



Westerberg, Fabienne Elina (2020) *Heavens, what a sound! The acoustics and articulation of Swedish Viby-i*. PhD thesis.

<http://theses.gla.ac.uk/81886/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

**Heavens, what a sound!**  
**The acoustics and articulation of Swedish Viby-i**

Fabienne Elina Westerberg

Submitted in fulfilment of the requirements for the  
Degree of Doctor of Philosophy

School of Critical Studies  
College of Arts  
University of Glasgow



University  
of Glasgow

December 2020

© Westerberg 2020



# Acknowledgements

For my family,  
who taught me to ask:  
“Says who?”

I would like to express my great gratitude to my supervisors, Jane Stuart-Smith and Eleanor Lawson, for their ongoing support, guidance, and expertise, without which this research would not have been possible.

I would also like to thank the many people who have given their advice and practical assistance over the course of this project, particularly Rachel Smith, Clara Cohen, Jim Scobbie, Alan Wrench, Michael Ashby, Márton Sóskuthy, and Stefano Coretta.

Many thanks to the staff and students at the University of Gothenburg, Stockholm University, and Uppsala University for facilitating my data collection, for participating in my research, and for providing a friendly and helpful support network.

A big thank you to all the labsters in GULP, especially Ewa Wanat, Julia Moreno, Jo Pearce, Edward Marshall, Ebtehal Asiry, and Robert Lennon, for their emotional support and commiseration.

Finally, to my friends, with honourable mentions to Alicia van Eden, Alexandra Palmquist, Sophie Kromholz, and Gaia Burgio, who have helped me hang on to the remaining shreds of my sanity: Thank you.

This work was supported by the Economic and Social Research Council, grant number 1515770. Ultrasound equipment was provided by Queen Margaret University Edinburgh.

# Abstract

This thesis provides a phonetic case study of the Swedish /i:/ variant known as *Viby-i* (sometimes also called *Lidingö-i*). This sound is characterised by an unusual ‘thick’, ‘buzzing’ vowel quality, but its articulation has long been disputed. Previous research suggests that this vowel may be subject to articulatory trade-off, whereby speakers can achieve the same sound using different articulatory strategies. There are also indications that *Viby-i* may be subject to sociolinguistic variation, as it appears to be spreading across Sweden, and it is frequently used as a prestige marker in urban dialects. This thesis addresses the issues of how speakers produce *Viby-i*, how its acoustic properties relate to its articulation, and how it is used across different social and linguistic contexts.

The study presents data from 34 Swedish speakers from Gothenburg, Stockholm, and Uppsala, recorded with simultaneous audio, ultrasound tongue imaging, and lip video. The speakers’ /i:/ productions are analysed acoustically with regard to their formant values, dynamic properties, and frication. Linguistic and social variation is also explored using a word list and a demographic questionnaire. The articulatory analysis establishes the tongue gestures and lip positions used to produce *Viby-i*, and investigates the link between acoustics and articulation using a set of normalised articulatory measurement points, which are compared to the first two formants.

The acoustic analysis shows that *Viby-i* is characterised by a low acoustic F2, which is usually lower than [e:], and a relatively high F1, which is usually similar to [e:]. Linguistic context has a small but reliable effect on formant values, and also affects the fricated offglide of the vowel. All speakers in the sample are found to use *Viby-i* rather than standard [i:], but regional differences exist in both acoustics and articulation.

The articulatory analysis reveals that *Viby-i* can be produced with a variety of tongue shapes, most of which involve a low, fronted tongue body, high tongue tip, and retraction of the post-dorsal part of the tongue. Surprisingly, the low F2 is not produced by lip-rounding or overall tongue backing, but appears to be caused by a combination of tongue lowering and post-dorsal retraction. Young speakers retract more than old speakers, with no difference in acoustics. Frication during the vowel is associated with a raised, fronted tongue tip, or in velar contexts, by a raised tongue body.

There are several implications of these findings for the wider literature. Firstly, *Viby-i* appears to be more widespread in Central Sweden than previously assumed, and it may already

have replaced the standard variant [i:], although further research is needed to confirm this. Secondly, the ‘mismatch’ between acoustics and articulation demonstrates the benefit of including articulatory data in the analysis of vowel sounds. Finally, the study shows that traditional frameworks of vowel analysis may be too simplistic to adequately describe the articulation of complex vowel sounds, and that new methods may be required as articulatory data becomes increasingly available.

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research motivation . . . . .	1
1.2 Aim and research questions . . . . .	2
1.3 Approach . . . . .	2
1.4 Thesis structure . . . . .	3
<b>2 Phonetic analysis of vowels</b>	<b>4</b>
2.1 Chapter overview . . . . .	4
2.2 What is a vowel? . . . . .	5
2.3 Current approaches to vowel analysis . . . . .	6
2.3.1 The vowel quadrilateral . . . . .	6
2.3.2 Acoustic models of vowel production . . . . .	7
2.4 Historical methods for investigating vowel articulation . . . . .	12
2.4.1 Auditory, proprioceptive, and simple instrumental methods . . . . .	12
2.4.2 X-ray imaging . . . . .	14
2.4.3 Spectrography and acoustic analysis . . . . .	16
2.4.4 New articulatory methods . . . . .	17
2.5 Necessity of articulatory vowel data . . . . .	20
2.5.1 Critique of the vowel quadrilateral . . . . .	20
2.5.2 Vocal tract parameters contributing to vowel acoustics . . . . .	23
2.5.3 Articulatory trade-off . . . . .	25
2.6 Chapter summary . . . . .	28

<b>3</b>	<b>Swedish vowels</b>	<b>29</b>
3.1	Chapter overview . . . . .	29
3.2	The Swedish language . . . . .	29
3.3	Swedish dialects . . . . .	30
3.3.1	Dialect regions . . . . .	30
3.3.2	Regional vs. Standard Swedish . . . . .	33
3.3.3	Swedish multiethnolect . . . . .	34
3.4	Phonology . . . . .	35
3.4.1	Vowels . . . . .	35
3.4.2	Consonants . . . . .	36
3.5	Linguistic influences on vowel production . . . . .	37
3.5.1	Consonant context . . . . .	37
3.5.2	Vowel duration . . . . .	37
3.6	Sociolinguistic influences on vowel production . . . . .	39
3.6.1	Age . . . . .	39
3.6.2	Gender . . . . .	40
3.6.3	Socioeconomic class . . . . .	40
3.6.4	Lifestyle, identity and attitudes . . . . .	41
3.6.5	Dialect and dialect contact . . . . .	42
3.6.6	Social pressures . . . . .	43
3.7	Description of studied cities . . . . .	44
3.7.1	Overview of cities . . . . .	44
3.7.2	Stockholm . . . . .	46
3.7.3	Göteborg . . . . .	48
3.7.4	Uppsala . . . . .	49
3.8	Chapter summary . . . . .	50
<b>4</b>	<b>Viby-i</b>	<b>51</b>
4.1	Chapter overview . . . . .	51
4.2	What is Viby-i? . . . . .	51
4.2.1	Descriptions in the literature . . . . .	51
4.2.2	Working definition and points of interest . . . . .	52
4.2.3	Phonetic notation . . . . .	53
4.3	Perception . . . . .	53
4.3.1	Viby-i vs. standard [i:] . . . . .	53
4.3.2	Viby-i vs. Turkish [i] . . . . .	54
4.3.3	Viby-i vs. /y:/ . . . . .	55
4.3.4	Gradient Viby-colouring . . . . .	56
4.4	Acoustics . . . . .	57

4.4.1	The first three formants . . . . .	57
4.4.2	Variability in standard [i:] . . . . .	59
4.4.3	Reported formant values for Viby-i . . . . .	59
4.4.4	Beyond formant frequencies . . . . .	62
4.5	Articulation . . . . .	62
4.5.1	Impressionistic theories . . . . .	62
4.5.2	Speech synthesis . . . . .	63
4.5.3	Tongue gesture dynamics (EMA) . . . . .	63
4.5.4	Static tongue contours (UTI) . . . . .	65
4.6	Geographic spread and variation . . . . .	68
4.6.1	Rural vs. urban Viby-i . . . . .	68
4.6.2	Regional variation . . . . .	70
4.6.3	Chronological spread . . . . .	71
4.7	Social significance . . . . .	73
4.7.1	Prestige associations . . . . .	73
4.7.2	Change in progress . . . . .	74
4.7.3	Negative associations . . . . .	74
4.7.4	Viby-i in popular culture . . . . .	76
4.8	Similar vowels in other languages . . . . .	78
4.8.1	Occurrence in the world's languages . . . . .	78
4.8.2	Mandarin apical vowels . . . . .	79
4.8.3	Other Scandinavian languages . . . . .	81
4.9	Chapter summary . . . . .	81
<b>5</b>	<b>Method</b>	<b>83</b>
5.1	Chapter overview . . . . .	83
5.2	Method summary . . . . .	83
5.3	Participants . . . . .	84
5.3.1	Recruitment . . . . .	84
5.3.2	Sampled cities . . . . .	85
5.3.3	Speaker selection . . . . .	85
5.4	Word list design . . . . .	86
5.4.1	Word list items . . . . .	86
5.4.2	Linguistic factors for analysis . . . . .	87
5.4.3	Carrier phrase . . . . .	91
5.4.4	Word list evaluation . . . . .	91
5.5	Spontaneous speech recordings . . . . .	93
5.5.1	Interview questions . . . . .	93
5.5.2	Target words . . . . .	94

5.6	Questionnaire . . . . .	94
5.6.1	Factor: City . . . . .	95
5.6.2	Factor: Distance from city centre . . . . .	95
5.6.3	Factor: Age . . . . .	95
5.6.4	Factor: Gender . . . . .	96
5.6.5	Factor: Social interaction . . . . .	96
5.6.6	Factor: Prescriptivism . . . . .	97
5.6.7	Excluded factors . . . . .	97
5.7	Equipment . . . . .	99
5.7.1	Audio recordings . . . . .	99
5.7.2	Ultrasound recordings . . . . .	99
5.7.3	Recording set-up . . . . .	101
5.7.4	Recording procedure . . . . .	102
5.7.5	Preliminary evaluation of set-up and procedure . . . . .	104
5.8	Acoustic data processing and analysis . . . . .	107
5.8.1	Overview of segmentation . . . . .	107
5.8.2	Segmentation protocol . . . . .	108
5.8.3	Formant measurement . . . . .	113
5.8.4	Bandpass-filtered zero-crossing rate . . . . .	114
5.8.5	Bandwidth measurement . . . . .	116
5.9	Articulatory data processing and analysis . . . . .	116
5.9.1	Tongue splines . . . . .	116
5.9.2	Qualitative tongue shape analysis . . . . .	118
5.9.3	Normalised articulatory measures . . . . .	119
5.9.4	Lip data . . . . .	120
5.10	Statistical methods . . . . .	120
5.10.1	Mixed-effects modelling . . . . .	120
5.10.2	Generalised additive mixed models . . . . .	122
<b>6</b>	<b>Acoustic results</b>	<b>125</b>
6.1	Chapter overview . . . . .	125
6.2	Auditory impressions and acoustic parameters . . . . .	125
6.3	Position of /i:/ in the Swedish vowel space . . . . .	126
6.4	Individual speakers' use of Viby-i . . . . .	129
6.5	Overlap between /i:/ and other vowels . . . . .	131
6.5.1	Formant distributions . . . . .	131
6.5.2	Bark scaling . . . . .	133
6.6	Relationship between F1 and F2 within Viby-i . . . . .	135
6.7	Linguistic effects on Viby-i formants . . . . .	136

6.7.1	Variable selection . . . . .	136
6.7.2	Plosive set: Duration, preceding voicing, following backness . . . . .	138
6.7.3	Fricative set: Duration, preceding voicing, following complexity . . . . .	140
6.7.4	Full word list set: Preceding manner . . . . .	141
6.7.5	Summary of linguistic results . . . . .	142
6.8	Social effects on Viby-i formants . . . . .	144
6.8.1	Variable selection . . . . .	144
6.8.2	Initial model . . . . .	145
6.8.3	Step model . . . . .	145
6.8.4	Summary of social results . . . . .	150
6.9	Viby-i dynamics . . . . .	152
6.9.1	Spectrograms and formant measurements . . . . .	152
6.9.2	Auditory transcriptions . . . . .	155
6.9.3	Zero-crossing rate . . . . .	159
6.9.4	Summary of Viby-i dynamics . . . . .	162
6.10	Summary of acoustic results . . . . .	163
<b>7</b>	<b>Articulatory results</b>	<b>165</b>
7.1	Chapter overview . . . . .	165
7.2	Qualitative description of /i:/ tongue gesture . . . . .	166
7.2.1	Tongue body position . . . . .	168
7.2.2	Tongue tip . . . . .	172
7.2.3	Tongue retraction . . . . .	175
7.2.4	Tongue shape . . . . .	177
7.2.5	Summary of mean tongue gestures . . . . .	179
7.3	GAMM analysis of high front vowels . . . . .	181
7.3.1	/i:/ vs. /y:/ . . . . .	181
7.3.2	/i/ vs. /e/ . . . . .	183
7.3.3	Summary of GAMM analysis . . . . .	183
7.4	Normalised tongue point measures . . . . .	187
7.4.1	Highest point of the tongue . . . . .	187
7.4.2	Most constricted point of the tongue . . . . .	189
7.4.3	Retraction measures . . . . .	191
7.4.4	Tongue tip measure . . . . .	193
7.5	Lip posture . . . . .	194
7.6	Effect of tongue point measures on F1 and F2 . . . . .	195
7.6.1	Correlation between highest point and F1/F2 for all vowels . . . . .	195
7.6.2	Correlation between highest point and F1/F2 for Viby-i . . . . .	197
7.6.3	Correlation between multiple tongue measures and F1/F2 for Viby-i . . . . .	198



7.6.4	Sociolinguistic variation in tongue gesture . . . . .	202
7.7	Tongue dynamics of Viby-i . . . . .	206
7.7.1	Visual inspection of dynamic plots . . . . .	206
7.7.2	Dynamic influences of articulation on formants . . . . .	210
7.7.3	Dynamic influences of articulation on vowel frication . . . . .	210
7.8	Chapter summary . . . . .	212
<b>8</b>	<b>Discussion</b>	<b>213</b>
8.1	Chapter overview . . . . .	213
8.2	Discussion of findings . . . . .	214
8.2.1	Viby-i formant structure . . . . .	214
8.2.2	Viby-i dynamics and frication . . . . .	215
8.2.3	Viby-i articulation . . . . .	216
8.2.4	Effect of linguistic environment on Viby-i . . . . .	217
8.2.5	Sociolinguistic variation of Viby-i . . . . .	219
8.3	General discussion . . . . .	221
8.3.1	What counts as a Viby-i? . . . . .	221
8.3.2	Vowel or consonant? . . . . .	222
8.3.3	How exotic is Viby-i? . . . . .	223
8.3.4	Change in progress? . . . . .	225
8.3.5	Mismatch between acoustics and articulation . . . . .	227
8.3.6	Contribution to ultrasound methodology . . . . .	229
8.4	Future directions . . . . .	232
<b>9</b>	<b>Conclusion</b>	<b>234</b>
	<b>References</b>	<b>236</b>
<b>A</b>	<b>Questionnaire (English)</b>	<b>253</b>
<b>B</b>	<b>Questionnaire (Swedish)</b>	<b>260</b>
<b>C</b>	<b>Information and consent forms (English)</b>	<b>267</b>
<b>D</b>	<b>Information and consent forms (Swedish)</b>	<b>271</b>
<b>E</b>	<b>Historical formant values for Swedish vowels</b>	<b>275</b>
<b>F</b>	<b>Individual formant values</b>	<b>280</b>
<b>G</b>	<b>Individual mean tongue splines</b>	<b>289</b>

<b>H</b>	<b>Individual tongue GAMMs</b>	<b>296</b>
<b>I</b>	<b>Dynamic mean splines for /i:/</b>	<b>308</b>
<b>J</b>	<b>Individual dynamic GAMMs</b>	<b>321</b>
<b>K</b>	<b>Lip images</b>	<b>344</b>

# List of Tables

3.1	Summary of differences between Swedish dialect regions (after Bruce, 2010; Gårding, 1977). Phonemes listed in order of use. . . . .	33
3.2	Long and short vowel pairs (after Engstrand, 1999: 140). . . . .	35
3.3	Offglides for Central Swedish vowels (after Elert, 1995: 40). . . . .	36
3.4	Central Swedish consonants and common realisations (after Engstrand, 2004: 167). . . . .	37
3.5	Summary of demographic information for Gothenburg, Stockholm, and Uppsala municipalities, compared to Sweden as a whole (Statistics Sweden, 2019a, 2019b). . . . .	46
4.1	Formant values for standard [i:] in previous Swedish studies, compared to Standard Southern British English. Entries with both a higher F1 and a lower F2 than SSBE are shaded. . . . .	58
4.2	Reported F1 to F4 values for Viby-i (grey) and /e:/ (white). . . . .	60
5.1	Participants by city, gender, and age group (age in parentheses). . . . .	86
5.2	Word list tokens applicable to all nine long vowels. . . . .	87
5.3	Word list tokens applicable to /i:/ only. Duplicates from previous table in grey. . . . .	88
5.4	Schematic representation of how the word list was randomised. . . . .	91
5.5	Target words for the spontaneous speech task. . . . .	94
5.6	Segmentation criteria for vowels based on preceding and following segment. . . . .	108
6.1	Mean F1 to F4 values (Hz) for all nine vowels across the sample, divided by gender. . . . .	128
6.2	Individual mean F1 and F2 values for /i:/ and /e:/ (Hz), with calculated difference. . . . .	130
6.3	LMER output showing the difference in distribution between different vowels (F1 to F4). . . . .	132
6.4	LMER output showing the difference in distribution of different vowels (Bark). . . . .	134
6.5	LMER output showing the relationship between normalised F1 and normalised F2. . . . .	136
6.6	Summary of word list items and their assigned categories. . . . .	137

6.7	LMER output showing the effect of vowel duration, preceding consonant voicing, and following consonant backness on normalised F1 and F2 within the plosive set. . . . .	138
6.8	LMER output showing the effect of vowel duration, preceding consonant voicing, and following consonant manner on normalised F1 and F2 within the fricative set. . . . .	140
6.9	LMER output showing the effect of preceding consonant manner on normalised F1 and F2. . . . .	142
6.10	Summary of linguistic effects on normalised F2, categorised by word list item. .	143
6.11	LMER output showing the effect of age, gender, city, distance from city centre, social interaction, and prescriptivism on normalised F1 and F2 (Gothenburg). .	146
6.12	LMER output showing the effect of age, gender, city, distance from city centre, social interaction, and prescriptivism on normalised F1 and F2 (Stockholm). . .	147
6.13	LMER output showing the effect of age, gender, city, distance from city centre, social interaction, and prescriptivism on normalised F1 and F2 (Uppsala). . . .	148
6.14	LMER output (from step modelling) showing the effect of city on normalised F1 and F2, and the effect of gender on normalised F1. . . . .	149
6.15	Viby-i offglide types. Entries in grey are theoretically possible, but were not observed in this sample. . . . .	157
6.16	Mean vowel durations across word list items. . . . .	157
7.1	Summary of mean tongue spline patterns for /i:/ compared to /e:/ by individual speaker. . . . .	180
7.2	Summary of GAMM tongue patterns for /i:/ compared to /y/, e:/. . . . .	186
7.3	LMER output showing the effect of the highest point of the tongue on normalised F1 and F2 for the full set of vowels. . . . .	196
7.4	LMER output showing the effect of the highest point of the tongue on normalised F1/F2 for /i:/. . . . .	197
7.5	LMER output showing the effect of normalised tongue point measures on F2 for Viby-i. . . . .	201
7.6	LMER output showing the effect of city on the height of the tongue peak. . . .	203
7.7	LMER output showing the effect of age and city on the backness of the main tongue constriction. . . . .	204
7.8	LMER output showing the effect of tongue tip position and maximum tongue constriction height on bpZCR over the course of /i:/. . . . .	211

# List of Figures

2.1	Vowel quadrilateral currently used by the International Phonetic Association (2015).	7
2.2	Correspondence between the vocal tract and a tube model (Ladefoged, 1996: 123).	8
2.3	Front and back resonances of the vocal tract based on a two-tube model (K. Johnson, 1997: 94).	9
2.4	Tube models for the vowels [ʊ], [a], [y], [i] and their corresponding formant patterns (Fant, 1960: 66).	9
2.5	Velocity nodes (N) and antinodes (A) for F1 to F4 according to perturbation theory (Chiba & Kajiyama, 1941).	10
2.6	Promotional image of Atkinson's Mouth Measurer (in Panconcelli-Calzia, 1994: 65).	12
2.7	Ancient vowel model (Wood, 1982: 4).	13
2.8	Bell-Sweet vowel model (Wood, 1982: 4).	13
2.9	Jones' cardinal vowel model, including English vowels in grey (D. Jones, 1917: frontispiece).	13
2.10	X-ray of speaker with metal chain outlining the tongue contour (Russell, 1928: 261).	15
2.11	X-ray of Stephen Jones producing cardinal [a] with two metal chains outlining the tongue and velum (S. Jones, 1929: 151).	15
2.12	Ultrasound image of the tongue surface, speaker facing right.	18
2.13	MRI image of the vocal tract, with added tongue outline (Zhou et al., 2008: 4,467).	18
2.14	Tongue measurement points from EMA (solid line represents the palate) (Hoole & Nguyen, 1997: 182).	19
2.15	Three-dimensional reconstructions of the tongue surfaces for English vowels and schwab (Stone & Lundberg, 1996: 3,732).	22
3.1	Map of Swedish provinces and dialect regions (after Bruce, 2010: 27). Map adapted from Wikimedia Commons (2019).	31
3.2	Dialectal variation in old vs. young Swedish speakers. Similar colours represent similar linguistic features (Leinonen, 2011: 86).	34

3.3	Central Standard Swedish vowel space (Engstrand, 1999: 140). . . . .	36
3.4	S-curve model of language change (University of Duisburg-Essen, 2019). . . .	44
3.5	Location of Gothenburg, Stockholm and Uppsala in Sweden (map adapted from Wikimedia Commons, 2019). . . . .	45
4.1	Listeners' acceptance of vowels as 'damped /i:/' by stimulus (Björsten & Engstrand, 1999: 1,958). . . . .	55
4.2	Correlation between Viby-i 'strength' and a high normalised F1 (data from Westerberg, 2016). . . . .	56
4.3	Correlation between Viby-i 'strength' and a low normalised F2 (data from Westerberg, 2016). . . . .	56
4.4	Viby-i strength ratings (black) and normalised F2 values (white, shifted up for visibility) for each speaker (data from Westerberg, 2016). . . . .	57
4.5	Position of standard [i:] vs. Viby-i in the F1/F2 space using data from previous studies. . . . .	59
4.6	Viby-i and Standard Swedish [i:] plotted against cardinal vowels from D. Jones (1917) (Schötz, Frid, & Löfqvist, 2011: 1,768). . . . .	61
4.7	Vowels in the Kräklinge dialect (black) plotted against Standard Swedish vowels (white) from Eklund and Traunmüller (1997) (Engstrand, Björsten, Lindblom, Bruce, & Eriksson, 1998: 86). . . . .	61
4.8	Vowels from Gothenburg speakers with Viby-i (black) and Gothenburg multi-ethnolect speakers without Viby-i (grey) (Gross, 2018: 329). . . . .	62
4.9	Vertical movement of the tongue body and tip in two Stockholm and two Gothenburg speakers (Frid, Schötz, Gustafsson, & Löfqvist, 2015: 2). Viby-i users in bottom row. . . . .	64
4.10	Horizontal movement of the tongue body and tip in two Stockholm and two Gothenburg speakers (Frid et al., 2015: 2). Viby-i users in bottom row. . . . .	64
4.11	EMA measurement points indicating the tongue gesture for standard [i:] vs. Viby-i (Schötz et al., 2011: 3). . . . .	65
4.12	UTI tongue splines for three Scottish speakers producing /i:, ʉ:, a:, o:/ and three Swedish speakers producing /i:, e:, a:, u:/ (data from Westerberg, 2013). . . . .	65
4.13	Ultrasound splines of four Swedish speakers producing /i:, e:, ʉ:, a:, u:/ (data from Westerberg, 2016). . . . .	67
4.14	Example of lip postures for /e:/ and /i:/ in Westerberg (2016). Note the fronted tongue position for /i:/. . . . .	68
4.15	Map of locations where Viby-i has been documented (after Elert, 1995: 45). . .	69
4.16	Exaggerated Viby-i facial posture demonstrated by Swedish comedian Jonas Fagerström (Fagerström, 2019). Sign reads "super rich millionaire from Stockholm". . . . .	78

4.17	Tongue contours for [a, i, ɿ] in one Mandarin speaker (Faytak & Lin, 2015: 2).	80
4.18	Tongue contours of two Mandarin speakers, comparing [i, ɿ, ʊ] to preceding [ɛ, s, ʂ] (Faytak & Lin, 2015: 3).	80
5.1	Ultrasound headset with profile-view lip camera.	100
5.2	Bite plate with measurements.	100
5.3	UTI image of the tongue pushing against the bite plate.	100
5.4	Image from side-facing lip camera mounted on the ultrasound headset.	101
5.5	Set-up for recording and synchronising audio, ultrasound, and lip video.	102
5.6	Example of how end-frication was segmented.	108
5.7	Illustration of vowel segmentation with preceding and following plosives.	109
5.8	Illustration of vowel segmentation with preceding /v/ and following liquid.	109
5.9	Word-final exhale overlapping with epenthetic vowel for /i:/.	111
5.10	Word-final end-frication of /i:/.	111
5.11	Loss of amplitude before a liquid.	112
5.12	Boundary between end-frication and /s/.	113
5.13	Frication superimposed on periodic waveform.	113
5.14	Waveform of modal /i:/ before and after filtering.	115
5.15	Waveform of /i:/ with superimposed frication before and after filtering.	115
5.16	Example of tongue spline in AAA.	117
5.17	Example of palate trace in AAA.	118
5.18	Example of fiducial (bite plate) spline in AAA.	118
5.19	Measurement points used for the normalised articulatory analysis.	120
5.20	Example of static GAMM plot.	123
5.21	Example of corresponding difference smooth, comparing /i:/ and /e:/.	123
5.22	Example of relationship between GAMMs and difference smooths, comparing /i:/ and /e:/ splines.	124
5.23	Example of erroneous GAMM.	124
6.1	Example of frication and formant weakening in the middle of a vowel segment, followed by schwa.	126
6.2	All speakers' vowels on the F1/F2 plane (Lobanov normalised).	127
6.3	Historical F1/F2 values (Hz) for standard /i:/ and /e:/ from the literature.	128
6.4	Position of Central Standard Swedish vowels on the vowel quadrilateral (Engstrand, 1999: 140).	129
6.5	F1 to F4 distributions (Hz) for each vowel, based on all speakers.	131
6.6	All speakers' vowels on the F1/F2 plane (Bark Difference Metric).	133
6.7	Distribution of normalised F1 against normalised F2 for /i:/ for all speakers.	135
6.8	Correlation matrix of linguistic factors.	137

6.9	Significant effects of duration and preceding consonant voicing on normalised F2 in the plosive set. . . . .	139
6.10	Significant effects of following consonant place on normalised F1 and F2 in the plosive set. . . . .	139
6.11	Significant effects of vowel duration and preceding consonant voicing on normalised F1 in the fricative set. . . . .	141
6.12	Significant effects of preceding consonant voicing and following consonant manner on normalised F2 in the fricative set. . . . .	141
6.13	Significant effects of preceding consonant manner on normalised F1 and F2. . .	142
6.14	Word list items in order of F2, from high to low. . . . .	143
6.15	Correlation plot of social factors. . . . .	144
6.16	Distribution of scores for ‘social interaction’ and ‘local rootedness’ respectively, with an uneven distribution in the latter. . . . .	145
6.17	Statistical effects on F1 by city and gender. . . . .	150
6.18	Statistical effects on F2 by city. . . . .	150
6.19	Distribution of normalised F1 and F2 for /i:/ by individual speakers. . . . .	151
6.20	Velar pinch during Viby-i in speaker GF1. . . . .	152
6.21	Modal Viby-i in speaker UF2. . . . .	152
6.22	Final formant weakening during Viby-i in speaker SF1. . . . .	153
6.23	Medial formant weakening during Viby-i in speaker GF5. . . . .	153
6.24	Low frequency (velar) end-frication during Viby-i in speaker SF1. . . . .	153
6.25	High frequency (alveolar) end-frication during Viby-i in speaker SF1. . . . .	153
6.26	Frication and formant weakening during Viby-i in speaker GF4. . . . .	154
6.27	Frication and formant weakening during Viby-i in speaker GF1. . . . .	154
6.28	Time-normalised formant trajectories (Hz) across Viby-i by gender. F1 dark grey, F2 light grey. Blue lines represent overall means for each formant. . . . .	154
6.29	Mean F1 and F2 trajectories (Hz) for VIRA. . . . .	155
6.30	Mean F1 and F2 trajectories (Hz) for BITA. . . . .	155
6.31	Zero-crossing rate across Viby-i by speaker. . . . .	155
6.32	Zero-crossing rate across Viby-i by word. . . . .	156
6.33	Frication superimposed on waveform for /i:/ during VILA. . . . .	156
6.34	Viby-i offglides by word across the sample. . . . .	157
6.35	Viby-i offglides by speaker across the sample. . . . .	158
6.36	Overall difference in bpZCR between all vowels in the sample. . . . .	159
6.37	Overall difference in bpZCR between auditorily ‘buzzy’ vs. ‘not buzzy’ /i:/. . .	160
6.38	Zero-crossing rate across Viby-i by word. . . . .	161
6.39	Zero-crossing rate across Viby-i by preceding consonant /v, b, f, p/ . . . . .	162
7.1	Illustration of how the UTI data is orientated. . . . .	166



7.2	Example of an arched tongue shape for /i:/ (red).	167
7.3	Example of a front-bunched tongue shape for /i:/ (red).	167
7.4	Example of a back-bunched tongue shape for /i:/ (red).	168
7.5	Example of a double-bunched tongue shape for /i:/ (red).	168
7.6	Low mean tongue height for /i:/ compared to /e:/.	169
7.7	Very low mean tongue height for /i:/ compared to /e:/.	169
7.8	Similar mean tongue heights for /i:/ and /e:/.	170
7.9	Overlapping mean tongue gestures for /i:/ and /e:/.	170
7.10	Distribution of tongue body fronting of /i:/ against /e:/ across the sample.	172
7.11	Distribution of tongue body lowering of /i:/ against /e:/ across the sample.	172
7.12	High tip-down gesture for /i:/.	173
7.13	High tip-up gesture for /i:/.	173
7.14	Similar tip gestures between /i:/ and /e:/.	174
7.15	Lower tip gesture for /i:/ than for /e:/.	174
7.16	Distribution of tongue tip height for /i:/ compared to /e:/ across the sample.	175
7.17	Distribution of tongue tip gesture for /i:/ across the sample.	175
7.18	Post-dorsal constriction for /i:/.	176
7.19	Post-dorsal backing for /i:/ compared to /e:/.	176
7.20	Post-dorsal constriction and backing for /i:/ compared to /e:/.	177
7.21	Distribution of post-dorsal constriction for /i:/ across the sample.	178
7.22	Distribution of post-dorsal backing for /i:/ compared to /e:/ across the sample.	178
7.23	Distribution of tongue shapes for /i:/ across the sample.	178
7.24	Comparison between mean spline plot and GAMM plot.	181
7.25	GAMM plot and difference smooth showing no significant difference between /i:/ and /y:/.	182
7.26	GAMM plot and difference smooth showing an interval of significant difference between /i:/ and /y:/.	182
7.27	GAMM plot and difference smooth showing significant tongue retraction and lowering for /i:/ compared to /e:/.	184
7.28	GAMM plot and difference smooth showing significant tongue retraction, lowering, and tip raising for /i:/ compared to /e:/.	184
7.29	GAMM plot and difference smooth showing no significant difference between /i:/ and /e:/.	185
7.30	GAMM plot and difference smooth showing significant fronting/lowering of /i:/ compared to /e:/.	185
7.31	Measurement points used for the normalised articulatory analysis.	187
7.32	Normalised F1 and F2 values for all vowels and speakers.	188
7.33	Normalised highest point of the tongue for all vowels and speakers.	188

7.34	X-rays showing highest point of the tongue for Cardinal Vowels (after D. Jones, 1972: frontispiece). Note the slight forward tilt. . . . .	189
7.35	Normalised most constricted point of the tongue for all vowels and speakers. . .	190
7.36	Normalised backest point of the tongue for all vowels and speakers. . . . .	191
7.37	Normalised midback tongue point for all vowels and speakers. . . . .	192
7.38	Normalised tongue tip measures for all vowels and speakers. . . . .	193
7.39	Lip postures for /i:, y:, e:/ in a female speaker with a high F2. . . . .	194
7.40	Lip postures for /i:, y:, e:/ in a male speaker with a high F2. . . . .	195
7.41	Lip postures for /i:, y:, e:/ in a female speaker with a low F2. . . . .	195
7.42	Lip postures for /i:, y:, e:/ in a male speaker with a low F2. . . . .	195
7.43	Significant negative correlation between tongue lowering (highest point) and F1 for the full vowel set. . . . .	196
7.44	Significant positive correlation between tongue backing (highest point) and F2 for the full vowel set. . . . .	196
7.45	Significant negative correlation between tongue lowering (highest point) and F1 for /i:/ . . . . .	198
7.46	Significant negative correlation between tongue backing (highest point) and F2 for /i:/ . . . . .	198
7.47	Correlation plot of tongue point measures. . . . .	199
7.48	Horizontal tongue point measures significantly correlated with F2 for Viby-i. .	202
7.49	Vertical tongue point measures significantly correlated with F2 for Viby-i. . .	202
7.50	Significant correlation between city of origin and tongue peak height. . . . .	203
7.51	Significant correlation between city of origin, age, and post-dorsal constriction. .	205
7.52	Mean dynamic tongue splines for /i:/ before a front consonant. . . . .	207
7.53	Mean dynamic tongue splines for /i:/ before a back consonant. . . . .	207
7.54	GAMM plot and difference smooths showing gesture dynamics for /i:/ before a front consonant. . . . .	208
7.55	GAMM plot and difference smooths showing gesture dynamics for /i:/ before a back consonant. . . . .	209
7.56	Velar pinch during Viby-i in speaker GF1. . . . .	210
7.57	Low frequency (velar) end-frication during Viby-i in speaker SF1. . . . .	210
7.58	Significant effects of tip height, tip frontness, and max constriction height on bpZCR. . . . .	211
8.1	Acceptability ratings of Turkish [i] and Swedish Viby-i as ‘damped /i:/’ (Björsten & Engstrand, 1999). . . . .	224
8.2	Tongue contours of two Mandarin speakers, comparing [i, ɿ, ʅ] to preceding [ɕ, s, ʂ]. . . . .	225
8.3	Schematic s-curves of how Viby-i might have changed over time. . . . .	227

8.4 Velocity nodes (N) and antinodes (A) for F2 according to perturbation theory  
(Chiba & Kajiyama, 1941). . . . . 228

# Chapter 1

## Introduction

### 1.1 Research motivation

Most vowel research relies on acoustic information, since audio recordings are generally easier to collect and analyse than articulatory data. Based on what we know about the acoustic properties of the vocal tract, acoustics are often regarded as a shortcut to articulation: For example, the first vowel formant is usually associated with tongue height, and the second formant with tongue backness (e.g. Delattre, 1951). However, this approach provides a simplified picture of vowel articulation that does not consider the complex relationship between vocal tract settings and acoustic output. Previous research has shown that speakers can use articulatory trade-off to achieve equivalent acoustic targets using different articulatory strategies, and that these strategies can vary between different regions or social groups (e.g. Lawson, Scobbie, & Stuart-Smith, 2014; Lawson, Stuart-Smith, & Rodger, 2019). Thus, acoustic methods may fall short when it comes to explaining complex articulatory behaviour.

The Central Swedish vowel *Viby-i* provides an interesting case study of variable articulatory strategies. *Viby-i* is an /i:/ variant that has been remarked upon for its unusual “thick”, “dark”, “damped”, “buzzing” quality (Engstrand et al., 1998: 1-2). However, its underlying articulation has long been disputed. Based on proprioceptive and acoustic data, a number of potential strategies have been suggested (Björsten & Engstrand, 1999; Borgström, 1913; Ladefoged & Lindau, 1989; Lundell, 1878; Noreen, 1903), but some of these suggestions are conflicted, and due to a lack of articulatory data, we do not yet know how speakers actually produce *Viby-i*, or how their articulatory gestures relate to its vowel quality.

In addition, *Viby-i* production may vary based on regional, social, or linguistic factors, as this sound appears to be part of an ongoing vowel shift. According to Bruce (2010: 216), it is rapidly replacing the standard variant [i:] in many parts of Sweden, but the extent of this shift has not previously been studied. *Viby-i* also has a complex set of social functions: In urban dialects, particularly Stockholm, it is a well-known prestige marker (Kotsinas, 2007), while in rural dialects, it is often stigmatised and subject to dialect levelling (Elert, 1995). Despite the

great number of unanswered questions about Viby-i, there have only been a handful of studies investigating this vowel since it was first documented in urban speech in the 1950s (Björseth, 1958).

## 1.2 Aim and research questions

The aim of this research is to address some of the gaps in the literature about Viby-i, and in doing so, to contribute to the broader research areas of articulatory phonetics, vowel production, sociolinguistics, and Swedish phonology. The main focus points of the thesis will be to provide a phonetic description of Viby-i in both articulation and acoustics, to explore the relationship between the two, to investigate its social and linguistic variation, and to test and evaluate a mixed methodology for studying vowels. The study is guided by the following research questions:

- What acoustic properties characterise Viby-i?
- How is Viby-i articulated by the tongue and lips?
- Is Viby-i subject to articulatory trade-off?
- Is there regional, social, or linguistic variation in the use of Viby-i?
- What methods are best used to describe and quantify the articulatory behaviour of Viby-i?

## 1.3 Approach

The study uses a newly collected corpus of acoustic, articulatory, and demographic data from 34 Central Swedish speakers, aged 20-80, from Gothenburg, Stockholm, and Uppsala. The data consists of audio recordings, ultrasound tongue imaging (UTI) video, lip video, and questionnaire responses.

The acoustic analysis investigates the formant values, dynamics, and levels of fricative noise associated with Viby-i, in order to identify acoustic sources of its ‘thick’, ‘buzzing’ vowel quality. Frication is quantified using bandpass-filtered zero crossing rate (bpZCR) (Gordeeva & Scobbie, 2013). Linguistic and social variation is explored using a structured word list and self-reported demographic information. Statistical analyses using linear mixed effects regression models (LMERs) are used to evaluate whether patterns observed in the data are likely to be consistent across the wider population.

The articulatory analysis describes the midsagittal tongue shapes and positions used to produce Viby-i, linking these findings to the acoustic analysis through qualitative descriptions, generalised additive mixed models (GAMMs), and LMERs of normalised tongue measures. The latter is an innovative approach introduced by Lawson et al. (2019). Images of the lips are also used to account for non-lingual effects on the acoustic signal.

## 1.4 Thesis structure

The thesis begins with a review of the literature: Chapter 2 provides a background to historical and current methods used to study vowels phonetically, discussing their potential benefits and drawbacks in relation to articulatory information. Chapter 3 describes the Swedish vowel system, as well as the regional and social context in which Viby-i is studied in this thesis. Chapter 4 provides a comprehensive account of what is currently known about Viby-i, summarising both research findings and anecdotal evidence.

The remaining chapters outline the experimental phonetic research carried out for this thesis: Chapter 5 describes the methods of data collection, processing, and analysis. Chapter 6 presents and discusses the acoustic results. Chapter 7 presents and discusses the articulatory results. Chapter 8 summarises the findings, provides a detailed discussion of the results, and addresses the research questions. Finally, Chapter 9 concludes the thesis.

# Chapter 2

## Phonetic analysis of vowels

### 2.1 Chapter overview

This chapter contextualises the phonetic analysis of Viby-i by describing the frameworks commonly used to study vowel sounds, and examining potential reasons why these methods have not been successful in resolving how Viby-i is produced. As a solution, the chapter proposes a mixed methodology, combining acoustic and articulatory data to gain a fuller understanding of vowel production.

The chapter begins by discussing the phonetic definition of a vowel, and why Viby-i may not fully fit this description. The complexity of Viby-i is the starting point for examining how current approaches to vowel analysis could be improved.

Section 2.3 provides a brief summary of the auditory and acoustic models which currently influence how we conceptualise and analyse vowels. This section describes the relationship between the impressionistic vowel space and the acoustic formants, and explores two acoustic theories that can be used to explain this relationship. This section also discusses the use of static formant data for vowel analysis.

Section 2.4 outlines the historical development of the field, showing that articulatory data was used to inform the acoustic theories we use today, but that acoustic analysis gradually became the prevalent method, causing articulatory research to fall behind. Due to recent technological developments, however, articulatory methods are once again gaining popularity, and this opens up new possibilities for the study of vowel sounds.

Finally, Section 2.5 argues for the necessity of articulatory data, pointing out the simplified aspects of the vowel quadrilateral and the assumed equivalence between formant values and the highest point of the tongue. This section also discusses the phenomenon of articulatory trade-off, demonstrating that speakers can achieve the same acoustic output using different articulatory strategies.

## 2.2 What is a vowel?

The phonetic definition of a vowel is a pulmonic, voiced, periodic sound created without obstruction in the vocal tract (Catford, 1994; Ladefoged, 2001; Laver, 1994). Ogden adds that vowels are also usually oral with a central airflow, and produced “with a convex tongue shape, and without friction” (Ogden, 2009: 56). In addition, vowels are usually distinguished from other sounds which fulfil the same phonetic criteria, such as [w], [j] or [ɹ], by including their ability to function as the nucleus of a syllable (Catford, 1994; Ladefoged, 2001; Laver, 1994). The definition is thus not only based on phonetic criteria, but also on the phonological function of the sound.

In some cases, however, the line between vowel and consonant can be blurred. For example, “schwar” [ɐ̯], also known as “r-coloured schwa”, phonologically represents a merging between a vowel and a consonant (e.g. in Am. Eng. ‘learn’ [lɜ̃n]) (Kuecker, Lockenvitz, & Müller, 2015). Phonetically, this sound has both vocalic and consonantal features: On the one hand, it is voiced, periodic and created without contact between the articulators; on the other hand, the constriction is somewhat greater than for [ə], and it is often produced with a “bunched” tongue shape (Lawson, Scobbie, & Stuart-Smith, 2013). In cases like this, it is not always easy to determine if a sound should qualify as a vowel or not. Instead, classification may depend on the context. In its phonological function, *schwar* can be considered both a vowel and a consonant, while in a phonetic analysis, the methods used to study this sound would probably be most reminiscent of those used for vowels.

A similar case can be made for *Viby-i*. This vowel is known to have some consonantal qualities; for example, it is often described as being fricated or “buzzy” (Engstrand et al., 1998; Kotsinas, 2007; Schötz et al., 2011), and previous studies suggest that it can be produced with a number of different tongue gestures, most of which seem to lack the stereotypical convex tongue shape usually associated with vowel sounds (Catford, 1994; Ogden, 2009). Part of what this thesis aims to investigate is whether *Viby-i* is as “exotic” as some sources have suggested (Schötz et al., 2011). In other words, does this sound qualify as a ‘true’ (phonetic) vowel, or does it defy traditional definitions? In what ways does it deviate from our expectations of a vowel sound, and what implications does this have for phoneticians who wish to study this vowel, or others like it? This research will address these questions by providing an acoustic and articulatory analysis of *Viby-i*, which can be used to form a basis for a broader understanding of vowels, particularly in terms of their articulation.



