanio, C. (2021a). *Spectral moments of fricative spectra script in Praat* (Version 4.0) [Computer software].

DiCanio, C. (2021b). *Vowel acoustics script for corpus data analysis* (Version 2.0) [Computer software].

Dissassa, M. (1980). *Some aspects of Oromo phonology* [Master's thesis, Kansas Wesleyan University].

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.005

Erwin, S. (1996). Quantity and moras: An amicable separation. *UCLA Occasional Papers in Linguistics*, *The Structure of Malagasy, Volume 1*.

Esposito, A. (2002). On Vowel Height and Consonantal Voicing Effects: Data from Italian. *Phonetica*, *59*, 197–231. https://doi.org/10.1159/000068347

Fagyal, Z., & Moisset, C. (1999). *Sound change and articulatory release: Where and why are high vowels devoiced in Parisian French?* 309–312.

Faust, N., & Pike, E. G. (1959). The Cocama Sound System. In *Série Lingüistica Especial* (Vol. 1, pp. 10–55).

Ferrand, M. G. (1909). Note sur l'alphabet arabico-malgache. *Anthropos*, *4*(1), 190–206.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115–123. https://doi.org/10.1121/1.396977

Gafos, A. I. (2002). A Grammar of Gestural Coordination. *Natural Language & Linguistic Theory*, *20*(2), 269–337.

Goldsmith, J. (1976). *Autosegmental Phonology* [Doctoral dissertation, Massachusetts Institute of Technology].

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to Language via* *the Mirror Neuron System* (1st ed., pp. 215–249). Cambridge University Press. https://doi.org/10.1017/CBO9780511541599.008

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory* (pp. 111–120). https://www.research.ed.ac.uk/en/publications/learning-ot-constraint-rankings-using-a-maximum-entropy-model

Gordon, M. (1998). The Phonetics and Phonology of Non-modal Vowels: A Cross-Linguistic Perspective. *Proceedings of the Twenty-Fourth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Phonetics and Phonological Universals*, 93–105.

Gordon, M. (2007). Typology in Optimality Theory. *Language and Linguistics Compass*, *1*(6), 750–769. https://doi.org/10.1111/j.1749-818X.2007.00036.x

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174. https://doi.org/10.1017/S0025100302001020

Griffiths, D. (1854). *A Grammar of the Malagasy Language: In the Ankova Dialect*. E. Pite.

Gu, Y., & Durvasula, K. (2024). *The Relation between Sonority and Gestural Timing* (SSRN Scholarly Paper 4695254). https://doi.org/10.2139/ssrn.4695254

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, *23*(3), 387–429. https://doi.org/10.1017/S0952675706000996

Hall, N. (2010). Articulatory Phonology. *Language and Linguistics Compass*, *4*(9), 818–830. https://doi.org/10.1111/j.1749-818X.2010.00236.x

Hall, N. (2017). Articulatory Phonology. In *The Routledge Handbook of Phonological Theory*. Routledge.

Hawkins, S. (1992). An introduction to task dynamics. In D. R. Ladd & G. J. Docherty (Eds.), *Gesture, Segment, Prosody* (pp. 9–25). Cambridge University Press. https://doi.org/10.1017/CBO9780511519918.002

Howe, P. (2017). *Tonogenesis in Central dialects of Malagasy: Acoustic and perceptual evidence with implications for synchronic mechanisms of sound change*. https://hdl.handle.net/1911/96031

Howe, P. (2019). Central Malagasy. *Journal of the International Phonetic Association*, *51*(1), 103–136. https://doi.org/10.1017/S0025100319000100

Idsardi, W. J. (1998). Tiberian Hebrew Spirantization and Phonological Derivations. *Linguistic Inquiry*, *29*(1), 37–73. https://doi.org/10.1162/002438998553653

Iskarous, K., McDonough, J., & Whalen, D. H. (2012). A gestural account of the velar fricative in Navajo. *Laboratory Phonology*, *3*(1), 195–210. https://doi.org/10.1515/lp-2012-0011

Jannedy, S. (1995). *Gestural Phasing as an Explanation for Vowel Devoicing in Turkish* [Working Paper]. Ohio State University. Department of Linguistics. https://kb.osu.edu/handle/1811/81970

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–1263. https://doi.org/10.1121/1.1288413

Jun, S.-A. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press.