## 2.3 Current approaches to vowel analysis

### 2.3.1 The vowel quadrilateral

Because vowels tend to be voiced and produced without obstruction in the vocal tract, they are usually not described in terms of their voicing, place of articulation, and manner of articulation, in the way that consonants are. Instead, the most important vocal tract settings that influence vowel production are the position of the tongue within the oral cavity, and the position of the lips (Catford, 1994: 124). These articulatory parameters are somewhat more fluid than those of consonants, partly because the tongue is not making contact with any of the other articulators, and partly because there may be individual variation in the shape and size of the oral cavity (Beck, 2010: 156). For these reasons, it can be challenging to provide detailed articulatory descriptions of vowel sounds, as there are no clear landmarks that the tongue can be compared to. Instead, vowels are usually described in more abstract terms, using the concept of the “vowel space” (e.g. Catford, 1994: 133).

The vowel space refers to a conceptual area in the mouth, within which the tongue is thought to move to give rise to different vowel sounds. In this system, the height of the highest point of the tongue curve is (inversely) correlated with the first formant (F1), while the advancement of the same point is correlated with the second formant (F2) (Ladefoged, 1993: 196). The boundaries of the vowel space are marked by frication, i.e. when the tongue comes so close to another articulator that the airflow becomes turbulent. This system establishes three “corner” vowels at the extreme edges of the vowel space, [i, ʌ, u], which, if they were constricted any further, would become the fricated consonants [j, ɣ, ɰ] (Catford, 1994: 130-131). Since the corner vowels are the only ones that can be defined articulatorily in this way, they are often used as anchor points to describe the position of other vowels in the system. The fact that a lack of frication is part of their definition renders the concept of a ‘fricated vowel’ somewhat self-contradictory from a phonetic standpoint, although fricated vowels have been documented in some languages (Connell, 2000: 233). It is nevertheless worth noting that most of the world’s languages contain some version of the three corner vowels, as their distance from each other provides a clear contrast between different vowel qualities (Maddieson, 1984; Stevens, 1972).

One of the earliest and most successful systems used to visualise the vowel space is the vowel quadrilateral, originally conceived by Bell (1867), popularised by D. Jones (1917) through his cardinal vowels, and later adopted by the International Phonetic Association (1949) (Fig. 2.1). The history of the quadrilateral will be discussed further in section 2.4.1.

The quadrilateral was originally developed using impressionistic auditory and proprioceptive methods (Ladefoged, 1967: 67-68), and shows the corner vowels as well as a number of intermediate vowels, which Jones believed to be equidistant in both auditory quality and in the highest point of the tongue (Ladefoged, 1976: 70). However, later work suggests that these vowels are only “auditorily equidistant” (Ladefoged & Johnson, 2015: 229), and that their articulatory re-

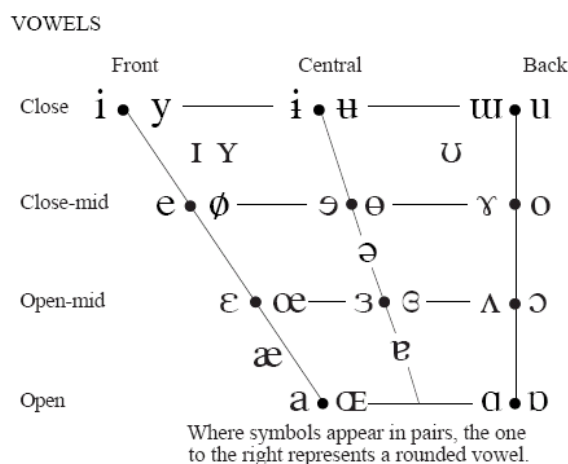


Figure 2.1: Vowel quadrilateral currently used by the International Phonetic Association (2015).

relationships are more complex. For example, Ladefoged, referring to a set of X-ray images of S. Jones (1929) producing the cardinal vowels, asserts that “the tongue does not move in a series of even approximately equidistant steps”, and that it “has such a different shape for the front and for the back vowels that it is meaningless to compare [[a]] with [[ɑ]]” (Ladefoged, 1967: 1).

As articulatory vowel research progressed, it became increasingly clear that the vowel quadrilateral is a schematic representation of vowel production, which by necessity omits some articulatory detail. Although it remains an extremely useful set of auditory and acoustic reference points, it is not possible for any model to map “cavity-formant relations [to] satisfy the requirements of both simplicity and general validity” (Fant, 1960: 123).

Because of its schematic nature, some phoneticians have expressed criticism against the vowel quadrilateral, notably Russell, who suggests that “Phoneticians are thinking in terms of acoustic<sup>1</sup> fact, and using physiological fantasy to express the idea” (Russell, 1928, in Ladefoged & Johnson, 2015: 208). Similarly, Lindblad states that it is an “unfortunate practice” that phoneticians frequently use articulatory terminology such as ‘high, low, front, back’ when referring to acoustic formant values (Lindblad, 2010: 54). This practice, however, likely stems from the fact that a better system has yet to be introduced (Ladefoged, 1967: 51-52).

The lack of detailed articulatory methodologies for studying vowels can, in part, be attributed to the historical context in which the field developed. This issue will be discussed further in section 2.4.

### 2.3.2 Acoustic models of vowel production

In addition to the vowel quadrilateral, there are more detailed models of acoustic theory which describe how the formants relate to different configurations in the vocal tract. These models rely on the acoustic theory of speech production, also known as source-filter theory (Fant, 1960).

<sup>1</sup>Note: Russell’s use of the word ‘acoustic’ corresponds to ‘auditory’ in current terminology.

Source-filter theory posits that speech sounds can be divided into a source (usually the voicing produced by the vocal folds) and a filter (the acoustic chamber of the vocal tract, which, depending on its size and shape, amplifies certain frequencies, and dampens others).

### Tube models

One way of conceptualising the oral cavity during vowel production is by comparing it to a series of tubes (Fant, 1960), as illustrated in Fig. 2.2. The length of each tube determines its natural resonance; long tubes accommodate lower frequencies, while short tubes accommodate higher ones. Since formants represent groups of amplified frequencies, it is possible to calculate the lengths of the ‘tubes’ that produced them, and thus arrive at an approximation of the vocal tract shape.

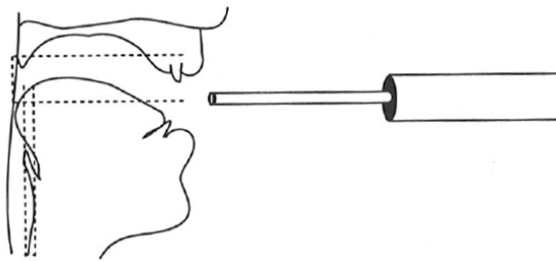


Figure 2.2: Correspondence between the vocal tract and a tube model (Ladefoged, 1996: 123).

Based on the vowel quadrilateral, as the tongue moves to produce different vowel sounds, the highest point of the tongue creates a constriction between the front and the back part of the oral cavity. In a simplified tube model, this division of the vocal tract can be represented by two tubes, one front and one back, which resonate separately. As the tongue shifts back and forward, the relative length of each tube changes; for a high front vowel like [i], the front tube is short and the back tube is long, meaning that the back tube accommodates the lower frequencies of F1, while the front tube accommodates the higher frequencies of F2. For the low back vowel [ɑ], this relationship is reversed. For [y], the constriction is the same as for [i], but the vocal tract is lengthened by lip-rounding, resulting in lower formants overall, but F2 is particularly affected because the length is added to the front tube.

Knowing the length of these respective tubes, it is possible to calculate their resonating frequencies, and vice versa. Fig. 2.3 provides a visual representation of the resonances of a two-tube model with a set length of 16 cm. For example, when the back tube is 6 cm long (and the front tube is 10 cm long), F1 and F3 are produced by the front tube, while F2 is produced by the back tube. Some examples of two-tube models and their corresponding formant outputs are also shown in 2.4.

Although the resonances of a tube are mainly predicted by its length (K. Johnson, 1997: 84), more sophisticated tube models are also able to account for the aerodynamic effects of constriction, as well as acoustic coupling, since the tubes are connected. For example, K. Johnson (1997:

95) introduces a three-tube model where the third tube represents a short and narrow section between the front and back cavity. This narrowing both affects the resonances of the back tube, and introduces an additional resonance of its own. For these calculations, the cross-sectional areas of the tubes become important, which more closely reflects the three-dimensional nature of the vocal tract. However, K. Johnson (1997: 102) points out that while tube models are very helpful, they are best used for sounds with a single, relatively narrow constriction, while more complex, open articulations can be better explained by perturbation theory.

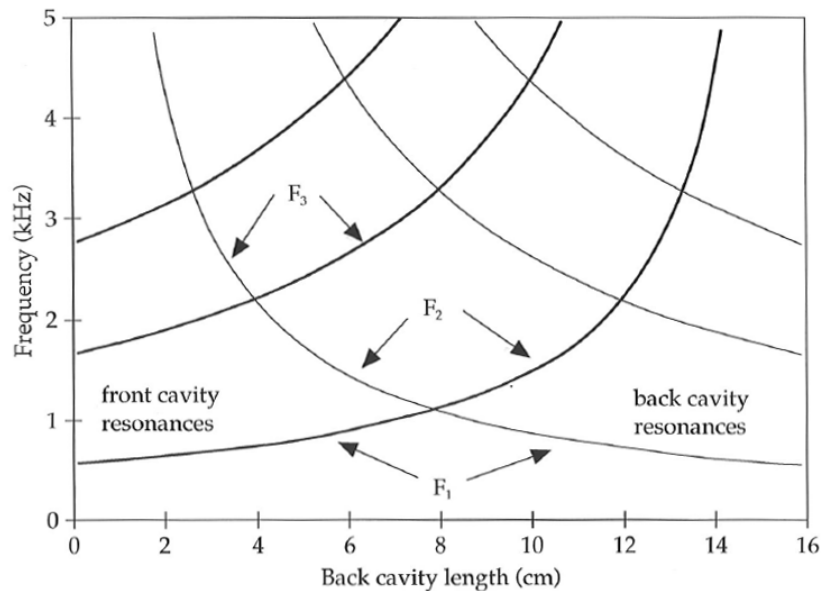


Figure 2.3: Front and back resonances of the vocal tract based on a two-tube model (K. Johnson, 1997: 94).

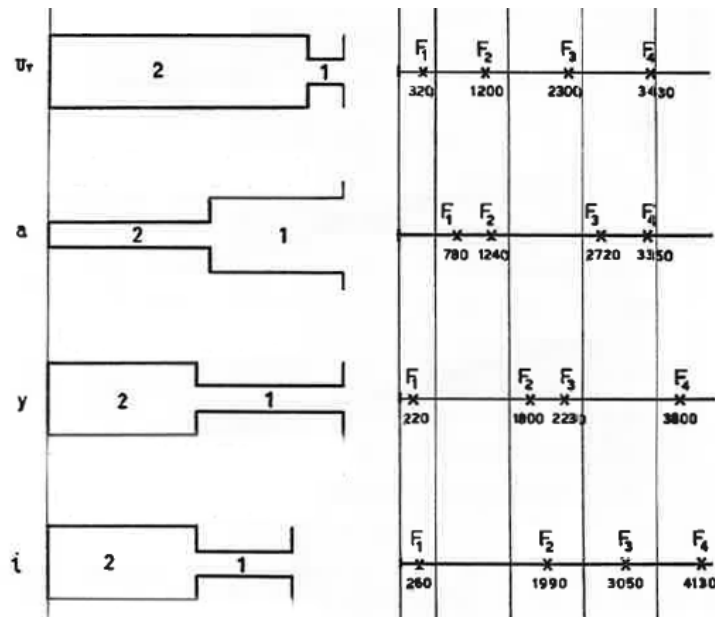


Figure 2.4: Tube models for the vowels [u], [a], [y], [i] and their corresponding formant patterns (Fant, 1960: 66).

### Perturbation theory

Another popular way of explaining the acoustic consequences of vowel articulation is perturbation theory (Chiba & Kajiyama, 1941). This theory considers the movement of standing waves in the vocal tract, and the relative influence of constriction in places where either the pressure or the velocity of the air particles are at their most extreme.

Standing waves occur when soundwaves resonate back and forth within a confined space, creating specific points where the waves' peaks, troughs, and nodes (points of no movement) reoccur. Frequencies that 'fit' the resonator and can create standing waves are amplified, since energy is repeatedly added to the same points, similar to pushing a swing at the exact right time.

Standing waves are particularly affected if they are manipulated at points where the air pressure is high (i.e. where the particles are maximally crowded together), or where the air velocity is high (i.e. where the particles are maximally spaced apart). Pressure and velocity have a complementary relationship, in that the most extreme velocity points (positive or negative) correspond to zero pressure, and vice versa. The points where velocity at its most extreme are usually called 'antinodes', while points of zero velocity are called 'nodes'.

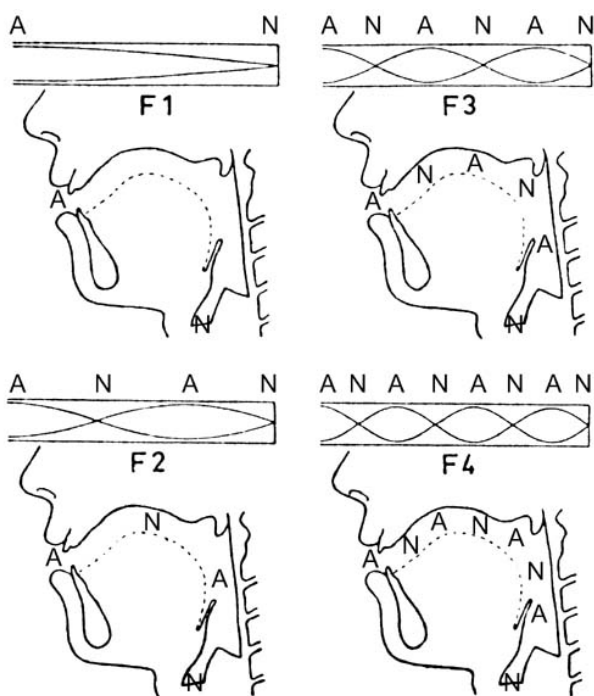


Figure 2.5: Velocity nodes (N) and antinodes (A) for F1 to F4 according to perturbation theory (Chiba & Kajiyama, 1941).

The natural resonances of the vocal tract mean that the standing waves always have an antinode near the source (the vocal folds) and a node near the end of the resonator (the lips). The number of times the waves travel through the vocal tract depends on the frequency of the formant; F1 has time to complete one cycle, F2 two cycles, etc. The shape of the vocal tract interacts with the nodes and antinodes of the waves, creating different formant patterns for different con-

figurations. Fig. 2.5 illustrates the approximate locations of velocity nodes and antinodes in the vocal tract for F1 to F4. Constriction at an antinode (A) lowers the formant frequency, while constriction at a node (N) increases it (K. Johnson, 1997: 100). For example, constriction at the lips lowers all four formants, while constriction at the palate raises F2, and lowers F3 and F4.

According to K. Johnson (1997: 101-102), perturbation theory is good at explaining sounds with complex articulations. For example, he describes American [ɪ] as being characterised by simultaneous constriction at the lips, palate, and pharynx, all three of which contribute to its characteristic low F3. This phenomenon is easier to explain with perturbation theory than with tube models.

### **Application of acoustic models**

Tube models and perturbation theory are useful acoustic frameworks, which can be used either to work backwards from acoustic values to potential articulations, or to explain the effect that a particular articulatory gesture will have on acoustics. However, both frameworks, in the form that they have been explained here, are quite simplified, and to some extent they still rely on the same assumptions as the vowel quadrilateral. As later sections will point out, articulatory data can contain additional complexity, for instance in the fact that vowels are not always produced with an arched tongue shape, making it more difficult to estimate the vocal tract area, or that sometimes, multiple articulatory strategies can result in the same output. These issues will be explored further in Section 2.5.

### **Dynamic vowel analysis**

It should be mentioned that most vowel analyses rely on static, rather than dynamic, formant measurements (although recently, the number dynamic studies has increased). In other words, formants are either measured at a single point, usually 50% of the vowel duration, or average formant values are calculated over the course of the vowel. This approach is useful when studying monophthongs, where the vowel quality does not substantially change, but when studying diphthongs, or sounds which may in other ways be dynamic, it is relevant to include multiple formant measures to account for changes in resonances over time. This is the case for Central Swedish vowels, which are canonically described as monophthongal, but often contain dynamic elements, such as diphthongised or fricativised offglides (Engstrand, 1999). The dynamic aspects of Swedish vowels will be discussed further in Chapter 3, and this study will include some dynamic elements to explore whether this type of analysis is relevant to Viby-i.

## 2.4 Historical methods for investigating vowel articulation

### 2.4.1 Auditory, proprioceptive, and simple instrumental methods

The earliest methods used for phonetic vowel research mainly relied on the phonetician's ability to describe what they could hear, feel, and see in the vocal tract. With a very limited view of the tongue, descriptions of vowel sounds were primarily auditory, although some phoneticians have argued that it is possible to feel changes in the tongue's position through extensive training (e.g. Catford, 1994; Sweet, 1877). Even so, the lack of contact between the tongue and other articulators made the analysis of vowels particularly difficult (Darwin, 1804: 119), and researchers recognised that instrumental data was required to verify if their impressions were correct. This led to the invention of a number of simple phonetic tools made from e.g. tin foil, metal wire, cardboard, and whalebone (Ashby, 2016: 71-74), as well as an instrument known as Atkinson's Mouth Measurer (Atkinson, 1897), which consisted of a metal handle and a sliding wooden block (Fig. 2.6). Researchers also used palatograms and linguograms to map lateral tongue contact using ink or charcoal (Gósy, 2011: 173). Unfortunately, these instruments often relied on inserting objects into the mouth, with the risk of affecting speech production. It also appears that this kind of equipment was fairly rare, at least for studying vowels; descriptions of early phonetic work (e.g. Catford, 1981; Ladefoged, 1967) suggest that phoneticians predominantly used a combination of sensory information and auditory cues when describing vowel sounds.

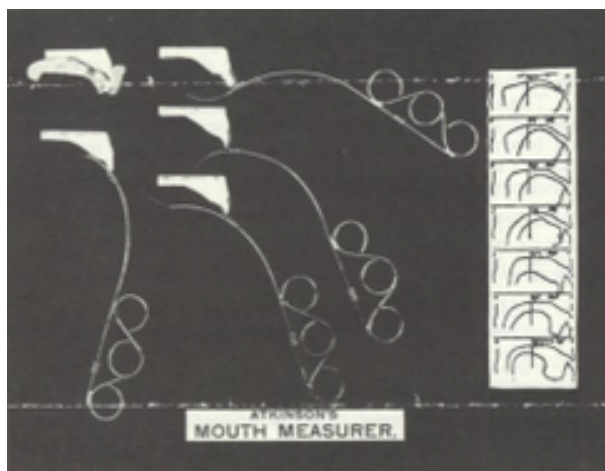


Figure 2.6: Promotional image of Atkinson's Mouth Measurer (in Panconcelli-Calzia, 1994: 65).

Auditory and proprioceptive methods gave rise to the vowel quadrilateral, which was originally designed by Bell (1867) and further developed by Sweet (1877), before being adopted by D. Jones (1917) for the cardinal vowels. The Bell-Sweet model (Fig. 2.8) had the advantage over the previous 'ancient' model (Fig. 2.7) that it allowed for a continuum of vowel sounds, while the ancient model only included "three basic tongue manoeuvres (palatal, velar, pharyngeal), jaw position and lip position" (Wood, 1982: 4-6). The Bell-Sweet model also corresponded better to new acoustic theory: The ancient model had suggested that vowels were produced by a sin-

gle resonating cavity delimited by the tongue, like a bottle filled with water, but von Helmholtz (1863) disputed this, stating that vowels were formed by two resonating cavities, one on each side of the tongue arch (Wood, 1982: 7). This theory fit well with the Bell-Sweet model, as the ‘highest point of the tongue’ implied tongue arching, and the sliding parameters of height and backness could be seen as gradually manipulating the size of the front and back cavities respectively.

When D. Jones (1917) introduced the cardinal vowel system (Fig. 2.9), three additional factors contributed to its popularity: Firstly, it provided a simple, standardised framework for describing vowel qualities in any language (Ladefoged, 1993: 219). Secondly, the cardinal vowels provided set reference points, which were passed down auditorily from Jones to his students, as well as being available on audio record. Thirdly, and perhaps most importantly, it was later discovered that the positions of the cardinal vowels on the quadrilateral were roughly equivalent to the frequencies of the first two formants if plotted on a two-dimensional grid. This point will be returned to in Section 2.4.3.

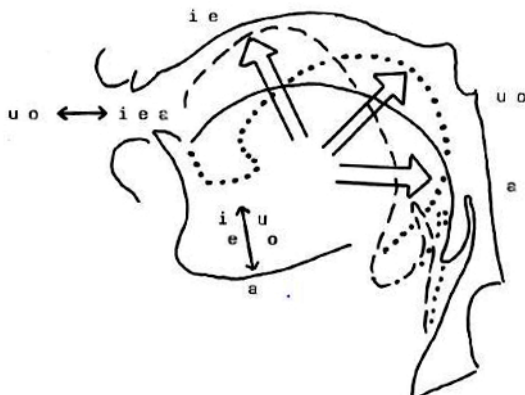


Figure 2.7: Ancient vowel model (Wood, 1982: 4).

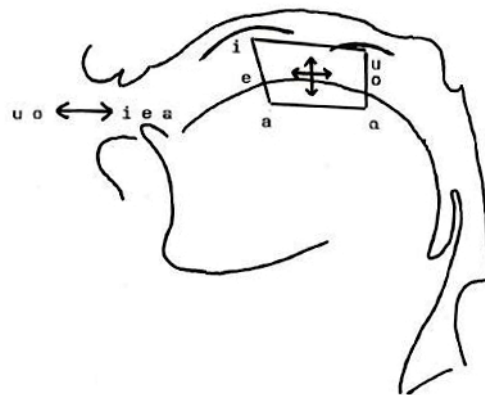


Figure 2.8: Bell-Sweet vowel model (Wood, 1982: 4).

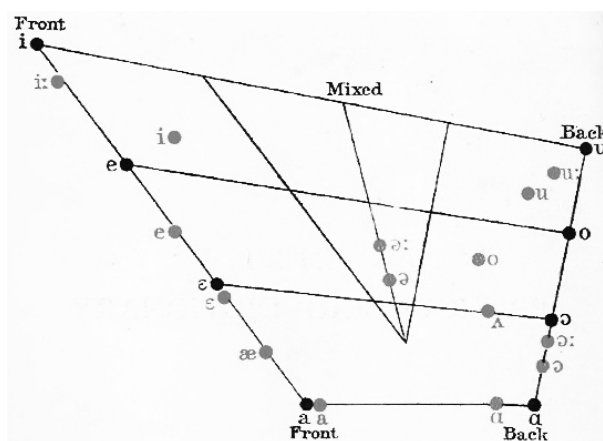


Figure 2.9: Jones' cardinal vowel model, including English vowels in grey (D. Jones, 1917: frontispiece).



### 2.4.2 X-ray imaging

The invention of X-ray (Röntgen, 1895) had a great impact on phonetics, as it made it possible for the first time to obtain images of the (living) vocal tract in cross-section. X-ray images are created by projecting a radiation beam onto the body, which passes through the tissue, creating an image on a plate on the other side (Stone, 2010: 10). One of the pioneers of this method for phonetic research was Scheier (1897), who “in his activity throughout 10 years, clarified practically every major problem of this method of investigation, so far as the technique of his time permitted” (Macmillan & Kelemen, 1952: 672). Among other things, he managed to obtain “an ideal cross section of the organs of the head and neck”, and gradually reduced the exposure time of the X-ray from over 10 minutes to around half a second (Macmillan & Kelemen, 1952: 671-672).

Scheier’s work led to a golden age of phonetic X-ray research, beginning in the late 1920s (Macmillan & Kelemen, 1952: 675), with cinematic X-ray appearing around the same time (e.g. Gutzmann, 1930). One of the most important insights from this period was that natural speech consists of continuous movements, rather than a series of articulatory segments (Menzerath & de Lacerda, 1933). X-ray research also contributed greatly to anatomical knowledge of the vocal tract, including the fact that individual speakers vary in the shape and size of their vocal organs (e.g. Russell, 1928: 142).

However, there were a number of issues associated with X-ray imaging. Firstly, there was a significant health risk. Several researchers from this time report burns and hair loss from prolonged exposure (e.g. Russell, 1928: 44), but despite this, the nature of the danger was not well understood, and researchers continued to expose themselves and their participants to ionising radiation. Secondly, there was the issue of visibility. Soft vocal organs, such as the tongue and velum, were frequently obscured by denser bony tissue, which appeared much brighter on the X-ray images. Researchers were often quite creative in circumventing this problem: For example, Scheier purposefully recruited a toothless participant, creating “the best photographs thus far published” of the vocal tract (Russell, 1928: 52), but this strategy later called the usefulness of his data into question (Macmillan & Kelemen, 1952: 672). Russell describes experimenting with various metal-based pastes, lead strips glued to the roof of the mouth, silk thread coated in gold foil and dipped in gelatine, and eventually a very thin, gold-plated lead chain with a wooden handle, which the participant could hold onto while swallowing the loose end (Russell, 1928: 67-69) (Fig. 2.10). The chain would then lie along the central line of the tongue, creating a visible contour of its shape on the X-ray image. Similar methods are also reported by Barth (1907), Grunmach (1907), and Meyer (1907). Having overcome this practical obstacle, however, Russell points out that participants were not always happy with this solution:

But the link chain gags the subject. The links pinch and tickle and lead to a most unfortunate distortion of the cavity and tongue position. The moment the subject starts such a process we might just as well give up. [...] I say [the chain] can hardly

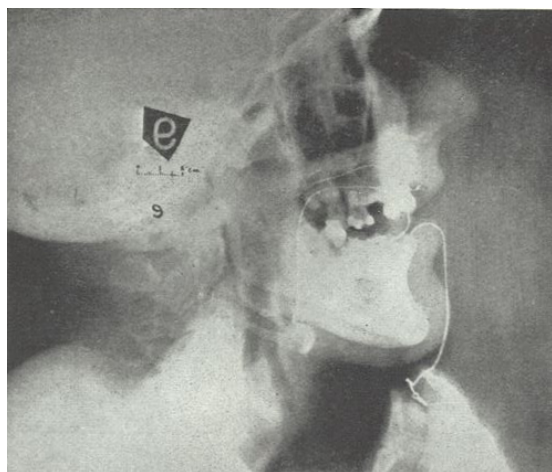


Figure 2.10: X-ray of speaker with metal chain outlining the tongue contour (Russell, 1928: 261).

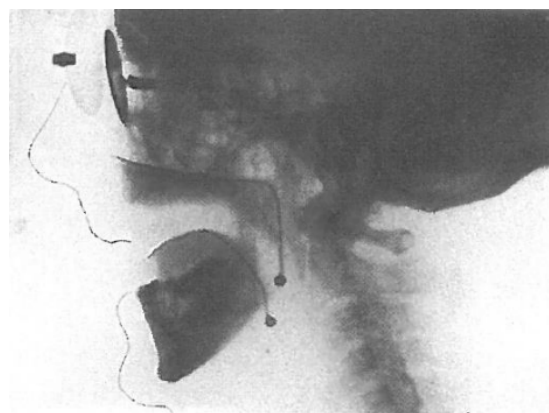


Figure 2.11: X-ray of Stephen Jones producing cardinal [a] with two metal chains outlining the tongue and velum (S. Jones, 1929: 151).

be felt, but of course that statement [overlooks] the psychological or imagination factor. And the operator must acquire a ‘bag of tricks’ in order to side-step these. As stated before, some people are so ‘finicky’ that it is actually impossible to open their mouths without disturbing them. [...] Then [the participant] likes to elaborate on his experience, and he spreads the word among his friends. In that event, a later subject’s imagination sometimes runs riot. (Russell, 1928: 67-69)

Russell does however commend his colleague Stephen Jones on his ability to pass a second chain down through the nasal cavity and into the pharynx, to simultaneously image both the tongue and the velum (Fig. 2.11).

The third major issue that X-ray imaging faced was standardisation. Researchers frequently produced series of images which were not internally consistent, e.g. where the position of the head varied between images (Macmillan & Kelemen, 1952: 678). As mentioned previously, the method also revealed individual variation in vocal tract anatomy, making it difficult to provide a unified description of the results. In addition, X-rays were often unaccompanied by audio recordings, meaning that the vowel shown in the image could only be specified as closely as transcription allowed (Catford, 1981: 28).

Thus, although X-ray imaging provided a wealth of information which propelled the phonetic sciences forward, the method suffered from a number of practical and safety concerns (as well as high running costs), and the data was in some senses too detailed and variable to allow for a general theory of vowel articulation. These issues became significant drawbacks in comparison to acoustic analysis, which became available in the 1940s. Although X-ray technologies continued to develop, with several new methods becoming available in the 1960s and 1970s (e.g. digital fluoroscopy, computed tomography, and X-ray microbeam) (Stone, 2010), increased awareness of the radiation risk led to a gradual decline in X-ray imaging for phonetic purposes, and new data does not seem to have been collected since the 1990s (Ericsson, 2005).

### 2.4.3 Spectrography and acoustic analysis

With the invention of the sound spectrograph (Koenig, Dunn, & Lacey, 1946; Potter, Kopp, & Green, 1946), acoustic information rapidly became widely available, and for the first time it was possible to carry out acoustic analyses on a large scale (Ladefoged, 1967: 74). Although several studies benefited from combining spectrograms with X-ray data (e.g. Chiba & Kajiyama, 1941; Delattre, 1951), the relative ease and low cost of spectrography meant that it quickly overtook articulatory methods, and it has remained the prevalent method of vowel analysis since around the 1950s (Wood, 2019).

An important benefit of spectrography, mentioned earlier in this chapter, was that it aligned with existing vowel theory. Shortly after the introduction of the spectrogram, Essner (1947) and Joos (1948) discovered that if vowels were plotted according to their first two formant values, their positions closely resembled those on the vowel quadrilateral. To some extent, articulatory studies also supported the mapping of F1 onto tongue height, and F2 onto tongue advancement (e.g. Delattre, 1951), although Delattre acknowledges that it is not as simple as a one-to-one-relationship (Delattre, 1951: 865). The complexity of articulatory data led to tensions between the acoustic and articulatory approaches, and although most phoneticians were aware of the articulatory inaccuracies of the vowel quadrilateral “most just continued to use it (with perhaps and apologetic disclaimer)” (Wood, 1982: 18). This preference was sometimes so strong that X-ray images contradicting the vowel quadrilateral went unpublished, because “the authors either could get nothing out of them, or thought it best not to make them public” (Russell, 1928: 53). The consistency and conceptual clarity of the vowel quadrilateral, its relationship to the vowel formants, and rapid technological advances in acoustic methods, thus made spectrography highly accessible both practically and theoretically, which led to a move away from articulatory methods.

As acoustic research progressed, a number of important discoveries were made about the acoustic properties of speech. For example, instead of viewing the oral cavity as having two separate resonance chambers, phoneticians began to have a more holistic view of this space. Experiments involving speech synthesis and vocal tract modelling (e.g. Fant, 1960; Stevens & House, 1955) resulted in the development of one of the cornerstones of present-day phonetics: source filter-theory. This is the theory that “speech involves a source function and a vocal-tract filtering process, i.e. a raw material and a sound shaping” (Fant, 1981: 21), and that these two stages can be analysed separately. This theory allowed for simpler descriptions of the vocal tract, which could be reverse-engineered from the acoustics to arrive at possible articulatory settings.

Another major development, which articulatory phonetics had not managed to achieve, was inter-speaker normalisation. Peterson and Barney (1952) note that, while absolute formant values may vary greatly from speaker to speaker, there is a systematic pattern for how formants are distributed, causing us to perceive different speakers’ vowels as ‘the same’. Source-filter theory made it possible to develop a number of normalisation methods which remove physiological

variation (usually from the source), while retaining important dialectal and social information (usually from the filter), e.g. disregarding differences in voice pitch, while retaining the relationships between the formants created by similar constrictions in the vocal tract. Normalisation was an important focus from the 1950s onwards (Strange, 1989), and gave rise to a number of normalisation methods (e.g. Labov, Ash, & Boberg, 2006; Lobanov, 1971; Nearey, 1977; Syrdal & Gopal, 1986; Watt & Fabricius, 2002), all of which became increasingly available, and increasingly automated, as technology improved.

Today, acoustic vowel analysis still relies on these foundations: the correspondence between the vowel quadrilateral and the frequencies of the first and second formants; the theoretical knowledge that speech sounds can be divided into a sound source and a vocal tract filter; and the separation of source and filter in comparing acoustic materials across speakers. These frameworks and theories were popularised by prominent phoneticians such as Fant (1960) and Ladefoged (1962), and acoustic analysis gradually became widely available to both students and professionals through free software such as Praat (Boersma & Weenink, 1995). Meanwhile, articulatory methods have continued to develop, but at a much slower pace, and without the same degree of advancement in vowel theory.

#### **2.4.4 New articulatory methods**

Due to the difficulties associated with recording and analysing articulatory data, there has been a tendency for phonetic research to use acoustic analysis as a proxy for articulatory information (Ladefoged, 1967: 70). This practice has created the problem that articulatory behaviour, particularly with regard to vowels, is still not very well understood, and there are still no established frameworks for studying vowels articulatorily. At present, however, articulatory vowel study appears to be undergoing a revival, as new imaging and tracking methods are becoming more available (Stone, 2010). Some of the most frequently used techniques for studying vowel sounds include ultrasound tongue imaging (UTI), magnetic resonance imaging (MRI), and electromagnetic articulography (EMA).

##### **Ultrasound tongue imaging**

Ultrasound tongue imaging has been available since the 1980s (e.g. Shawker, Sonies, Stone, & Baum, 1983), and is based on the same technology as pre-natal ultrasound imaging (Donald, Macvicar, & Brown, 1958). UTI creates a video image of the tongue surface (Fig. 2.12) using echolocation. An ultrasound probe, placed under the chin, projects high-frequency soundwaves which travel through the soft tissue and are reflected by the air barrier on top of the tongue (Stone & Lundberg, 1996: 3,729). This technique is frequently used to study both vowels and consonants, and has also been used for speech therapy (e.g. Cleland, Scobbie, Heyde, Roxburgh, & Wrench, 2017; Preston et al., 2017; Shawker et al., 1983).

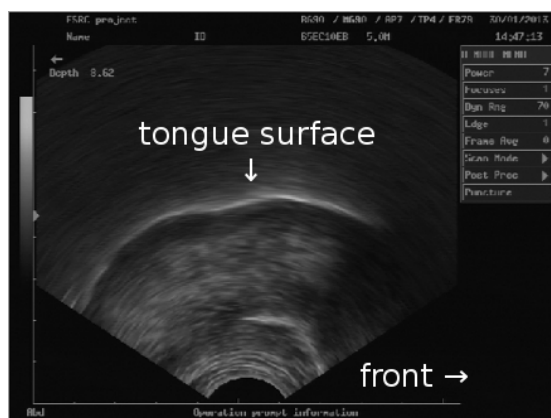


Figure 2.12: Ultrasound image of the tongue surface, speaker facing right.



Figure 2.13: MRI image of the vocal tract, with added tongue outline (Zhou et al., 2008: 4,467).

UTI rose in popularity during the 1990s and is currently undergoing rapid development, as ultrasound equipment is becoming increasingly affordable and portable, making it suitable for phonetic fieldwork. The method is particularly useful for studying vowels, as it produces an image of the entire tongue contour. It also has the advantage of being non-invasive, while producing a sufficiently high frame rate to study natural speech. Since UTI is the method used to collect articulatory data in this thesis, a fuller description and evaluation of this method will be provided in Chapter 5.

### Magnetic resonance imaging

Magnetic resonance imaging emerged as a technique in the 1970s (Lauterbur, 1973) and began to be used for speech research in the 1980s (e.g. Baer, Gore, Boyce, & Nye, 1987)). MRI relies on electromagnetic fields to manipulate and detect the presence of hydrogen atoms, using this information to create an image based on the distribution of water in the body (Stone, 2010: 14-15). MRI images are similar to X-rays in the sense that they can provide a cross-section of the whole vocal tract (Fig. 2.13), and the technique has therefore “replaced X-ray for many research applications” (Stone, 2010: 15). For speech research in particular, MRI is useful because it images both soft and hard tissue, and is able to provide three-dimensional information through incremental ‘slices’ of the imaged object. MRI is currently a popular tool for phonetic research, but is less accessible than UTI and EMA, primarily due to its high cost. MRI also has a relatively low frame rate, rendering it less suitable for studying natural speech, which may be too fast (Stone, 2010: 16), but this problem is likely to be resolved within the next few years. MRI also has the added difficulties that participants are required to lie down, which affects tongue posture (Kitamura et al., 2005). The method also produces a great deal of noise while imaging, which makes it difficult to capture simultaneous speech recordings.

### Electromagnetic articulography

Electromagnetic articulography was developed for speech research in the 1980s (Schönle et al., 1987) and is a point-tracking method, as opposed to an imaging technique. EMA uses alternating electromagnetic fields to track the movements of metal coils, attached to the articulators, within a magnetic field (Stone, 2010: 25). One or several ‘reference coils’ are usually attached to a fixed point, such as the bridge of the nose, while other coils are attached to different parts of the tongue and lips. EMA has mostly been used to study dynamic movement in the vocal tract, which it is able to do with great accuracy, and at a fast sampling rate (Stone, 2010: 25). This method is frequently used for both phonetic research and speech therapy, and produces an output of single points (Fig. 2.14), from which “the behavior of the entire articulator is largely inferred” (Stone, 2010: 27). This inference is often cited as one of the largest drawbacks of EMA. In addition, EMA is also more invasive than UTI or MRI, in that the metal coils must be attached to the inside of the mouth, which may affect articulation.

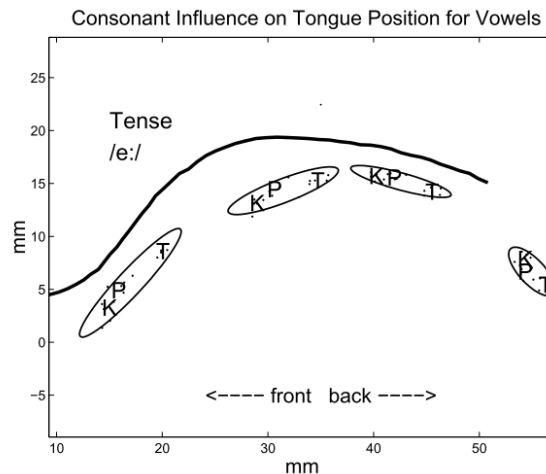


Figure 2.14: Tongue measurement points from EMA (solid line represents the palate) (Hoole & Nguyen, 1997: 182).

### Summary of new articulatory methods

As this section has shown, UTI, MRI, and EMA produce very different types of data, with their own sets of benefits and drawbacks. For this reason, the choice of technique usually depends on the resources available, and the specific focus of the research. Nevertheless, all three methods are useful for studying vowel articulation, as they do not rely on contact between the tongue and other articulators, and are mostly non-intrusive. These methodologies provide a great deal of previously unavailable information, which can provide new insights into vowel production, particularly if used in combination with acoustic data.

## 2.5 Necessity of articulatory vowel data

Due to the rough correspondence between tongue position and formant values (e.g. Delattre, 1951; Joos, 1948; Stevens & House, 1955), as well as the overlap in acoustic and articulatory terminology, there has been a tendency for phonetic studies to make assumptions about vowel articulation from acoustic data. However, in the case of vowels like Vi-by-i, where articulation has been disputed, or where the acoustic values could be achieved in more than one way, acoustic data alone cannot disambiguate how the vowel is produced. Beyond this issue, there are two important theoretical reasons why articulatory vowel data is useful: Firstly, we know that there is a complex relationship between articulatory gesture and acoustic output, not only in terms of the tongue, but in the combination of multiple parameters, yet this issue remains fairly under-researched. Secondly, in the articulation of any sound, there is the possibility of individual variation (in anatomy or speech gesture), which may be masked by articulatory trade-off. Combining acoustic and articulatory data allows us to investigate these phenomena in more detail, contributing to a better fundamental understanding of vowel production.

### 2.5.1 Critique of the vowel quadrilateral

The vowel quadrilateral has been described as “lack[ing] [...] physiological realism” (Wood, 1982: 1) and having a “weak predictive capability” of overall tongue position (Wood, 1982: 22), despite the established correlation between the highest point of the tongue and the frequencies of the first two formants. In reality, the correspondence between tongue gesture and acoustics is complex, and Delattre (1951) encourages phoneticians to study the overall acoustic effects of each articulatory setting, as “it will probably not be possible to examine exactly to what extent a certain formant can be assigned to a certain cavity” (Delattre, 1951: 865). Using F1 and F2 to predict the height and backness of the tongue may be mostly correct, but over-reliance on this model to describe vowel articulation can be misleading. This section will outline some of the more problematic aspects of the vowel quadrilateral, demonstrating the need for more detailed articulatory data.

#### Tongue shape

Since there is “no handy landmark on the tongue to serve as point of reference” (Clark, Yallop, & Fletcher, 2007: 22), the vowel quadrilateral tends to use the highest point of the tongue as an indicator of the point of greatest constriction in the oral cavity. The issue with this approach is that it assumes an arched tongue shape, and disregards the possibility that constriction could occur elsewhere, either instead or simultaneously. For example, due to the slope from the hard palate down to the alveolar ridge, a raised tongue tip may constitute the main constriction, while still being lower than the highest point of the tongue body. Similarly, vowels may be produced with tongue tip raising or root retraction in addition to their original vowel quality. Wood (1982)

states that pharyngeal contraction is a common and important component of low vowels such as [a] (Wood, 1982: 20), and Björsten and Engstrand find that Viby-i can be produced with or without tongue tip raising (Björsten & Engstrand, 1999: 1,959).

The assumption that the tongue is arched can also be problematic; non-arched tongue shapes can be seen e.g. in the X-ray images of S. Jones (1929) producing the cardinal vowels, an issue which is referred to in Ladefoged's comments about [a] and [ɑ] having different tongue shapes (Ladefoged, 1967: 71). In my own previous work, I have also observed that Viby-i can sometimes be produced with a “double-bunched” (saddle-shaped) tongue gesture (Westerberg, 2016), reminiscent of the “bunched” shapes used by some speakers of American or Scottish English for /r/ (e.g. Delattre & Freeman, 1968; Lawson et al., 2014), or the tongue shapes for pharyngealised sounds in Arabic (e.g. Altairi, Brown, Watson, & Gick, 2017).

Thus, although the highest point of the tongue has long been canonical shorthand for overall tongue position, in terms of articulation, “the position of the highest point of the tongue is not a valid indicator of vowel quality” (Ladefoged & Johnson, 2015: 230), as it disregards potentially important information about the effect of overall tongue shape on vowel production.

### **Midsagittal view of the tongue**

Most models of vowel articulation only represent the tongue in two dimensions – height and backness – but “the tongue is a mobile and polymorphous mass” (Catford, 1994: 132), and vowels are also affected by three-dimensional behaviour, such as bracing, grooving, and lateralisation. ‘Bracing’ refers to lateral contact between the tongue and the upper molars, ‘grooving’ refers to a central groove along the length of the tongue, and ‘lateralisation’ refers to the release of air on one or both sides of the tongue. One of the reasons why we know very little about these behaviours is that the early work which laid the foundations for our understanding of vowel production (preliminarily using X-ray) tended to rely on midsagittal images. Although three-dimensional tongue models have been presented by e.g. Gick (2019) and Wrench and Balch (2015), to my knowledge, there are no articulatory studies directly investigating the effect of three-dimensional tongue behaviour on acoustics.

A few studies have however demonstrated that vowels can display various kinds of both bracing and grooving. Stone and Lundberg (1996) find that all American English vowels, but particularly /i, e, æ/, are produced with some degree of lateral contact with the upper teeth (Stone & Lundberg, 1996: 3,733-3,734). The vowels also displayed different types of grooving, which sometimes appeared to be deeper at the back of the tongue (Fig. 2.15). Grooving in American English vowels has also been shown in manual palatograms produced by Russell (1928: 318-348).

Despite this three-dimensional behaviour, Russell (1928) nevertheless claims that the midsagittal dimension is the most important aspect of vowel articulation (Russell, 1928: 119), perhaps alluding to the tongue's hydrostatic properties. The mass of the tongue can be compared



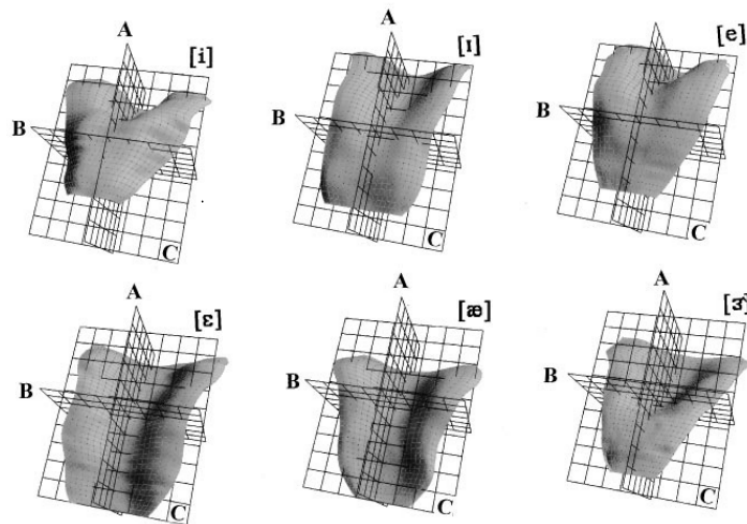


Figure 2.15: Three-dimensional reconstructions of the tongue surfaces for English vowels and schwa (Stone & Lundberg, 1996: 3,732).

to a water balloon, in that when one part is displaced or manipulated, the overall shape of the tongue is affected (Gick, Wilson, & Derrick, 2013: 156). To some extent, it may therefore be possible to predict the three-dimensional shape of the tongue from two-dimensional data (see Wrench & Balch, 2015).

Although the issue of three-dimensional tongue behaviour is addressed here, it was not possible to provide coronal ultrasound data for the present study, as there is still no standardised way of locating the appropriate position for the ultrasound probe in this dimension. Since the available ultrasound equipment is only able to collect two-dimensional ‘slices’ of data, the researcher has to determine where along the length of the tongue the ultrasound image should be taken, and this location may vary depending on the participant and the vowel studied. For this reason, coronal ultrasound data was not collected, but I look forward to seeing this aspect of ultrasound research developing in the future.

### Lip-rounding

Although the vowel quadrilateral focuses primarily on the tongue (an issue which will be discussed further in Section 2.5.2), it also displays the vowels according to lip-rounding. However, it only provides two options (rounded vs. unrounded), which is problematic for a language like Swedish, which distinguishes between three degrees of lip-rounding. For example, a gradient difference in lip protrusion is sometimes said to be the main factor distinguishing the Swedish vowels /i:/, y:/, ʉ:/ (Engstrand, 1999: 141). The low F2 of /ʉ:/ could thus be caused by lip-rounding rather than tongue backing (or a combination of the two), meaning that its placement as a central vowel on the quadrilateral is potentially misleading. Similarly, Ladefoged and Johnson find that back vowels tend to become increasingly rounded as their height increases (Ladefoged

& Johnson, 2015: 230), which may be why the articulatory distances between the back vowels is smaller than those of the front vowels (Ladefoged & Maddieson, 1996: 283-284).

### **2.5.2 Vocal tract parameters contributing to vowel acoustics**

While the interpretation of vowel formants relies to a great extent on tongue posture, the acoustic properties of any sound “are based on the entire articulatory space” (Scobbie, Stuart-Smith, & Lawson, 2012: 105), which may include the effects of other articulators, such as the lips, jaw, velum, tongue root, and larynx, as well as the oral and pharyngeal cavities. These factors affect the overall resonances of the vocal tract and shape the acoustic output, but are often ignored even in articulatory work. This section will briefly outline some of the articulatory factors besides the tongue which may affect vowel acoustics.

#### **Lip position**

Lip protrusion changes the acoustic properties of the vocal tract by lengthening it, and by reducing the size of the mouth opening (Rosner & Pickering, 1994: 22). Acoustically, this results in a lowering of the formants, particularly F2 (Rosner & Pickering, 1994: 42). As mentioned above, the effects of lip-rounding may sometimes be misattributed to other factors, and articulatory analysis therefore benefits from combining tongue and lip data. By doing so, previous studies (e.g. Lawson et al., 2019) have been able to reveal instances of articulatory trade-off, as will be discussed in Section 2.5.3.

#### **Oral cavity shape**

A great portion of the oral cavity, e.g. the teeth, alveolar ridge, and hard palate, are static, and do not change the resonances within an individual speaker’s vocal tract. However, the size and shape of the oral cavity may still vary between speakers. For example, Moisik and Dediu (2017) observed that many speakers do not have a prominent alveolar ridge, and that this could influence the way that they produce certain speech sounds. In addition, ultrasound images frequently reveal that the shape of the palate can vary broadly from speaker to speaker (e.g. Bakst, 2016). This kind of data indicates that individual speakers have unique vocal tract resonances, and in some cases, that they can adapt their articulatory gestures to compensate for these differences (Bakst, 2016).

#### **Jaw opening**

The size of the oral cavity can also be affected by the degree of jaw opening. Although the jaw “cooperates with the tongue in producing the desired area function” (Lindblom & Sundberg, 1971: 1166), this nevertheless means that vowels with more jaw opening use a fundamentally

different acoustic chamber than vowels with less jaw opening. Opening the jaw has the effect of enlarging the front part of the oral cavity, while reducing the pharyngeal cavity (Rosner & Pickering, 1994: 24). Low vowels such as [a], which are more prone to jaw opening, may therefore use different types of articulatory gestures than high vowels such as [i]. This difference is perhaps also related to the varying degree of lip-rounding seen for back high vowels, as mentioned above.

### **Velar constriction and nasalisation**

The oral and pharyngeal cavities are both affected by velar constriction, which narrows this part of the vocal tract. Velar constriction has been observed in high back vowels (Rosner & Pickering, 1994: 42), and may be one of the reasons why back vowels have a smaller articulatory space than front vowels (Ladefoged & Maddieson, 1996: 283-284).

Vowels can also be nasalised by lowering the velum and allowing air to escape through the nasal cavity. Nasalisation adds a second resonating chamber, which mainly affects the bandwidths and amplitudes of the formants (K. Johnson, 1997: 158-159).

### **Tongue root retraction and pharyngealisation**

Tongue root retraction and pharyngeal constriction both reduce the size of the pharyngeal cavity, which affects the acoustic resonances in this space. Root retraction has the additional effect that it pulls on the tongue, which can affect its overall shape (e.g. Altairi et al., 2017). Although pharyngeal constriction is mainly associated with languages that use this feature for phonemic contrast (e.g. Arabic), tongue root retraction has also been observed in low back vowels in Swedish (Lindblom & Sundberg, 1971: 1,167).

### **Larynx movement**

Larynx lowering has a similar effect to lip rounding, in that it lengthens the vocal tract, while larynx raising shortens the vocal tract (Sundberg & Nordström, 1976). Since the change mainly affects the pharynx, the acoustic consequences are stronger for formants that rely on the back cavity for their resonances.

### **Summary of vocal tract parameters**

As these examples have shown, the vocal tract is a dynamic space, with a great number of potential articulatory settings. Thus, the assumption of a one-to-one relationship between formant values and the highest point of the tongue is misleading. This phenomenon is demonstrated by Fant (1960), stating in relation to tube models that “specific shifts in a specific formant for a given vowel quality may occur when either the front or the back cavity changes” (Fant, 1960, in Rosner & Pickering, 1994: 39). In other words, a tube model may be constricted at the front

or at the back, and produce a similar acoustic result. Furthermore, “alterations in cavity characteristics will not affect all vowels identically” (Fant, 1960, in Rosner & Pickering, 1994: 39), because their baseline settings may be different.

In summary, it is difficult to predict exactly what combination of articulatory gestures give rise to a set of formant values. From an articulatory perspective, it is also difficult to decide which kinds of articulatory data should be collected to answer this question. For example, in the case of *Viby-i*, the low F2 could be created by several factors, such as lip rounding, tongue backing, or a lowered larynx. Personal correspondence has also raised formant bandwidth and amplitude as potential causes of this vowel’s ‘damped’ quality. Although this thesis focuses primarily on the articulatory settings of the tongue, and to a lesser extent the lips, it is important to note that there may be additional factors contributing to the production of *Viby-i*, which are beyond the scope of this research, but which may be relevant for future study.

### **2.5.3 Articulatory trade-off**

As the previous sections have indicated, “the relation between articulatory parameters and acoustic output is not linear” (Stevens, 1972, in Heinz, 2011: 12), and this issue gives rise to the phenomenon of articulatory trade-off (e.g. Guenther et al., 1999), also known as articulatory compensation (e.g. Wright & Riordan, 1980). These terms describe the ability of speakers to produce equivalent acoustic outputs using different articulatory strategies. Speakers may use this ability to compensate for individual differences in the vocal tract, for specific vocal tract settings, or sometimes as a consequence of sociolinguistic factors, as will be demonstrated below. The extent to which trade-off occurs in speech is not yet known, and the issue remains under-researched, but this section will provide a few examples of cases where articulatory data has revealed trade-off to be present.

#### **Individual vocal tract differences**

Much of the research on articulatory trade-off comes from speech and language therapy, as speakers sometimes have to adopt strategies to deal with anatomical differences which may make speech difficult. For example, instances of compensation have been observed in speakers with cleft palate (Trost, 1981), tongue-tie (Block, 1968), and patients who have undergone glossectomy (Barry & Timmermann, 1985). In these cases, trade-off may be more relevant to producing sufficient contrast between speech sounds, rather than recreating the acoustic output exactly, and the extent to which it is successful may depend on the severity of the condition. However, all speakers have a unique vocal tract anatomy, and the extent to which we use articulatory trade-off to compensate for this is not known, as articulatory data has largely been unavailable.

### Specific vocal tract settings

Speakers are also able to use articulatory trade-off to adapt to short-term contexts, or to minimise variability in their own speech. For example, ‘true’ trade-off has been observed in opera singers, who are trained to modify their tongue positions to compensate for the greater degree of jaw opening associated with this singing technique (Nair, Schellenberg, & Gick, 2015). This phenomenon can partly be explained by Lindblom and Sundberg’s theory that the tongue co-operates with the jaw “in producing the desired area function” (Lindblom & Sundberg, 1971: 1,166).

Wright and Riordan (1980) make similar findings in their study where they measure the formants, as well as the distance between the tongue and palate, in a number of speakers producing vowels with and without bite blocks. However, these authors dispute Lindblom and Sundberg, claiming that “speakers invoke a set of strategies, more complex than simply attempting to maintain the area function, in order to achieve an acoustic target” (Wright & Riordan, 1980: 31). Unfortunately, as these findings were presented orally, further details about what strategies speakers might have used are not currently available.

Finally, Guenther et al. (1999) found that speakers of American English could produce equivalent acoustic outputs for /r/ using either a “bunched” (dorsal) or “retroflex” (apical) tongue shape (Guenther et al., 1999: 6), and that both these shapes could occur within the same speaker (Guenther et al., 1999: 11). The authors suggest that speakers used an “acoustic target” to minimise acoustic variation when producing /r/ in different linguistic contexts. In other words, speakers used different kinds of /r/ gestures to compensate for coarticulatory changes to the vocal tract. Instead of choosing from a set of “canonical vocal tract shapes”, Guenther et al. state that speakers used a “continuum of tongue shapes for /r/ across contexts”, suggesting that they were aiming for a particular acoustic output (Guenther et al., 1999: 17).

### Sociolinguistic variation

Other studies investigating /r/ in rhotic varieties of English have found that the articulatory gesture used for this sound can also be geographically or socially stratified. Delattre and Freeman (1968) found that English and American /r/ could be divided into at least eight articulatory categories, and that speakers’ articulation of this sound was correlated with their regional accent (Delattre & Freeman, 1968: 66). However, most accents displayed a range of tongue gestures, meaning that this relationship was not entirely categorical (Delattre & Freeman, 1968: 57).

Lawson et al. (2013) similarly found that Scottish speakers were socially stratified in their choice between a “bunched” and a “tip-up” gesture for /r/. In this study, middle-class speakers were more likely to use a bunched articulation, while working-class speakers were more likely to use a tip-up articulation. Both Lawson et al.’s and Delattre and Freeman’s findings are surprising, as the difference between bunched and tip-up /r/ “has been claimed to be inaudible in other varieties of English” (Lawson, Scobbie, & Stuart-Smith, 2011: 268). While some acoustic

differences exist in the relationship between F4 and F5 for bunched and tip-up /r/ (Zhou et al., 2008), this information is not readily accessible to listeners, raising the question of how speakers learn these strategies.

Regional variation has also been documented in the production of fronted /u/ in the British Isles. Lawson et al. (2019) found that English and Irish speakers produced [u] with a fronted tongue gesture, while Scottish speakers achieved an equivalent vowel with a low tongue gesture but spread lips. The fact that this sound is regionally stratified is perhaps less surprising, as the lips provide a visual cue for speakers when learning their preferred strategy. Nevertheless, as all speakers were participating in the phenomenon of /u/-fronting, this finding supports the notion that speakers produce vowels with an acoustic target in mind, and that this target can be achieved in multiple ways.

### **Unexplained variation**

In addition to the aforementioned studies, recent articulatory work is also beginning to uncover patterns of articulatory variation which have not yet been explained. For example, Noiray, Iskarous, and Whalen (2014) discovered that the American English vowels /ɪ/ and /e/ sometimes had “flipped” articulations, i.e. /ɪ/ had a lower tongue gesture than /e/ (Noiray et al., 2014: 272). According to Ladefoged et al., “the highest point of the tongue for the vowels /ɪ/ and /e/ is flipped by many speakers”, but the correct phonemic contrast is usually still maintained (Ladefoged et al., 1972, in Noiray et al., 2014: 274). In their analysis, Noiray et al. investigate whether this articulatory flipping is accompanied by equivalent acoustic changes, and find that indeed, “the individual variability in articulation is directly retrieved in the acoustic signal and presumably available to listeners” (Noiray et al., 2014: 280). However, “both the speakers with flips and those without produced vowels that were recognizable as the intended vowel”, meaning that the reversed F1 did not seem to impact listener perception (Noiray et al., 2014: 281). The authors attribute this result to the reversed F1 being overridden by other linguistic or contextual cues (Noiray et al., 2014: 281), but further research on this issue is yet to be carried out. The study also finds indications that vowel flipping is regionally stratified, but the small sample size makes this difficult to establish.

### **Summary of articulatory trade-off**

This section has illustrated the phenomenon of articulatory trade-off, showing that articulatory gestures may not be as uniform as phoneticians have previously assumed. As articulatory methods are becoming increasingly available, we are beginning to discover that speech sounds, perhaps particularly vowels and approximants, may rely on unexpected articulatory strategies. These strategies may be idiosyncratic, context-dependent, or socially meaningful, but the small scale of many articulatory studies makes this difficult to evaluate. These issues highlight the

value of combining acoustic and articulatory techniques, both to make new discoveries, and to test assumptions from acoustic theory.

## 2.6 Chapter summary

As this chapter has shown, the description and classification of vowel sounds based solely on acoustic properties is problematised by a number of factors, such as variation in tongue shape and vocal tract settings, as well as the potential for articulatory trade-off. These factors contribute to the complex relationship between articulation and acoustic output. Due to historical limitations, we still know very little about the articulatory settings of vowels, and the potential variation that may occur below the acoustic signal.

The unusual nature of *Viby-i* brings some of these issues to the forefront, as methods commonly used to analyse other vowels have been unsuccessful in describing how this sound is produced. Specifically, *Viby-i* seems to defy the common assumption that the highest point of the tongue is related to the first two formants. This question can now be investigated using articulatory techniques that have only recently become available.

# Chapter 3

## Swedish vowels

### 3.1 Chapter overview

This chapter provides an overview of Swedish phonology, focusing on Central Swedish vowels, and how they are produced by speakers from different geographic and social backgrounds. The role of Viby-i in this system is addressed briefly, but will be explored further in Chapter 4. The chapter also provides linguistic and social context for the three cities studied in this thesis: Gothenburg, Stockholm, and Uppsala.

The first part of the chapter focuses on the Swedish language as a whole. Section 3.2 provides a general linguistic background to Swedish, summarising some of its distinct features. Section 3.3 outlines the Swedish dialect regions and the main differences between them, as well as their relation to different sociolects. Section 3.4 describes the phonology of Central Standard Swedish, explaining how Viby-i fits into this system.

The second part of the chapter focuses more closely on factors that could influence vowel production in the present sample. Section 3.5 discusses linguistic factors that have been shown to affect vowel quality in Swedish and other languages. Section 3.6 provides a brief overview of sociolinguistic factors that can affect vowel production, as well as different processes of sound change. Finally, Section 3.7 provides a sociolinguistic profile of Gothenburg, Stockholm, and Uppsala, describing relevant aspects of vowel and consonant production in these cities.

### 3.2 The Swedish language

Swedish is the official language of Sweden and one of the official languages of Finland, spoken as a first language by around 9 million people in Sweden and 300,000 people in Finland (Network to Promote Linguistic Diversity, 2016). These numbers represent around 90% of the Swedish population, and around 5% of the Finnish population (Network to Promote Linguistic Diversity, 2016; Statistics Sweden, 2019a). There is also a very small community of native Swedish speakers in Estonia (Network to Promote Linguistic Diversity, 2016).



Swedish belongs to the North Germanic branch of Indo-European languages, and is closely related to Danish and Norwegian, with which it is more or less mutually comprehensible (Boyd, 2010: 282). Phonologically, Swedish is mainly noted for its densely populated vowel system and its use of tonal word accents (sometimes called pitch accents) (e.g. Bye, 2004; Riad, 2006). These features are also selectively found in other Scandinavian languages (Haugen, 1976).

According to Bruce (2010: 24), Swedish is somewhat less standardised than Danish, but more so than Norwegian. As such, spoken Swedish displays a great deal of dialectal variation, but there is still no single officially recognised standard variety in the spoken language. Instead, there are a number of regional standards, towards which speakers are increasingly converging (Leinonen, 2011: 85-86). Leinonen shows that spoken Swedish has recently seen “large-scale on-going levelling” (Leinonen, 2011: 85), which will be addressed in Section 3.3.2. Grönberg (2004) also finds indications that young speakers in rural locations may replace local dialect features with equivalents from nearby urban centres. Since Stockholm and Gothenburg are the two largest cities in Sweden, located in the most densely populated Central Swedish region, they are likely to exert a great influence on Swedish speech in general. Uppsala is located close to Stockholm, and while it is smaller than Stockholm, it has been a historically important location for Central Standard Swedish, as will be discussed in Section 3.7.

### 3.3 Swedish dialects

#### 3.3.1 Dialect regions

Swedish is commonly divided into three broad dialect areas: South Swedish (*sydsvenska*), Central Swedish (*centralsvenska*), and Finland Swedish (*finlandssvenska*) (Bruce, 2010: 25-26). Finland Swedish is primarily spoken on the Finnish West and South coast, e.g. Österbotten, Åland, Åboland, and Nyland (M. Lindberg, 2016). South Swedish is spoken in areas that were historically Danish, e.g. Skåne, Blekinge, and parts of Småland and Halland (Bruce, 2010: 25). The remaining part of Sweden, constituting the vast majority of its landmass, speaks Central Swedish. This grouping reflects a “shared sound identity” for this region (Bruce, 2010: 26), however, it also encompasses a lot of phonetic variation. For this reason, Central Swedish is usually further divided into the following sub-varieties (after Bruce, 2010: 27):

- North Swedish (*norrländska*), spoken in Lappland, Norrbotten, Västerbotten, Ångermanland, Jämtland, Härjedalen, Medelpad, Hälsingland, Gästrikland;
- Dalaberg Swedish (*dalabergslagska*), spoken in Dalarna;
- East Central Swedish (*östmellansvenska*), spoken in Västmanland, Uppland, Södermanland, Närke;

- West Central Swedish (*väst mellansvenska*), spoken in Värmland, Dalsland, Bohuslän, Halland, Västergötland, Östergötland, Småland, Öland;
- Gotland Swedish (*gotländska*), spoken on Gotland.

A map of these dialect regions is provided in Fig. 3.1. This thesis focuses on the East and West Central varieties, which will be described in further detail in Section 3.7.

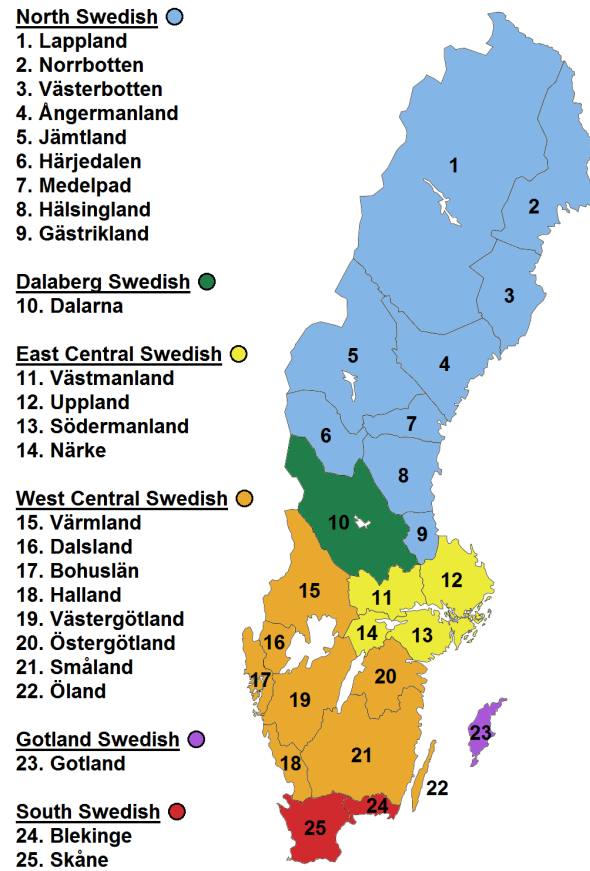


Figure 3.1: Map of Swedish provinces and dialect regions (after Bruce, 2010: 27). Map adapted from Wikimedia Commons (2019).

Some of the most easily classified differences between the Swedish dialect regions (summarised in Table 3.1) include:

- **Realisation of pitch accent.** Swedish has two pitch accent types, commonly referred to as “acute” and “grave” (Garlén, 1988: 138). All words have pitch accents, but the number of minimal pairs is relatively small. A common example is /ánden/ ‘the duck’ (acute accent) and /ànden/ ‘the spirit’ (grave accent) (Garlén, 1988: 138). The pitch contours of the word accents vary from region to region, as shown in Table 3.1. It is not known whether the choice of pitch accent influences vowel quality.
- **Complementary vowel/consonant length.** Many Swedish dialects have the phonological rule that stressed syllables must contain either a phonologically long vowel, or a geminated

consonant (Schaeffler, 2005: 7-8), e.g. /mim/ vs. /mm:/ . Consonant gemination does not differentiate meaning in the same way that vowel length does, but it is nevertheless common in many dialects, perhaps as an additional cue to vowel length, or as a prosodic indicator of e.g. syllable structure. South Swedish does not use this rule, as it does not have consonant gemination (Bruce, 2010: 174).

- **Difference in long/short vowel quality.** In most Swedish dialects, phonemically long vowels have different qualities from their phonemically short counterparts. For example, the letter ‘e’ has two phonemes: long /e:/ and short /ɛ/. These sounds are differentiated not only by duration, but also by vowel quality. As a rule, short vowels tend to be more centralised than their long counterparts (see Engstrand, 1999: 140). In Finland Swedish, long and short vowels are not distinguished by quality, but only by duration (Bruce, 2010: 181).
- **Diphthongal vowel system.** With the exception of South Swedish and Gotland Swedish, most Swedish vowel systems are (at least canonically) monophthongal (Bruce, 2010: 120). A few regions have rural dialects with diphthongal systems, as shown in Table 3.1, but these will not be covered in detail here. However, many systems do use diphthongal vowel offglides, which will be explained further in Section 3.4.
- **Production of /r, l, ʃ/.** The consonant phonemes that differ the most between Swedish dialects are /r, l, ʃ/ (the latter is commonly known as ‘sje’) (Bruce, 2010: 145). For /r/, most Swedish dialects use a trill [r] or tap [ɾ], but many also use other variants, e.g. approximant [ɹ], fricative [ʀ], or uvular [ʁ]. For /l/, speakers may use a clear [l], dark or velarised [ɫ], or flapped [ɾ]. Flapped [ɾ] is usually regarded as rural and low prestige (Bruce, 2010: 160). Finally, ‘sje’ has several realisations, and can be used as both a geographic and social marker. The velarised variant [ʃ] is common in most regions, but the fronted allophone [ɕ] (sometimes [ç]) is usually considered more prestigious (Bruce, 2010: 166). Fronted ‘sje’ creates a merger with another consonant ‘tje’ /tʃ/, causing words like *stjärna* /ʃjɛ:ɳa/ ‘star’ and *kärna* /çɛ:ɳa/ ‘core’ to become homophones.

While prosody, vowels, and consonants all vary regionally, “vowels display the largest degree of variation, and contribute the most to the impression of local colour” in Swedish (Bruce, 2010: 34-35). Even within the dialect regions listed above, there is still a great deal of variation, but detailed accounts of vowel realisations at more local levels are rare. While there are some available dialect resources for Swedish, e.g. the SweDia corpus (A. Eriksson, 2004), the survey by Leinonen (2011), or teaching materials such as Bruce (2010) and Pamp (1978), these do not provide detailed accounts of vowel productions, and their source data is usually restricted. Thus, for the time being, there is no readily accessible dialect atlas of Swedish vowel sounds.

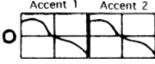
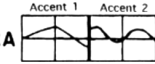
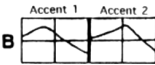
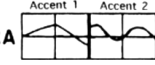
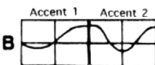
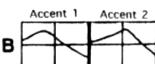
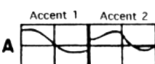
Dialect region	Pitch accent	Compl. V/C length	Long/short V qual.	Diphthongs	/r/	/l/	/c/
Finland		Yes	No	No	[r]	[ɫ]	[ç]
North		Yes	Yes	Variable	[r]	[ɫ] [ɮ]	[ɧ] [ʂ]
Dalaberg		Yes	Yes	No	[r]	[ɮ]	[ɧ] [ʂ]
East C.		Yes	Yes	Variable	[r] [ɹ] [z]	[ɫ] [ɮ]	[ɧ] [ʂ]
West C.		Yes	Yes	Variable	[r] [ɹ]	[ɫ] [ɮ]	[ɧ] [x]
Gotland		Yes	Yes	Yes	[ɹ]	[ɫ]	[ɧ] [ʂ]
South		No	Yes	Yes	[ɹ]	[ɫ]	[ɧ]

Table 3.1: Summary of differences between Swedish dialect regions (after Bruce, 2010; Gårding, 1977). Phonemes listed in order of use.

### 3.3.2 Regional vs. Standard Swedish

As mentioned previously, Swedish has no regionally neutral standard variety comparable to e.g. RP in British English (Leinonen, 2010: 6). Instead, speakers converge towards a number of regional standards. The variety that is most often described in the literature is Central Standard Swedish, which is primarily based on East Central varieties, since this region exerts most of Sweden's social and political power. More or less levelled varieties of East Central Swedish, representing the national standard, are prevalent in broadcast media, and are often described as coming from Mälardalen or Uppsala (Leinonen, 2010: 6).

Central Standard Swedish is associated with higher social status than most other varieties, particularly those from more rural locations, or from other dialect areas (Bruce, 2010: 19). Regional language in Sweden is often described as belonging to one of four ordered levels (e.g. Grönberg, 2004: 30):

- Traditional local dialect;

- Levelled local dialect;
- Regional standard;
- National standard.

Overall, older speakers tend to use more local varieties, middle-aged speakers are more levelled, and young speakers are closer to the regional or national standard (Norrby & Håkansson, 2015: 77). In terms of vowel production, Leinonen (2011: 86) has shown that Swedish is currently ongoing rapid dialect levelling in apparent time, with young people displaying much less dialectal variation than previous generations (Fig. 3.2). This finding is also supported by Svahn and Nilsson (2014).

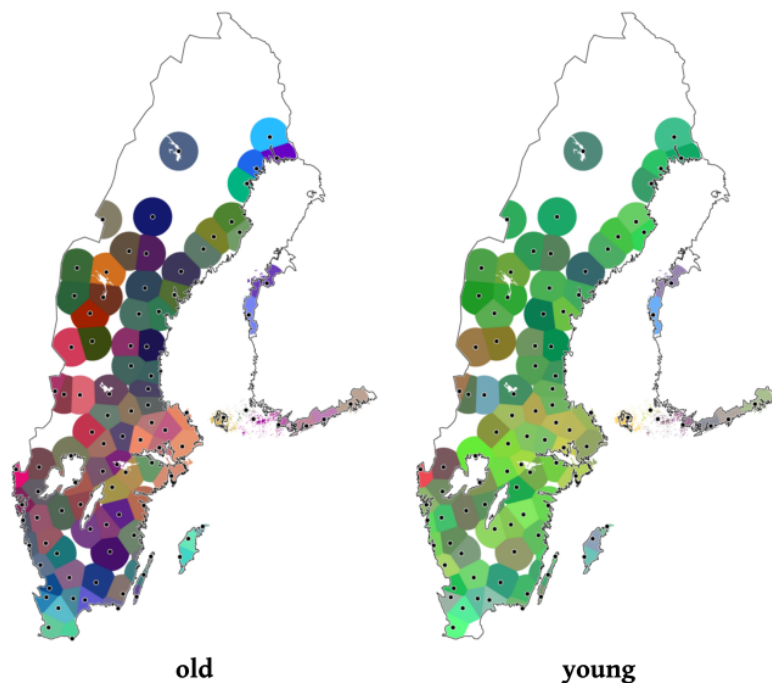


Figure 3.2: Dialectal variation in old vs. young Swedish speakers. Similar colours represent similar linguistic features (Leinonen, 2011: 86).

### 3.3.3 Swedish multiethnolect

In addition to regional dialects, Swedish also has several varieties of urban multiethnolect. The most well-known of these are found in Sweden’s three largest cities: Stockholm, Gothenburg, and Malmö. Although the multiethnic varieties in these cities have some underlying similarities, they do not stem from contact with any one language (Bodén, 2010: 76), and are not the result of incomplete language learning (Gross, Boyd, Leinonen, & Walker, 2016: 228). Instead, they constitute their own varieties of Swedish, which tend to retain “regionally coloured features” from the cities where they are spoken (Bodén, 2010: 77). Multiethnolect is also not exclusively

nor exhaustively used by speakers with immigrant backgrounds (Bodén, 2010: 68). The biggest differences between multiethnolects and local Swedish varieties have been found to be prosody and vowel production (Kotsinas, 1988). However, since multiethnolect has been claimed to not feature Viby-i (Bruce, 2010: 225), this thesis does not analyse vowels from these varieties.

## 3.4 Phonology

### 3.4.1 Vowels

Central Standard Swedish has 18 vowel phonemes, divided into nine long/short pairs (e.g. Engstrand, 1999; Riad, 2014). The orthographic and phonemic use of these vowels is summarised in Table 3.2. The long and short vowels are distinct from each other in both duration and vowel quality, with the exception of [ɛ], which represents the short version of both /e:/ and /ɛ:/ in most varieties.

Figure 3.3 also shows a schematic representation of the Central Standard Swedish vowel space. As the figure shows, there is a great deal of crowding in the mid front and high front part of this system. To help distinguish between the many vowel qualities, Swedish speakers use additional cues, such as lip-rounding and dynamic offglides, described below. Note that Viby-i is not included on the vowel chart, as it is not a recognised feature of Central Standard Swedish.

Orthogr.	Long	Short
i	i: <i>sil</i> ‘strainer’	ɪ <i>sill</i> ‘herring’
y	y: <i>syl</i> ‘awl’	ʏ <i>syll</i> ‘sill’
u	u: <i>ful</i> ‘ugly’	ʊ <i>full</i> ‘full’
e	e: <i>hel</i> ‘whole’	ɛ <i>hell</i> ‘hail’
ö	ø: <i>nöt</i> ‘nut’	œ <i>nött</i> ‘worn’
ä	ɛ: <i>häl</i> ‘heel’	ɛ <i>häll</i> ‘rock face’
a	ɑ: <i>mat</i> ‘food’	a <i>matt</i> ‘weak’
å	o: <i>gråt</i> ‘crying’	ɔ <i>grått</i> ‘grey’
o	u: <i>bot</i> ‘penance’	ʊ <i>bott</i> ‘resided’

Table 3.2: Long and short vowel pairs (after Engstrand, 1999: 140).

### Lip-rounding

Swedish vowels distinguish between three degrees of lip-rounding: spread /i:, e:, ɛ:/, inrounded (or exolabial) /u:, ʊ:/, and outrounded (or endolabial) /y:, ø:, o:/ (Engstrand, 1999: 141). The /ɑ:/ vowel also tends to be slightly rounded (Engstrand, 1999: 141). Lip-rounding is often said to be the main difference between /i:, y:, u:/, but recent articulatory work by Frid et al. (2015) reports that these vowels can also differ from each other in tongue height. This finding suggests that the tongue and lips may have complementary roles in Swedish vowel production.

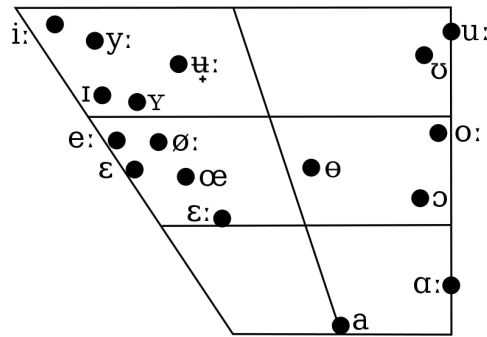


Figure 3.3: Central Standard Swedish vowel space (Engstrand, 1999: 140).

### Dynamic offglides

Although many Swedish varieties are described as monophthongal, most North and Central Swedish speakers use dynamic vowel offglides which manifest themselves differently depending on the vowel (Bruce, 2010: 126-128). Table 3.3 shows the most common offglide patterns in Central Swedish, where high vowels become fricated through progressive narrowing of the vocal tract, and non-high vowels diphthongise towards schwa (Elert, 1995: 40). As the table shows, diphthongisation can also occur in high vowels, either instead of, or in addition to, end-frication.

Dynamic offglides tend to occur at the very end of the vowel, as Swedish vowels reach their target early, and maintain it throughout most of the production (Bruce, 2010: 127). The offglides are believed to help listeners disambiguate vowels, since the crowded vowel space leaves room for perceptual overlap (Bruce, 2010; Elert, 1995).

Vowel	Diphthongised	End-fricated	Both
i:	i:ə	i:j, i:ʃ	i:jə, i:ʃə
y:	y:ə	y:j, y:ʃ	y:jə, y:ʃə
u:	u:ə	u:β	u:βə
e:	e:ə		
ø:	ø:ə		
ɛ:	ɛ:ə		
ɑ:	ɑ:ə		
o:	o:ə		
ʊ:	ʊ:ə	ʊ:β	ʊ:βə

Table 3.3: Offglides for Central Swedish vowels (after Elert, 1995: 40).

### 3.4.2 Consonants

The consonants of Central Standard Swedish, and their most common realisations, are shown in Table 3.4. The consonant system is similar to that of many other European languages, with the exception of the more unusual consonants /ç/ ‘tje’ and /ʃj/ ‘sje’ (Engstrand, 1999: 140). As

mentioned above, the most variable consonants across Swedish regions are /r, l, ʃ/.

Since this thesis does not specifically investigate consonants, but only uses consonant context as a way of studying vowel production, the Swedish consonant system will not be discussed in detail. However, a brief overview of the influence of consonants on vowel production will be provided in Section 3.5.

	Bilabial		Labio-dental		Dental			Alveolar		Palatal		Velar		Glottal	
Plosive	p	b			t	d						k	g		
Approx.							l				j				
Fricative			f	v	s					ç		ʃ		h	
Trill/tap								r							
Nasal		m			n							ŋ			

Table 3.4: Central Swedish consonants and common realisations (after Engstrand, 2004: 167).

## 3.5 Linguistic influences on vowel production

### 3.5.1 Consonant context

Consonant context is known to affect Swedish vowel quality in two ways: Firstly through the phonological rule that /ɛː, ɛ, øː, œ/ are lowered before /r/ and the rhotic clusters /rt, rd, rn, rs, rl/ (realised as [t, d, ŋ, ʃ, l]) (Engstrand, 1999: 141). An example can be seen in the realisation of /øː/ in *öga* [øːga] ‘eye’ vs. *öra* [œːra] ‘ear’.

Secondly, there are coarticulatory effects, whereby the vowel formants (and tongue gesture) assimilate towards the consonant context. This phenomenon has been shown in both CVC and VCV syllables in Swedish (Lindblom, 1963; Öhman, 1966), indicating that consonant coarticulation can affect the vowel from either direction. However, in general, coarticulatory effects tend to be stronger from the preceding consonant than from the following consonant (Recasens, 1999: 98). Short vowels are also more likely to be affected by coarticulatory effects than long vowels (Rosner & Pickering, 1994: 92-94).

### 3.5.2 Vowel duration

This study focuses on stressed, phonemically long vowels in word list speech, meaning that the vowel duration is inherently long. However, even within phonemically long vowels, there may still be differences in raw duration, which could in turn affect vowel quality. As mentioned above, vowels with shorter durations are more likely to undershoot their articulatory target and become reduced or centralised (Rosner & Pickering, 1994: 342-343). This phenomenon has also been documented in Swedish (Lindblom, 1963: 1,780).



There are many factors known to affect raw vowel duration in Swedish, many of which also hold true for other languages (see van Santen, 1992). Some of these factors include:

- **Phonological length.** This definition includes the distinction between long and short vowel phonemes, as well as complementary length between vowels and consonants. In Swedish, “a [phonemically] long vowel is always followed by a short [non-geminated] consonant ... and a [phonemically] short vowel is followed by a long [geminated] consonant or a consonant cluster” (Schaeffler, Wretling, & Strangert, 2002: 3). Naturally, the phonological length of the vowel (e.g. /ε:/ vs. /ε/) also affects its duration.
- **Vowel height and backness.** Back vowels generally have longer durations than front vowels (Garlén, 1988: 120), and low vowels generally have longer durations than high vowels (Lindblom, 1967: 1-3).
- **Consonant context.** Vowels are durationally longer before voiced consonants than before voiceless consonants (Lindblom, 1967: 1), and likewise longer before fricatives than before plosives (Garlén, 1988: 120). Vowel duration is also affected by consonant place, being longest before alveolars, and increasingly shorter before dentals, labials, and palatals respectively (Garlén, 1988: 120).
- **Stress and syllable structure.** Vowel duration is longer in stressed syllables than in unstressed ones, and longer in monosyllables than in polysyllables (Engstrand, 2004: 207). However, the number of syllables has been found to have a very small effect on Swedish vowel duration compared to the effects of phonological vowel length and consonant voicing (Erikson & Alstermark, 1972: 56). Vowels also undergo word-final lengthening, with longer durations in word-final contexts than in word-medial ones (Engstrand, 2004: 207-208).
- **Word and sentence context.** Vowels have long durations when produced in isolation, shorter when produced in a word, and even shorter when produced in a carrier sentence (Rosner & Pickering, 1994: 367). Within words, vowel duration is longest in phrase-final positions (van Santen, 1992: 535). Content words also tend to have longer vowel durations than function words (Rosner & Pickering, 1994: 367).
- **Speech rate and style.** Slow speech naturally produces longer vowel durations than fast speech (Garlén, 1988: 119). Accordingly, speech styles associated with a slow speech rate (e.g. reading, clear speech, or formal/careful speech) tend to have longer vowel durations than styles associated with fast speech (e.g. spontaneous, conversational, or informal/emotionally involved speech) (Rosner & Pickering, 1994: 342-343).

In this study, the influences on vowel duration and consonant context on Viby-i production is investigated through a word list, which will be described further in Chapter 5.

## 3.6 Sociolinguistic influences on vowel production

Beyond dialectal and linguistic variation, there are also a number of social factors that can affect speech. Some of these factors include age, gender, socioeconomic class, identity and attitudes, dialect contact, and speech styles (Chambers, 2003; Eckert, 2012; Labov, 1994, 2001). Observations made about these factors in the general sociolinguistic literature have been found to also hold true for Swedish (Bruce, 2010: 19-20). This section will describe some of the common patterns pertaining to the effect of these sociolinguistic factors on speech, and discuss how they might apply to a Swedish context.

### 3.6.1 Age

As mentioned in Section 3.3, a speaker's age can affect the linguistic features that they use. In Swedish, as in many other languages, older speakers are usually closer to the local dialect, while younger speakers are closer to the standard (Norrbj & Håkansson, 2015: 77). Leinonen (2011) has shown that young Swedish speakers are becoming increasingly levelled, meaning that they display less dialectal variation than previous generations, and thus behave more similarly to each other.

Age-based differences in linguistic behaviour can be interpreted in different ways. For example, a common approach to studying language change is to investigate linguistic variables in “apparent time” (Labov, 1972). This approach assumes that speakers' language use does not change much over the course of their life. Thus, an older speaker can provide insight into what the language looked like when they were young. This type of research is especially useful when real-time data (i.e. historical recordings) are not available.

The other interpretation of age-based differences would be to assume that speakers *do* change their language use over time. For example, each generation of speakers may go through similar patterns of moving closer to and further away from standard language over the course of their lives; a phenomenon known as “age-grading” (Hockett, 1950). Age-grading is usually described as having three stages:

First, in childhood, the vernacular develops under the influence of family and friends. Second, in adolescence vernacular norms tend to accelerate beyond the norms established by the previous generation, under the influence of dense networking. Third, in young adulthood standardization tends to increase, especially for the sub-set of speakers involved in language sensitive occupations in the broadest sense of the term. (Chambers, 2003: 171)

According to Chambers, speakers' language use stabilises in adulthood, and older speakers are usually “impervious” to linguistic innovation (Chambers, 2003: 203). However, in the same way that working-age speakers may adjust their language towards the standard to make themselves

more marketable (e.g. Macaulay, 1977), elderly speakers who have left the job market may have a more relaxed attitude to language conventions, and thus use more non-standard features (Eckert, 1997: 165). Older speakers can also participate in “communal change” (Labov, 2001: 76), whereby they simultaneously undergo age-grading and adopt new features from the speech community. Research on this topic is relatively scarce, but evidence of communal change in elderly speakers has been shown by e.g. Sankoff and Blondeau (2007) in Canadian French, and Harrington (2006) in RP.