Jun, S.-A., & Beckman, M. E. (1993). *A gestural-overlap analysis of vowel devoicing in Japanese and Korean*. 1993 Annual Meeting of the Linguistic Society of America, Los Angeles, CA.

Jun, S.-A., Beckman, M., Niimi, S., & Tiede, M. (1997). Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korean. *Journal of Speech Sciences*, *1*(1), 153–200.

Kampstra, P. (2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software*, *28*(Code Snippet 1). https://doi.org/10.18637/jss.v028.c01

Kari, E. E. (2004). *A reference grammar of Degema*. R. Köppe. https://cir.nii.ac.jp/crid/1130000794550421120

Kariņš, A. K. (1995). Vowel deletion in Latvian. *Language Variation and Change*, *7*, 15–34.

Keating, P., & Lahiri, A. (1993). Fronted Velars, Palatalized Velars, and Palatals. *Phonetica*, *50*(2), 73–101. https://doi.org/10.1159/000261928

Keenan, E. L. (1976). Remarkable Subjects in Malagasy. In C. Li (Ed.), *Subject and Topic* (pp. 247–301). Academic Press, Inc. https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=12402662

Keenan, E. L., & Polinsky, M. (1998). Malagasy. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology*. Blackwell. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405166348.ch28

Kim, Y.-J. (2023). *Phonetic faithfulness in counterfeeding opacity*. The 54th Annual Meeting of the North East Linguistics Society, Massachusetts Institute of Technology.

Klatt, D. H. (1975). Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters. *Journal of Speech and Hearing Research*, *18*(4), 686–706. https://doi.org/10.1044/jshr.1804.686

Kochetov, A. (2017). Acoustics of Russian voiceless sibilant fricatives. *Journal of the International Phonetic Association*, *47*(3), 321–348. https://doi.org/10.1017/S0025100317000019

Koopman, H. (2005). Malagasy Imperatives. *UCLA Working Papers in Linguistics*, *12*.

Kuo, J. (forthcoming). Phonological reanalysis is guided by markedness: The case of Malagasy weak stems. *Phonology*.

Kurowski, K., & Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *The Journal of the Acoustical Society of America*, *76*(2), 383–390. https://doi.org/10.1121/1.391139

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press.

Leben, W. R. (1973). *Suprasegmental Phonology* [Doctoral dissertation, Massachusetts Institute of Technology].

Lehiste, I. (1970). *Suprasegmentals* (pp. viii, 194). Massachusetts Inst. of Technology P.

Lombardi, L. (1999). Positional Faithfulness and Voicing Assimilation in Optimality Theory. *Natural Language & Linguistic Theory*, *17*(2), 267–302. https://doi.org/10.1023/A:1006182130229

Malagasy. (n.d.). Ethnologue. https://www.ethnologue.com/language/mlg/

Manuel, S. Y., & Krakow, R. A. (1984). *Universal and language particular aspects of vowel-to-vowel coarticulation* (Collected Works SR-77/78; Status Report on Speech Research, pp. 69–78). Haskins Laboratories.

Martin, A. (2005). LOANWORDS AS PSEUDO-COMPOUNDS IN MALAGASY. *UCLA Working Papers in Linguistics*, *12*(Proceedings of AFLA XII).