In this thesis, speaker age will be used to analyse the use of *Viby-i* in apparent time. In other words, if older speakers use this sound, it will be taken as an indication that *Viby-i* has existed in their variety since they were young. Although it is possible that speakers have adopted this feature later in life, real-time data would be required to confirm this.

### 3.6.2 Gender

Another factor that can influence linguistic behaviour is gender. In the sociolinguistic literature, women are often found to gravitate towards standard and prestige forms to a greater extent than men (Chambers, 2003: 139). However, Eckert (1989: 253) points out that these gender differences may be greater or smaller depending on other social contexts, such as age and socioeconomic class.

At the same time, women are also known for “leading” when it comes to sound change (e.g. Eckert, 1989; Labov, 2001). This phenomenon has also been demonstrated in Swedish (e.g. Gross et al., 2016; Kotsinas, 2007; Sundgren, 2001). As mentioned above, gender often intersects with other variables, such as age and social class, in demonstrating change. For example, Kotsinas (2007) finds that young upper-class women are the most likely to use innovative language variants, including *Viby-i*. The gendered use of *Viby-i* in this study will be used as an indication of its high prestige, as well as possible change in progress.

Recent developments in gender theory have informed the way that sociolinguistics investigate gender differences today. Firstly, gender is not the same thing as sex, but refers to the social role performed by the speaker (West & Zimmerman, 1987). Secondly, gender is not only limited to male and female, but may include a spectrum of different identities (e.g. Zimman, 2017). This thesis allowed participants to self-report their gender, with the options ‘female’, ‘male’, and ‘non-binary’. However, as all participants identified as either male or female, a more detailed investigation of gender differences was not conducted.

### 3.6.3 Socioeconomic class

Traditionally, sociolinguistic studies tend to divide speakers into groups based on socioeconomic status. In some works (e.g. Labov, 2001) these groupings are fine-grained, and may involve six or more categories. In other cases, participants may simply be divided into working-class and

middle-class (Chambers, 2003: 42). Socioeconomic group has been found to correlate with language use in the sense that working-class speakers often use more low-prestige or local variants, while middle-class speakers use more high-prestige or standard variants (e.g. Trudgill, 1974b). However, there can also be differences along the class spectrum; for instance, working-class speakers may align themselves with middle-class speech norms to express social aspiration (e.g. Labov, 1966). More recent work has also shown that upwardly mobile speakers may change social class over the course of their lives (e.g. Dickson & Hall-Lew, 2017).

In Sweden, social class is a challenging socio-indexical variable, because the subject of class is regarded as taboo (Norrby & Håkansson, 2015: 93). As a result, many speakers do not acknowledge or even necessarily recognise social class. This is not to say that Sweden does not have a class system, or that the Swedish language does not display class differences; such differences have been demonstrated by e.g. Kotsinas (2007) and Sundgren (2001). However, the Swedish class system is less pronounced than in e.g. the UK or USA, and social disparity appears to be smaller (Organisation for Economic Co-Operation and Development, 2020). Rather than measuring social class, Swedish sociolinguistic studies instead tend to use other methods to access this social information.

One available social index is SEI (*socioekonomisk indelning* ‘socioeconomic categorisation’), created by Statistics Sweden to investigate socioeconomic status. This index is not widely used however, as it requires fairly detailed data about the speakers’ professional roles, and has a very large number of categories (Norrby & Håkansson, 2015: 93). A simpler approximation of social class, which has been described as the most informative variable for adult Swedish speakers, is education (Norrby & Håkansson, 2015: 93-94). For younger speakers, their parents’ education may also be relevant. In addition, the choice of *gymnasieprogram* (roughly ‘upper secondary education programme’, usually undertaken between the ages of 16-19) can be useful in indicating both the speaker’s family background and their lifestyle ambitions (Norrby & Håkansson, 2015: 94). Grönberg (2004) found that this variable, combined with gender, was able to predict the linguistic behaviour of young Swedish speakers.

In this thesis, speakers’ highest education level was used to indicate socioeconomic class, but the sample was fairly uniform. Thus, the speakers in this sample could probably be described as middle-class (more fine-grained groupings were not possible with the available social information). This group is predicted to have a high occurrence of *Viby-i*, since it is generally regarded a high prestige feature, as will be explored in Chapter 4.

### 3.6.4 Lifestyle, identity and attitudes

In addition to the factors mentioned above, recent sociolinguistic work has shifted its focus more towards the way that speakers socially project their individual identities and attitudes, which may involve an intersection between many different group affiliations (Eckert, 2012). Some of the factors which can influence speech patterns in this way include social network (Milroy, 1987),

gender presentation or sexuality (Zimman, 2017), geographic mobility (Vilhelmson, 1994), personal “taste and style” (Bjurström, 1997), as well as the aspirational factors described above.

The identity and attitudinal factors investigated in this thesis are mainly informed by Grönberg (2004), who divided speakers into different groups based on a combination of their choice of educational programme, taste in music and clothes, spare time activities, level of social interaction, and attitudes towards their home town. These factors were found to influence the speakers’ linguistic behaviour, including which allophone they used for /i:/. A similar analysis may thus be relevant to *Viby-i*, since this vowel has been said to serve as both a social and a regional marker. For example, speakers may use *Viby-i* to signal social aspiration, or as an expression of local identity. A more detailed description of the lifestyle and attitude factors investigated in this thesis will be presented in Chapter 5.

### 3.6.5 Dialect and dialect contact

As mentioned previously, speakers’ relationship to the local dialect may vary depending on their age, with older speakers often using more regional language than younger speakers. Regional features also tend to be associated with low social status, and as such, they are often used more frequently by men and working-class speakers, while women and middle-class speakers tend to gravitate towards the standard.

However, in urban environments, the nature of non-standard language is slightly different, since linguistic innovation which is not part of the standard language can still be regarded as high status. In large cities, linguistic changes can spread quickly, “not only because there is a greater number of speakers who can potentially spread the change, but the phenomenon is also connected to the [cities’] prestige” (Norrby & Håkansson, 2015: 66). This process can also affect nearby cities; for example, Grönberg (2004) found that speakers who lived closer to Gothenburg were more likely to adopt innovative features from the city compared to those who lived further away. At the same time, some speakers may resist the influences of nearby cities, particularly if they have strong ties to their local dialect. For example, Uppsala is strongly associated with Standard Swedish, and even though it is geographically close to Stockholm, speakers may resist adopting urban features such as *Viby-i*, which are considered non-standard.

There are several sociolinguistic models explaining how linguistic features may spread from area to area, outlined in e.g. Kerswill (2003). The basis of these models is that “countless individual acts of short-term accommodation over a period of time lead to long-term accommodation in those same speakers” (Kerswill, 2003: 223, after Trudgill, 1986). Two models which might be relevant to *Viby-i* are the ‘wave’ model and the ‘gravity’ model.

The wave model suggests that linguistic changes travel outward from an epicentre like ripples on the water, becoming weaker the further they travel (Chambers & Trudgill, 1988). If the epicentre is a large city, the change will occur faster in nearby smaller towns, and may eventually reach other urban environments. In the case of *Viby-i*, this development may have taken place

in Stockholm, with the vowel later spreading to Uppsala.

The gravity model argues that features can ‘hop’ between urban centres, before diffusing to smaller locations (Trudgill, 1974a). This development tends to be hierarchical, with the change spreading to other large cities first, and then gradually to increasingly rural towns (Britain, 2009). The gravity model approach also seems applicable to *Viby-i*, since this vowel is found in Gothenburg, at the opposite end of the country from Stockholm, but not necessarily in the smaller cities in between. Theories surrounding how *Viby-i* may have spread across Sweden will be discussed further in Chapter 4.

### 3.6.6 Social pressures

Language change is not only driven by contact between different varieties, but also by social pressures within communities. On an individual level, speakers are not entirely stable in their language use, but may use different speech styles depending on the social context. This phenomenon is known as “style-shifting” (Chambers, 2003: 253), and tends to occur when speakers become self-aware. For example, this thesis investigates word list speech, which constitutes a more formal setting than casual conversation. As a result, speakers may (consciously or subconsciously) regulate their speech to conform to standard norms.

On a wider societal level, Labov (2001) describes two different processes which can affect how a linguistic feature spreads through a community: Change from above, and change from below. These terms refer “simultaneously to levels of social awareness and positions in the socioeconomic hierarchy” (Labov, 1994: 78).

Change from above refers to a process that is usually driven by the dominant social classes, and which lies above the level of consciousness. This kind of change is relatively rare (Labov, 2001: 31-32), and appears to be driven mainly by prescriptive norms, such as a wish to resist or reverse change, or to borrow features from another speech community (Labov, 1994: 78). Linguistic variants driven by a change from above are therefore associated with overt prestige, and tend to be widely known, commented upon, and subject to style-shifting. Speakers who are upwardly mobile are prone to adopting these changes, which operate as social capital (although it should be noted that non-standard features can carry *covert* prestige in contexts where they are used to signal solidarity with the local community).

Changes from below usually lie below the level of consciousness, and are driven by speakers who are central rather than peripheral in the socioeconomic hierarchy, particularly those who are upwardly mobile, but who still maintain strong links to the vernacular (Labov, 2001: 409-411). Because changes from below tend to go unnoticed, they are not prone to style-shifting, but tend to “represent the operation of internal, linguistic factors” rather than social attitudes about how the language ‘should’ be spoken (Labov, 1994: 78). For linguists and community members alike, changes from below are therefore often difficult to spot until they are nearly complete (Labov, 1994: 78). Strangely, they nevertheless tend to be subject to social stratification

during the process of change, as “it would appear that some speakers have learned to produce a distinction that they cannot perceive” (Labov, 1994: 391). This may be particularly true for vowels, which often lie below the level of consciousness (Holmes, 2013: 211).

Both change from above and change from below follow similar trajectories in terms of development over time. Generally, language change follows an “s-shaped curve” (Bailey, 1973: 77), which starts off slow, progresses very quickly over a short period of time, and finally slows down and stabilises (Fig. 3.4). Since the period of greatest change is so short, most new features are observed during the initial or final stage (Chambers & Trudgill, 1988: 163). This timing can greatly affect distribution within the community; for example, a new variant that is observed at the start of the curve may be much more common in young people and women, who usually adopt new features early, while these differences may disappear once the change is complete. However, depending on the speakers’ awareness of and attitude towards the change, patterns of social stratification may still persist even after the curve stabilises (Labov, 2001: 81-82).

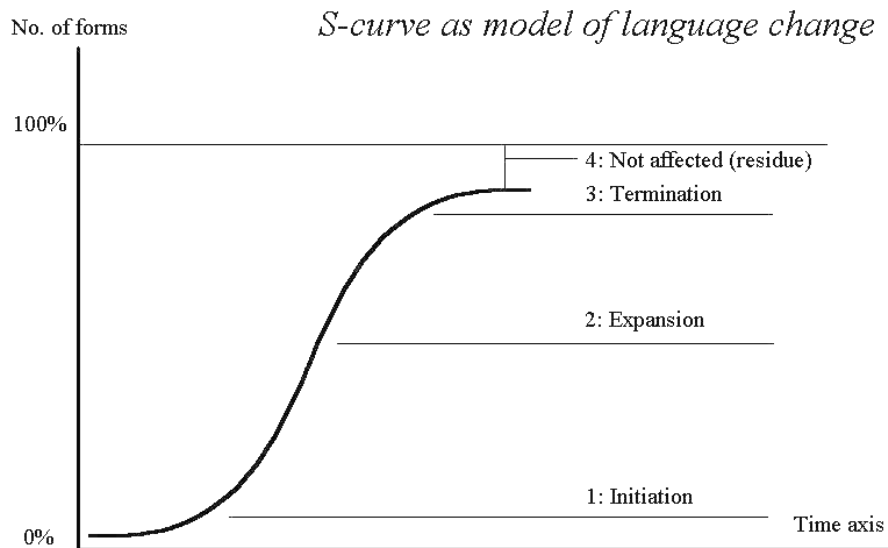


Figure 3.4: S-curve model of language change (University of Duisburg-Essen, 2019).

## 3.7 Description of studied cities

### 3.7.1 Overview of cities

Gothenburg and Stockholm were selected for this study because they both represent urban environments where Viby-i is known to occur. Uppsala was chosen because, while the use of Viby-i has not been investigated there before, the city is located close to Stockholm (see Fig. 3.5), and has similar demographic characteristics to both Stockholm and Gothenburg, with the exception

of its smaller population. A summary of potentially relevant demographic information for these three cities compared to the rest of Sweden is provided in Table 3.5.



Figure 3.5: Location of Gothenburg, Stockholm and Uppsala in Sweden (map adapted from Wikimedia Commons, 2019).

Stockholm is the capital of Sweden, and the most populated city, followed by Gothenburg, Malmö, and Uppsala in descending order (Statistics Sweden, 2019a). Malmö is not described here, as it is not located in the Central Swedish dialect region. The size of these cities means that they have a relatively strong social (and thus linguistic) influence, both on their immediately surrounding regions, but also potentially through the broadcast media. They are also attractive locations to live, work, and study, as shown by an increasing number of Swedish people moving from the countryside into larger cities (Statistics Sweden, 2015).

As Table 3.5 shows, all three cities have a younger population than the national average, unemployment is lower, income is higher, and education levels are higher. Many people commute into these cities to work, and student populations are large, as well as the proportion of foreign-born speakers compared to the rest of Sweden. All of these features create an environment where a large number of different dialect and language backgrounds coexist in the same space, which may promote language levelling, linguistic innovation, and, in some contexts, the use of linguistic in-group markers to signal e.g. identity, social status, or heritage.

Beyond these similarities, each city also has its own unique traits that might influence how *Viby-i* is used and propagated. The sections below will provide a brief description of each city to contextualise the social and regional distribution of *Viby-i*, which will be discussed further in Chapter 4.



	Gothenburg	Stockholm	Uppsala	Sweden
Population (2018)	571,868	962,154	225,16	10,230,185
Foreign-born (2018)	2.7%	2.5%	2.1%	1.9%
Univ. students (2017-18)	8.7%	7.5%	19.8% <sup>a</sup>	1.1%
Unemployed (2018)	5.9%	5.6%	5.2%	6.3%
Mean age (2018)	38.9 yrs	39.1 yrs	38.9 yrs	41.2 yrs
Mean income (2017)	311,700 SEK	369,200 SEK	311,200 SEK	308,700 SEK
<i>Highest edu. level (2018)</i>				
Mandatory	17%	13%	14%	20%
Secondary	51%	48%	56%	56%
Higher	29%	35%	34%	22%
<i>Commuting (2017)</i>				
Live & work	41%	39%	37%	N/A
Commute in	21%	33%	10%	N/A
Commute out	9%	14%	11%	N/A

<sup>a</sup>Student numbers in Uppsala may be inflated, as the Swedish University of Agricultural Sciences, which is based in Uppsala, also has campuses in other cities.

Table 3.5: Summary of demographic information for Gothenburg, Stockholm, and Uppsala municipalities, compared to Sweden as a whole (Statistics Sweden, 2019a, 2019b).

### 3.7.2 Stockholm

Stockholm is located in the East Central Swedish region, and is the capital of Sweden. The city has a great deal of political and social influence, acting as the seat of Parliament, the Royal Family, and much of the national media; the public television network SVT, the public radio network SR, Sweden's largest film company SF, and several major newspapers all have their main offices in Stockholm. Historically, the city has also exerted a lot of linguistic influence, as levelled Stockholm (and East Central) dialects were more or less the only represented varieties in Swedish broadcast media until the 1960s (Norrbý & Håkansson, 2015: 87). These varieties are still the most prevalent today, although different regional accents are now more frequently represented on TV and radio.

By nature of being a large metropolitan area, Stockholm is also home to a lot of social diversity, and is described by Kotsinas as a “split city” (Kotsinas, 2007: 26). While social disparity in Central Stockholm appears to have diminished over the years, there are still notable differences between the city's suburbs, with the North-East region being one of the richest areas in Sweden, while the South-West is associated with working- and lower middle-class populations. In particular, the peripheral South-West is associated with speakers from immigrant backgrounds (Kotsinas, 2007: 26-28). Within these areas, there are many different sociolects that can all be described as Stockholm Swedish, including working-class, upper-class, levelled standard, and multiethnolect varieties.

Because of this varied picture, Stockholm also has multiple social stereotypes associated with it, ranging from high-society socialites to urban street gangs. The most salient, however, is probably the image of young, wealthy, fashionable professionals, taking part in the extravagant leisure culture of Stureplan in Central Stockholm (Borneskog, 2011: 6-7). This group of people are particularly associated with the use of Lidingö-i (the local name for Viby-i), as well as a number of other linguistic markers, such as /fj/-fronting and /ɑ:/-fronting (Melin & Melin, 2005: 20-23).

### Stockholm vowels

The main vowel feature associated with Stockholm is an /e:/-/ɛ:/ merger, where both phonemes are realised as [e:] (Bruce, 2010: 216). This is a stereotypical feature of Stockholm speech, recorded as far back as the 1600s (Gross et al., 2016: 232). Today, it is mainly used by older speakers, although Kotsinas (2007) found that the feature still persisted in some working-class boys in the 1990s (Kotsinas, 2007: 113-114).

In recent years, both /ɛ:/ and /ø:/ in Stockholm have been shifting down, erasing the difference between pre-rhotic and non-rhotic contexts (Engstrand, 2004; Kotsinas, 2007). For example, *ära* (St. [æ:ra]) ‘honour’ and *äta* (St. [ɛ:ta]) ‘eat’ would both be realised with the vowel quality [æ:]. Lowering of these vowels also affects the short phonemes /ɛ, œ/, lowering them to [æ, ɐ]. The phenomenon of /ɛ:/- and /ø:/-lowering currently appears to be quite widespread in Sweden, and has been documented in the area around Stockholm (Leinonen, 2010; Nordberg, 1975; Wenner, 2010), as well as in Gothenburg (Gross et al., 2016).

The /ɑ:/ vowel in Stockholm has a somewhat complex history, but the phenomenon has not been studied in great detail. The prestige realisation of this vowel in East Central Sweden is [a:], a form which is frequently heard in old TV and radio recordings (Bruce, 2010: 215). At the same time, language authorities in the early 1900s complained about the [ɒ:] -like realisation of this vowel, most likely in working-class varieties (Kotsinas, 2007: 112). For this reason, it can probably be assumed that these two variants have existed alongside each other. Indeed, Kotsinas (2007: 112-113) found that Stockholm girls were more likely to use the prestige variant [a:], while boys used [ɑ:] or [ɒ:], although she did not find a class distinction.

### Stockholm consonants

In Stockholm, the phonemes /fj/ ‘sje’ and /ɕ/ ‘tje’ are frequently merged into /ɕ/, which is usually perceived as a prestige feature (Elert, 1995: 78). This merger is particularly common in women and middle-class speakers, while the more velar variants /fj, x/ are common in male and working-class speakers (Kotsinas, 2007: 111).

The /r/ and /l/ phonemes in Stockholm follow (or perhaps inform) the standard pronunciation: /r/ is usually either trilled or tapped, but is also commonly realised as [z] (Elert, 1995: 65),

which is less common in other parts of Sweden. /l/ is usually clear and not flapped, in contrast to many rural dialects in the surrounding area (Bruce, 2010: 201).

### 3.7.3 Gothenburg

Gothenburg is located in the West Central Swedish region, and has historically played an important part in the shipping and fishing industries, with a large portion of the population working in shipyards and factories (Olsson, 1996). In the 1970s, Sweden experienced an industrial crisis, and although Gothenburg is still well-known for fishing, shipping, and manufacturing, there has been a shift towards other industries, e.g. IT and environmental technology (Mehner, 2017).

The Gothenburg dialect is readily recognised by most Swedish speakers, but despite being fairly dominant in the West Central region – “it is unlikely that any other city in the Nordic countries has had such a large linguistic influence on its surroundings without being the capital” (Svahn & Nilsson, 2014: 17) – it is usually regarded as less prestigious than East Central Swedish. This could be because it is further from the national standard, or possibly because it is associated with working-class industries.

In popular culture, Gothenburg is known for its distinct sense of humour, and characters from Gothenburg are often played for comic relief, stereotypically portrayed as jolly, pun-loving, salt-of-the-earth factory workers (Jönsson, 2014: 16). There is a playful rivalry between Gothenburg and Stockholm, and many Gothenburg speakers express a sense of local patriotism, which can be reflected in their speech. According to Wenner (in interview with Ekstrand, 2016), “most Gothenburg speakers are proud of their accent, although this pride may not exist in everyone”.

#### Gothenburg vowels

As mentioned previously, the lowering of /ɛ:/, ɛ/ and /ø:/, œ/ which is ongoing in Stockholm has also been observed in Gothenburg (Gross et al., 2016). However, the traditional dialect variant is a relatively high [e̞], which remains high in pre-rhotic contexts (Björseth, 1958: 2-3). This high realisation of /ɛ:/ is “largely absent” in Gothenburg speech today (Gross et al., 2016: 233).

The traditional pronunciation of /ø:/ is somewhat mixed. On the one hand, the low and sometimes central realisation [œ:] is a “long-standing dialect feature” of Gothenburg, which is often written with the non-standard symbol ‘ô’ instead of ‘ö’ (Grönberg, 2004: 111). However, this variant seems to be lexically or contextually restricted in ways that are not yet understood (Grönberg, 2004: 111-112). On the other hand, traditional Gothenburg dialect also uses the raised variant [ø̞:] (Grönberg, 2004: 110-111). Either variant occurs categorically in the surrounding regions (Grönberg, 2004: 111), making it difficult to tell which is the ‘true’ dialect feature for Gothenburg.

The Gothenburg /ɑ:/ vowel is known for being a rounded [ɒ:], particularly in working-class speech (Bruce, 2010: 138). Unlike in Stockholm, there is no competing [a:] variant, but Elert

(1995: 50) points out that, while mild rounding is accepted in standard speech, heavy rounding is still associated with low status in Gothenburg as well as in Stockholm.

A traditional dialect feature of the whole West Central region is the use of [ɛ, æ] to signify short /ɪ, ʏ/ respectively (Norrby & Håkansson, 2015: 52). Since this constitutes local dialect, it is usually associated with low prestige, but the feature is occasionally retained in standard speech. One example is the name of the well-known fish market *Feskekôrka* [fɛskɛçœrka] ‘fish church’ (St. [fɪskɛçyrka]). The endorsement of this name on maps and street signs in Gothenburg could be an indicator of pride in the local dialect.

Similarly to above, the centralisation of several short vowels, including /ʏ, ɵ, ɔ/ to [œ] is a stereotypical dialect feature of Gothenburg, seen in words like *kôrka* (St. *kyrka*) ‘church’, *gôbbe* (St. *gubbe*) ‘old man’, and *gôtt* (St. *gott*) ‘good’ (Elert, 1995: 48).

### Gothenburg consonants

Gothenburg /r/ is stereotypically realised as a trilled [r], but in natural speech it is more likely to be lenited into a tap or approximant, as in many other parts of Sweden (Engstrand, 2004: 167).

Traditional Gothenburg dialect does not produce the ‘rhotic’ consonants [ɾ, ɖ, ɳ, ʂ, ʈ], but instead uses their non-rhotic counterparts [t, d, n, s, l] (Gross et al., 2016: 233). For example, *bort* (St. [bɔɾt]) ‘away’ would traditionally be produced [bɔt].

In the broader West Central region, /r/ is sometimes realised as [ɾ] in word-initial position, in a phenomenon known as the “Göta rule” (Norrby & Håkansson, 2015: 52). However, this is generally a rural feature, and appears to be uncommon in Gothenburg (Bruce, 2010: 194-195).

The /l/ phoneme in Gothenburg is usually clear, as in Standard Swedish. Flapped /l/ occurs in rural dialects in the West Central region, but not in the city (Grönberg, 2004: 129).

‘Sje’ is usually realised as [fj] or [x], and is distinct from ‘tje’ in the whole West Central region (Elert, 1995: 78).

### 3.7.4 Uppsala

Uppsala is located in East Central Sweden, about one hour’s drive north of Stockholm. The city has historically been an important location for learning and religion, containing both Sweden’s oldest university and its largest cathedral, where the Church of Sweden is also based. As such, Uppsala is mainly associated with education, and has a very large student population despite its relatively small size.

Although Uppsala appears to have a clear sense of identity, it is regarded as having relatively few local language markers (Källskog et al., 1993: 68). The Uppsala dialect lacks many of the traditional features from the surrounding area, and despite close ties with Stockholm, remains quite levelled. Because of this, Uppsala is sometimes regarded as a reference point for Standard Swedish.

### Uppsala vowels

There is very little literature on Uppsala vowels, presumably because realisations are relatively close to the standard. Most dialectal literature refers to grammar and vocabulary (e.g. Källskog et al., 1993; Wenner, 2010), where dialectal forms are more noticeable.

Among the few known phonetic features is the same /e:/-/ɛ:/ merger that is found in Stockholm (Källskog et al., 1993: 68). Lazić (2012: 12) found that young Uppsala speakers tended to be aware of this feature, but most reported that they did not use it themselves. The study does not discuss the use of this merger in older speakers, but it could perhaps be assumed to follow the same pattern as in Stockholm.

Uppsala speakers have also been shown to lower the short versions of the vowels /ɛ:/, ø:/, similarly to Stockholm and Gothenburg, but the long phonemes have not been investigated. The lowering in Uppsala nevertheless results in a merger between words like *lugn* (St. [lœɲn]) ‘calm’ and *lögn* (St. [lœɲn]) ‘lie’, producing both vowels as [œ]. This merger was found to be more common in men and highly educated speakers (Wenner, 2010).

### Uppsala consonants

The pronunciation of ‘sje’ and /ɾ/ in Uppsala appear to be the same as in Stockholm, or at least in line with most literature on Standard Swedish, given that no alternate variants are reported for this dialect. However, the use of flapped /l/ has been reported in the broader East Central region (Källskog et al., 1993: 69), and may appear in Uppsala. A news article encouraging the preservation of local dialect reports the occasional use of flapped /l/ by older speakers (Hårdstedt, 2009), but even in this context, it does not appear to be particularly common. Overall, this feature is likely to be rare in Uppsala, as it is primarily associated with rural speech (Bruce, 2010: 160).

## 3.8 Chapter summary

This chapter has provided an overview of the Swedish language and its dialects, focusing on vowels in the three varieties studied in this thesis: Gothenburg, Stockholm, and Uppsala. The chapter also explored factors which may affect Swedish vowel production, describing a number of potentially important linguistic and social factors, including vowel duration, consonant context, age, gender, class, identity, dialect, and speech style.

While the literature on vowel variation is fairly scarce, there are indications that the vowel inventories in Gothenburg, Stockholm and Uppsala are currently undergoing change. The downward shifts of /ɛ:/ and /ø:/ resemble Viby-i in that they are currently reported to occur in both Gothenburg and Stockholm, at opposite ends of the country. The following chapter will now provide a more detailed description of the geographic and social distribution of Viby-i.

# Chapter 4

## Viby-i

### 4.1 Chapter overview

This chapter brings together the many different portrayals of Viby-i in the literature and popular culture, in order to create a comprehensive phonetic profile of this vowel, as well as the linguistic and social context in which it occurs.

Section 4.2 introduces Viby-i, discusses the lack of previous research on this vowel, and summarises the main points of interest for this thesis.

The chapter then discusses the phonetic properties of Viby-i. Section 4.3 describes its auditory quality, addressing how listeners perceive this vowel, and the possibility of a gradient scale of ‘Viby-colouring’. Section 4.4 summarises previous acoustic work on this vowel and its position within the Swedish vowel system. Section 4.5 discusses the articulatory gestures that have been suggested for Viby-i, and presents articulatory data from three previous studies.

The next part of the chapter concerns the sociolinguistic aspects of Viby-i. Section 4.6 outlines the geographic distribution of this vowel, how it is realised in different dialects, and how it might have spread across Sweden. Section 4.7 discusses the social functions of Viby-i, and the attitudes which Swedish speakers might have towards this vowel.

Finally, Section 4.8 compares Viby-i to similar sounds in other languages in both acoustics and articulation.

### 4.2 What is Viby-i?

#### 4.2.1 Descriptions in the literature

Viby-i is an allophone of the long Swedish /i:/ vowel, used by some speakers instead of the standard variant [i:]. Viby-i is known for its unusual vowel quality, which has been described as “damped” (Björsten & Engstrand, 1999: 1,957), “thick”, “dark”, and “buzzing” (Borgström, 1913: 33). The vowel is named after Viby parish in Central Sweden, partly because this sound

appears in the local dialect, and partly because the place name contains both of the vowels commonly affected by “Viby-colouring”: /i:/ and /y:/ (Bruce, 2010: 132). The same vowel is also known as *Lidingö-i*, named after the island Lidingö in Stockholm (Kotsinas, 2007: 109), or, less commonly, *Göteborgs-i* in Gothenburg (Björseth, 1958: 7). In addition, speakers may refer to Viby-i without naming it, e.g. by producing it, or by using innovative spelling. For example, the town Lysekil (near Gothenburg) is sometimes jokingly referred to as “Lyyysekiil” (e.g. Johansson, 2015), mimicking the use of Viby-i in this dialect. Similar examples can be found for Lidingö and Viby (Berzelius, 2012; S. Eriksson, 2008). Speakers may also refer to the use of Viby-i as “i-ing” (*att i:a*) (Engstrand, Bruce, Elert, Eriksson, & Strangert, 2000) or “speaking on i” (*tala på i*) (af Klintberg & Ripås, 2003), but these expressions appear to be more common when referring to rural dialects.

## 4.2.2 Working definition and points of interest

Despite the fact that Viby-i is said to be fairly common in Sweden, it is rarely discussed in works on Swedish phonetics and phonology, and the literature does not establish any phonetic criteria that distinguish Viby-i from a standard [i:]. Different sources may therefore use different terminology to refer to the same sound, or use the same terminology to refer to different sounds. What appears to underpin Viby-i across the literature is its categorisation as ‘not standard [i:]’ by native listeners. However, since listeners’ perceptions are affected by their own phonological systems, different instances of Viby-i may be more or less salient to different people. For the purposes of this thesis, I have therefore chosen to define Viby-i as a sound which phonemically functions as /i:/, but has a lower acoustic F2 than /e:/ within the same speaker or sample. A prototypical Viby-i will also have an F1 that is similar to, or higher than, /e:/.

Due to the lack of literature on Viby-i, there are many unanswered questions about this vowel, including:

- How is it produced? Can it be produced in more than one way?
- Are there different types of Viby-i? If so, do speakers vary depending on their dialect, sociolect, or idiolect?
- Is Viby-i categorical within speakers, or does it change depending on the linguistic context or speech style?
- How common is Viby-i? Where does it occur geographically, and how did it arise there?
- How aware are speakers and listeners of the use of Viby-i? How is this sound evaluated socially?
- Is Viby-i unique to Swedish?

This thesis will primarily focus on the first three points: the articulation of Viby-i; variation in acoustics and articulation; and the effects of linguistic, dialectal, and social factors. However, this chapter, and the thesis as a whole, will also touch upon the other questions posed here, attempting to start a discussion which will hopefully continue as linguists grow increasingly aware of this unusual vowel.

### 4.2.3 Phonetic notation

The phonetic symbol used for Viby-i in this thesis is [iː]. This symbol is primarily chosen for convenience, partly because it has been used in previous literature (Björsten & Engstrand, 1999; Gross & Forsberg, 2019), and partly because it corresponds to expectations of how this vowel is produced based on its formant structure. However, since this study seeks to problematise the idea of a one-to-one relationship between acoustics and articulation, the use of this symbol should be regarded as tentative. Viby-i has other properties which are not well-represented by [iː], and to some extent, it may be more suitable to use the symbol originally devised for Viby-i in the dialectological literature: [ɿ] (Malmberg, 1971: 59). This symbol was also traditionally used for one of the apical vowels in Mandarin Chinese (Karlgren, 1915: 295), which will be discussed further in Section 4.8. However, the symbol [ɿ] is no longer recognised by the IPA, and using it would bring with it its own assumptions of equivalence between apical vowels and Viby-i, which is problematic. There have also been instances of Viby-i being denoted [iːʔ] (Elert, 1995; Grönberg, 2004; Kotsinas, 2007) or even [ø] (Bruce, 2010: 132) due to its characteristic ‘buzzing’ quality. However, as Section 4.6 will discuss, the fricative part of Viby-i appears to be optional, at least in some dialects. For the sake of consistency, and to avoid confusion, I will use the notation ‘Viby-i’ as widely as possible, rather than using a phonetic symbol. However, [iː] may be used in transcriptions where there is no other way of distinguishing Viby-i from a standard [iː].

## 4.3 Perception

### 4.3.1 Viby-i vs. standard [iː]

As mentioned previously, Viby-i is usually described as having a “damped” vowel quality (Björsten & Engstrand, 1999: 1957), often with an added element of “buzziness” or frication (e.g. Borgström, 1913: 33). During the course of this project, phoneticians who have heard instances of Viby-i have remarked that it sounds /l/-like, /ə/-like, /z/-like, nasalised, like the larynx is raised or the tongue root retracted, like the lips are rounded, or like the tongue is protruded. To most non-native speakers, Viby-i is clearly auditorily distinct from [iː], and many do not recognise that it could operate as the same phoneme.



On the other hand, based on my own experience as a native Swedish speaker, many Swedes do not appear to perceive the difference between Viby-i and standard [i:], unless the Viby-i is extremely salient, or conforms to an expectation of the speaker as a Viby-i user. This phenomenon has not been investigated in Swedish, but a similar effect has been observed in listeners' perception of Canadian vowel raising in Detroit speech (Niedzielski, 1999). An indication that Swedish speakers may not hear the difference between Viby-i and standard [i:] is that many Swedish speakers also use Viby-i in their English speech (Norstedt, 2019: 24), e.g. when producing words like 'Sweden' [swi:dən].

There are some indications that Viby-i may vary in its auditory quality. For example, several researchers have suggested that some versions of Viby-i may be “buzzier” than others (Björseth, 1958; Björsten & Engstrand, 1999; Kotsinas, 2007). Furthermore, Björseth (1958) states that Viby-i in Gothenburg is different from nearby rural dialects, because it “lacks the peculiar throat tension” associated with this vowel, and only retains its fricated aspects (Björseth, 1958: 7). These accounts suggest that the damped vowel quality and the buzziness may operate independently of one another. It should be noted, however, that it is not always clear from these descriptions whether “buzziness” refers to a quality superimposed on the vowel itself, or to the fricative offglide which is common at the end of many high vowels in Swedish (see Chapter 3).

It is also possible that certain linguistic contexts encourage Viby-i to become stronger or weaker, rather than excluding this vowel quality altogether. Such contexts could include differences in duration, surrounding segments, or prosodic context. These aspects have not previously been researched, but will be investigated in this thesis.

### 4.3.2 Viby-i vs. Turkish [i]

The ‘damped’ impression of Viby-i appears to have some connection to its formant values, particularly F2. In a cross-linguistic perceptual experiment, Björsten and Engstrand (1999) tested whether Swedish and Turkish listeners would accept instances of Viby-i and Turkish [i] (collectively referred to as “damped /i/”) depending on the degree of F2-lowering in the vowel. The researchers do not specify what aspects of the acoustics cause them to describe these vowels as “damped”, but they seem to be referring to the low F2 (Björsten & Engstrand, 1999: 1,957).

The study included 15 Swedish and 5 Turkish listeners, listening to 54 different stimuli. The stimuli consisted of Viby-i from three Swedish dialects (Gothenburg, Kräklinge, Orust), as well as Turkish [i], and, representing non-damped high front vowels, standard [i:] in Swedish, Polish and Turkish, as well as Kräklinge [e:], which has similar formant values to standard [i:]. The study found a statistically significant pattern whereby Swedish listeners were increasingly likely to accept vowels as Viby-i when F2 was lower, regardless of the language it was produced in (Björsten & Engstrand, 1999: 1,960). The acceptance rates for each vowel are shown in Fig. 4.1 below. Note that this graph also indicates some degree of regional variation in Swedish, which will be returned to in Section 4.6.

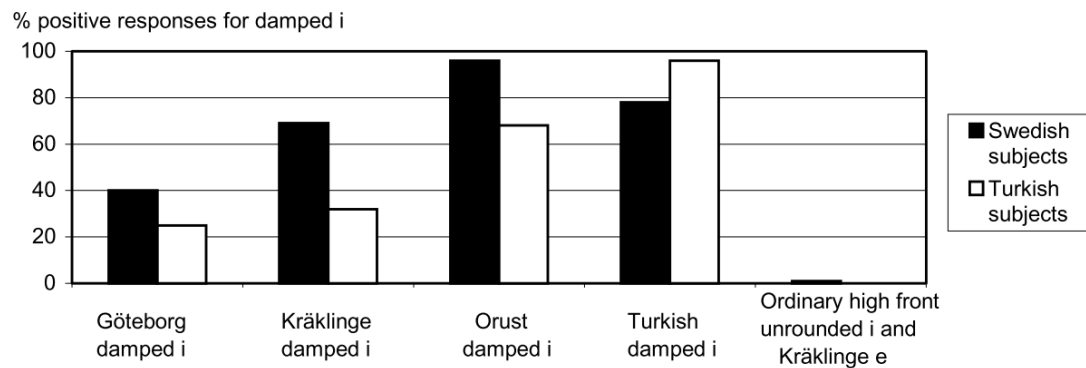


Figure 4.1: Listeners' acceptance of vowels as 'damped /i:/' by stimulus (Björsten & Engstrand, 1999: 1,958).

### 4.3.3 Viby-i vs. /y:/

Another perceptual study was carried out by Gross and Forsberg (2019), investigating the ability of native Swedish listeners to disambiguate standard Swedish [i:] and [y:] from Viby-i and Viby-y. The study was based on the finding that Viby-i and Viby-y tended to overlap in the F1/F2 space, in a way that the standard phonemes did not, as well as the claim by Björsten and Engstrand (1999) that Viby-y in Kräklinge does not appear, from its acoustics, to use lip-rounding.

The authors used 60 vowel stimuli of 19 young Gothenburg speakers producing a balanced number of tokens for [i:], [y:], Viby-i and Viby-y. These tokens were presented in the form of an online survey, where listeners were prompted to pick which grapheme, 'i' or 'y', corresponded to the sound they heard, either within a word or in isolation. The survey was completed by 203 listeners, 26 of whom were from Gothenburg. All listener groups produced equivalent results.

The study found that listeners were consistently able to recognise both standard [i:] and Viby-i as allophones of /i:/ (91% and 71% respectively in isolation, 98% and 92% respectively within a word), but the results for /y:/ were more variable. The standard [y:] phoneme was correctly identified 40% of the time in isolation, and 95% within a word, while Viby-y was only correctly identified 12% of the time in isolation, and 68% within a word. The large difference between the isolated and in-word contexts can be attributed to the fact that there are very few minimal pairs for /i:/-/y:/ in Swedish, meaning that the word more or less entirely disambiguates the vowel.

The fact that variants of /y:/ were frequently mistaken for /i:/, but not the other way around, supports the theory that the Gothenburg /y:/ vowel (particularly Viby-y) is becoming de-labialised (Gross & Forsberg, 2019: 15). The authors also point to the fact that the acoustic centralisation of Viby-i likely brought it into the perceptual territory of /y:/ in the first place (Gross & Forsberg, 2019: 17), suggesting that the merger has arisen from perceptual confusion. However, the authors do not believe that this merger has been fully established yet, due to the amount of individual variation in both production and perception (Gross & Forsberg, 2019: 16).

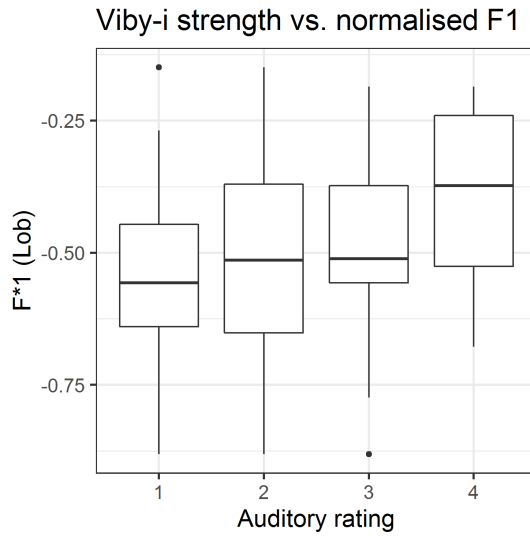


Figure 4.2: Correlation between Viby-i ‘strength’ and a high normalised F1 (data from Westerberg, 2016).

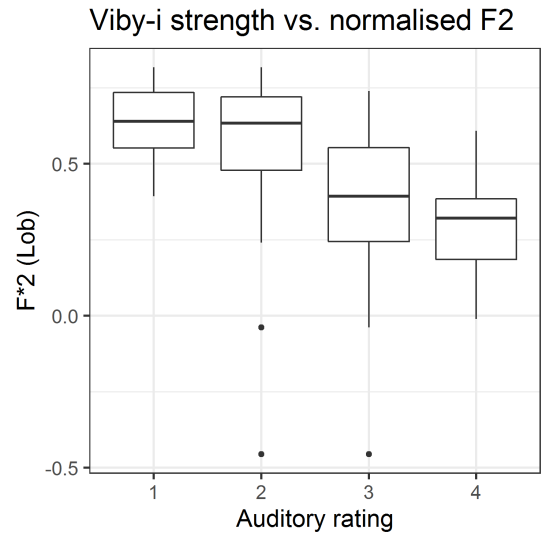


Figure 4.3: Correlation between Viby-i ‘strength’ and a low normalised F2 (data from Westerberg, 2016).

#### 4.3.4 Gradient Viby-colouring

Finally, as part of the pilot study for the current project (Westerberg, 2016), I conducted a small perceptual experiment to investigate whether Viby-colouring could be quantified on a gradient scale, ranging from a weaker, more [i:]‑like vowel, to a stronger, more ‘damped’ Viby-i. The experiment was based on personal observation during the segmentation process, whereby some instances of Viby-i appeared to be more perceptually salient than others.

The experiment consisted of four phonetically trained, non-Swedish speakers being asked to rate 39 instances of /i:/ in the word BIBEL (‘bible’), produced by 13 speakers from various parts of Central Sweden (this dataset will be described further in Sections 4.4 and 4.5). The auditory ratings consisted of a four-point scale, where 1 represented “normal /i:/” and 4 represented a “very thick” /i:/. Listeners were asked to ignore any fricative or diphthongal offglides and focus only on the main vowel quality. To some extent, these instructions also indirectly discouraged listeners from acknowledging frication *during* the vowel. In part, this focus was intended to make the experiment more straightforward, but it was also based on the observation that a damped vowel quality did not always co-occur with intra-vowel frication.

The results showed a relationship between perceived Viby-i “thickness” and a lowering of acoustic F2, as well as a raising of F1 (Figs. 4.2 and 4.3). Statistical testing using linear regression revealed a significant relationship between Viby-i strength and a low normalised F2 ( $r = -5.3$ ,  $p < 0.01$ ), as well as a marginally significant relationship between Viby-i strength and a high normalised F1 ( $r = 2.1$ ,  $p < 0.1$ ). It should be noted that this analysis relied on a very small dataset, and did not include random factors, meaning that the statistical effects may be exaggerated.

Individual speakers could also be categorised as having an overall ‘weaker’ or ‘stronger’

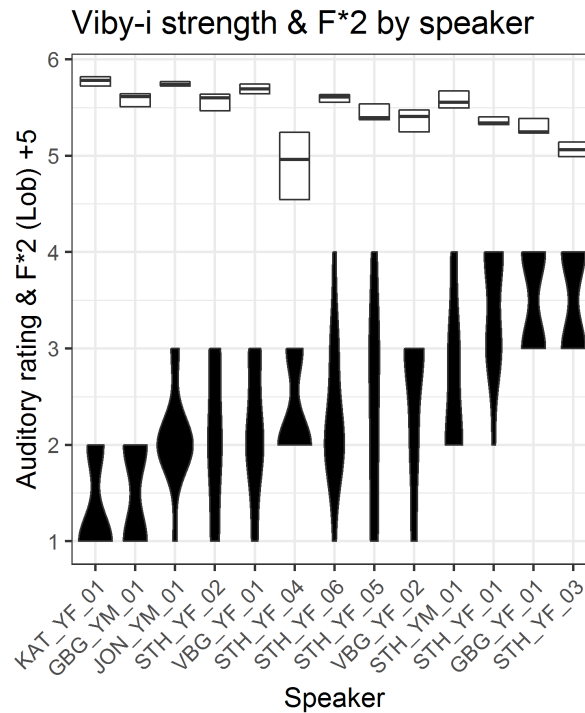


Figure 4.4: Viby-i strength ratings (black) and normalised F2 values (white, shifted up for visibility) for each speaker (data from Westerberg, 2016).

Viby-i, although ratings were fairly variable even within the same speaker (see Fig. 4.4). With the notable exception of STH\_YF\_04, speakers with a lower normalised F2 were generally rated as having a stronger Viby-i, and vice versa.

It is possible that some of the variability in the auditory ratings was caused by other factors than formant values, e.g. voice quality, formant bandwidth, formants above F2, or fricative noise. Although it would be informative to conduct a larger and more detailed study on the perception of Viby-i, it is beyond the scope of this thesis. The main factor to bear in mind is that the perceptual phenomenon of ‘Viby-colouring’ appears to be driven by a lowering of F2, as previous literature has also suggested (e.g. Björsten & Engstrand, 1999; Gross & Forsberg, 2019).

## 4.4 Acoustics

### 4.4.1 The first three formants

The main acoustic parameter associated with Viby-i is a low F2, often accompanied by a high F1 (Björsten & Engstrand, 1999; Bruce, 2010; Frid et al., 2015; Gross, 2018; Gross & Forsberg, 2019; Ladefoged & Lindau, 1989). Ladefoged and Lindau (1989) additionally report that Viby-i has a high F3, but this has not yet been confirmed. It nevertheless seems likely that F3 could play a role in the perception of Viby-i, as it is said to contribute to the vowel quality of standard [i:] (Malmberg, 1971: 29).

The low F2 appears to be particularly important for Viby-i, as it seems to differ more from a standard [i:] in this dimension than in F1 (Björsten & Engstrand, 1999; Gross & Forsberg, 2019; Schötz et al., 2011; Westerberg, 2016). As mentioned above, F2 also seems to be driving the perception of this vowel as ‘damped’.

Unfortunately, the incomplete acoustic context in which Viby-i is often presented makes it difficult to say how this vowel corresponds to others within the same system. In many cases, acoustic information is absent, or alternatively, it is used to describe Viby-i in isolation, meaning that it is not always clear what the low F2 and high F1 are in comparison to. So far, no studies have compared Viby-i to standard [i:] within the same speaker, as it appears that speakers’ use of this vowel is categorical (i.e. speakers either use Viby-i or standard [i:]). However, a few studies do compare speakers with Viby-i to speakers without, as discussed below. The main issue with this practice is that Swedish /i:/ is variable, and the line between standard [i:] and Viby-i can be fuzzy depending on the criteria used by researchers.

Source	Variety	Female [i:]			Male [i:]		
		F1	F2	F3	F1	F2	F3
Deterding (1997)	SSBE	303	2,654	3,203	280	2,249	2,765
Eklund and Traunmüller (1997)	Stockholm	351	2,455	3,500	291	2,107	3,135
Engstrand (2004)	Unspecified	430	2,700	3,875			
Engwall and Badin (1999)	Stockholm				320	1,520	2,900
Ericsson (2005)	Unspecified	300	2,510	3,370	220	2,180	3,100
Fant (1959)	Unspecified	278	2,520	3,460	255	2,065	2,960
Fant (1972)	Unspecified	345	2,060	3,100	230	2,000	3,000
Fant et al. (1969)	Unspecified				255	2,190	3,150
Kuronen (2000)	Nyköping				275	2,363	3,304
Malmberg (1956)	Unspecified				300	2,400	3,500
Malmberg (1971)	Unspecified				280	2,400	3,400
Nord (1986)	Stockholm				300	2,087	N/A
Schötz et al. (2011)	South Swe.				332	2,017	2,685
Stålhammar et al. (1973)	Stockholm				285	2,000	3,015
Traunmüller and Öhrström (2007)	Stockholm	450	2,100	2,845	328	1,963	2,625
Min.	Swedish	278	2,060	2,845	220	1,520	2,625
Max.	Swedish	450	2,700	3,875	332	2,400	3,500
Median	Swedish	348	2,483	3,415	285	2,087	3,058
Mean	Swedish	359	2,391	3,358	282	2,099	3,065

Table 4.1: Formant values for standard [i:] in previous Swedish studies, compared to Standard Southern British English. Entries with both a higher F1 and a lower F2 than SSBE are shaded.

### 4.4.2 Variability in standard [i:]

Table 4.1 presents the mean values for the first three formants of standard [i:] from a number of studies on Swedish vowels, with equivalent values from Southern Standard British English (SSBE) as a point of reference. As the table shows, the formant values, and the relationships between them, are highly variable, with many displaying both a relatively low F2 and high F1 compared to SSBE (these entries are shaded). The shaded values do not necessarily represent instances of Viby-i, but they demonstrate the difficulties of drawing a line between Viby-i and standard [i:] using the relative criteria of a ‘low F2’ and ‘high F1’.

It should also be noted that, although these values are all presented as standard [i:], F2 is generally quite low, and a few entries – particularly those of Engwall and Badin (1999) and Stålhammar et al. (1973) – show fairly strong indications of Viby-colouring (see Fig. 4.5). Although Stålhammar et al. (1973) do not mention Viby-i, they state that “It is a general characteristic of Swedish that [i:] has a lower F2 than [e:] and that the phonetic distance between these sounds is not very great” (Stålhammar et al., 1973: 6). This comment could be interpreted as a general acceptance of (weaker) forms of Viby-i as normal for Standard Swedish as early as the 1970s.

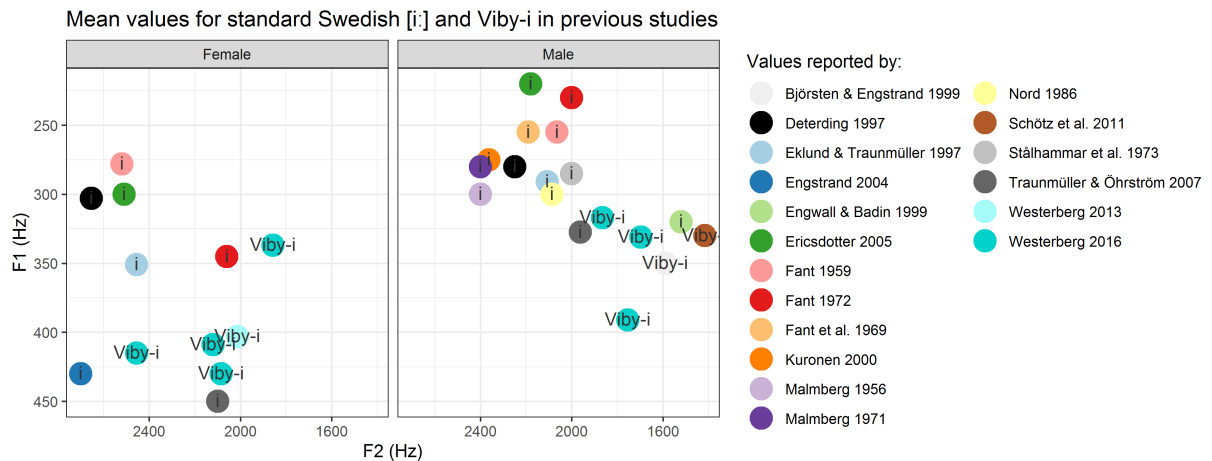


Figure 4.5: Position of standard [i:] vs. Viby-i in the F1/F2 space using data from previous studies.

### 4.4.3 Reported formant values for Viby-i

Studies reporting formant values for Viby-i are fairly rare, and sample sizes tend to be small. Björsten and Engstrand (1999) give values for the first four formants of one older male speaker from Kräklinge (near Viby), and Schötz et al. (2011) of one male speaker from East Central Sweden. These values are provided in Table 4.2, alongside values for /e:/ where available. The table also contains formant values for Viby-i and /e:/ from previous stages of the current project. Westerberg (2013) presents values from three young female speakers: 2 from Gothenburg (West Central) and 1 from Varberg (West Central, 2f). Westerberg (2016) presents values from 13 young Swedish speakers: 2 from Gothenburg (West Central; 1 male, 1 female), 2 from Varberg

(West Central; both female), 1 from Jönköping (West Central, male), 7 from Stockholm (East Central; 1 male, 6 female), and 1 from Katrineholm (East Central, female).

Source	Variety	Female				Male			
		F1	F2	F3	F4	F1	F2	F3	F4
Björsten and Engstr. (1999)	Kräklinge					350	1,590	2,860	3,590
						340	2,220	2,730	3,670
Schötz et al. (2011)	East C.					330	1,416	2,254	3,521
Westerberg (2013)	Gothenburg	424	1,965	3,052	4,177				
		491	2,253	2,918	4,034				
	Varberg	362	2,110	3,185	4,144				
		395	2,331	2,933	3,743				
Westerberg (2016)	Gothenburg	391	1,754	2,921	4,116	339	1,857	2,841	3,563
		393	2,406	2,837	3,928	344	2,225	2,585	3,548
	Stockholm	427	2,088	3,262	4,244	331	1,696	2,716	3,748
		412	2,591	3,069	3,971	314	2,071	2,784	3,550
	Jönköping					318	1,882	2,746	3,862
						338	2,188	2,783	3,448
	Katrineholm	415	2,427	3,551	4,396				
		404	2,798	3,474	4,023				
	Varberg	407	2,117	3,324	4,500				
		415	2,500	3,183	3,994				
Min.	Swedish	362	1,754	2,921	4,116	318	1,416	2,254	3,521
		393	2,253	2,837	3,743	314	2,071	2,585	3,448
Max.	Swedish	427	2,427	3,551	4,500	350	1,882	2,860	3,862
		491	2,798	3,474	4,034	344	2,225	2,784	3,670
Median	Swedish	411	2,099	3,224	4,211	331	1,696	2,746	3,590
		418	2,480	3,069	3,949	334	2,176	2,721	3,554
Mean	Swedish	404	2,077	3,216	4,263	334	1,688	2,683	3,657
		408	2,453	3,001	3,983	339	2,204	2,757	3,549

Table 4.2: Reported F1 to F4 values for Viby-i (grey) and /e:/ (white).

Combining these four studies, the mean F1 for Viby-i is around 400 Hz for women and 330 Hz for men, while the mean F2 is around 2,080 Hz for women and 1,700 Hz for men. Compared to /e:/, Viby-i has a very similar F1 (the difference is around 5 Hz), but its F2 is much lower, with a difference of around 380 Hz for women, and around 500 Hz for men. Compared to the standard [i:] values of SSBE presented in Table 4.1, Viby-i is around 50 Hz higher in F1 regardless of gender, but its F2 is around 300 Hz lower for women, and around 400 Hz lower for men. This large difference in F2 is likely to also create an audible difference in vowel quality between Viby-i, standard [i:], and [e:] respectively.

Notably, the samples from Björsten and Engstrand (1999) and Schötz et al. (2011) have much lower F2 values than the rest of the entries in Table 4.2, although their F1 values are

similar. Likewise, Bruce (2010) refers to (male) /i:/ productions with an F2 of around 1,500 Hz as Viby-i, while he refers to productions with an F2 of around 2,000 Hz as standard [i:] (Bruce, 2010: 134). This is perhaps an indication that previous studies have selected more prototypical instances of Viby-i to be studied, and have thus set a higher ‘threshold’ for what qualifies as a Viby-i than the current study does. In other words, they have only considered the most extreme examples to constitute Viby-i, while this study takes a more gradient approach.

As both Björsten and Engstrand (1999) and Schötz et al. (2011) point out, Viby-i is acoustically centralised, mainly in the F2 dimension, rendering it fairly similar to a cardinal [i] (Björsten & Engstrand, 1999: 1,959), while the highest and frontest vowel in the system (/i:/ in the case of standard speakers, and /e:/ in the case of Viby-i users) seems to be similar to a cardinal [e] (Schötz et al., 2011: 1,768). This phenomenon is demonstrated in Figs. 4.6 and 4.7 respectively. There are some discrepancies between the two datasets, but overall, Viby-i seems to reside somewhere behind [e:], and behind and/or below [i:] in the acoustic space.

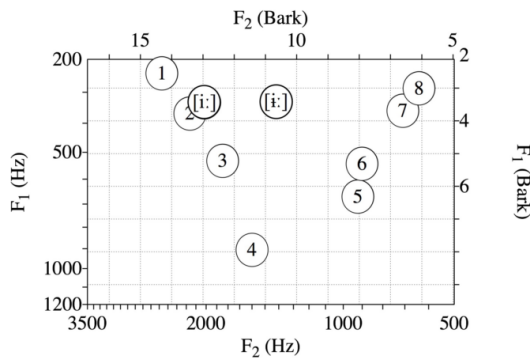


Figure 4.6: Viby-i and Standard Swedish [i:] plotted against cardinal vowels from D. Jones (1917) (Schötz et al., 2011: 1,768).

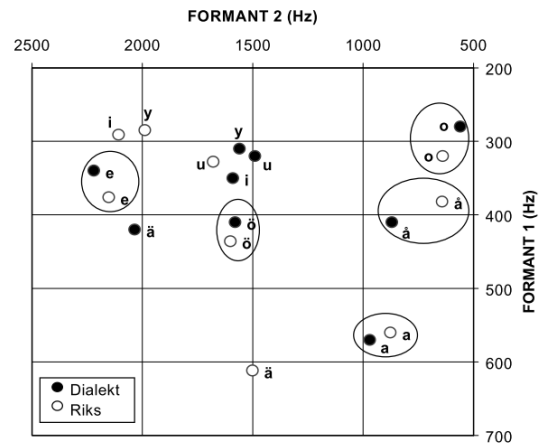


Figure 4.7: Vowels in the Kräklinge dialect (black) plotted against Standard Swedish vowels (white) from Eklund and Traunmüller (1997) (Engstrand et al., 1998: 86).

Gross (2018) produces a similar illustration of Viby-i’s position in the vowel system compared to standard [i:], shown in Fig. 4.8. Unfortunately, [e:] is not included in Gross’ dataset, and raw formant values are not provided. The axes instead represent Principal Components, corresponding roughly to F1 and F2. The figure nevertheless shows the same relationship between Viby-i and standard [i:] as Björsten and Engstrand (1999), and also illustrates the overlap between Viby-i and Viby-y mentioned previously. The same overlap was found in Björsten and Engstrand’s Kräklinge speaker (Björsten & Engstrand, 1999: 1957), despite Viby-i and Viby-y looking fairly far apart in Fig. 4.7.



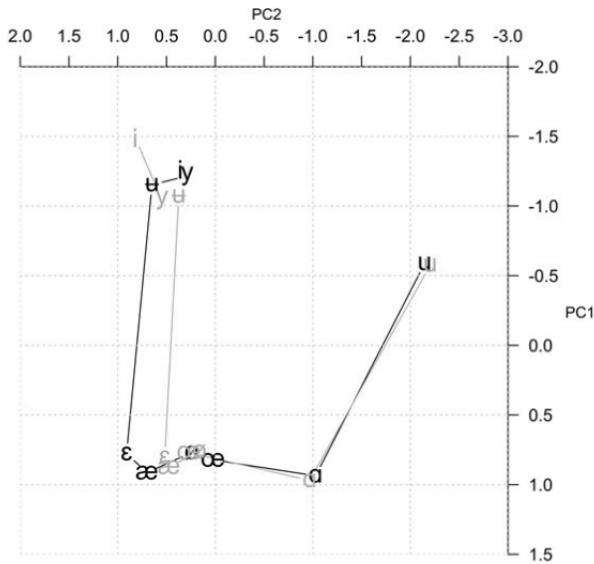


Figure 4.8: Vowels from Gothenburg speakers with Viby-i (black) and Gothenburg multiethnolect speakers without Viby-i (grey) (Gross, 2018: 329).

#### 4.4.4 Beyond formant frequencies

Additional acoustic information beyond formant values is scarce in the literature. Gross and Forsberg (2019) found that Viby-i does not display much formant movement over the course of the vowel, but other dynamic aspects (such as frication) have never been investigated. Since fricative offglides are common in all high Swedish vowels (see Chapter 3), it is possible that some of the ‘buzzing’ of Viby-i has been misattributed to this phenomenon. However, it is also possible that Viby-i undergoes frication during the vocalic portion of the vowel, which will be investigated further in this thesis.

The ‘damped’ impression of Viby-i could also arise from formants having wider, less defined bands, or from some formants being amplified or weakened in relation to others. The timing and interaction between ‘damping’ and frication may also affect the overall impression of Viby-i. This study originally intended to investigate formant bandwidth and intensity, but was not able to do so, as it was not possible within the given timeframe to obtain accurate measurements from the available data, as discussed in Chapter 5.

### 4.5 Articulation

#### 4.5.1 Impressionistic theories

Much of the articulatory phonetic literature on Viby-i focuses on trying to determine the articulatory settings that give rise to its unusual vowel quality. This literature is marked by a number of different theories, some of which contradict each other. Until very recently, this literature did not make reference to experimental data, but seems to have been based on impressionistic

information alone.

Three of these early sources provide different theories for how Viby-i could be produced: Lundell (1878) suggests that it is achieved with a high back tongue gesture and spread lips (Lundell, 1878: 91), rendering it essentially the same as [uɪ]. Noreen (1903), on the other hand, believes that Viby-i is an extremely fronted (apical) [i:], produced either with a tip-up or a tip-down gesture (Noreen, 1903: 495). Finally, Borgström (1913) describes a complex articulation, where

the tongue tip rests against the lower teeth [...] the anterior sides of the tongue rest against the upper teeth [...] [and] between its raised anterior and posterior parts, the tongue is somewhat lowered, creating a confined resonance chamber, which brings to listeners its observed ‘thick’, ‘dark’, ‘buzzing’ quality. (Borgström, 1913: 33)

Many years later, Ladefoged and Lindau (1989) propose a theory similar to that of Borgström, stating that Viby-i is achieved “by slightly lowering the body of the tongue while simultaneously raising the blade of the tongue” (Ladefoged and Lindau, 1989, in Ladefoged & Maddieson, 1996: 292). None of these claims have been either verified or falsified in the subsequent literature, but will be addressed in the current work.

### 4.5.2 Speech synthesis

To date, there have only been two experimental articulatory studies (outwith the current project) focusing on Viby-i. Björsten and Engstrand (1999) (see also Engstrand et al., 1998) conducted an acoustic and perceptual study of this vowel as produced by an elderly male speaker from Kräklinge (near Viby), and used the speech synthesis model APEX (Stark, Lindblom, & Sundberg, 1996) to identify different potential articulations based on the vowel’s formant values. They found that Viby-i in this speaker was likely produced with a centralised, tip-down tongue gesture. Their model also showed that the damped quality of the vowel, which was correlated with F2 lowering, could be exaggerated further by raising the tongue tip (Björsten & Engstrand, 1999: 1,959). Based on these results, the authors treat Viby-i as a central [i:], with optional tip-raising in some speakers. They suggest that the choice between a tip-down and tip-up gesture could depend on the speaker’s dialect (Björsten & Engstrand, 1999: 1,959).

### 4.5.3 Tongue gesture dynamics (EMA)

The second study, by Frid et al. (2015) (see also Frid, Schötz, & Löfqvist, 2011; Schötz, Frid, Gustafsson, & Löfqvist, 2013, 2014; Schötz et al., 2011), as part of the project ‘Exotic Vowels in Swedish’, used electromagnetic articulography (EMA) to investigate the tongue gesture dynamics of /i:, y:, u:/ in 27 speakers from Stockholm, Gothenburg and Malmö (roughly evenly divided between males and females). The study does not focus on Viby-i directly, but found that

this vowel was connected to tongue gesture variation in Stockholm and Gothenburg. Based on the observed differences in tongue gestures, the authors divided these speakers into subgroups (GS1, GS2, SS1, SS2), and subsequently found that most GS2 and SS2 speakers used Viby-i, whereas most (but not all) speakers in the other groups used standard [i:]. In the dynamic results, GS2 and SS2 speakers had an overall more backed and lowered tongue body than the others, but there was little to no difference in tongue tip behaviour (Figs. 4.9 and 4.10) (Frid et al., 2015: 3). However, in a previous paper, Schötz et al. (2011) used steady-state data from two male speakers, one with Viby-i and one without, and found that Viby-i was produced with a tip-up gesture, while standard [i:] was tip-down (Fig. 4.11) (Schötz et al., 2011: 3).

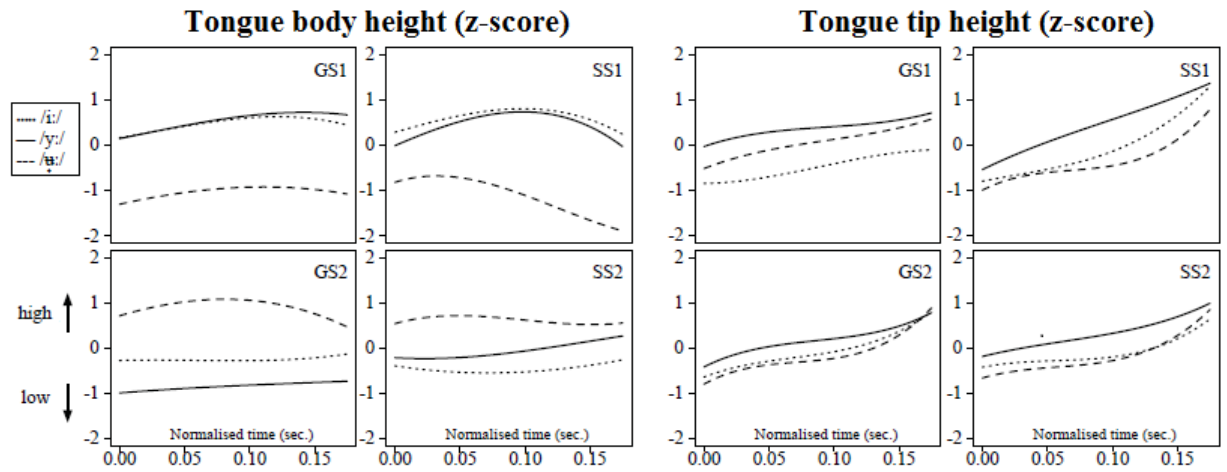


Figure 4.9: Vertical movement of the tongue body and tip in two Stockholm and two Gothenburg speakers (Frid et al., 2015: 2). Viby-i users in bottom row.

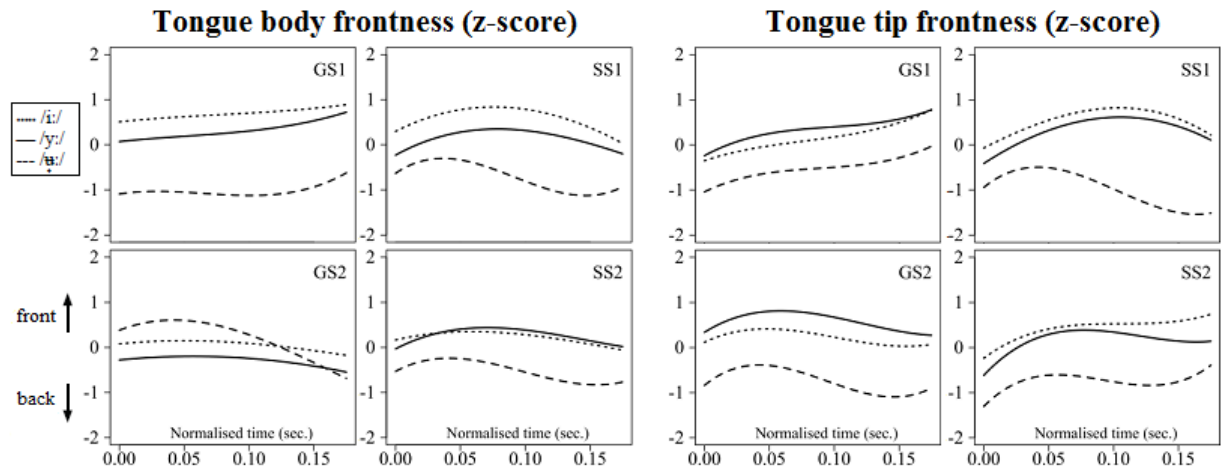


Figure 4.10: Horizontal movement of the tongue body and tip in two Stockholm and two Gothenburg speakers (Frid et al., 2015: 2). Viby-i users in bottom row.

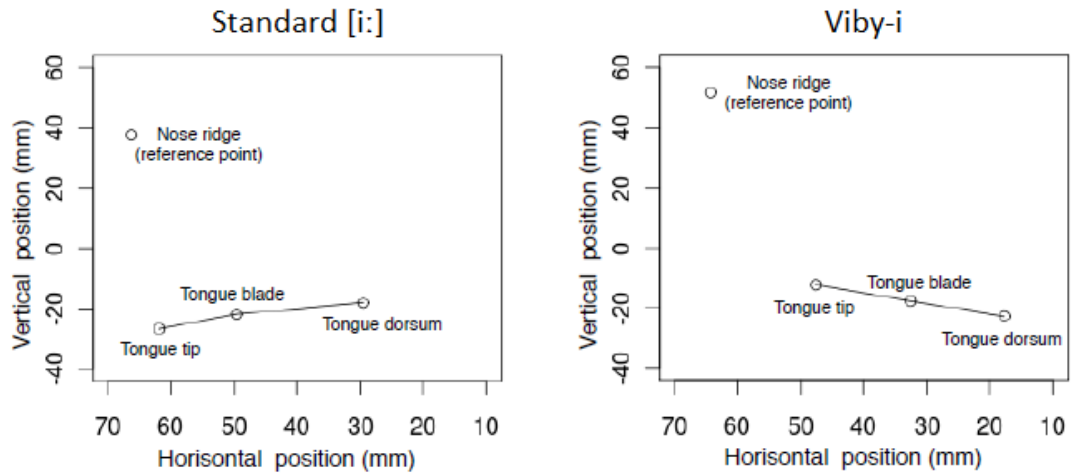


Figure 4.11: EMA measurement points indicating the tongue gesture for standard [i:] vs. Viby-i (Schötz et al., 2011: 3).

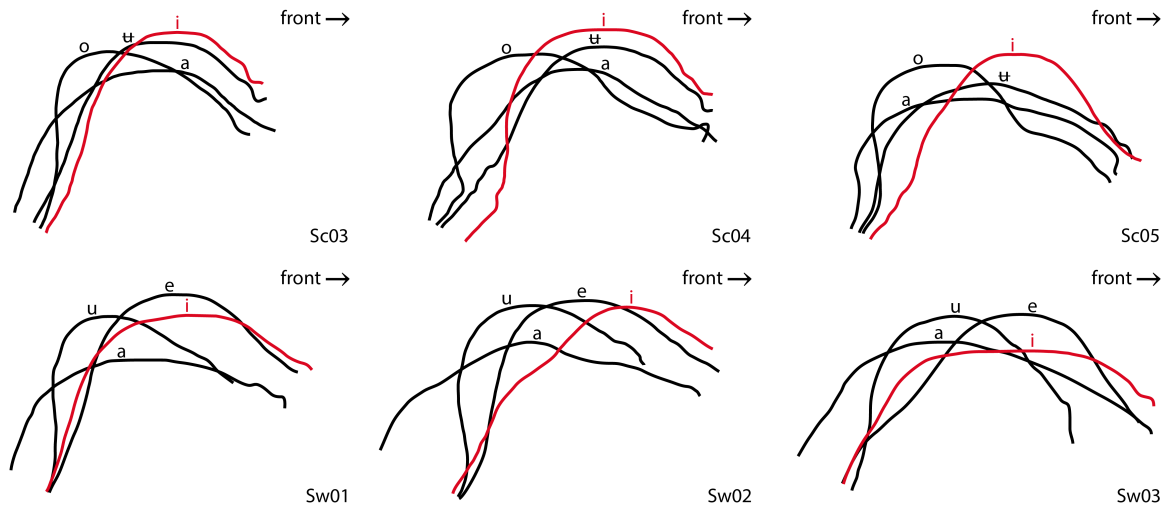


Figure 4.12: UTI tongue splines for three Scottish speakers producing /i:, ʉ:, a:, o:/ and three Swedish speakers producing /i:, e:, a:, u:/ (data from Westerberg, 2013).

#### 4.5.4 Static tongue contours (UTI)

As a precursor to the current project, I conducted two small-scale articulatory studies investigating the tongue and lip gestures used for Viby-i (Westerberg, 2013, 2016). The first study (Westerberg, 2013) focused on data from three young female speakers from West Central Sweden (2 from Gothenburg, 1 from Varberg), and used ultrasound tongue imaging to compare the tongue gestures for Viby-i to /e:, a:, u:/. These results were then compared to three female Scottish speakers producing /i:, ʉ:, a:, o:/. The mean tongue shapes for each speaker are shown in Fig. 4.12.

Compared to the other vowels in the system, Scottish /i:/ was consistently the highest and frontest vowel, whereas Swedish /i:/ was more variable. The Varberg speaker (Sw02) had a relatively high front tongue gesture, similar in height to /e:/, but fronter. The two Gothenburg

speakers produced their /i:/ with a much lower, flatter tongue shape; in both cases lower than /e:/, and in one case lower than /a:/. The speaker with this low tongue gesture (Sw03) was also described as having an auditorily stronger Viby-i (Westerberg, 2013: 25). The difference in tongue gesture between the two cities was taken as an indicator of regional variation, with the reservation that social factors could also be present, since the two Gothenburg speakers produced Viby-i with different vowel qualities from each other (Westerberg, 2013: 34).

In the front-back dimension, /i:/ was classified as a front vowel for all participants, but differences in tongue shape hindered any further comparison. The flattened tongue shape used by two of the Swedish speakers rendered it difficult (and perhaps less meaningful) to measure the highest point of the tongue, since the blade of the tongue was fronted, while the root was backed. Thus, the highest point of the tongue would not necessarily capture the main constriction in the vocal tract.

Westerberg (2013) did not have an acoustic component (its acoustic values are provided for the first time in Section 4.4) but it briefly discusses the relationship between the Swedish speakers' tongue shapes and the resulting vowel qualities, as I perceived them. As previous literature has indicated (e.g. Björsten & Engstrand, 1999; Faytak & Lin, 2015), this relationship appears to be non-linear, in that similar auditory outputs are not always achieved by the same tongue gesture. Surprisingly, the effect of Viby-colouring mainly appeared to be produced, not by tongue backing, as might be expected from the low F2, but by tongue *lowering*.

Following this study, Westerberg (2016) collected a larger dataset of young male and female speakers from different parts of Central Sweden (Stockholm, Gothenburg, Varberg, Jönköping, Katrineholm), combining ultrasound tongue imaging, acoustic analysis, and a small listening experiment (see Section 4.3) to investigate possible regional or social variation in Viby-i.

A selection of the tongue spline data is shown in Fig. 4.13. The data showed a great deal of variation in the midsagittal tongue gestures used to produce Viby-i. Again, Viby-i consistently had a lower tongue gesture than /e:/, and the blade of the tongue tended to be fronted, while the tongue root was backed. In this study, backness was measured at half the vertical height of the tongue (rather than the highest point), in order to account for constriction at the back of the oral cavity. This measure was found to correlate well with perceived Viby-i strength, but not with F2. At the same time, tongue height was found to correlate well with F1, but not with Viby-i strength (Westerberg, 2016: 83-85).

The study suggests that the mismatch between F2 and tongue backing could be attributed to the backness measure not accurately capturing the tongue constriction (Westerberg, 2016: 91); an issue which will be addressed in the current methodology. However, the study also raises the possibility that speakers might use compensatory behaviours to lower F2 without backing the tongue body. For example, many speakers used a “double-bunched” tongue shape, similar to that described by Borgström (1913), where the front and the back were simultaneously constricted (see VBG\_YF\_02 and STH\_YF\_01 in Fig. 4.13). This shape was not associated with

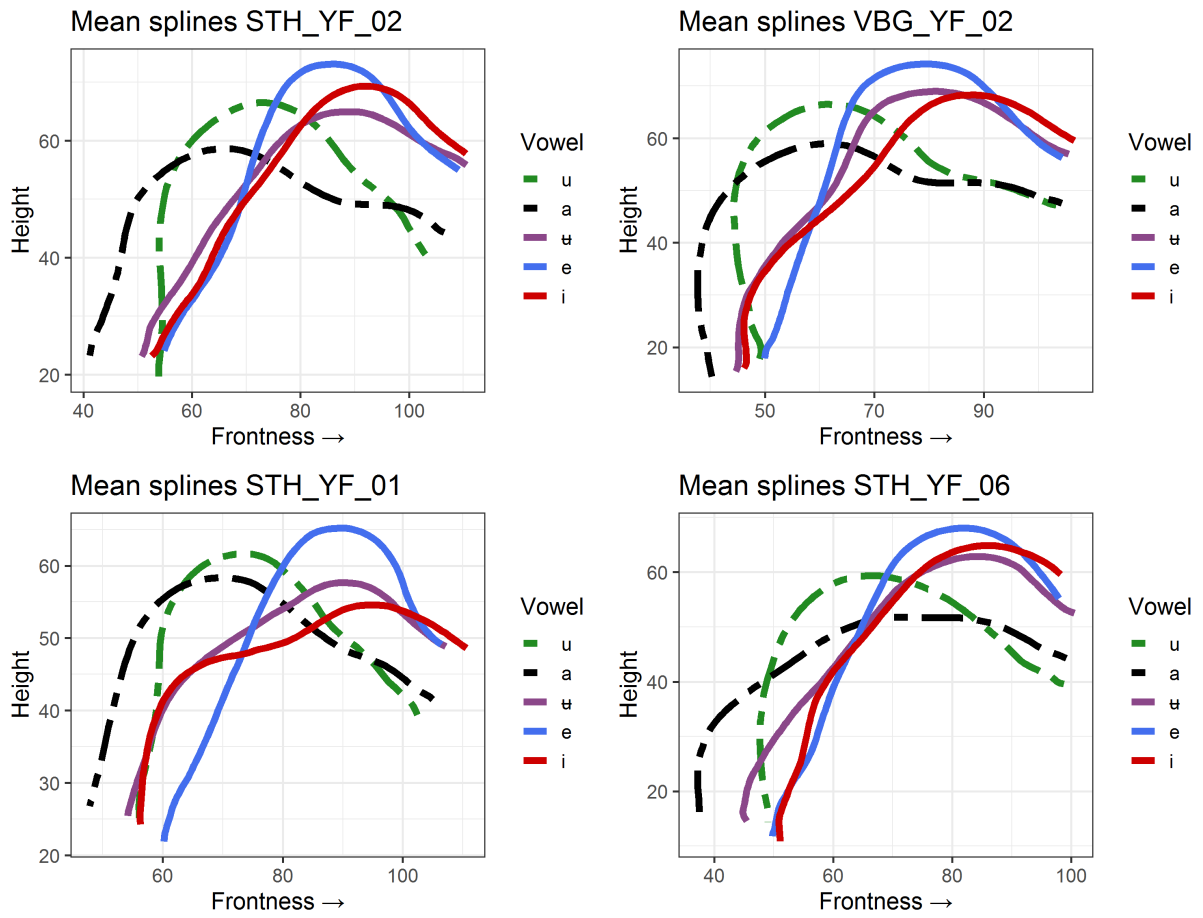


Figure 4.13: Ultrasound splines of four Swedish speakers producing /i, e, ɐ, a, u/ (data from Westerberg, 2016).

any difference in auditory ratings, but the study suggests that tongue shape and tongue position may have worked together to achieve the desired acoustic output (Westerberg, 2016: 84).

As suggested by Björsten and Engstrand (1999), tongue tip raising could also influence F2, but since the tongue tip is commonly obscured in ultrasound images, it was not possible to investigate this aspect further. However, some information about tongue tip behaviour was provided by the lip data.

Westerberg (2016) collected images of the lips to study their position, as lip rounding is known to lower the formants, particularly F3, but also F1 and F2 (Engstrand, 2004: 99). However, the size and shape of the mouth opening for Viby-i was found to be similar to /e:/ (Fig. 4.14). Thus, the lip posture for Viby-i was not rounded, but also did not appear to be actively spread; instead, it was described as “lax” (Westerberg, 2016: 96). What the lip videos did reveal, however, was that many speakers produced Viby-i with an extremely fronted tongue, to the point where the tip could sometimes be seen protruding between the front teeth (see Fig. 4.14). This phenomenon was not investigated quantitatively, but could be related to F2 lowering, as Björsten and Engstrand (1999) indicate that tongue tip raising could lower F2. The current study further investigates the relationship between the teeth and the tongue tip on the ultrasound image by using a bite plate, as described in Chapter 5.

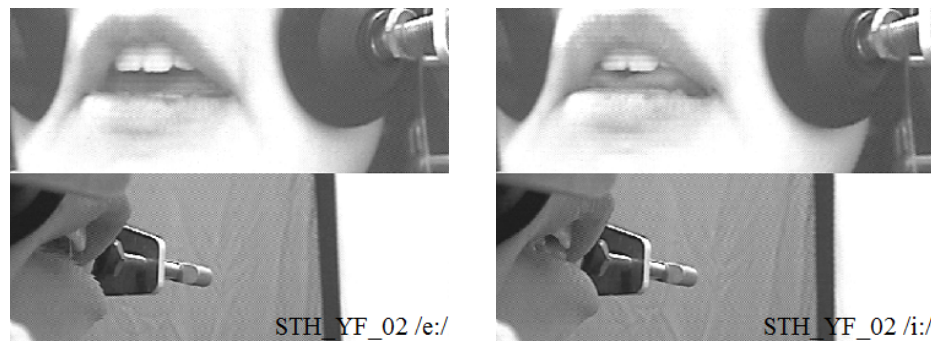


Figure 4.14: Example of lip postures for /e:/ and /i:/ in Westerberg (2016). Note the fronted tongue position for /i:/.

The fact that Viby-i simultaneously has a high F1 and low F2 speaks against the idea that lip-rounding is responsible for its peculiar vowel quality; if this was the case, we would expect both formants to be lowered. Similarly to Westerberg (2016), Gross and Forsberg (2019) also show that Viby-i is produced with spread or ‘lax’ lips. To account for the acoustic overlap between Viby-i and Viby-y, Gross and Forsberg conducted a perceptual experiment where three phonetically trained listeners classified audio tokens of Viby-y according to their perceived lip-rounding. They found that most speakers seemed to produce Viby-y with either spread or lax lips, and that the ‘lax’ category was essential to classifying the data (Gross & Forsberg, 2019: 10). However, the study also found a great deal of individual variation in perceived lip posture.

In summary, the number of experimental studies on Viby-i remain few, with small sample sizes and often a narrow investigative focus, leaving room for further acoustic and articulatory investigation of Viby-i, which this thesis seeks to provide.

## 4.6 Geographic spread and variation

### 4.6.1 Rural vs. urban Viby-i

Viby-i is traditionally associated with a number of rural dialects, notably those of Viby in Närke; Orust, Tjörn, and other parts of southern Bohuslän; Liden, Sundsvall and other parts of the Indalsälven valley in Medelpad; southern parts of Dalsland; Norra Ny and other parts of northern Klarälven valley in Värmland; Mariannelund and other parts of eastern Östergötland, as well as bordering parts of north-eastern Småland (Björseth, 1958; Bruce, 2010; Elert, 1995; Götlind, 1940; Pamp, 1978; Wessén, 1945). In addition to these rural locations, Viby-i is also known to occur in Sweden’s two largest cities: Stockholm and Gothenburg (Björseth, 1958; Bruce, 2010; Elert, 1995; Kotsinas, 2007; Malmberg, 1971; Wessén, 1945). A map of these locations is provided in Fig. 4.15.

It is not clear what data the map from Elert (1995) is based on, how old it is, or which criteria were used to distinguish Viby-i from standard [i:], but this list of locations seems to persist in

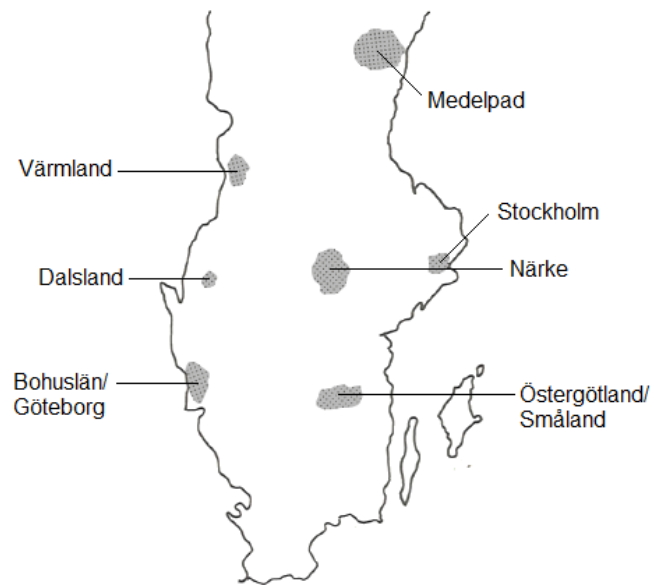


Figure 4.15: Map of locations where Viby-i has been documented (after Elert, 1995: 45).

the literature. It is possible that this information has simply been passed down anecdotally, as a “collected, comprehensive description of vowel variation in Swedish does not yet exist” (Bruce, 2010: 103).

Interestingly, several sources report that, while Viby-i can be found in Stockholm and Gothenburg, nearby suburban or rural areas do not use it. According to Holmberg (1976), the areas directly around Gothenburg use “neither hissing nor damped ‘i’”, although he posits that it may have existed on the island of Hisingen, which is a suburb of Gothenburg (Holmberg, 1976: 10). Furthermore, Holmberg claims, Viby-i is “not known to have existed in neighbouring parishes in Västergötland or Northern Halland” (Holmberg, 1976: 10). Similarly, Frid et al. (2015) find that speakers from the outskirts of Stockholm and Gothenburg are less likely to use Viby-i than speakers from the city centre (Frid et al., 2015: 3), and these results were supported by Westerberg (2016: 105). However, Grönberg (2004) finds that Viby-i can occur in the speech of young people from Alingsås, a suburb of Gothenburg, although it is less common than other /i:/ variants, and is largely limited to speakers who use few local features (Grönberg, 2004: 228-231).