McCarthy, J. J. (2000). Harmonic Serialism and Parallelism. *Proceedings of the 30th Annual Meeting of the North East Linguistics Society (NELS 30)*, *2*. https://roa.rutgers.edu/files/357-1099/roa-357-mccarthy-1.pdf

McCarthy, J. J., & Prince, A. S. (1995). *Faithfulness and Reduplicative Identity*. https://doi.org/10.7282/T31R6NJ9

Meneses, F., & Albano, E. (2015). From Reduction to Apocope: Final Poststressed Vowel Devoicing in Brazilian Portuguese. *Phonetica*, *72*(2–3), 121–137. https://doi.org/10.1159/000439599

Munson, B. (2001). A method for studying variability in fricatives using dynamic measures of spectral mean. *The Journal of the Acoustical Society of America*, *110*(2), 1203–1206. https://doi.org/10.1121/1.1387093

Namasivayam, A. K., Coleman, D., O'Dwyer, A., & van Lieshout, P. (2020). Speech Sound Disorders in Children: An Articulatory Phonology Perspective. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02998

Nearey, T. M., & Rochet, B. L. (1994). Effects of Place of Articulation and Vowel Context on VOT Production and Perception for French and English Stops. *Journal of the International Phonetic Association*, *24*(1), 1–18. https://doi.org/10.1017/S0025100300004965

Nirgianaki, E. (2014). Acoustic characteristics of Greek fricatives. *The Journal of the Acoustical Society of America*, *135*(5), 2964–2976. https://doi.org/10.1121/1.4870487

Ohala, J. (1981). The listener as a source of sound change. *Chicago Linguistic Society*, *17*, 178–203. https://doi.org/10.1075/cilt.323.05oha

Öhman, S. E. G. (1966). Coarticulation in VCV Utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America*, *39*(1), 151–168. https://doi.org/10.1121/1.1909864

O'Neill, T. (2015). *The phonology of Betsimisaraka Malagasy* [Ph.D., University of Delaware]. https://www.proquest.com/docview/1767229982/abstract/F579508122134DECPQ/1

O'Shaughnessy, D. (1981). A study of French vowel and consonant durations. *Journal of Phonetics*, *9*, 385–406.

Parker, G. W. (1883). *A Concise Grammar of the Malagasy Language*. BoD – Books on Demand.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, *36*(1), 55–90. https://doi.org/10.1016/j.wocn.2007.09.003

Pater, J. (1999). Austronesian Nasal Substitution and other NC effects. In R. Kager, H. van der Hulst, & W. Zonneveld (Eds.), *The Prosody Morphology Interface* (pp. 310–343). Cambridge University Press.

Paul, I. (2001). Concealed pseudo-clefts. *Lingua*, *111*(10), 707–727. https://doi.org/10.1016/S0024-3841(00)00044-9

Pearson, M. (1994). *Stress and vowel devoicing in Malagasy*. Unpublished manuscript, UCLA.

Pearson, M. (2018). Predicate raising and perception verb complements in Malagasy. *Natural Language & Linguistic Theory*, *36*(3), 781–849. https://doi.org/10.1007/s11049-017-9388-6

Perkell, J. S., & Chiang, C. (1986). Preliminary support for a "hybrid model" of anticipatory coarticulation. *Proceedings of the 12th International Congress of Acoustics*, A3–A6.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Doctoral dissertation, Massachusetts Institute of Technology].

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German*. *Journal of Phonetics*, *13*(4), 455–471. https://doi.org/10.1016/S0095-4470(19)30797-1

Rahajarizafy, R. P. A. (1960). *Essai sur la grammaire malgache*. Imprimerie catholique.

Rahidy, B. (1895). *Cours pratique de langue malgache: 1. partie: Grammaire*. Librairie africaine et coloniale.

Rakotofiringa, H. (1969). *Contributions à l'étude de la phonétique malgache* (Vol. 3).

Raoniarisoa, N. (1990). *Accent and intonation in a Malagasy dialect*. Bangor University.

Rasoloson, J., & Rubino, C. (2004). Malagasy. In K. A. A. Himmelmann Nikolaus (Ed.), *The Austronesian Languages of Asia and Madagascar* (pp. 456–488). Routledge. https://doi.org/10.4324/9780203821121

Renwick, M. E. L., & Cassidy, C. N. (2015). Detecting palatalization in spontaneous spoken English. *Proceedings of Meetings on Acoustics*, *23*(1), 060001. https://doi.org/10.1121/2.0000052

Riverside. (2024). [Online podcasting software]. riverside.fm

Saltzman, E. (1986). *Task dynamic coordination of the speech articulators: A preliminary model* (Research Report SR-84; Status Report on Speech Research, pp. 9–26).