According to Elert (1995), the rural dialects that do feature Viby-i are likely to avoid this sound, as it is often stigmatised:

Whether speakers in these areas use [Viby-i] depends on how closely they associate with the genuine [local] dialect. Viby-i and Viby-y are dialectal features that tend to be put aside when speakers move closer to standard speech. (Elert, 1995: 45)

It may thus be the case that Viby-i is gaining popularity in large cities and nearby areas, while it is waning in rural speech. However, it is difficult to predict how this change will manifest itself if Viby-i becomes accepted as the standard /i:/ variant. If rural and urban Viby-i are realised slightly differently, it may be the case that speakers will simply exchange one realisation of this



vowel for another.

## 4.6.2 Regional variation

As mentioned in Section 4.3, there are some accounts that suggest that Viby-i may be subject to regional variation, or at least that the degree to which this vowel is either fricated or damped may vary. For example, much of the literature on Gothenburg Viby-i describes this vowel as uniquely fricated, buzzing, or hissing (e.g. Grönberg, 2004: 226). Björseth has the following to say about the sound which he calls “Gothenburg-i” (IPA symbols inserted for clarity):

The long /i:/ sound, and occasionally also its short counterpart, shows a tendency towards “buzzing” in Gothenburg Regional Standard (and Gothenburg Dialect), in a way that is reminiscent of the so-called Viby-i. This tendency is also noticeable for /y:/, but not for /æ:/ or /e:/, as is the case in the Viby[-i] area of Bohuslän. It would be inaccurate to denote this Gothenburg /i:/ as a true Viby-i. In the Gothenburg pronunciation, the vowel is slightly buzzy, due to a higher tongue position and a narrower air passage than in Standard Swedish. The peculiar throat tension associated with Viby-i is missing. One may ask whether the buzzing /i:/ in Gothenburg has any historical connection to the Viby-i used in Bohuslän. [...] One should probably not count on such a connection. Instead, one would probably be more correct in regarding this buzzing as a general trend within the Gothenburg vowel system, which could be described as centrifugal. In other words, the back vowels have tendency to shift back, and the front vowels have a tendency to shift forward in the oral cavity. (Björseth, 1958: 7)

This evaluation is shared by Holmberg (1976), who describes Gothenburg /i:/ as “hissing” and “reminiscent of /z/” (Holmberg, 1976: 10). Both authors also mention that some speakers who use the fricated Gothenburg variant diphthongise their productions by inserting an /e/- or /ɛ/-like vowel before it, e.g. [fæi:nt] (*fint*, ‘fine’, St. [fi:nt]) (Björseth, 1958; Holmberg, 1976). From personal experience, I have not observed this kind of diphthongisation together with Viby-i, but I have heard the two separately. I have also witnessed both rural and urban speakers using Viby-coloured /y:/, but have only ever heard rural speakers use Viby-coloured short /ɪ/ and /ʏ/.

The use of Viby-colouring for short vowels and other vowels than /i:, y:/ appears to be dialect-specific; for example, Pamp (1978) describes /e:/ as sometimes being Viby-coloured in Närke and Östergötland (Pamp, 1978: 94), and states that /e:/ in Medelpad takes on a quality close to standard [i:]. Furthermore, Pamp specifies that Viby-coloured short vowels in Närke and Östergötland are linguistically constrained, in that they never occur before /k/, /g/, or /ŋ/ (Pamp, 1978: 94). In contrast, Bruce (2010) claims that Swedish Viby-i is not subject to any linguistic constraints, and also appears to endorse this vowel being used categorically by speakers (Bruce, 2010: 136).

Some indications of dialectal variation is also found by Grönberg (2004) in her study of Alingsås, a suburb of Gothenburg. Here, speakers either use a standard [i:], a “lowered” version of this vowel, or a “fricativized” version (Grönberg, 2004: 223-224). Grönberg does not provide acoustic data for these vowels, but her description brings to mind two different kinds of Viby-i, of which one may be ‘damped’ and the other fricated. The lowered vowel was common in speakers who used more local variants, and the fricated vowel was more common in speakers who adopted more urban (Gothenburg) features (Grönberg, 2004: 248).

In addition to Viby-colouring affecting different vowels, Bruce (2010) notes that this quality is also sometimes applied to the consonant /j/ in frequent words like *hej* ‘hello’ [hɛj] or *okej* ‘okay’ [ɔkɛj], especially in East Central Sweden (Bruce, 2010: 201). I too have noticed this phenomenon, and I have seen the word *hej* spelled ‘heiii’ in personal correspondence. One source, cited in Section 4.7, extends this claim by stating that Viby-colouring can in fact be applied to the whole alphabet.

### 4.6.3 Chronological spread

The use of Viby-i in Gothenburg is documented in the linguistic literature since at least the 1950s (Björseth, 1958), but given that it appears to already have been established by then, it has likely existed there much longer. It is possible that the Gothenburg Viby-i originates from the nearby rural dialects of Bohuslän, but according to Björseth, this is unlikely, since the vowel qualities differ between the two locations (Björseth, 1958: 7). A similar observation is made by Grönberg (2004: 224). As mentioned earlier, Bohuslän Viby-colouring also affects short phonemes, which is not the case in Gothenburg (Björseth, 1958: 7).

According to Elert (1995), Viby-i is likely older in Gothenburg than in Stockholm (Elert, 1995: 45), but there are also records of this vowel occurring in Stockholm in the 1950s (Langenfelt, 1953: 134). However, in the 1970s, Viby-i was still considered an innovation there. For example, Holmberg (1976) writes that a “hissing /i:/” has “in recent decades, and under mysterious circumstances, appeared among Standard speakers from Djursholm and Stockholm” (Holmberg, 1976: 10).

The Stockholm vowel appears to be similar to the Gothenburg one, in that it too is usually described as highly fricated (e.g. Holmberg, 1976; Kotsinas, 2007), or, at the very least, as having an equivalent vowel quality (e.g. Björsten & Engstrand, 1999; Bruce, 2010; Elert, 1995). It also seems that Viby-colouring only affects long vowels, and is associated with prestige, in both cities. The latter point will be discussed further in Section 4.7.

The similarity between Viby-i in East and West Central Sweden, as well as evidence pointing to this vowel quality being older on the West Coast, has led to speculations that the Stockholm Viby-i originally came from rural Bohuslän, the province just north of Gothenburg. This suggestion was first introduced by Stähle (1981), who writes:

[N]ew features, likely adopted from other parts of the country, include the buzzing /i:/ commonly used by youths in e.g. Djursholm and Lidingö (the high-society language of Marstrand summers?). (Ståhle, 1981: 86-87)

Djursholm and Lidingö are both wealthy suburbs of Stockholm, the latter giving Viby-i its local name in Stockholm, *Lidingö-i*. The quote refers to the fact that, since the early 1900s, it has been popular for affluent Stockholm residents to spend their summers on the West Coast, particularly in Bohuslän (e.g. Smögen, Marstrand, Lysekil). Ståhle's comment suggests that these holidaymakers adopted Viby-i from the local population, and used it as an in-group marker to display their wealth once they returned home. This idea has recently been popularised by TV personality Fredrik Lindström through a series of linguistic TV programmes (af Klintberg & Ripås, 2003; Hellberg, 2012), as well as an article written for the popular language magazine *Språktidningen* (Lindström, 2014). Lindström posits that Stockholm children would have been particularly likely to adopt Viby-i from their local playmates, and used it to signal group identity in schools (Lindström, 2014).

While this scenario is technically possible, the only mention of it in the academic literature besides Ståhle (1981) is Kotsinas (2007), who writes that “this is an uncertain matter, and the evidence is vague” (Kotsinas, 2007: 109). To my knowledge, there is no linguistic research to either support or refute that Viby-i spread from Bohuslän to Stockholm. My main objection to this theory would be that Viby-i in Stockholm appears to be more similar to the Gothenburg variant than to the rural one. In addition, it seems unusual, although not impossible, that a low-status rural feature would find its way into urban high-society language, especially given the limited periods of contact between the two groups. It seems more likely that Viby-i has appeared in Stockholm as a result of dialect contact within the city itself (e.g. Norrby & Håkansson, 2015: 77), or that it has “leapfrogged” to Stockholm from Gothenburg, as suggested by the gravity model of language transmission (Chambers & Trudgill, 1988: 166). Björsten and Engstrand (1999) also suggest that, since [i] has developed independently in many widespread language groups, “the Swedish situation, where [i] occurs in several widely separated dialects, might be a mirror image of [areal tendencies in other language families]” (Björsten & Engstrand, 1999: 1,959). In other words, there may be underlying structures in the Swedish vowel system that has led Viby-i to arise independently in multiple locations.

In recent years, there have been several reports of Viby-i spreading rapidly both within and around Gothenburg and Stockholm. Data from the 1990s indicates that Viby-i was becoming increasingly fricated in younger generations in Stockholm (Kotsinas, 2007: 109), and shows the usage patterns of a sound change in progress (discussed further in Section 4.7). Support for this ongoing change is also voiced by Riad (2014: 21). In Gothenburg, Viby-i appears to be well-established in speakers from Swedish cultural backgrounds, but can now also be seen in some speakers of Gothenburg multiethnolect (Gross, 2018; Westerberg, 2016), despite previous literature stating that multiethnolect speakers do not use this feature (Bruce, 2010: 225). Viby-i

has also been documented in higher sociolects of South Swedish, where it appears to be new (Bruce, 2010: 136). Based on personal observation, Viby-i currently seems more or less ubiquitous throughout Central Sweden, as it is frequently heard in everyday speech, as well as more formal language contexts, e.g. news broadcasts, and in the speech of the Royal Family (Melin & Melin, 2005: 21). Regarding the rapid spread of Viby-i, Bruce (2010) writes:

This Viby-colouring appears to be particularly contagious and difficult to resist, to the extent that one should wish to do such a thing. It would not be surprising if this pronunciation of long /i:/ and /y:/ was established as standard within a few decades. (Bruce, 2010: 216)

## 4.7 Social significance

### 4.7.1 Prestige associations

Viby-i is well-known, both in the literature and in Swedish culture, to be associated with specific social signals. However, these signals vary depending on the context in which the sounds are used. According to Elert (1995), “the social evaluation of [Viby-i] is completely different in Stockholm and Gothenburg compared to [...] areas where they form part of the genuine dialect” (Elert, 1995: 45). He further states that Viby-i in urban areas is perceived as “somewhat snobbish” (Elert, 1995: 45), while rural speakers are only likely to use it if they feel a strong connection to the local dialect. Bruce mirrors these comments, stating that “In some dialects, [Viby-i] is perceived as provincial with relatively low prestige, while, in urban varieties, it is instead perceived as prestigious and trendy” (Bruce, 2010: 135). It is not clear whether this contrast is caused by a difference in how Viby-i is produced, or simply the expectations of the listener, but is nevertheless in line with sociolinguistic expectations of urban speech (which may be closer to the standard) having higher social status than rural speech (Bruce, 2010: 19).

In Stockholm in particular, the use of the name ‘Lidingö-i’ for this vowel is telling, since Lidingö is one of the wealthiest municipalities in Sweden (Kederstedt, 2007). Although Viby-i is stereotypically perceived as a high-status feature in Stockholm (as supported by e.g. Kotsinas, 2007), recent work by Westerberg (2016) and Norstedt (2019) have found that speakers of lower class backgrounds use it too.

As in Stockholm, Gothenburg Viby-i has also been explicitly described as a prestige marker. For example, Björseth (1958) states that “This linguistic feature should not be regarded as vulgar, but rather the opposite” (Björseth, 1958: 7). To some extent, Gross (2018) supports the link between higher social status and the use of Viby-i in Gothenburg, although in Gross’ study, social class overlaps somewhat with cultural heritage and integration.

As might be expected for a prestige form, observations of Viby-i indicate that this is a feature used primarily by women (Elert, 1995; Kotsinas, 2007). In Gothenburg, Björseth describes Viby-

i as distinctly gendered, used by women and girls regardless of social class, but used only by men “who make an effort to speak clearly and politely, e.g. priests and teachers” (Björseth, 1958: 7). On the other hand, older male speakers, especially those who tend towards traditional Gothenburg dialect, do not use this sound (Björseth, 1958: 7).

### 4.7.2 Change in progress

Apart from signalling gender, it seems that Viby-i is mainly used by younger speakers, and this aspect tends to be stressed more in recent literature (e.g. Bruce, 2010; Elert, 1995; Kotsinas, 2007). There are thus indications of Viby-i spreading or undergoing change, at least in some areas.

In her Stockholm data from the 1990s, Kotsinas (2007) shows that Viby-i was primarily used by young women from higher class backgrounds, a sociolinguistic profile commonly associated with linguistic change in progress (e.g. Eckert, 1989; Labov, 2001). The feature was also used, to a lesser degree, by young middle-class men. Working-class speakers were less likely to use Viby-i, but girls did so to a greater extent than boys. In working-class boys, Viby-i was virtually absent (Kotsinas, 2007: 109-110). Kotsinas concludes that Viby-i in Stockholm is mainly associated with high status, but that stronger versions of this vowel are particularly indexical of young women (Kotsinas, 2007: 111). Kotsinas also describes Stockholm Viby-i as a feature which speakers generally notice and comment upon (Kotsinas, 2007: 108).

Since the 1990s, there have likely been further developments of the /i:/ vowel in both Stockholm and Gothenburg, which have not yet been studied. In Gothenburg in particular, we may expect some change, since its sociolinguistic picture has not been updated since the 1970s (Holmberg, 1976). According to Riad (2014), as Viby-i is spreading, it “is now in the process of losing its social charge” (Riad, 2014: 21). This theory is supported by more recent data from Westerberg (2016), who did not find any auditory, acoustic, or articulatory differences in Viby-i production based on social factors, except a weak trend towards an auditorily stronger Viby-i in speakers from higher social backgrounds (Westerberg, 2016: 68). These findings could be indicative of Viby-i becoming socially ‘neutralised’ in urban areas, but further research with more diverse samples would be required to confirm this. The relationship between Viby-i production and social factors will also be investigated in this thesis.

### 4.7.3 Negative associations

Although Viby-i is used as a prestige marker in cities, Elert (1995) nevertheless remarks that “those who notice these speech sounds do not usually appreciate them” (Elert, 1995: 45). Thus, there is a strange contrast in how Viby-i is perceived: On the one hand, it is regarded as ‘posh’, but on the other hand, speakers are often reluctant to admit that they use it. This could be because drawing attention to social class, and particularly flaunting wealth, is taboo in Sweden (Norrby &

Håkansson, 2015: 93). However, there could also be a lingering association of Viby-i with rural dialects and non-standard language, reflected in the following comments from older literature:

Strangely enough, the hissing pronunciation of /i/ has long appeared even in Standard Gothenburg speech. [...] That this language feature, as Björseth points out, is not regarded as vulgar but rather ‘refined’, seems to be confirmed by the fact that this sound has, in recent decades, and under mysterious circumstances, appeared among Standard speakers from Djursholm and Stockholm. Well, ‘refined’ or not, hissing /i/ should surely still be regarded as dialectal. (Holmberg, 1976: 10)

This so-called Viby-i [...] appears to be spreading in several places. It is an articulation, which in the interest of well-sounding and clear speech, should be counteracted. (Malmberg, 1971: 60)

To some extent, this attitude is also (albeit perhaps jokingly) reflected in the comment that Viby-i “appears to be particularly contagious and difficult to resist, to the extent that one should wish to do such a thing” (Bruce, 2010: 216).

To investigate this phenomenon further, a recent study by Norstedt (2019) looked at attitudes and style-shifting of Viby-i in 16 young Stockholm speakers (8 male, 8 female). The study was interested in whether speakers were aware of their own use of Viby-i, and whether they would style-shift either away from, or towards it, when the topic of Viby-i was brought up.

Norstedt found that all speakers in the sample were highly aware of Viby-i and its use as a status marker, but only three speakers (2 female, 1 male) considered themselves to be ‘Viby-i users’. In interview extracts, several participants expressed that they thought Viby-i sounded “ridiculous” (Norstedt, 2019: 23-24), one speaker stated that they “sometimes use it a little bit, but definitely not on purpose” (Norstedt, 2019: 24), and another stated that she “is trying to get rid of it” (Norstedt, 2019: 23). Meanwhile, these speakers’ acoustic values for /i:/ were highly reminiscent of previous Viby-i data from Westerberg (2013, 2016), with F2 ranging roughly between 1,500-2,000 Hz in both men and women (Norstedt, 2019: 22-23).

In addition, Norstedt did not find any correlation between social class and the use of Viby-i (Norstedt, 2019: 20). Note that the speakers in this study were not recruited on the basis of having Viby-i in their speech, but were selected simply on the basis of being from Stockholm.

In terms of style-shifting, Norstedt found that regardless of their group affiliation, most speakers either did not style-shift, or used a stronger Viby-i when the interviewer asked them about this vowel. Only three speakers used a weaker Viby-i in the metalinguistic context: two females who claimed to be averse to this vowel, and one male speaker who identified as a Viby-i user. Norstedt suggested that the reason why most participants continued (or exaggerated) their use of Viby-i, even when they did not affiliate with this sound, was because they were trying to align with the researcher’s expectations (Norstedt, 2019: 26, referencing Schilling-Estes, 1998). According to Norstedt, “When I introduced the speakers to the second topic of my study,

Viby-i, all speakers carried out a speech performance [of] the variant, without being asked” (Norstedt, 2019: 26). At the same time, it is possible that speakers considered this performance of a “strong” Viby-i as the true form of this variant, while disregarding the form that they normally used.

The present study will not be investigating style-shifting of Viby-i, as the speech material analysed here consists only of word list data. Spontaneous speech was recorded, but could not be analysed due to time constraints. This material will nevertheless be available for future research.

#### 4.7.4 Viby-i in popular culture

The dual attitude towards Viby-i is reflected by an association in the public consciousness between this sound and its recurring description as *fint* ‘fine’ (commonly [fɪ:nt]). This word can either be translated as ‘beautiful, refined’, or as ‘pretentious, posh’. In other words, Viby-i appears to be a good example of a sound that is “overtly despised but covertly imitated” (Wells, 1994: 205). An example of this can be seen in the humorous popular linguistics book *Fiint språk* (‘refined language’) (Melin & Melin, 2005). The spelling of the title references Viby-i in a way that is immediately clear to most Swedish speakers, but never directly addressed by the authors. Within the book, the alternate spelling “fzznt” is also introduced (Melin & Melin, 2005: 17–20). In this chapter, the authors describe Viby-i as “the most well-known status marker in the Swedish language” (Melin & Melin, 2005: 20), and also state that “there are different degrees of buzziness, ranging from barely noticeable to somewhat insect-like” (Melin & Melin, 2005: 20). The association with insects again suggests that this sound carries negative connotations.

Viby-i is also a frequent feature in satirical depictions of the upper class and the Royal Family, notably in the popular comedy programs *Hey Baberiba* (S. Lindberg, 2005) and *SNV News* (Ericstam & Nilsson, 2013). However, some actors appear to use this feature more than others, possibly due to different levels of awareness or linguistic flexibility. Notably, however, *Hey Baberiba* picks up on the distinction between Princess Madeleine’s use of Viby-i, and its absence in other members of the Royal Family (Melin & Melin, 2005: 21).

Comical portrayals of rural Viby-i can also be found in popular culture, e.g. in an interview with actor Peter Flack, known for his comedy character Hjalmar Berglund, who is portrayed as being from Viby (Berzelius, 2012). The article (from a Närke newspaper), makes reference to the character’s pronunciation of /i:/ through spellings like “tiining” (‘newspaper’), “Viiby” (Viby), “viitsipper” (‘wood anemones’), “preciis” (‘precisely’), and, impressively, “sajtsiing” (‘sightseeing’) (Berzelius, 2012).

In my own experience, it is usually only the stronger forms of Viby-i that are met with these social evaluations. Weaker versions seem, for most speakers, to lie below the level of consciousness, illustrated e.g. by the fact that Swedish speakers sometimes use the same vowel quality when speaking English (Norstedt, 2019). Learners of Swedish have also commented on not being taught how to produce this sound:

I have listened to a lot of native speakers of Swedish, particularly people with a Stockholm accent. This accent has a name I have now forgotten. This Stockholm accent has a long i that sounds nothing like what is heard in English. I can only describe it as being augmented to the rest of the pronunciation. It is very distinctive. (Anonymous, 2015)

Similarly, a participant in Norstedt (2019) remarks:

I don't think everyone can hear it, I think it depends on who's listening. [...] I don't think everyone in Stockholm can hear it. (Norstedt, 2019: 24)

This gives the impression that there is a perceptual 'threshold' for Viby-i, and that this threshold may be particularly high in Swedish speakers, perhaps because Viby-i becoming standardised. It seems that Swedish listeners associate the term 'Viby-i' with highly salient versions of this vowel, while weaker Viby-i is not recognised, at least not unless the speaker is expected to use it. This could also be one reason why Viby-i is so rarely mentioned in phonetic descriptions of Swedish.

Another phenomenon that has begun to appear in the portrayal of Viby-i is a particular facial posture where the tongue is protruded, as shown in Fig. 4.16. This facial expression recurs in satirical portrayals of the upper class in Swedish popular media. The association between Viby-i and this facial expression can be seen from the satirical opinion piece below:

#### **Who the hell wants to see your floppy tongues?**

When certain types of imagined-posh young girls from the capital speak, you can see their tongues. The whole time. The tongue sits like a lid on the lower jaw, all the way up to the inside of the lower lip. Like this: Take the infamous Lidingö-i with its strange tongue position. Then transfer that to as many of your other speech sounds as possible (if you're good, you can do almost the whole alphabet). Voila! There it is. Since people think the aforementioned Lidingö-i signals status and sophistication (even though we've inherited it from Närke farmers), it makes sense that it would be even more sophisticated to use the same style throughout all of your speech. It misses the mark, though. Because when your tongue flops around at the bottom of your mouth you don't look very sophisticated. You look dumb. Like a village idiot. Or like you have some kind of annoying illness. *Paresis lingualis* or *villagus imbecillus*. Or something. Take my advice and put your tongue back in your stupid mouth. (And take that Canada Goose jacket off while you're at it.) (Anonymous, 2007)

Similar pop-cultural connections between vocal tract settings and socially stereotyped facial postures have also been documented in Californian English (in the 'valley girl/surfer dude' persona) (Pratt & D'Onofrio, 2017) and in upper-class Southern Standard British English (in the



‘gap yah’ persona) (Holmes-Elliott & Levon, 2017). In both cases, speakers parodying this style produce both the acoustic characteristic of a compressed vowel space, and the visual cue of an “open-mouth, protruded jaw setting” (Pratt & D’Onofrio, 2017: 288). Holmes-Elliott and Levon (2017) anecdotally report that this setting sends the social signal, “I’m too posh to move my mouth”.



Figure 4.16: Exaggerated Viby-i facial posture demonstrated by Swedish comedian Jonas Fagerström (Fagerström, 2019). Sign reads “super rich millionaire from Stockholm”.

## 4.8 Similar vowels in other languages

Some of the debate surrounding the articulation of Viby-i has to do with how this sound should be classified in terms of the world’s languages. The question is whether this sound should be regarded as “odd” (Engstrand et al., 1998) or “exotic” (Schötz et al., 2011), or if it belongs to a pre-existing category of sounds that are fairly common in other languages. Given the limited articulatory data, as well as possible variation in the production of Viby-i across Sweden, this classification has been difficult to establish.

### 4.8.1 Occurrence in the world’s languages

According to Björsten and Engstrand (1999), Viby-i can be classified as a high central unrounded [ɨ], equivalent in auditory quality and acoustics to the Turkish phoneme /i/. By extension, they believe that Viby-i is part of a vowel category that is

fairly wide-spread among the world’s languages. It can be assumed that this vowel type is produced with [additional] apicalization in some languages since [...] this would further enhance its damped quality. (Björsten & Engstrand, 1999: 1,960)

In other words, Viby-i may not be realised exactly the same as [i], but there may be sufficient similarity to group them into the same category. The authors are nevertheless tentative in this conclusion, given the limited amount of data in their study.

In a separate paper, Engstrand et al. (1998) also point out that it is unusual for languages to feature [i:] as an allophone of /i:/, rather than as a separate phoneme (Engstrand et al., 1998: 92). Furthermore, while the formant patterns of Viby-i suggest that this vowel is centralised, there may be other acoustic characteristics which render Viby-i different from [i], e.g. formant intensity, bandwidth, or frication. The latter will be investigated in more detail in this thesis.

## 4.8.2 Mandarin apical vowels

The fricative aspects of Viby-i in particular have led it to be compared to apical, fricative, or strident vowels, found in various dialects of Chinese (e.g. Faytak & Lin, 2015; Shao & Ridouane, 2019), as well as some Bantoid and Tibeto-Burman languages (Faytak, 2014: 1). These sounds are relatively rare across the world, and defy the common phonetic definition of vowels as periodic sounds produced without obstructions in the vocal tract (e.g. Ladefoged, 2001). Faytak (2014) points out that these sounds present an interesting mismatch between phonological use and phonetic form (Faytak, 2014: 1), in that they do not conform to traditional ideas of how vowels ‘should’ behave.

Since the current research will provide ultrasound data of the tongue gestures associated with Viby-i, it is relevant to briefly review the main points from Faytak and Lin (2015) on the articulation of apical vowels in Mandarin Chinese, so that the articulatory characteristics of these vowels can be compared. Faytak and Lin (2015) investigated the articulation and dynamics of alveolar [ɿ] and post-alveolar [ʮ], produced by 4 female and 1 male speaker of Standard Mandarin. In Mandarin, these sounds only occur directly after [s] and [ʃ] respectively, and “are typically either weakly fricated or free of frication” (Faytak & Lin, 2015: 1). The authors further report that the vowels in their sample “exhibit[ed] essentially no fricative noise” (Faytak & Lin, 2015: 3).

The authors find that the tongue gesture for [ɿ] was similar in height and retraction to [a], but the blade and tip were raised close to the palate (Fig. 4.17). The vowel [ɿ] is reported to behave similarly (Faytak & Lin, 2015: 2). The lowered tongue body, with a characteristic ‘dip’ in the middle of the tongue, accompanied by tongue tip raising, bears close resemblance to descriptions of Viby-i, both from the proprioceptive literature (Borgström, 1913), the articulatory simulation run by Björsten and Engstrand (1999), and the articulatory data of Westerberg (2013, 2016).

Faytak and Lin also note that the apical vowels were extremely similar in tongue position and shape to the preceding fricative consonants [s] and [ʃ] (Faytak & Lin, 2015: 2). Fig. 4.18 shows the speakers with the greatest (S1) and smallest (S8) displacement between the preceding consonant and the vowel respectively. The similarity between the apical vowels and their preceding consonants was unexpected since, as mentioned before, the vowels themselves contained

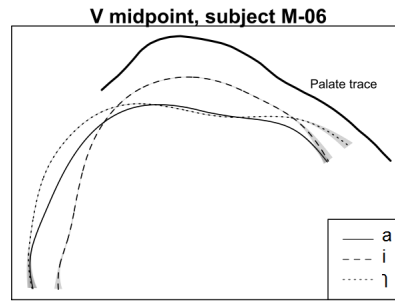


Figure 4.17: Tongue contours for [a, i, ɿ] in one Mandarin speaker (Faytak & Lin, 2015: 2).

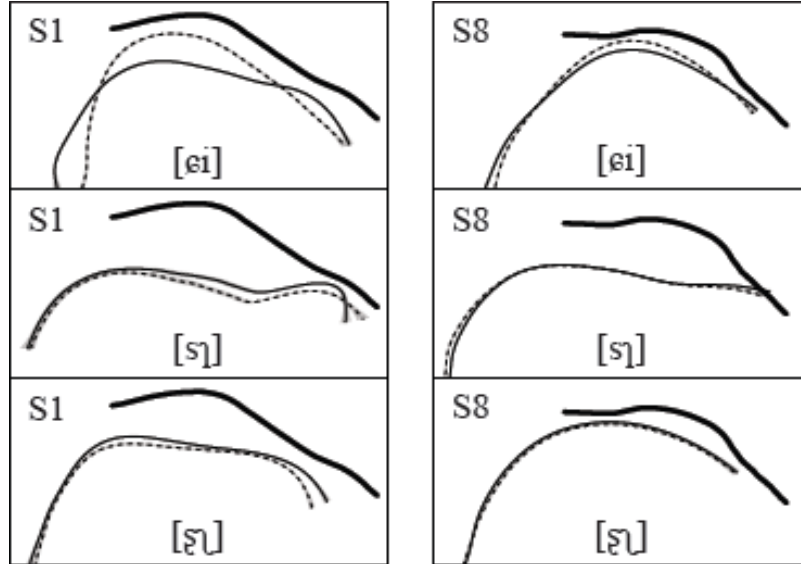


Figure 4.18: Tongue contours of two Mandarin speakers, comparing [i, ɿ, ʊ] to preceding [ɛ, s, ʂ] (Faytak & Lin, 2015: 3).

virtually no frication. The authors postulate that there may have been a widening of the vocal tract, not visible in the midsagittal orientation, which might have reduced the turbulent airflow. To investigate this, they recorded additional coronal ultrasound images, but could not identify any release or widening of the constriction (Faytak & Lin, 2015: 4). They suggest that the lack of fricative noise could be explained by reduced airflow, or reduced intra-oral pressure achieved by expanding the pharynx or lowering the velum (Faytak & Lin, 2015: 4). The latter two strategies could also have an effect on the quality of the vowel, and are reminiscent of the description of Viby-i as having “throat tension” (Holmberg, 1976: 10).

Finally, Faytak and Lin write that “there is a substantial amount of articulatory variation that does not result in correspondingly substantial acoustic variation” (Faytak & Lin, 2015: 4), suggesting that speakers may be using compensatory strategies to create their desired output. This description too bears similarities to the description of Viby-i, as compensatory strategies have been suggested by e.g. Björsten and Engstrand (1999), Bruce (2010), and Westerberg (2016). The present work will address this issue by providing a more in-depth analysis of the relationship between articulatory gesture and acoustic output for Viby-i.

### 4.8.3 Other Scandinavian languages

There have been some reports that *Viby-i* can also be found in other Scandinavian languages, but there does not seem to be any research on this topic. Bruce (2010) states that a similar vowel quality can be found in Copenhagen Danish, but only before [ð] (Bruce, 2010: 136), which could be a case of coarticulatory assimilation. In my impression, however, the ‘damped’ quality of *Viby-i* is fairly reminiscent of some sounds in Danish, and it would be worth investigating the acoustic and articulatory similarities between them. In interactions with other phoneticians, I have also been told that *Viby-i* can be heard in some dialects of Norwegian, but there does not appear to be any records of this in the literature.

## 4.9 Chapter summary

This chapter has summarised the phonetic and sociolinguistic properties of *Viby-i*, showing that this vowel is characterised by a low acoustic F2, which appears to be driving the perception of its unusual vowel quality. In terms of its formant frequencies, *Viby-i* is similar to [i], but there are other factors that problematise this classification, particularly the description of *Viby-i* as ‘buzzy’ or ‘damped’. This thesis will address formant values, formant dynamics and intra-vowel friction, but further work on other aspects of *Viby-i* acoustics are encouraged in the future.

The thesis will also provide articulatory data to address the long-standing question of how *Viby-i* is produced. Based on previous articulatory data, there appears to be some support for the occurrence of articulatory trade-off in this vowel. For this reason, a core part of the analysis will investigate the relationship between articulatory gesture and acoustic output.

In addition, this chapter has demonstrated that *Viby-i* is commonly used as a status marker in urban environments, and that it displays some of the signs of a sound change in progress. There is also the possibility of regional variation. As well as investigating its acoustic-articulatory relationship, this thesis will therefore conduct an analysis of the linguistic and social factors which may influence *Viby-i* production, with a particular focus on how established *Viby-i* appears to be in different varieties of Swedish.

Although this chapter has tried to summarise the available literature on *Viby-i* in a comprehensive way, the research on this vowel is nevertheless characterised by many gaps. Particularly, previous studies have not provided a formal definition of this vowel, and there seems to be an issue concerning what is perceived as a *Viby-i*, and what is not. This study takes the approach that *Viby-i* can be gradient, and that *Viby*-colouring ‘begins’ when the F2 of /i:/ is lower than that of /e:/. However, further research is needed into the perception of *Viby-i* by native Swedish listeners.

Overall, research on *Viby-i* has long been held back by a lack of experimental data. Most likely, our limited understanding of this vowel has caused the difficulties we now face when trying to classify it phonetically. The fact that *Viby-i* appears to be highly variable in acoustics,

articulation, and social evaluation, has only added to this problem. By investigating this variation in more detail, this study aims to bring the field one step closer to an accurate description of *Viby-i*, which has broader implications for our understanding of vowel production as a whole.

# Chapter 5

## Method

### 5.1 Chapter overview

This chapter outlines the methodology of the experimental study undertaken for this thesis, describing the method of data collection, data processing, and analysis. Section 5.2 provides a brief summary of the chapter, before more details are given in subsequent sections.

The first part concerns data collection. Section 5.3 describes the participant sample, including the recruitment process, the cities sampled for the study, and selection of the current sample. Section 5.4 describes the word list used to elicit the speech materials, and outlines the factors to be used in the linguistic analysis. Section 5.5 briefly outlines the collection of spontaneous speech data, which was not used in this study. Section 5.6 discusses the design of the demographic questionnaire, explores speaker demographics, and explains the factors used in the sociolinguistic analysis. Section 5.7 specifies the recording equipment and experimental set-up, reviews the use of ultrasound tongue imaging to study vowel articulation, and describes and evaluates the recording procedure.

The second part of the chapter focuses on data processing and analysis. Section 5.8 illustrates how the acoustic data was segmented and measured to investigate formant values, zero-crossing rate, and bandwidth. Section 5.9 explains the splining of the ultrasound data, qualitative analysis of the tongue and lip gestures, as well as the use of normalised tongue measures to analyse the statistical relationship between acoustics and articulation. Finally, Section 5.10 describes the use of mixed-effects modelling throughout the thesis, and the use of generalised additive mixed models to plot and quantitatively analyse the tongue data.

### 5.2 Method summary

The experiment collected word list recordings, spontaneous speech recordings (not used), and questionnaire data from a sample of Central Swedish speakers from Gothenburg, Stockholm, and Uppsala. The speech recordings consisted of simultaneous audio, ultrasound tongue imaging

(UTI), and lip video.

Static and dynamic measurements were taken of both the acoustic and articulatory data. From the acoustic data, I sampled formant frequencies, bandwidth (not used), and bandpass-filtered zero crossing rate (bpZCR), which was used to measure frication. The acoustic information was used to establish the position of Viby-i in the F1/F2 vowel space, to investigate linguistic environments which may condition Viby-i, and to explore the nature of Viby-i's characteristic 'buzziness'.

From the ultrasound recordings, I extracted contours of the midsagittal tongue surface, in order to compare tongue gestures for Viby-i both within and between speakers. Static tongue contours, sampled at a corresponding time to the acoustic measurements, were used to establish the position of Viby-i in the articulatory vowel space, and to investigate the relationship between tongue gesture and acoustic output. A dynamic analysis was also used to explore potential articulatory strategies that could result in frication or formant movement.

The lip video recordings were used to investigate whether lip position plays a role in Viby-i production, and to assess the possible impact of lip gesture on acoustics. Due to time constraints, the lip analysis does not have a quantitative component.

Linear mixed-effects modelling was used at several stages in the analysis to investigate linguistic and social effects on vowel production, and to investigate the relationship between tongue gesture and acoustics. The linguistic analysis focused on consonant environment and vowel duration, while the social analysis focused on age, gender, city, distance from city centre, social interaction, and prescriptivism. The acoustic-articulatory analysis measured the relationship between F1/F2 and multiple normalised points along the tongue surface, including the highest point of the tongue, the tongue back, and the tongue tip.

A dynamic statistical analysis using generalised additive mixed models (GAMMs) was also carried out on the articulatory data, in order to quantify differences between Viby-i and other vowels, and to investigate the dynamics of the tongue gesture used for this vowel.

## 5.3 Participants

### 5.3.1 Recruitment

Participant recruitment was carried out in parallel with data collection, taking place over a six-week fieldwork period in Sweden in 2016. Convenience sampling was used to recruit speakers both in person and online, advertising to nearby shops, cafés, and community spaces, as well as local interest groups, students, and professionals (e.g. speech therapists, language teachers, and researchers).

Since data collection took place at local universities, most of the participants were either university students or staff, or their friends and family members. The sample is therefore fairly

homogeneous in terms of the speakers' educational background, which may also reflect their socioeconomic class. This somewhat limits the conclusions that can be drawn about *Viby-i* in working-class speech. On the other hand, keeping education relatively stable allows for a more detailed analysis of other factors, such as gender or age.

All participants were pre-screened to ensure that they were over 18; that they were native Swedish speakers; that they had grown up and spent most of their life in either Gothenburg, Stockholm, or Uppsala; that they had no major issues with speech, reading, or writing; and that they did not have a large beard or dreadlocks (as these would interfere with the ultrasound equipment). Two speakers did not meet the language criteria: SM1, who spoke Greek as a first language, but had native proficiency in Swedish; and GM5, who had grown up in Småland (South Sweden), but had lived in Gothenburg for nearly 50 years and had adopted the local accent. Two speakers, GF5 and UF2, had grown up bilingual, speaking Finnish and Polish respectively. There were no multiethnolect speakers in the sample.

Originally, the sample was intended to only consist of speakers who had never lived outside their home city, and whose parents both came from there, but these criteria had to be widened in order to obtain enough participants in the limited recruitment time.

### 5.3.2 Sampled cities

The speaker sample used in this study consisted of 34 speakers in total: 12 from Gothenburg, 12 from Stockholm, and 10 from Uppsala. These three cities were chosen to represent regions where the use of *Viby-i* may vary: Gothenburg and Stockholm are both large cities where speakers are known to use *Viby-i* (Björsten & Engstrand, 1999: 1,957), but the cities are located in different dialect regions, approximately 400 km apart. *Viby-i* appears to have existed slightly longer in Gothenburg (Elert, 1995: 45), meaning that it could be produced with a different vowel quality than in Stockholm. Stockholm and Uppsala are located in the same dialect region, about 65 km apart, but Uppsala is a smaller city where *Viby-i* has not previously been documented. Uppsala is an important location for Standard Swedish, as it is a long-established seat of learning, but it is also likely to be influenced by Stockholm, where *Viby-i* is perceived as a high-status feature (Bruce, 2010: 135-136). If *Viby-i* has spread to Uppsala, it could be an indication that this vowel quality is also spreading to other parts of Sweden.

### 5.3.3 Speaker selection

The 34 speakers used for this analysis were selected from a total of 62 recorded participants, based on their city of origin, gender, and age. Analysis of all 62 speakers was beyond the scope of this thesis, but the unused data will be available for future research.

To create the speaker sample, I first divided the speakers by city and gender. I then constructed three age categories within each group: the 'young' category consisted of the two



youngest speakers; the ‘old’ category of the two oldest; and the ‘middle-aged’ category of the two closest to the group median age, with the caveat that they had to be within a similar age range as the middle-aged speakers from the other cities. These age groups were constructed to ensure a balanced and comprehensive range of speaker ages for each city and gender, so that an apparent-time analysis could be carried out. The age groups were only used in the sampling process however, as, for the statistical analysis, the raw age of the speaker was believed to be more informative.

An overview of the sample is provided in Table 5.1. Further details about participant demographics will be provided in Section 5.6. Note that, due to time constraints, no old Uppsala males could be recorded.

	Gothenburg		Stockholm		Uppsala	
	Female	Male	Female	Male	Female	Male
Young	GF1 (20)	GM1 (21)	SF1 (20)	SM1 (20)	UF1 (20)	UM1 (20)
	GF2 (23)	GM2 (23)	SF2 (22)	SM2 (22)	UF2 (22)	UM2 (22)
Middle-aged	GF3 (30)	GM3 (31)	SF3 (31)	SM3 (33)	UF3 (32)	UM3 (32)
	GF4 (41)	GM4 (34)	SF4 (33)	SM4 (37)	UF4 (37)	UM4 (33)
Old	GF5 (51)	GM5 (69)	SF5 (63)	SM5 (67)	UF5 (58)	
	GF6 (62)	GM6 (72)	SF6 (63)	SM6 (80)	UF6 (80)	

Table 5.1: Participants by city, gender, and age group (age in parentheses).

## 5.4 Word list design

Vowel productions were elicited using a word list. The purpose of the word list was to obtain tokens of Viby-i and the other long vowels of Swedish while controlling for their linguistic environment. Variations in consonant context and vowel duration were used in the analysis to investigate whether these factors would condition the use of Viby-i over standard [i:], or whether the realisation of Viby-i would be affected, e.g. in terms of its formant values or vowel frication.

### 5.4.1 Word list items

The word list consisted of 41 target words and 8 distractor words. For simplicity, the target words are represented in two separate tables: Table 5.2 shows the linguistic environments that could be applied across all nine vowels, and Table 5.3 shows those that could only be applied to /i:/, while still producing real lexical items. The distractors were loosely connected to recurring themes from the word list (e.g. housework, nature, food), but differed slightly in structure from the target words, in order to make the word patterns less predictable. The distractor words were:

FEJA ‘to work hard’, FETA ‘fat’, FÅ ‘to receive’, SAGA ‘fairytale’, SPÖKA ‘to haunt’, SUSÄ ‘to rustle’, TRO ‘to believe’, and ÄGG ‘egg’.

Vowel	Followed by /t, d/	Followed by /k, g/	Word-final
/i:/	BITA ‘to bite’	BIGA ‘carriage’ (arch.)	BI ‘bee’
/i:/	PITA ‘pitta bread’	PIGA ‘maid’	PI ‘pi’ (num.)
/y:/	BYTA ‘to swap’	BYKA ‘to wash’ (arch.)	BY ‘village’
/ʉ:/	BUDA ‘to bid’ (colloq.)	BUGA ‘to bow’	BU ‘boo’ (interj.)
/e:/	BETA ‘to graze’	PEKA ‘to point’	BE ‘to pray’
/ø:/	BÖTA ‘to fine’	BÖKA ‘to root around’	HÖ ‘hay’
/ɛ:/	VÄTE ‘hydrogen’	VÄGA ‘to weigh’	BÄ ‘baa’ (onom.)
/ɑ:/	BADA ‘to bathe’	BAKA ‘to bake’	HA ‘to have’
/o:/	BÅDA ‘both’	BÅGE ‘bow’	PÅ ‘on’
/u:/	BOTA ‘to cure’	BOKA ‘to book’	BO ‘to reside’

Table 5.2: Word list tokens applicable to all nine long vowels.

### 5.4.2 Linguistic factors for analysis

This section will describe how the word list was designed, and how the different linguistic environments were expected to affect vowel production. The headings represent factors that were later used in the analysis.

#### Factor: Vowel

The target vowel varied between the nine long Central Swedish vowels /i:/, y:/, ʉ:/, e:/, ø:/, ɛ:/, ɑ:/, o:/, u:/. These vowels were always long and stressed, as Vibry-colouring does not usually affect short unstressed vowels (Bruce, 2010: 135). Out of the 41 target words, there were 17 words containing /i:/, and 3 words containing each of the remaining vowels. This lack of balance was caused by difficulties in finding (near-) minimal pairs for the full vowel set, but it was also a time-saving measure. Since /i:/ was the focus of the study, there were a number of linguistic environments designed to explore variation within this vowel, and a smaller number designed to explore its relation to the vowel space as a whole.

Following consonant	Preceding consonant			
	/v/	/f/	/b/	/p/
(null)			BI 'bee'	PI 'pi' (num.)
/p, b/			BIBEL 'bible'	PIPA 'pipe'
/t, d/	VITA 'white' (pl.)		BITA 'to bite'	PITA 'pitta bread'
/k, g/	VIKA 'to fold'	FIKA 'to have coffee'	BIGA 'carriage' (arch.)	PIGA 'maid'
/l/	VILA 'to rest'	FILA 'to file down'		
/r/	VIRA 'to wind'	FIRA 'to celebrate'		
/s/	VISA 'to show'			
/n/	VINA 'to whine'			

Table 5.3: Word list tokens applicable to /i:/ only. Duplicates from previous table in grey.

The purpose of collecting the full set of long Swedish vowels was to contextualise the position of /i:/ in the (acoustic and articulatory) vowel space. My hypothesis was that at least some speakers would produce *Viby-i*, i.e. an /i:/ with a similar F1 to /e:/, but with a lower F2 than /e:/. From an articulatory perspective, I was expecting *Viby-i* to be produced with a centralised tongue gesture, where the tongue was either primarily retracted, or accompanied by some degree of lip-rounding, which would have the effect of lowering F2. I also hypothesised that /y:/ would undergo *Viby-colouring* in some speakers, as other studies have found this to be the case (e.g. Gross & Forsberg, 2019).

### Factor: Duration (syllable structure)

Target words had either a CV or CVCV(C) syllable structure. The first vowel was always the target, and was either word-final in a monosyllable (e.g. BI), or word-medial in a disyllable (e.g. BITA). The second vowel was always a short, unstressed /a/ or /ɛ/, and was not included in the analysis.

The purpose of varying the syllable structure was to manipulate the raw duration of the target vowel. Although all target vowels were *phonemically* long, my intention was to encourage longer durations in the CV context compared to the CVCV(C) context, by exploiting the effects of polysyllabic shortening and word-final lengthening (Engstrand, 2004: 207-208). My hypothesis was that, since a durationally longer vowel has more time to reach its target (Browman & Goldstein, 1992), these contexts might produce a stronger *Viby-i*, e.g. a vowel where F2 is lower, where there is more frication, or where the tongue gesture is more complex or peripheral.

It should be noted, however, that this analysis cannot disentangle the effects of vowel duration from the effects of syllable structure.

### **Factors: Place and manner of following consonant**

The word list systematically varied the consonant following the target vowel by place of articulation (for the full vowel set), and manner of articulation (for /i:/ only). Place of articulation varied between front (alveolar/dental) /t, d/ and back (velar) /k, g/. There was also a ‘null’ level for word-final contexts. Manner of articulation varied between ‘simple’ (plosives) /p, b, t, d, k, g/ and ‘complex’ (liquids, sibilants, nasals) /l, r, s, n/.

The following consonant was manipulated to investigate how coarticulation would affect the acoustic properties and tongue gesture of the target vowel. My hypothesis was that consonants with a similar articulatory strategy to Viby-i would encourage a more prototypical production of this vowel. For example, if Viby-i was produced with a backed tongue gesture, it would be stronger before back consonants.

For manner of articulation, I was particularly interested in the tongue gesture and secondary articulations. While /t, d, k, g/ use relatively simple gestures, /p, b/ do not use the tongue at all, and /l, r, s, n/ have more complex articulatory settings, which could encourage secondary articulations, e.g. velum lowering, lateralisation, tongue bunching, or frication. My hypothesis was that, since Viby-i has been described as having a complex articulation (Borgström, 1913), which may be similar to /l, r, s, n/, proximity to a sound with a similar secondary articulation would encourage a stronger Viby-i. At the same time, some secondary gestures might ‘block’ Viby-i if the articulatory gestures are conflicting.

Since the following consonant was used to investigate the effects of lingual coarticulation, all initial consonants were kept non-lingual, in order to minimise coarticulatory effects on the tongue from this direction. This design was necessary because, due to lexical and phonotactic constraints, it was not possible to generate symmetrical consonant environments for the word list. For example, /k, g/ lenite before front vowels, meaning that a word like KIKA ‘to peek’ is pronounced /çi:ka/, not /ki:ka/. Likewise, it would not be practical to use non-symmetrical environments with lingual consonants on both sides, e.g. TIGA ‘to be quiet’, as it would be impossible to determine the effect of each consonant on the tongue gesture. Thus, lingual effects (relevant to e.g. formant structure) were manipulated through the following consonant only, while non-lingual effects (relevant to e.g. frication) were manipulated through the preceding consonant.

The study originally intended to investigate formant bandwidths, but could not obtain accurate data to do so, as described in Section 5.8.5. However, for future studies, it may be relevant to investigate whether the ‘damped’ quality of Viby-i results from wider bandwidths, which could be achieved through nasalisation or lateralisation (Stevens, 2000: 312, 554).

### Factors: Voicing/aspiration and manner of preceding consonant

The preceding consonant within the /i:/ set was systematically varied by voicing (which conditions aspiration in plosives), and manner of articulation. Voicing varied between voiceless /p, f/ and voiced /b, v/. Manner varied between plosive /p, b/ and fricative /f, v/. The only aspirated consonant was /p/.

Voicing, aspiration, and frication were expected to affect the ‘buzziness’ of Viby-i. For example, a voiceless or aspirated context could contribute to partial devoicing of the vowel, as well as increased airflow, which could result in turbulence. A fricated context, on the other hand, might carry frication over into the vowel, particularly if the source of the frication is similar to the articulatory setting for Viby-i.

Note that the consonants that were intended to encourage frication *preceded* the vowel segment, to ensure that any fricative noise would affect the vocalic portion of Viby-i, rather than the offglide. This strategy made it possible to explore the relationship between intra-vowel frication and end-frication, since these two types are not clearly distinguished in the literature on Viby-i.

For non-/i:/ target vowels, the preceding consonant was kept as stable as possible, using /b/, or else /p, v, h/ if necessary to generate real lexical items and avoid taboo words.

### Summary of hypotheses

To summarise, my hypotheses for the effects of vowel duration and consonant context on Viby-i are listed below. A ‘stronger’ Viby-i is operationalised as a vowel with a lower F2, and optionally, a higher F1.

- Viby-i will be stronger in words like BI compared to words like BITA, since their duration is longer, giving the tongue more time to reach the vowel target.
- Viby-i will be stronger when flanked by non-lingual consonants on both sides, since there is no lingual coarticulation, giving the tongue more freedom of movement.
- Viby-i will be stronger before consonants that have a similar place of articulation to the vowel. For example, if Viby-i uses a back articulation, it will be stronger before a back consonant, and vice versa.
- Viby-i will be stronger before a consonant with a similar secondary articulation to the vowel, or with a similar tongue shape. Thus, if Viby-i uses a complex articulation, it will be stronger before a complex consonant.
- Viby-i will be more fricated after voiceless, aspirated, or fricated consonants, since the articulatory settings used to generate the frication may be similar.

### 5.4.3 Carrier phrase

The word list items were presented in the carrier phrase *Jag sa ...* ('I said ...') in order to encourage more natural productions, both with regard to intonation and to connected speech processes. The target word was always utterance-final, since this would help to increase the duration of the 'hyper-long' vowels (e.g. BI) (Engstrand, 2004: 207-208).

Each participant produced the word list three times in a pre-randomised order, which was the same for all speakers. True randomisation was not possible, firstly because it was not supported by the articulatory recording software, and secondly because it was necessary to ensure that, if the experiment ended early, each word had been repeated approximately the same number of times. A schematic representation of the randomisation process is presented in Table 5.4.

	Participant 1	Participant 2
Rep. 1	Word order: A, B, C, D	Word order: A, B, C, D
Rep. 2	Word order: B, D, A, C	Word order: B, D, A, C
Rep. 3	Word order: C, A, D, B	Word order: C, A, D, B

Table 5.4: Schematic representation of how the word list was randomised.

### 5.4.4 Word list evaluation

There were some limitations to the word list design, caused by e.g. lexical restrictions, word frequency, and tonal accent. The researcher also noticed some patterns in how participants interacted with the word list during recordings, which will be discussed here as a way of preliminarily addressing these issues.

#### Lexical restrictions

The word list only contained real lexical items with unambiguous spellings, to ensure that speakers produced the intended target. Issues with nonsense words had arisen in previous work (Westerbergh, 2016), as they tended to result in unnatural prosodic patterns. Due to these lexical restrictions, the consonant environments could not be fully controlled. For example, the voicing of the following consonant was variable, and this could affect vowel duration (House & Fairbanks, 1953). However, this voicing effect has been found to be relatively small in Swedish (Elert, 1964), and thus, following consonant voicing was not expected to exert a great influence on vowel duration. Even if this was the case, the effect would be accounted for by including 'duration' as a factor in the statistical analysis.

### Word frequency and ambiguity

Although BIGA, BYKA, PITA, and BUDA are real words, they are relatively infrequent in Swedish, and some participants were unfamiliar with them. This did not appear to affect their productions, as the pronunciation of these words is highly predictable from the spelling. The only word which posed ambiguity problems was HA, which could either be produced [ha:] ('to have'; this was the target vowel quality) or [hɑ] ('hah!'). Since most of the other target words were verbs, however, speakers generally pronounced this token correctly. In cases when they did not, the token was re-recorded.

To control for the possible effect of word frequency on vowel production, word frequency data was extracted from the Korp corpus (Borin, Forsberg, & Roxendal, 2012), a collection of 232 modern Swedish corpora from a wide variety of periods and text genres. The search used all 220 available corpora, containing approximately 11.6 billion words.

### Tonal word accent

All disyllabic words in the word list canonically have grave tonal word accents, with the exception of BIBEL, which has an acute accent. However, the monosyllabic words could be produced on either a rising or falling pitch, based on the speaker's preference. According to Bye (2004: 4), monosyllables in Scandinavian languages do not typically use tonal word accents, but there are exceptions to this rule. There may thus be some degree of variation in the pitch contours of these words, but it is not known whether variations in pitch are likely to affect Swedish vowel quality.

### /ɛ:/-lowering

During the recordings, I noticed that many speakers produced the /ɛ:/ in BÄ with a much 'lower' vowel quality than VÄTE or VÄGA. The likely reason is that BÄ ('baa') is onomatopoeic, and that speakers were modelling their pronunciation on the original sound, rather than producing the vowel as they would normally. Thus, although the /ɛ:/ vowel is already expected to be low based on e.g. Gross (2018), some tokens in this sample may be even lower.

### Unnatural prosody

Although the carrier phrase was intended to encourage more natural prosody, some speakers still produced somewhat unnatural utterances, e.g. using a breathier voice quality than normal; using list intonation; pronouncing sounds that are normally omitted in casual speech (e.g. [jɑ:g] instead of [jɑ]); pausing before the target word; or syllabifying the whole utterance (e.g. [jɑ:.sa:.bi:.ta]). This kind of variation was annotated in the phonetic transcriptions, but it is not clear how they might have affected vowel quality. Nevertheless, as speakers were fairly consistent

within themselves, the use of ‘speaker’ as a intercept factor in the statistical model should be able to account for some of these individual prosodic differences.

## 5.5 Spontaneous speech recordings

Short segments of spontaneous speech were elicited using a set of interview questions. This part of the experiment was optional, and was not completed by all participants. In the end, the spontaneous data was excluded from the current analysis, partly due to missing data, and partly due to time constraints, as segmenting and preparing this data would require a significant amount of time.

### 5.5.1 Interview questions

Each participant selected four out of eight possible questions to answer, and was prompted to speak freely for up to 60 seconds in response to each question. The maximum time limit was set due to the heavy processing load and space requirement of the articulatory recordings. The questions available to answer were:

- What did you want to become when you were little? Is that how things turned out?  
*Vad ville du bli när du var liten? Blev det så?*
- Does anything specific in [city] make you extra happy?  
*Blir du extra glad av något specifikt i [stad]?*
- Would you like to live in [city] for the rest of your life?  
*Vill du bo i [stad] hela livet?*
- What do you do in your spare time?  
*Vad gör du på fritiden?*
- Where do you see yourself in ten years?  
*Var ser du dig själv om tio år?*
- What is the most beautiful thing about [city]?  
*Vad är det finaste med [stad]?*
- Would you advise other people to move here?  
*Skulle du råda andra människor att flytta hit?*
- Are you interested in linguistics? If so, how did this come about?  
*Är du intresserad av lingvistik? I så fall, hur blev du det?*



The questions were designed to be easy for the participant to answer, while hopefully also evoking positive feelings which would allow them to speak in a relaxed manner for an extended period of time. This aspect of the design was inspired by the common sociolinguistic approach, pioneered by Labov (1972), of asking emotionally engaging questions to elicit a response that is less linguistically guarded. As a consequence, I hoped that the speakers would use vowel qualities that were more similar to their natural speech.

### 5.5.2 Target words

The interview questions were intended to elicit specific words containing /i:/, listed in Table 5.5. However, these words were only guaranteed to appear if the participant repeated the wording of the question. Whether speakers produced the target words or not, some /i:/ tokens were likely to appear in the data, but the number of tokens produced by each speaker, and the linguistic contexts in which they appeared, might be variable. The inclusion of the target words was thus a way of attempting to make the speakers' utterances slightly more uniform.

Target	Transcription	Translation
BLI	/bli:/	'become'
LITEN	/li:tən/	'little'
SPECIFIKT	/spɛsɪfi:kt/	'specific'
I	/i:/	'in'
LIVET	/li:vət/	'life'
FRITIDEN	/'fri:tɪ:dən/	'spare time'
TIO	/ti:u/	'ten'
FINASTE	/fi:nastɛ/	'most beautiful'
HIT	/hi:t/	'here'
LINGVISTIK	/lɪŋvisti:k/	'linguistics'

Table 5.5: Target words for the spontaneous speech task.

## 5.6 Questionnaire

A written questionnaire was used to collect demographic, linguistic, and social information about the participants, in order to compare their vowel productions to their sociolinguistic background. The full questionnaire is available in Appendix A (in English) and Appendix B (in Swedish). The three main areas investigated in the questionnaire were:

- Demographic information, e.g. age, gender, education;
- Linguistic information, e.g. language background, dialect contact, speech pathology;
- Socio-cultural information, e.g. attitudes towards dialects and places, lifestyle choices.

This section will outline how the questionnaire was used to generate the factors used in the sociolinguistic analysis, explaining why they were investigated, and how they were expected to affect Viby-i production. The section will first individually describe the factors that were kept for statistical analysis, followed by a summary of the factors that were not used.

### 5.6.1 Factor: City

City was an unordered categorical variable, based on the speakers' self-reported city of origin: Gothenburg, Stockholm, or Uppsala. Although all speakers were screened before the study, some later reported growing up in the outskirts or suburbs of the city, and one speaker (GM5) had grown up in a different part of the country. To account for possible variation within each city, an additional variable was created, measuring the distance to the city centre, as described in Section 5.6.2.

The 'city' variable was used to investigate dialectal variation in the production of Viby-i, which could arise from geographical distance, social prestige, or dialect history. My hypothesis was that Viby-i would be prevalent in Stockholm and Gothenburg, while being absent in Uppsala. I also expected the Gothenburg and Stockholm productions of Viby-i to differ with regard to formant values and/or fricative noise, possibly with the Gothenburg speakers having a lower F2 and more frication, as Viby-i appears to be more established in Gothenburg. Interactions were also expected between 'city' and other factors, as outlined in the following sections.

### 5.6.2 Factor: Distance from city centre

'Central (km)' was a continuous numerical variable, measuring the distance in kilometres from the speakers' reported childhood home (the place where they had lived the longest before the age of 18), to the centre of their coded city of origin. This factor was calculated using the 'Measure distance' function on Google Maps (2019), which calculates the distance as a straight line.

This factor was designed to account for previous reports that Viby-i is more common in urban centres, and less common in the outskirts (e.g. Frid et al., 2015). My hypothesis was that Viby-i would continue to follow this pattern by being more frequent, or having more exaggerated characteristics, in speakers who grew up close to the city centre. However, in the case of Uppsala, this effect might be reversed, since locations further from Uppsala could be closer to Stockholm.

### 5.6.3 Factor: Age

Age was a continuous numerical variable, denoting the speaker's age in years at the time of recording, based on their reported year of birth. Speakers were not divided into age groups, as the boundaries between these groups would have been arbitrary.

Age was included as part of an apparent-time analysis, to establish how long Viby-i was

likely to have existed in each city. My hypothesis was that Viby-i would be more frequent, or have exaggerated characteristics, in younger speakers. I also expected Viby-i to be more common in older Gothenburg speakers, less common in older Stockholm speakers, and absent in older Uppsala speakers. Age was expected to interact with gender, as outlined below.

#### 5.6.4 Factor: Gender

Gender was an unordered categorical variable, based on participants' self-reported gender. A non-binary option was offered on the questionnaire, but no participants used it.

Previous literature has shown that use of Viby-i may be affected by a combination of gender and social class (Kotsinas, 2007). Kotsinas' study only investigates Stockholm, but there are indications that the same pattern could also be found in Gothenburg (Björseth, 1958). My hypothesis was that Viby-i would be more frequent, or have more exaggerated characteristics, in the speech of women (particularly young women), since Viby-i is a prestige variant, which is also showing signs of change. However, in cities where the change is already established, Viby-i may be used in similar ways by both men and women.

#### 5.6.5 Factor: Social interaction

'Social score' was a continuous numerical variable, representing the sum of three questions. Two of these questions were on a four-point Likert scale, and one was multiple choice, where one point was allocated to each hobby that required the participant to leave the home:

- How centrally do you prefer to live?  
*Countryside = 1, Suburb = 2, City = 3, Centre = 4*
- Do you visit many different areas in [city], or do you tend to stay in the same area?  
*Barely mobile = 1, A little mobile = 2, Quite mobile = 3, Very mobile = 4*
- What interests do you have in your spare time?  
*Arts & crafts, Cooking, DIY, Family activities, Films/TV, Literature, Music, Technology, Video games = 0*  
*Cultural events, Gigs, Going out, Going out for food, Hiking, Shopping, Sports, Travel, Volunteering = 1*

This factor tried to capture the idea that speakers with a high level of social interaction may be more exposed to language variation, and may have a greater incentive to adopt innovative or prestige forms. This approach was inspired by Grönberg (2004), who found that adolescents from the same demographic background sometimes used different language forms, possibly because of different attitudes, lifestyles, or social ambitions (Grönberg, 2004: 319). My hypothesis was that Viby-i would be more frequent, or have more exaggerated characteristics, in speakers who interacted more socially.

### 5.6.6 Factor: Prescriptivism

Prescriptivism was a continuous numerical variable, representing the sum of two questions on a four-point Likert scale:

- Do you think it's important to speak a certain way?  
*No = 1, A little/Don't care = 2, Quite = 3, Very = 4*
- Do you get annoyed when people use "incorrect" language?  
*No = 1, A little/Don't care = 2, Quite = 3, Very = 4*

This factor was used to investigate the effect of prescriptivist attitudes on Viby-i use, since Viby-i diverts from the standard, and, despite being a prestige marker, tends not to be well-liked (Elert, 1995: 45). My hypothesis was that Viby-i would be less frequent, or have less exaggerated characteristics, in speakers who were more linguistically prescriptive. However, some participants indicated that they interpreted these questions differently, e.g. as measuring meta-linguistic awareness. To some extent, this awareness may still influence the use of Viby-i, as speakers may consciously shift towards or away from this vowel.

### 5.6.7 Excluded factors

#### Education (Socioeconomic class)

The questionnaire asked speakers about their highest level of education, to be used as a proxy for socioeconomic class (after Norrby & Håkansson, 2015: 93-94). Class was of interest because previous research had shown that middle-class speakers were more likely to use Viby-i than working-class speakers (Kotsinas, 2007), and Viby-i is usually regarded as a prestige marker associated with upper-class speech (Bruce, 2010: 135). However, 'education' had to be omitted from the statistical analysis because the sample was too uniform to render this factor meaningful. All participants, with the exception of SF6, SM1, and UF1, either had or were currently undertaking a university education. The remaining three speakers had finished upper secondary school. Most of the sample thus had a relatively high level of education, indicating that their data mostly reflects middle-class speech.

Additional questions intended to inform the speakers' social class, e.g. parents' profession and level of education, whether the speaker had entered the job market, and their career or education type, were asked on the questionnaire, but some speakers omitted these questions, and the answers were too difficult to quantify for the statistical analysis.

#### Local rootedness

This factor was intended to investigate the degree to which Viby-i was affected by the speakers' sense of local belonging. This approach was inspired by Grönberg (2004), who found that

speakers' choice of linguistic variables could be influenced by a combination of geography and attitude or lifestyle factors (Grönberg, 2004: 318-319). My hypothesis was that speakers from Gothenburg and Stockholm would be more likely to use a (strong) *Viby-i* if they had positive attitudes towards their city of origin, since this vowel is particularly associated with these cities. The opposite pattern was expected in Uppsala, since *Viby-i* may be regarded by these speakers as a non-standard feature. However, this factor was found to be correlated with 'social score', and thus had to be dropped from the statistical analysis.

### **Part of city**

This factor was intended to provide more fine-grained detail to 'Central (km)', since it was possible that not only the distance from the city centre, but also the direction, could be important for how speakers produced *Viby-i*. In addition, this variable could capture demographic differences (e.g. income levels in different parts of the city), as well as the proximity to other cities (e.g. whether speakers who lived far from Uppsala lived closer to Stockholm). However, based on the small sample size, and the complexity of this factor, it was not possible to include it in the statistical analysis.

### **Speech pathology**

Speakers were screened for speech pathology, as previous or ongoing language disorders could potentially influence the speakers' vowel productions. No speakers reported having more than mild issues (e.g. lisping) in childhood. It was therefore not necessary to include this factor in the analysis.

### **Dialect contact, language contact**

These factors measured the speakers' estimated exposure to different varieties of Swedish, as well as other languages. Since there are several dialect regions where *Viby-i* appears to be less common (e.g. North and South Swedish), extensive contact with these varieties, particularly through close friends or family members, was expected to be linked to a weaker (or absent) *Viby-i*. Similarly, extensive contact with languages that do not use [i] was expected to have the same effect. However, the answers provided for the language contact questions were difficult to quantify, and this factor was therefore excluded. 'Dialect contact' was excluded later, because it was found to correlate with gender.

## 5.7 Equipment

### 5.7.1 Audio recordings

Audio recordings were made with an Audio-Technica AT831b cardioid lapel microphone, sampling at 44,100 Hertz (16 bits, 2 channels). The microphone was attached to the ultrasound headset, about 10 cm from the participant's mouth. This set-up resulted in high-quality audio recordings that allowed for analysis of the first four formants, despite occasional interference from e.g. participants' mobile phone signals. Tokens were not analysed if noise interfered with the functioning of the formant tracker in the segmentation software.

### 5.7.2 Ultrasound recordings

#### Ultrasound and probe

Ultrasound video was recorded with a Telemed Echo-Blaster 128 ultrasound machine, using a 2-4 MHz convex probe (20 mm radius, 104 ° field of view). This probe is suitable for imaging a range of adults, as its low-frequency signal is able to travel further into the tissue than that of a higher frequency probe, with the trade-off of a slightly grainier image quality. The image depth of this probe was 75 mm. Recordings had a frame rate of 67.19 frames per second, which is fast enough to capture most speech gestures without excessive motion blur. As the ultrasound data was collected in raw format rather than as a video image, certain settings (e.g. brightness, contrast) could be adjusted post-hoc for better visibility.

#### Headset

A probe stabilisation headset (Articulate Instruments Ltd., 2008; Scobbie, Wrench, & van der Linden, 2008), was used to ensure that the ultrasound probe did not move in relation to the participant's head. This headset was an older model, made from aluminium and weighing around 0.5 kg. Although the headset itself is not uncomfortable, its weight can put strain on the participant's neck after prolonged wear. For this reason, recording sessions were limited to one hour, and participants could request breaks at any time. A picture of the headset is provided in Fig. 5.1.

#### Bite plate

The probe-to-cranium angle, which affects the rotation of the midsagittal ultrasound image, was standardised using a bite plate. The plates were custom-made from medical-grade plastic (sterilised before each use), and consisted of a flat surface measuring 4 cm wide and 10 cm long, with a small vertical protrusion roughly halfway along its length (Fig. 5.2). Participants were asked to insert the shorter half of the bite plate into their mouth, resting their front teeth against

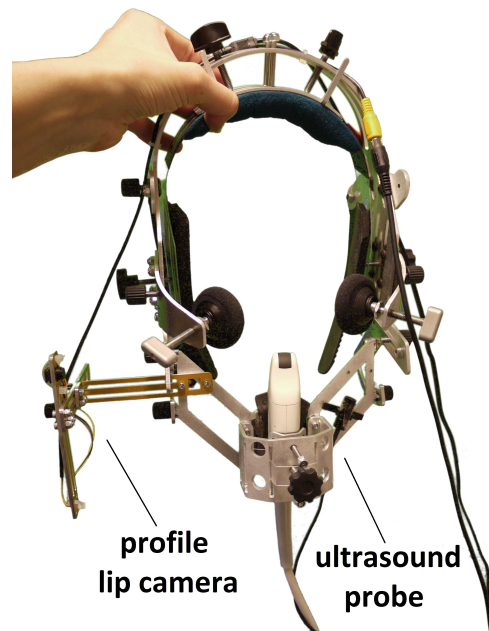


Figure 5.1: Ultrasound headset with profile-view lip camera.

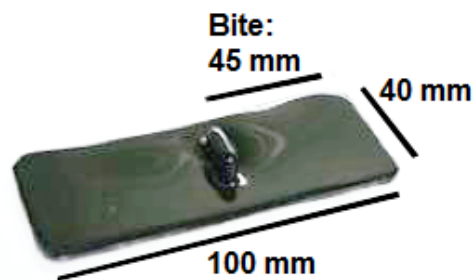


Figure 5.2: Bite plate with measurements.

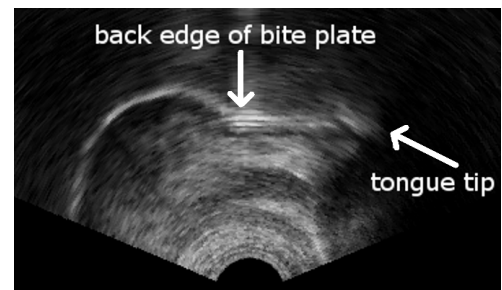


Figure 5.3: UTI image of the tongue pushing against the bite plate.

the protrusion, and bite down on the plate while pressing their tongue against its underside (Fig. 5.3). The image of the tongue against this flat surface provides a horizontal plane (the speaker's bite plane), which can be used to extrapolate the angle of the ultrasound probe in relation to the speaker's head. This allows the probe angle to be standardised between different speakers and sessions.

The length of the bite plate can also be used to approximate the position of the front teeth, which is useful because the frontest part of the oral cavity is sometimes obscured by the jaw bone shadow, making it difficult to gauge how fronted the tongue is.

### Palate trace

Although UTI can only image soft tissue, such as the tongue, it is possible to create an indirect image of other structures in the mouth, notably the hard palate and alveolar ridge, by pressing the tongue against them, or by swallowing. While doing this, the tongue surface remains visible, and

any liquids being swallowed appear as a bright mass on the ultrasound image. In UTI research, it is therefore common practice to make a palate trace recording, which can then be used as a reference point for the tongue. It can also be used to align the position of the probe between multiple recording sessions.

In this experiment, participants were instructed to swallow a few small sips of water, preferably in succession, since this provides the best chances of discerning a full palate trace, and also discourages large amounts of jaw opening. An alternative strategy used in previous studies (Westerberg, 2013, 2016) was for participants to drink through a straw. Regardless of method, some degree of jaw opening or probe displacement is relatively common, seen by the fact that the tongue contours sometimes overshoot the palate trace. To minimise this error, it is advisable to draw the palate trace from a timepoint near the end of the swallow, when the muscles underneath the chin are contracted and the jaw is less open.

### Lip video recordings

Images of the lips were recorded using a built-in micro-camera on the ultrasound headset. The camera collected a profile-view video of the participant's mouth from a distance of about 10 cm. Recordings consisted of analogue NTSC video with a frame rate of 29.97 frames per second (interlaced, with the possibility of de-interlacing for a doubled frame rate). The camera had an active pixel sensor and f/2.0 lens (55 ° angle of view), providing a good level of detail and focus at close range (Fig. 5.4). The videos were recorded in greyscale to reduce the file size.

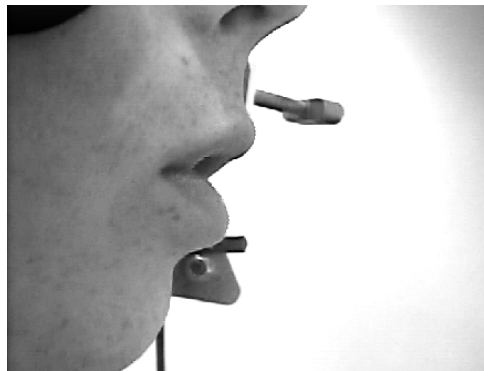


Figure 5.4: Image from side-facing lip camera mounted on the ultrasound headset.

### 5.7.3 Recording set-up

Audio, ultrasound, and lip video were recorded simultaneously using the software Articulate Assistant Advanced (AAA) (Articulate Instruments Ltd., 2019). The data was recorded and synchronised using the set-up illustrated in Fig. 5.5.

AAA automatically synchronised audio and ultrasound while recording, using information from audio channel 1 (microphone audio) and channel 2 (ultrasound synchronisation pulse).



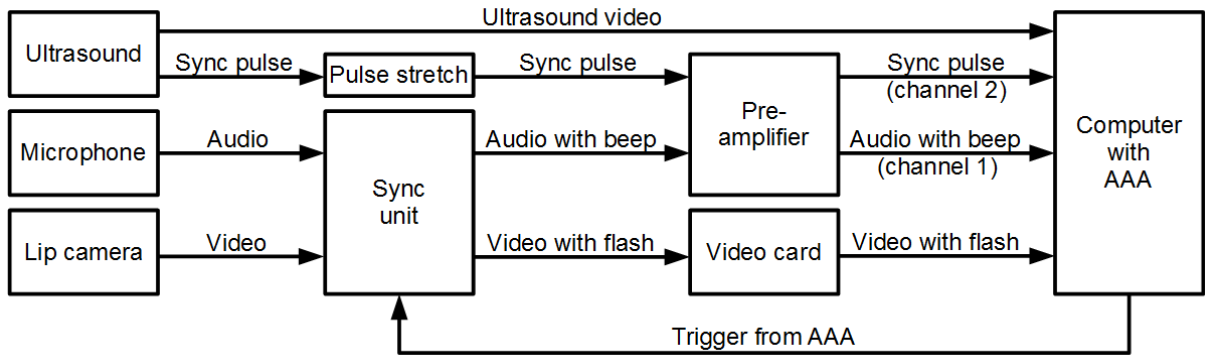


Figure 5.5: Set-up for recording and synchronising audio, ultrasound, and lip video.

Prior to this, the synchronisation pulse from the ultrasound was adjusted to AAA’s preferred frequency by a Pulse Stretch unit (Articulate Instruments Ltd., 2018). This unit is designed to enhance signals (synch pulses) produced by the ultrasound machine on completion of each scan to a level detectable by the software. Since the audio and ultrasound synch pulses were received by the same sound card, AAA could assign each completed ultrasound scan frame to the correct part of the audio recording.

Audio and video were synchronised using a SyncBrightUp unit (Articulate Instruments Ltd., 2010). This unit receives an audio trigger from AAA, which superimposes a timestamp on both audio and video (consisting of a beep on the audio, and a flash on the video), so that these can be aligned post-hoc.

The recordings were made on an HP Pavilion 15 Notebook (Intel Core i5-4200U processor, 8 GB RAM), running Windows 8 (64 bit). While this computer was fully capable of making the recordings, there was a lag of about 10-20 seconds from pressing the ‘stop recording’ button until the recording was saved. This meant that there was a fairly long pause between each prompt. The size of the files was also somewhat restrictive, amounting to around 25 GB per speaker, which required them to be stored on an external hard drive. The large file size was primarily caused by the inclusion of lip video.

#### 5.7.4 Recording procedure

Recordings took place in soundproofed studios at the University of Gothenburg, Stockholm University, and Uppsala University respectively. In a few cases when the studios were unavailable, recordings were carried out in a quiet room. Most participants were recorded individually, but a small number were accompanied by a friend. To avoid audio interference, the friend was seated in the next room, meaning that their effect on the participant’s speech is likely to be negligible. The study originally intended to recruit all participants in pairs in order to create a more relaxed recording setting, but due to time constraints, this criterion had to be dropped.

All participants were given written and verbal information about the study, and gave their

written informed consent to participate, with the option of withdrawing at any time. Copies of the information and consent forms are available in Appendix C (in English) and Appendix D (in Swedish). Ethical approval for the project was granted by the University of Glasgow.

Before recordings began, participants were fitted with the ultrasound headset, probe, and microphone. A small amount of water-based, hypo-allergenic ultrasound gel was applied to the probe surface in order to improve the ultrasound signal. The probe angle was adjusted to show as much of the tongue as possible, framing the tongue between the hyoid shadow (at the back) and the jaw shadow (at the front). The position of these shadows could potentially provide useful information about larynx raising and jaw opening, and can also be used as a method to measure and normalise some aspects of articulation (e.g. Zharkova, 2018), but they were not used in this study due to the large speaker sample.

Two short recordings were made of the participant's bite plane (see Section 5.7.2), and a palate trace was obtained by asking the participant to swallow a small amount of water (see Section 5.7.2). A test recording was also made to check the audio levels, and to ensure that the ultrasound synchronisation was working correctly.

The main part of the experiment consisted of the participant reading out the word list from the computer screen. The words were individually presented in AAA, and the participant was prompted to speak by a beep, as well as the prompt background turning green. Each recording was started and stopped manually by the researcher, who was seated next to the participant. Participants were instructed to produce the prompt in their normal voice, as if they were clarifying to someone who had misheard them. They were asked to emphasise the target word, which was presented in capital letters, e.g. "Jag sa BITA".

The participants were able to see their own lip movements, ultrasound video, and audio waveform displayed underneath the prompt, changing in real time as they spoke. This part of the screen could not be covered, as it was necessary for the researcher to ensure that the software was working correctly. Most participants did not seem distracted by their own speech movements, however, as the prompts directed their attention to the opposite side of the screen, and most people who are not familiar with ultrasound find it difficult to interpret these images, particularly if they are moving.

Due to the lag in saving each recording to the computer, the participant had a chance to chat to the researcher between prompts. This was encouraged, as it seemed to put the participant at ease. However, it became clear that many speakers style-shifted between conversational and word list speech, which is likely to have affected vowel production. Most conversations did not concern the experiment, but some participants asked about the more unusual lexical items on the word list. Given the nature of the study, some participants also initiated conversations about language use, in which a few mentioned Viby-i. Despite this, the speakers did not seem aware that the study focused on vowels, but only that it concerned Swedish dialects.

It is hard to estimate to what extent my own speech might have affected the participants;

my accent is a fairly levelled Western Central Swedish, with a weak Viby-i. Although I did not notice any accommodation effects in the participants' speech, it is possible that some fine-grained change may have occurred. Even so, all participants received approximately the same degree of exposure to my accent, even if the degree to which people accommodate the speech of others may vary from person to person.

After the word list recordings, some participants also produced four short segments of spontaneous speech, in response to the interview questions listed in Section 5.5. The questionnaire was completed either before or after the recordings, with the researcher present to answer any potential questions. At the end of the session, the participant was debriefed, and could choose to enter a lottery to win either a bookshop gift card or cinema tickets as thanks for their participation.

In total, each session took roughly 1.5 hours, including set-up, data collection, and debriefing. This was slightly longer than specified on the information sheet, mainly due to the computer working slowly, but participants were informed of this verbally at the start of the session. Participants were also allowed to leave early, or take breaks, if they wished.

### **5.7.5 Preliminary evaluation of set-up and procedure**

#### **Portability**

The ultrasound set-up was highly portable, and with the exception of the headset, all the equipment could be fitted inside a large backpack. Although fairly heavy, all the equipment could easily be carried by a single person. With practice, it was possible to set up and test the equipment in about 20 minutes, and dismantle it in about 5 minutes.

#### **Headset**

For a small number of speakers, the weight of the headset became uncomfortable after a short amount of time. This problem was addressed by taking the headset off, giving the participant a break, and then re-fitting the headset (and re-recording the bite plate), with the weight more carefully distributed towards the back of the head. Before doing this, the researcher ensured that the participant was happy to carry on with the experiment. One participant did not fit in the headset, and was recorded without it; however, since he visibly moved the probe during the recordings, his data had to be discarded.

A greater issue with the aluminium headset was the emotional response it evoked in participants. Although none objected to wearing it, many commented that it looked like torture instrument from a horror film, referencing popular films such as *Saw*, *Hellraiser*, and *A Clockwork Orange*. As stated, no participants were put off by this, but it might be an important issue to consider for future headset designs.

Some speakers were also worried that they might not be able to move their mouth or head normally while wearing the headset, but these worries were dispelled once the headset was fitted.

### **Jaw opening**

Although participants were able to speak normally with the probe fastened under the chin, it should be noted that the jaw cannot open fully without causing pitch movement of the probe, since the headset is designed to prevent the probe from moving. However, this issue is not likely to affect normal speech data, since jaw movements are normally quite small. Furthermore, the pitch movement of the probe, caused by the probe being pushed into the tissue under the chin, is more likely to affect low and back vowels, which have more jaw opening (Scobbie et al., 2008), and are thus unlikely to present problems for /i:/. Tracking jaw movement may still be relevant for future research, however. In this study, there were a few cases of the headset moving forward, or the probe moving to the side, as a result of jaw opening when the participant started speaking. In these cases, the probe position was adjusted and the affected tokens were re-recorded.

### **Probe contact**

This study used a convex ultrasound probe, with a slightly larger, flatter surface than the micro-convex probes used in previous experiments (Westerberg, 2013, 2016). This meant that the surface of the probe did not necessarily make full contact with the speaker's chin, and in some cases, adjusting the probe angle to improve contact would have been uncomfortable for the participant, as the probe would have pushed against the throat. For some speakers, the front of the mouth was therefore somewhat obscured in the ultrasound image. This issue is to be expected where large numbers of speakers, all with unique physiological characteristics, are being recorded. Other studies have avoided this issue by pre-screening and excluding participants for whom the complete tongue surface was not imaged (Gick, Campbell, Oh, & Tamburri-Watt, 2006: 54), but the current study could not pre-screen participants in the interest of time. In the future, poor imaging could perhaps also be addressed by having a range of different probes available.

### **Ultrasound safety**

Ultrasound equipment is generally regarded as safe for a variety of medical and research uses, including UTI (Preston et al., 2017). However, according to Joy, Cooke, and Love (2006: 222), prolonged exposure to high-intensity ultrasound signals could result in an increase of tissue temperature, which in the case of ultrasound tongue imaging could theoretically cause pain or skin irritation. To avoid this problem, I followed the advice of Preston et al. (2017: 2), setting the ultrasound to the lowest intensity (power) and increasing the amplification (gain) of the signal instead.

None of the participants reported any issues related to sensations of heat, pain, or skin irritation, despite many wearing the ultrasound probe in a fixed position for nearly an hour. As the ultrasound intensity was low, and the tissue under the chin appears to be able to carry away small amounts of excess heat, UTI did not seem to pose any health risks to participants.

### **Midsagittal tongue view**

Due to time constraints, this study only collected midsagittal images of the tongue. As a consequence, coronal movement, such as tongue grooving or lateralisation, was not visible from the ultrasound data. However, the inclusion of coronal data is not always informative, as coronal images are more difficult to interpret than midsagittal images. The main problem is that, since coronal images do not provide a full view of the tongue in the same way that midsagittal images do, it is difficult to know how far front or back the probe should be placed to record the relevant ‘slice’ of the tongue. It is also difficult to ensure that the probe is placed in the same position between different speakers or sessions.

Some of these issues also persist in midsagittal ultrasound. For example, since the positioning of the probe under the participant’s chin may vary, it is possible that some speakers’ images are taken closer to the midline of the tongue (i.e. where grooving might occur), whereas others’ are taken closer to the left or right side (i.e. where the tongue surface may appear to be higher if the sides of the tongue are raised) (Boyce, Tiede, Espy-Wilson, & Groves-Wright, 2015). This should be borne in mind when comparing the tongue splines presented in this study, as it is possible that the impression of e.g. tongue bunching in some speakers could be an artefact of probe positioning.

### **Ultrasound visibility**

With UTI, some speakers tend to produce higher quality images than others. For example, speakers with small heads usually image better, because the ultrasound signal only has to travel a short distance before it is reflected back to the probe, meaning that there is little attenuation of the ultrasound waves. From recording a wide variety of speakers, this study also found that younger speakers tend to image better than older ones, possibly because of the distribution of skin, fat, and muscle under the chin. Regardless of age, most speakers still produced ultrasound images of a high enough quality to be analysed, but visibility was generally better in women, as they tend to have smaller heads than men. Surprisingly, speakers with beards did not image any poorer than speakers without, perhaps because the ultrasound gel helped reduce the amount of air between the probe and the chin.

### Profile-view lip video

This study was only able to record lip video in profile view, since a headset with an additional front-facing lip camera was not available for fieldwork. This is an unfortunate limitation, given that the front-facing images from previous research (Westerberg, 2016) showed a pattern of tongue protrusion, which would benefit from further investigation (see Section 4.5.4). This study nevertheless hoped to gain useful information about lip posture from the profile view images, and by estimating the position of the front teeth on the ultrasound (see Section 5.7.2).

## 5.8 Acoustic data processing and analysis

### 5.8.1 Overview of segmentation

The audio recordings and their corresponding text prompts were exported from the recording software AAA (Articulate Instruments Ltd., 2019) using the ‘Export data’ function, and manually segmented, transcribed, and annotated in Praat (Boersma & Weenink, 2019). The annotations were later imported back into AAA, so that the ultrasound data could be sampled at the same timepoints as the acoustic data.

Segmentation was carried out at the word and phoneme level. Broad phonetic transcriptions were provided for consonants, and narrower transcriptions for vowels, with a particular focus on offglides. Viby-i was not transcribed differently from standard [i:], as the criteria for Viby-i used in this study were acoustic rather than auditory (as outlined in Chapter 4). However, my impression during transcription was that all speakers in the sample used some form of Viby-i.

Automatic segmentation was not possible, as none of the currently available systems support Swedish. Attempts were made to use custom settings with the Munich Automatic Segmentation System (MAUS, Ludwig Maximilian University of Munich, 2019), but the software was unable to cope with the unusual vowel realisations in the data. For example, most /i:/ tokens included fricative noise, either in the middle or at the end of the phoneme, e.g. [i:j], [i:jə] (Fig. 5.6). Since these offglides are common in many Swedish dialects (Elert, 1995: 40) and may be relevant to Viby-i, it was deemed important that they should be included in the analysis so that they could be investigated further.