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, *94*(1), 84–106. https://doi.org/10.1037/0033-295X.94.1.84

Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). *A task-dynamic toolkit for modeling the effects of prosodic structure on articulation*. 175–184.

Schwartz, M. F. (1969). Intraoral Air Pressures for /p/ in Oral and Whispered Vowel Environments. *The Journal of the Acoustical Society of America*, *46*(2B), 480–480. https://doi.org/10.1121/1.1911719

Shadle, C. H. (1985). *The acoustics of fricative consonants* [Doctoral dissertation, Massachusetts Institute of Technology].

Shadle, C. H. (1990). Articulatory-Acoustic Relationships in Fricative Consonants. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 187–209). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_8

Shadle, C. H. (2023). Alternatives to moments for characterizing fricatives: Reconsidering Forrest et al. (1988). *The Journal of the Acoustical Society of America*, *153*(2), 1412–1426.

Shadle, C. H., & Mair, S. J. (1996). Quantifying spectral characteristics of fricatives. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, *3*, 1521–1524. https://doi.org/10.1109/ICSLP.1996.607906

Silva, D. J. (1997). The variable deletion of unstressed vowels in Faialense Portuguese. *Language Variation and Change*, *9*(3), 295–308. https://doi.org/10.1017/S0954394500001939

Smith, C. (2003). Vowel Devoicing in Contemporary French. *Journal of French Language Studies*, *13*, 177–194. https://doi.org/10.1017/S095926950300111X

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, *25*(4), 471–500. https://doi.org/10.1006/jpho.1997.0053

Smith, C. M. (2018). *Harmony in Gestural Phonology* [Doctoral dissertation, University of Southern California]. https://www.proquest.com/docview/2128021312/abstract/E4F58B00A01E43CEPQ/1

Smolensky, P., & Prince, A. S. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. https://doi.org/10.7282/T34M92MV

Sole, maria-josep, & Ohala, J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. *Laboratory Phonology*, *10*.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, *70*(4), 976–984. https://doi.org/10.1121/1.387032

Stevens, K. N. (2000). *Acoustic Phonetics*. MIT Press.

Tsuchida, A. (1997). *Phonetics and phonology of Japanese vowel devoicing* [Ph.D., Cornell University]. https://www.proquest.com/docview/304340964/abstract/F053A4A7CA004E21PQ/1

Tsuchida, A. (2001). Japanese Vowel Devoicing: Cases of Consecutive Devoicing Environments. *Journal of East Asian Linguistics*, *10*(3), 225–245. https://doi.org/10.1023/A:1011221225072

van den Berg, J. (1958). Myoelastic-Aerodynamic Theory of Voice Production. *Journal of Speech and Hearing Research*, *1*(3), 227–244. https://doi.org/10.1044/jshr.0103.227

Vogel, R. (2021). A Unified Account of Two Vowel Devoicing Phenomena: The Case of Cheyenne. *Proceedings of the Annual Meetings on Phonology 2020*, *8*. https://doi.org/10.3765/amp.v9i0.4905

Vogel, R. C. (2022). Phonology of Vowel Devoicing: A Typological Perspective [Doctoral dissertation, Cornell University]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/docview/2726939746/abstract/267A724731BB4539PQ/1

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*(1), 145–166. https://doi.org/10.1017/S0022226700010598

Zsiga, E. C. (1997). Features, Gestures, and Igbo Vowels: An Approach to the Phonology-Phonetics Interface. *Language*, *73*(2), 227–274. https://doi.org/10.2307/416019

Zsiga, E. C. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, *28*(1), 69–102. https://doi.org/10.1006/jpho.2000.0109

Zygis, M., & Hamann, S. (2003). Perceptual and Acoustic Cues of Polish Coronal Fricatives. *Proceedings of ICPhS 15*, 395-398