The offglides were not clear or consistent enough to warrant further segmentation into e.g. ‘vowel’, ‘offglide’, and ‘epenthetic vowel’. Furthermore, since the dynamic properties of Viby-i are not very well understood, the researcher did not wish to impose categories on the data based on what a vowel ‘should’ be like, but instead chose to include everything between the flanking consonants, in order to carry out a more exploratory analysis. For this purpose, manual segmentation was more flexible, and allowed closer visual and auditory familiarisation with the data, although it limited the number of speakers that could be processed.

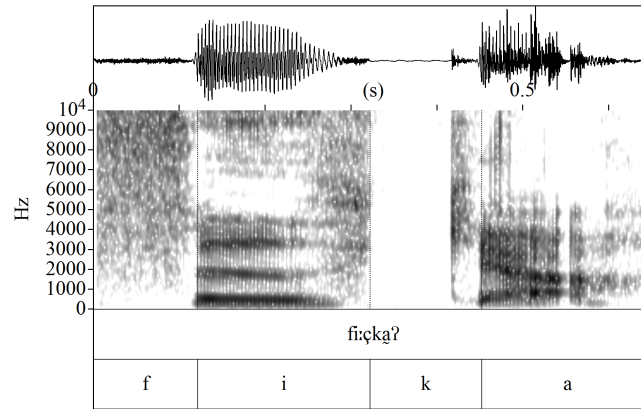


Figure 5.6: Example of how end-frication was segmented.

### 5.8.2 Segmentation protocol

A summary of the segmentation protocol is presented in Table 5.6. More detailed explanations of how these criteria were implemented, and what they were based on, are provided below. Due to the exploratory nature of this study, and the unusual offglides which are common in Swedish vowels, the segmentation criteria used in this study do not always conform with previously established methods for segmenting vowel sounds (e.g. Turk, Nakai, & Sugahara, 2006).

Preceding sound	Start of vowel segment
Plosive /p, b/	Waveform becomes periodic + F2 appears on spectrogram
Voiceless fricative /f, h/	Waveform becomes periodic + F2 appears on spectrogram
Voiced fricative /v/	Waveform complexity increases + F1 to F4 appear on spectrogram
Following sound	End of vowel segment
Plosive /p, b, t, d, k, g/	Modal: Waveform loses complexity + amplitude drops End-fric: Waveform loses high-frequency aperiodicity + amplitude drops
Liquid /r, l/	Waveform loses complexity <i>or</i> becomes aperiodic + amplitude drops
Nasal /n/	Waveform loses complexity + formants lose amplitude
Sibilant /s/	Waveform becomes aperiodic + high-frequency noise appears on spectrogram
Word-final	Modal: Waveform loses complexity + amplitude drops End-fric: Waveform loses high-frequency aperiodicity + amplitude drops

Table 5.6: Segmentation criteria for vowels based on preceding and following segment.

### Start of vowel segment

Vowels were annotated in a Praat interval tier. The start of the vowel segment was determined by the onset of periodic voicing on the waveform, combined with the appearance of F2 on the spectrogram (for /p, b, f, h/, see Fig. 5.7), and by increasing waveform complexity and a full F1 to F4 formant profile (for /v/, see Fig. 5.8). The separate criteria for /v/ were necessary because speakers often produced /v/ with formant structures which carried through the entirety of both the consonant and the vowel, making it difficult to place a boundary. In these cases, formants higher than F1 were a helpful landmark, as their appearance can signal that the vocal tract is open and unobstructed (Turk et al., 2006). The boundary between /v/ and the vowel could usually also be determined by an increased complexity in the waveform. In difficult cases, progressive listening was used to determine the boundary between the vowel and the initial consonant.

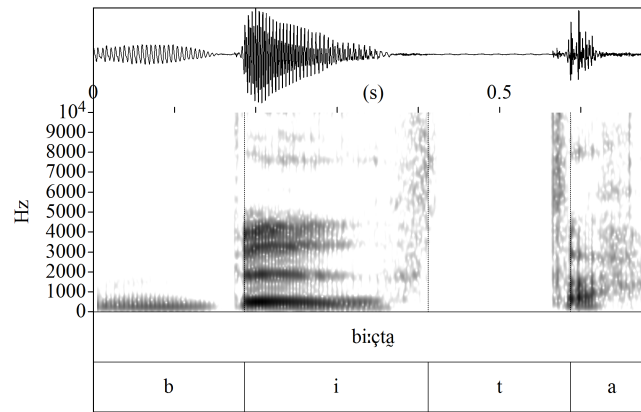


Figure 5.7: Illustration of vowel segmentation with preceding and following plosives.

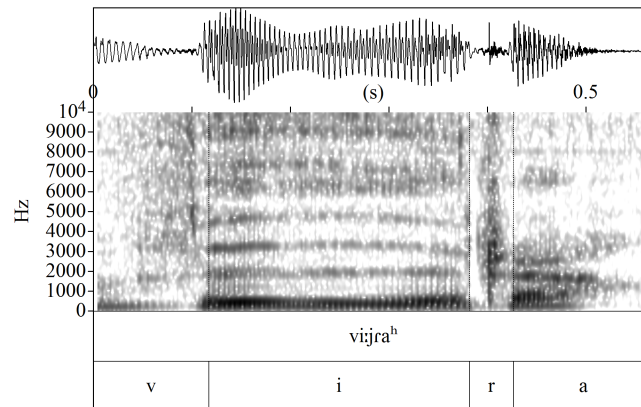


Figure 5.8: Illustration of vowel segmentation with preceding /v/ and following liquid.



Since Swedish vowels tend to reach their targets early (Bruce, 2010: 127) before moving into offglides which may vary greatly in their timing and realisation, it was determined that all static acoustic measurements should be taken at 10% of the vowel duration, rather than at the customary 50%. For this reason, annotation of the beginning of the vowel segment followed quite conservative criteria, to ensure that the formants could easily be measured at 10%, and would not be disproportionately affected by coarticulation with the preceding consonant. This aspect was particularly important to ensure consistency between vowel tokens with different preceding consonants; for example, vowels following plosives were not segmented from the burst, since this would have included aspiration from /p/, but not from /b/. This aspiration could then have been misinterpreted as ‘buzziness’ in the vowel itself.

The start of the vowel segment was thus based on acoustic consistency, rather than inferred articulatory criteria. From an articulatory point of view, if a vowel is preceded by a non-lingual consonant, e.g. *BI*, the tongue is usually in position to produce the vowel before the mouth opens for the consonant. Thus, for an articulatory analysis, it would be justifiable to include the initial consonant in the vowel segment as well. However, to make the acoustic and articulatory data comparable, it was necessary to use the same timepoints for both analyses, meaning that the more conservative acoustic criteria were also used for the articulatory data.

### **End of vowel segment**

The criteria for the end of the vowel segment differed somewhat depending on the following consonant, and to some extent depending on the vowel offglide. The final segmentation protocol was developed over several attempts to segment the data consistently, which was complicated by the degree of variation found in the data. In two environments (plosive and word-final), vowels were divided into the broad categories of ‘modal’ vowels (which ended in periodicity), and ‘end-fricated’ vowels (which had an aperiodic end-phase).

When followed by a plosive, the end boundary for modal vowels was set when there was a visible drop in the complexity and amplitude of the waveform, signalling closure of the vocal tract. The waveform was used, rather than the spectrogram, as it is more temporally accurate. The end-boundary criteria were the same for word-final modal contexts, but in these cases, it was usually more difficult to judge where to place the boundary, as the signal tended to diminish gradually. This issue was usually approached by zooming out in Praat to gain a wider view of the amplitude in relation to the rest of the signal, and by using progressive listening.

The most difficult cases of end-boundary annotation for word-final contexts were the few speakers that consistently exhaled at the end of their utterances, often heavily and over an extended duration (sometimes twice as long as the modal portion of the vowel). This aspirated phase usually overlapped with diphthongisation or epenthetic vowels, and was sometimes visually difficult to tell these apart from end-frication (see Figs. 5.9 and 5.10). Having tried a number of different strategies, the most consistent way of segmenting these tokens was to de-

termine auditorily whether the vowel was modal or end-fricated, and treating the two categories separately. The end boundary for modal vowels was set when the waveform became aperiodic. This sometimes meant that diphthongisation which overlapped with the exhale was excluded, but this cut-off rarely affected /i:/. The end boundary for end-fricated vowels was set when the waveform lost amplitude and aperiodic energy. Despite great effort, the end boundaries for word-final contexts are nevertheless likely to be less consistent than those of word-medial vowels. This issue only affects the dynamic part of the analysis, however, since the static measurements rely on the first 10% of the vowel duration.

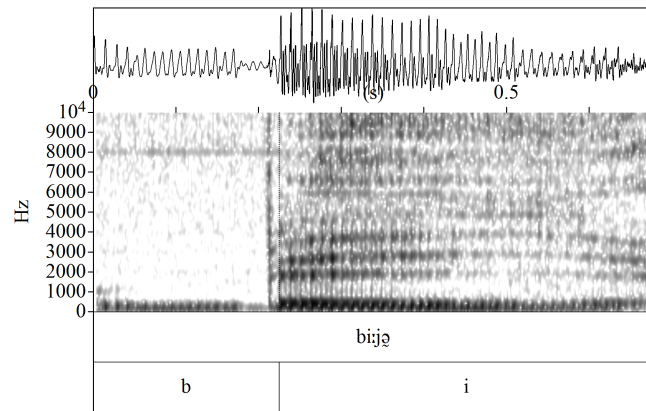


Figure 5.9: Word-final exhale overlapping with epenthetic vowel for /i:/.

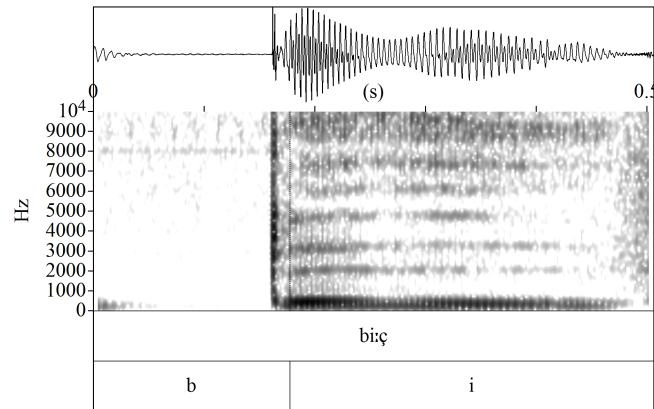


Figure 5.10: Word-final end-frication of /i:/.

End-frication before plosives was easier to distinguish visually, as the frication often had a high frequency and short duration (see Fig. 5.7). To some extent, end-frication seemed to be affected by coarticulatory processes, e.g. the frication before /t/ tended to be alveolar, with a higher frequency, while the frication before /k/ tended to be velar, with a lower frequency.

However, coarticulation only appeared to be part of the picture, as word-final contexts exhibited end-frication as well (Fig. 5.10), but the type and degree of frication varied between speakers.

The offglides occurring before /l, r, n, s/ also varied, making it necessary to use different segmentation criteria for these environments. Determining the boundaries between liquids, nasals, and vowels can be difficult, because all are usually periodic with formant structures, and there are not necessarily reliable acoustic cues indicating contact in the vocal tract. Some cues were observed, for example when speakers produced a tapped or trilled /r/, when the sides of the tongue were audibly released for /l/, or when there was visible formant transitions. However, in many cases, a loss of waveform complexity or amplitude were the most reliable criteria (Lawson et al., 2011) (see Fig. 5.11). Progressive listening was also used to ensure that the consonant could no longer be heard in the vowel segment.

When the vowel was followed by /s/, the end of the segment was determined by the onset of high-frequency aperiodic noise. Since the aperiodicity of /s/ closely resembled that of a fricated offglide, it was not always possible to fully separate the two, which may have implications for the frication analysis. However, the transition into /s/ was often gradual, beginning with frication superimposed on the vowel waveform (Fig. 5.13). These ‘hairy’ waveforms were included in the vowel segment, until they became fully aperiodic, when they were instead classified as /s/.

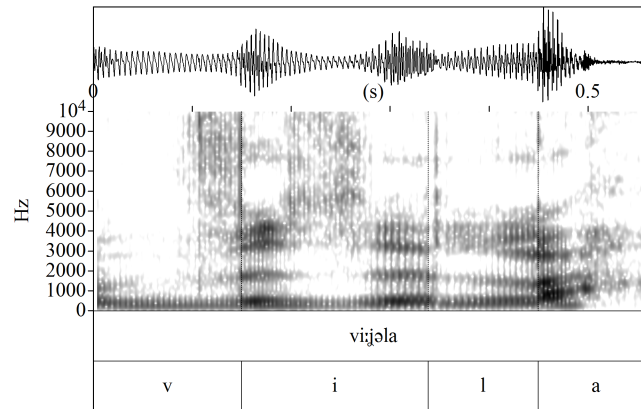


Figure 5.11: Loss of amplitude before a liquid.

### Additional annotations

The full target word was annotated in a Praat interval tier in order to preserve the consonant context of the vowels, and so that the words could be extracted for future projects. This process included additionally segmenting the preceding and following consonants, and the unstressed vowels.

The start boundary for plosives was set from the closure (signalled by a loss of waveform complexity and amplitude) until the onset of the following vowel. In cases where the participant

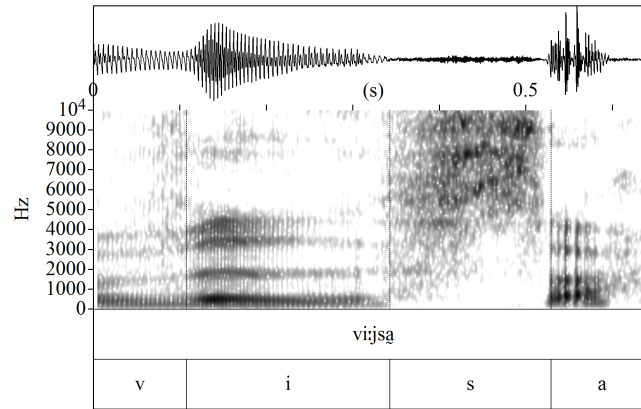


Figure 5.12: Boundary between end-frication and /s/.

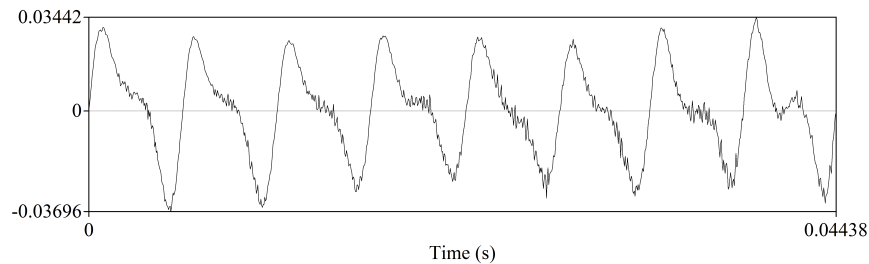


Figure 5.13: Frication superimposed on periodic waveform.

paused before the target word, the start boundary was set at the burst (signalled by a peak on the waveform). Fricatives were segmented from the onset of aperiodic noise, or, for /v/ (which often retained periodicity), from a drop in complexity and amplitude on the waveform. Unstressed (word-final) vowels used the same segmentation criteria as target vowels in plosive environments for both the start and the end boundary.

The segmentation also included a separate tier for comments about the speaker's voice quality, prosodic patterns (e.g. pausing before producing the target word), and other aspects of the recording.

After segmentation was complete, a custom-made Praat script was used to place 11 equidistant points at 10% intervals across the vowel duration (starting at 0%). These points were then used as landmarks when taking acoustic and articulatory measurements.

### 5.8.3 Formant measurement

The first four formants were measured using the 'Formant (to Burg)' function in Praat. The maximum frequency was set to 5,000 Hz for men, and 6,000 Hz for women. For other settings, Praat's default values were used (5 formants, 25 ms windows, 6.25 ms time-step, pre-emphasis

from 50 Hz). A customised Praat script based on Stuart-Smith and Lawson (2018) was used to extract average formant values across each 10% portion of the vowel. Average values were preferred over single-point measures, as they represented a larger sample of data points, which would minimise the risk of measurement errors. The values were taken in 10% blocks, rather than as an average across the whole vowel, to enable a dynamic analysis of the formant values.

For the static part of the formant analysis, only the first 10% of the vowel was used, representing the vowel ‘target’. Outliers (values  $\pm 1.5 \times$  the interquartile range) from the 10% timepoint were identified and manually re-measured in Praat. If they still remained outliers, with no apparent error in the formant measurement, the token was kept. If there were issues with the formant tracker, the frequency settings in Praat were adjusted, or, if there were still errors, the token was measured at an adjacent point with a similar formant structure.

Subsequent vowel portions (20% to 100%) were not checked for errors or outliers due to time limitations, but also because their confidence intervals could be informative to the dynamic analysis. Larger confidence intervals at specific points could indicate greater difficulties in measuring the vowel accurately, which could coincide with other dynamic properties, such as formant weakening or fricative noise.

#### 5.8.4 Bandpass-filtered zero-crossing rate

Bandpass-filtered zero-crossing rate (bpZCR) is a measure used to quantify the degree of aperiodicity in a signal which may contain voicing. The method was first used by Gordeeva and Scobbie (2010) to investigate pre-aspiration in Scottish vowels. The method works by filtering out periodic noise in the vowel, and measuring the number of times the waveform crosses the zero point in the remaining signal. A higher rate of zero-crossings indicates greater levels of aperiodicity in the signal, but it may also capture the frequency of the aperiodicity, as higher frequencies have shorter cycles, leading to more zero-crossings.

There were two kinds of aperiodicity of interest for this study: intra-vowel frication, i.e. fricative noise superimposed on the periodic waveform for /i:/, and end-frication, i.e. fricative offglides, which were usually entirely aperiodic. To ensure that bpZCR could capture both types of frication, as well as the difference between them, a number of different filtering frequencies were applied to a sub-sample of the data, and inspected manually to ensure that both types of frication were visible. Filtering out all frequencies below 1,000 Hz appeared to be the most effective way of reflecting what could be heard and seen in the recordings. This filter removed F0 and F1 for /i:/, but higher frequencies (including formants) were still present. After filtering, voiceless portions remained aperiodic (with more zero-crossings), while previously voiced portions were semi-periodic (with fewer zero-crossings). Fully periodic sounds were also visibly different from periodic sounds with superimposed frication (Figs. 5.14 and 5.15).

Filtering was carried out using the ‘Hann Band filter’ function in Praat, set to 100 Hz smoothing. After filtering, the ‘Zero Point Process’ function was used to identify all (falling and rising)

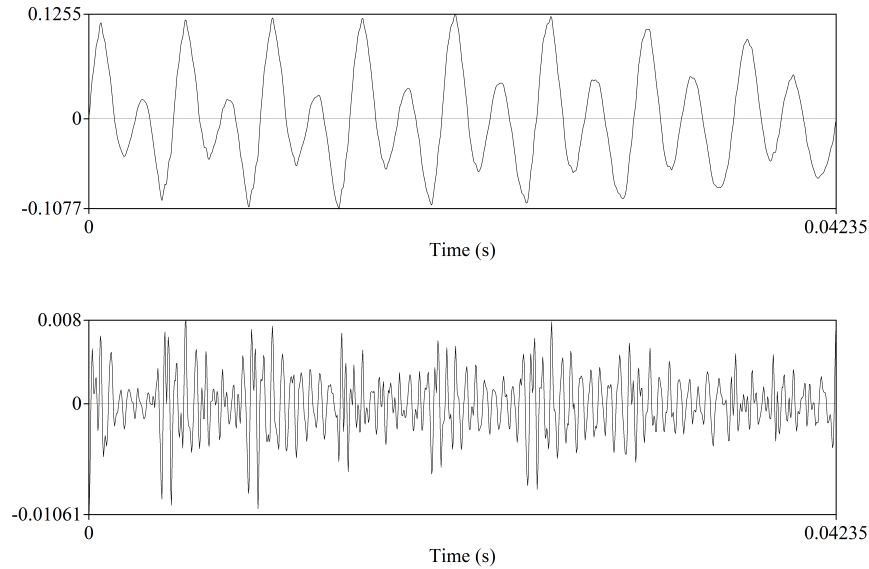


Figure 5.14: Waveform of modal /i:/ before and after filtering.

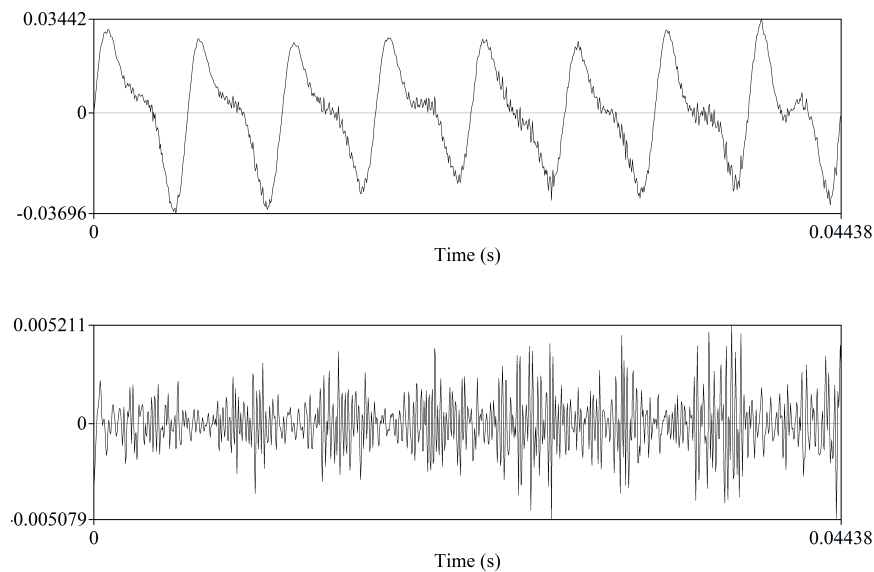


Figure 5.15: Waveform of /i:/ with superimposed frication before and after filtering.

zero-crossings. The zero-crossing rate was then calculated for each 10% block by dividing the number of zero-crossings by the duration of the vowel portion in seconds. The measure thus expresses the average number of zero-crossings per second for each 10% portion of the vowel.

The use of bpZCR was intended to illustrate frication dynamics throughout each vowel, but the measure may not be suitable for comparing noise levels across different vowel phonemes. This is because the degree of periodicity in the filtered sounds is likely to vary based on the formants. For example, a vowel like /a:/, which in this dataset has an average F2 of around 1,000 Hz, would lose most of its periodicity, whereas /e:/, whose F2 is around 2,000 Hz, would retain some periodicity even after filtering. The nearest baseline for frication in /i:/ would thus

need to be a non-fricated vowel with a similar formant distribution. In this dataset, the closest candidate was /ɛ:/, but this comparison is not ideal. From visible inspection, it is clear that /ɛ:/ has a lower zero-crossing rate, but it is difficult to know how much of this difference is due to frication, and how much is due to differences in spectral energy. There was also much more data for /i:/, which appears in 17 word contexts, compared to /ɛ:/, which only appears in 3. Nevertheless, a cautious comparison of the bpZCR of /i:/ and /ɛ:/ will be presented in Section 6.9.3. Since the analysis of fricative noise is only a small part of this thesis, the methods used to interpret the results are mainly qualitative.

### 5.8.5 Bandwidth measurement

One of the potential reasons why Viby-i has been described as having a “damped” vowel quality (Björsten & Engstrand, 1999) could be that the formants have wider, less defined bandwidths than a standard [i:]. To investigate this phenomenon further, bandwidth measurements were taken alongside the formant measurements, using the ‘Linear Hertz’ setting in Praat. Since Praat does not have an automatic averaging function for bandwidths, I instead sampled three single points per 10% vowel portion, and calculated their mean value. Upon examining these values, however, it became evident that the bandwidths were unreasonably variable (changing from single-digits to triple-digits from one vowel portion to the next). Previous researchers have also reported issues with Praat’s bandwidth measuring function, confirming that these values were probably inaccurate (Burris, Vorperian, Fourakis, Kent, & Bolt, 2014). Due to time constraints, no alternative method of obtaining bandwidth measurements was tested, but future work is encouraged in this area.

## 5.9 Articulatory data processing and analysis

### 5.9.1 Tongue splines

The UTI data was processed by drawing outlines of the tongue curve, known as ‘tongue splines’ (Fig. 5.16), in AAA, and exporting the spline coordinates for analysis. Each spline was first roughly drawn by hand, then adjusted using AAA’s ‘snap to fit’ edge-detection function, and finally hand-corrected to remove possible edge-detection errors. In difficult cases, the ultrasound frames could be played backwards and forwards at various speeds, which helped determine visually whether the splines were drawn correctly.

In some cases, the outline of the tongue was ambiguous, particularly near the root of the tongue. This problem can arise because of the placement of the ultrasound probe; if the probe is rotated sideways or not placed along the midline of the tongue, the image sometimes displays both the central groove and the raised sides of the tongue simultaneously. Similarly, certain artefacts of the ultrasound reflecting on the same surfaces multiple times can give the impression

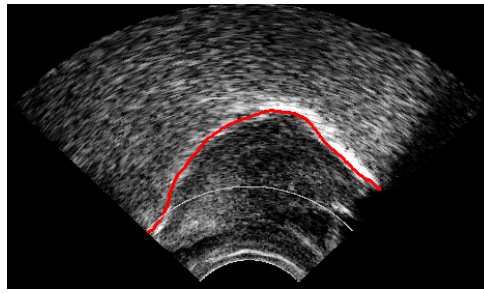


Figure 5.16: Example of tongue spline in AAA.

that there are two curves instead of one. The root of the tongue is also easily mistaken for the epiglottis. These issues are not likely to have a great effect on the articulatory analysis, as the large sample size should be able to absorb a small number of errors.

The timepoints annotated during acoustic segmentation were imported into AAA using the ‘import annotations’ function. Three splines were created for each vowel token, at 10%, 50%, and 90% of the vowel duration. The 10% splines were used for comparison with the static acoustic measurements in the analysis. All three timepoints were used to investigate dynamic tongue movement. Due to time constraints, it was not possible to create splines for all 11 timepoints. In total, the dataset consisted of around 12,200 splines.

The decision to use the acoustic timepoints, rather than trying to visually identify a steady state or maximum gesture in the articulation, was based on the fact that speakers can be highly variable, not only in their tongue gesture, but also in their timing (Lawson et al., 2014). Furthermore, it was not always possible to identify a point where the tongue was still, since participants were producing connected speech, where the tongue may be in constant gradual motion. For vowels with end-frication, it is also likely that the maximum gesture (the most peripheral or constricted gesture that the tongue achieves before reverting towards a neutral position) would occur during end-frication, rather than during the main vowel portion. In these cases, the tongue gesture would not be informative for an analysis of the vowel quality. Finally, using the acoustic timepoints meant that the acoustic and articulatory data could be directly compared, so that the effect of tongue gesture on acoustic output could be examined.

In addition to the tongue splines, a palate spline (Fig. 5.17) was also drawn from the recording(s) of the participant swallowing. A fiducial (straight) spline was also created for the bite plate (Fig. 5.18). The length of the fiducial spline was set to 45 mm, which is the length of the biting edge of the bite plate. The spline was aligned with the visible part of the bite plate from the recording. Since the spline had a fixed length, it could then extend into the part of the mouth obscured by the jaw shadow, showing the approximate position of the front teeth where the spline ended.

Once drawn, the tongue splines were exported into the AAA Workspace using the ‘batch export’ function. Roof and bite splines were exported individually. The Cartesian coordinates of the splines were then exported as plain text files, with the settings ‘Interpolation = 1’ and



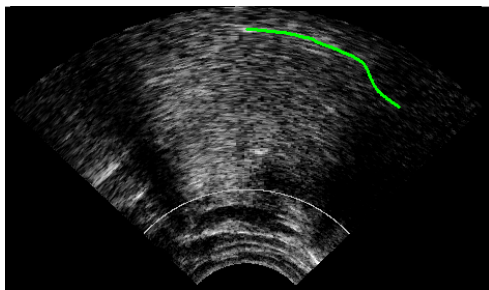


Figure 5.17: Example of palate trace in AAA.

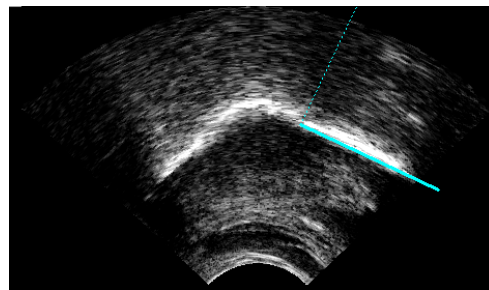


Figure 5.18: Example of fiducial (bite plate) spline in AAA.

‘Confidence threshold = 1’. Interpolation represents the number of data points per fan line; if a higher number is selected, AAA will add additional points between the original 42 fan lines to create a smoother curve. The confidence threshold is used to filter out parts of the spline where the edge-detection function had a low estimated accuracy. The degree of certainty is shown by the solidity of the line. However, hand-drawn splines do not have this feature, and therefore the confidence threshold was kept as low as possible, to ensure that all parts of the spline were exported.

The coordinate output files were converted into CSV using a Python script (Python Software Foundation, 2019) written by Cohen (2019). Spline measurements were taken using an R script (R Foundation for Statistical Computing, 2019) written by Leplâtre and Lawson (2019). R was also used to add additional data categories, and to plot and analyse the data, as outlined below.

## 5.9.2 Qualitative tongue shape analysis

Mean tongue splines for each speaker and vowel were calculated in R, and plotted using the ‘ggplot2’ package (Wickham, 2016). A selection of these plots will be presented in Chapter 7, with the full set available in Appendix G.

The splines were visually categorised according to their tongue body height and frontness, tongue tip gesture and height, tongue back or root retraction, tongue back or root constriction, and overall tongue shape. The tongue position for /i:/ was compared to the nearest high front vowel, /e:/ (after Westerberg, 2016). However, unlike in my previous work, I did not quantify tongue positions as a ratio of /e:/ (e.g. highest point of /i:/ divided by highest point of /e:/), since the splines showed that the tongue position for this ‘anchor vowel’ was highly variable. This variability could result from dialectal differences, as well as individual differences in vocal tract anatomy. The two vowels also differed greatly in dynamics, since /i:/ usually has a closing, fricated offglide, while /e:/ usually has a diphthongal offglide towards schwa.

The classification of tongue shapes used four different categories: arched, front-bunched, back-bunched, or double-bunched. Arched tongue shapes had no clear constriction, but took the shape of a convex curve. For arched shapes, the highest point of the tongue tended to coincide with the point of maximum narrowing in the vocal tract. The term ‘bunching’ was used to refer

to visible constriction, signalled by a protrusion on the tongue surface. Front-bunched shapes thus had their main constriction at the front, back-bunched shapes had their main constriction at the back, and double-bunched shapes were constricted at both the front and the back.

The qualitative classifications were not used in the statistical analysis, as the normalised measurement points described in Section 5.9.3 provided more objective representations of the same information. However, a summary of the qualitative results is provided in Section 7.2.

### 5.9.3 Normalised articulatory measures

Since there is no established way to normalise articulatory data between speakers, articulatory sample sizes tend to be small, and researchers are often required to classify speakers or vowels one by one. However, given the amount of data available in this thesis, I wished to apply a method that allowed quantification of key aspects of the tongue gesture, such as tongue body height and backness, and combine data from multiple speakers into a single statistical model. As mentioned above, I had previously used anchor vowels for this purpose, but found that it would be preferable if the measurements were modelled on a more reliable representation of the articulatory space.

Vocal-tract normalisation is unusual for ultrasound data, as there are many parameters to consider in the vocal tract, and ultrasound is not able to image any other articulators except for the tongue and palate. However, an approximation of the articulatory space can be obtained by combining information about the minimum and maximum values for specific tongue spline coordinates. In this way, the correspondence can be investigated between certain parts of the tongue and the resulting acoustic output. This approach to UTI normalisation was first described in Lawson et al. (2019).

As mentioned previously, the articulatory measurements for the current study were taken using an R script by Leplâtre and Lawson (2019). The script takes the following articulatory measures from all vowel tokens at the 10% timepoint:

- (A) Highest point of the tongue – ‘peak’
- (B) Frontest part of the tongue – ‘tip’
- (C) Backest part of the tongue – ‘backest’
- (D) Backest point at half tongue height – ‘midback’
- (E) Most constricted point (highest perpendicular point when a line is drawn between root and tip) – ‘max constriction’

A schematic of how these measurement points relate to the tongue curve is shown in Fig. 5.19. Note that ‘backest’ could occur at the tongue root, or at a similar height to ‘midback’ depending on the shape of the tongue.

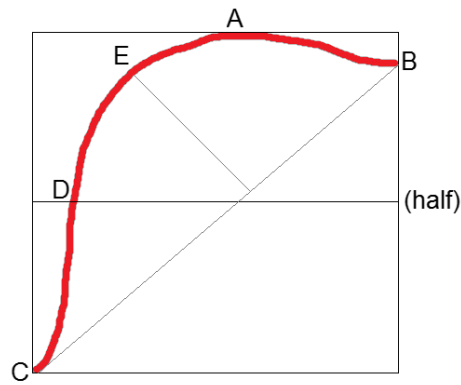


Figure 5.19: Measurement points used for the normalised articulatory analysis.

The measurements were z-score normalised using the Lobanov (1971) method, following Blackwood Ximenes, Shaw, and Carignan (2017) and Lawson et al. (2019). Normalisation was implemented through the ‘vowels’ package in R (Kendall & Thomas, 2018). Each measurement category (highest, midback, etc.) was normalised separately, entering the y-values into the ‘F1’ column, and the x-values into the ‘F2’ column. The data included all nine vowels, measured at the 10% timepoint. Since the acoustics showed little evidence of formant movement over the course of the vowel, dynamic articulatory measurements were not used for the acoustic-articulatory analysis. Instead, Section 7.7 will provide a qualitative analysis of the effect of dynamic tongue movement on the formants, and Section 7.7.3 will provide a statistical analysis of tongue gesture dynamics and frication.

### 5.9.4 Lip data

The thesis originally intended to present quantitative measurements of lip protrusion and mouth opening, in order to investigate the cumulative effect of tongue and lip gestures on acoustics. However, due to time constraints, the lip data could not be processed quantitatively. Instead, a qualitative analysis of the lip data is provided in Section 7.5, and the full set of lip images is presented in Appendix K.

## 5.10 Statistical methods

### 5.10.1 Mixed-effects modelling

Linear mixed-effects regression (LMER) was used for statistical testing of the acoustic and articulatory data. The tests were carried out using the R packages ‘lme4’ (Bates, Maechler, Bolker, & Walker, 2015) and ‘lmerTest’ (Kuznetsova, Brockhoff, & Christensen, 2017), and plotted using ‘effects’ (Fox & Hong, 2009). Mixed-effects modelling uses a combination of fixed and random effects to account for variation in the data. Fixed effects are factors which

can be predicted to affect the data in a certain way, e.g. raw F1 can be predicted to be higher in women, due to differences in vocal tract size. Random effects are factors whose direction cannot be predicted, but which are nevertheless likely to influence the data, e.g. individual variability.

In sociolinguistic research, the use of random effects usually helps to eliminate any disproportionate ‘pull’ on the results caused by outliers or individual speakers (Hay, 2011: 212). Such problems can arise in models that only use fixed effects, as they are built on the assumption that all data points are independent of one another (D. E. Johnson, 2009: 363). However, sociolinguistic data usually consists of multiple data points from the same speaker, meaning that they are not independent. Using ‘speaker’ as a random intercept in the model thus ensures that the data from each speaker is treated as a separate category, e.g. if one particular female speaker produces vowels with a low F1, her vowels are not mistaken for a general trend in the female data.

Mixed-effects models are more conservative than fixed-effects models, meaning that they are likely to give fewer significant results. However, they are also more robust, meaning that significant results are more reliable (Hay, 2011: 212-213). The main issue with mixed-effects modelling for linguistic data is that these models are less reliable when the dataset is small (Eager & Roy, 2017) or highly unbalanced (Roy & Levey, 2014). Also, unlike fixed-effects models, they do not provide an estimation of the total amount of variance explained by the model.

Since the statistical analyses in this thesis are mostly exploratory, i.e. less interested in falsifying testable hypotheses, and more interested in testing the effects of a number of different factors, some models were subjected to a variable selection process known as ‘step modelling’. This process used the ‘step’ function from the ‘lmerTest’ package, which uses backward elimination to determine which factors to keep in the model. Backward elimination works by removing factors from the original model one by one, assessing whether each of them contributes to the explanatory power of the model (Field, Miles, & Field, 2012: 265). For fixed variables, ‘step’ assesses each variable’s contribution using the F-statistic (calculated using the Satterthwaite Approximation method), and for random variables, it uses the Chi-Squared value (calculated using the Likelihood Ratio). It then generates a final model that only keeps factors that contribute to explaining the data.

Before the mixed-effects models were used, factors were tested for multicollinearity with the ‘cor’ function in base R (Pearson method), and the ‘rcorr’ function from the ‘Hmisc’ package (Harrell, 2019) to generate p-values. Factors were considered to be collinear at  $p < 0.05$ . Correlation plots were generated using ‘corrplot’ (Wei & Simko, 2017). Output for the lmer models was formatted using the ‘stargazer’ package (Hlavac, 2018).

### 5.10.2 Generalised additive mixed models

To quantify the visually observed patterns in the ultrasound splines, and to investigate tongue gesture dynamics, I used generalised additive mixed models (GAMMs) (e.g. Sóskuthy, 2017). This method is used to create tongue gesture visualisations that are more statistically informative than mean splines. Since mean values are easily skewed by outliers, mean splines sometimes create unfeasible tongue shapes that do not accurately reflect the data. GAMM plots are based on non-linear mixed-effects regression modelling, meaning that they can capture the distribution of data points using a curved trajectory, as opposed to linear regression, which uses a straight line. GAMMs can also account for the fact that each spline has its own predicted trajectory.

When investigating tongue curves, however, the statistical output from the GAMM model itself is not particularly informative, as it only tests if the trajectories are statistically different from a straight line. While there are a number of different statistical tests that can be applied to test whether two GAMM curves are significantly different from each other, many of them suffer from issues with either false positives or diminished statistical power (Sóskuthy, 2017: 20). I therefore used a combination of GAMM plots to visualise the data, accompanied by ‘difference smooths’ to investigate portions of significant difference between the curves. The latter has been described as “the preferred option” for identifying significant parts of the tongue curve (Sóskuthy, 2017: 20).

Difference smooths calculate the difference between two curves (e.g. /i:/ and /e:/) at each point along their trajectory. This difference is represented as a ‘smooth’ (i.e. a curvy line), which fluctuates around zero. When the smooth is at zero, the two curves overlap perfectly. Positive and negative values represent variation in different directions, i.e. one curve being higher or lower than the other. The difference smooth also has a 95% confidence interval, which can be used for visual significance testing: When the confidence interval no longer touches the zero line, the two curves are significantly different from each other at  $p < 0.05$ . For this analysis, I used the ‘rticulate’ package (Coretta, 2019) to create the GAMMs, and ‘itsadug’ (van Rij, Wieling, Baayen, & van Rijn, 2017) and ‘mgcv’ (Wood, 2017) to create the difference smooths.

For tongue curves, this kind of ‘continuous’ significance testing can be more informative than a single significance test, as it provides valuable information about the exact locations where the two curves differ. To some extent, similar judgements can be made from the confidence intervals shown on the GAMM plots, but Sóskuthy (2017) advises against this approach, because, while non-overlapping confidence intervals in these models can show that two curves are significantly different, overlap does not guarantee that two curves are ‘the same’. Although difference smooths are also unable to prove ‘sameness’ (only difference), this kind of misinterpretation is more likely for the GAMMS given their visual similarity to the tongue splines. For this reason, the GAMM plots shown here should only be taken as illustrative of patterns in the data, while the difference smooths provide a clearer indication of where the two curves are statistically different.

The output from the static GAMM analysis is discussed in Section 7.3 and provided in full in Appendix H. Three figures are presented for each speaker: The first is the predicted GAMM trajectories for /i:/, y:/, ʉ:/, e:/ for each speaker at 10% of the vowel duration. The palate trace (dotted line) and position of the front teeth (cross) were added manually for reference (Fig. 5.20). These graphs correspond to the mean spline plots in Appendix G, but provide a more accurate statistical representation of the tongue curves. The second and third figures are difference smooths exploring the difference between /i:/-/e:/ and /i:/-/y:/ respectively (Fig. 5.21). The black line represents the estimated difference between the two vowels, and the shaded area is a 95% confidence interval. Portions where two vowel curves are significantly different from each other ( $p < 0.05$ ) are marked by a red zero line.

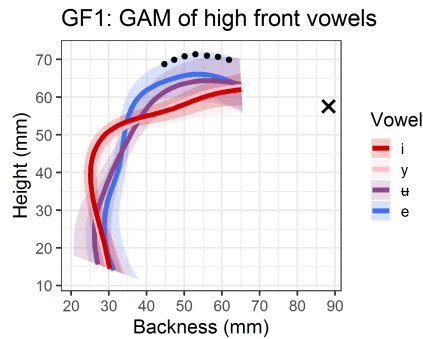


Figure 5.20: Example of static GAMM plot.

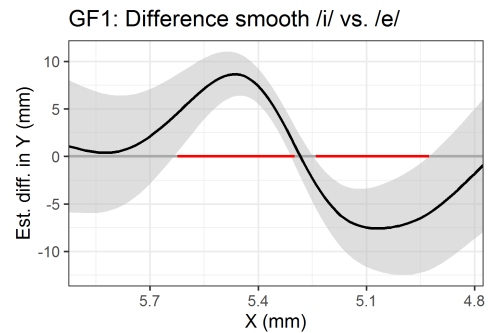


Figure 5.21: Example of corresponding difference smooth, comparing /i:/ and /e:/.

Unfortunately, the GAMM plots and the difference smooths do not use the same scaling. This is because the GAMM model converts the coordinates from a Cartesian format ( $x$  and  $y$ ) to a polar format (origin, distance, and angle). For the GAMM plots, the coordinates are automatically converted back into Cartesian to maintain their original dimensions, while the difference smooths remain polar. Although attempts were made to convert the difference smooths into Cartesian coordinates, issues with scaling remained, as the tongue curve could have multiple different  $y$ -values for the same  $x$ -value (e.g. when the tongue back was vertical). The difference smooth is not able to process these tokens accurately, and thus in this context, Cartesian difference smooths are no more informative in their scaling than polar smooths are.

Instead, the easiest way to interpret the difference smooths is by finding points where the smooth crosses the zero line. These points correspond to places where the two vowel curves cross or overlap. For ease of reading, the axes of the difference smooths have been reversed, so that the curve is presented in the same orientation as the tongue splines. Values below the zero line (positive values) represent an /i:/ curve that is nestled within the vowel it is being compared to, while values above the zero line (negative values) represent /i:/ ‘breaking out’ from the curve it is being compared to (Fig. 5.22). Since the tongue curves are arched while the difference smooth is flat, this ‘breaking out’ has slightly different interpretations depending on where on the smooth it occurs: At lower  $x$ -values, it represents tongue backing, while at higher

x-values, it represents tongue raising or fronting.

In a few cases, the GAMM function was not successful in processing the data, resulting in clearly erroneous trajectories that could not be analysed (Fig. 5.23). These problems seemed to have been caused by errors when the data was exported from the recording software, specifically related to the correspondence between the Cartesian coordinates and the ultrasound fan lines. Speakers who exhibited these data errors were excluded, and the GAMM analysis was thus carried out without the following eight speakers: GF5, SF4, SM1, SM3 (no ultrasound data), UF3, UF4, UM1, UM3.

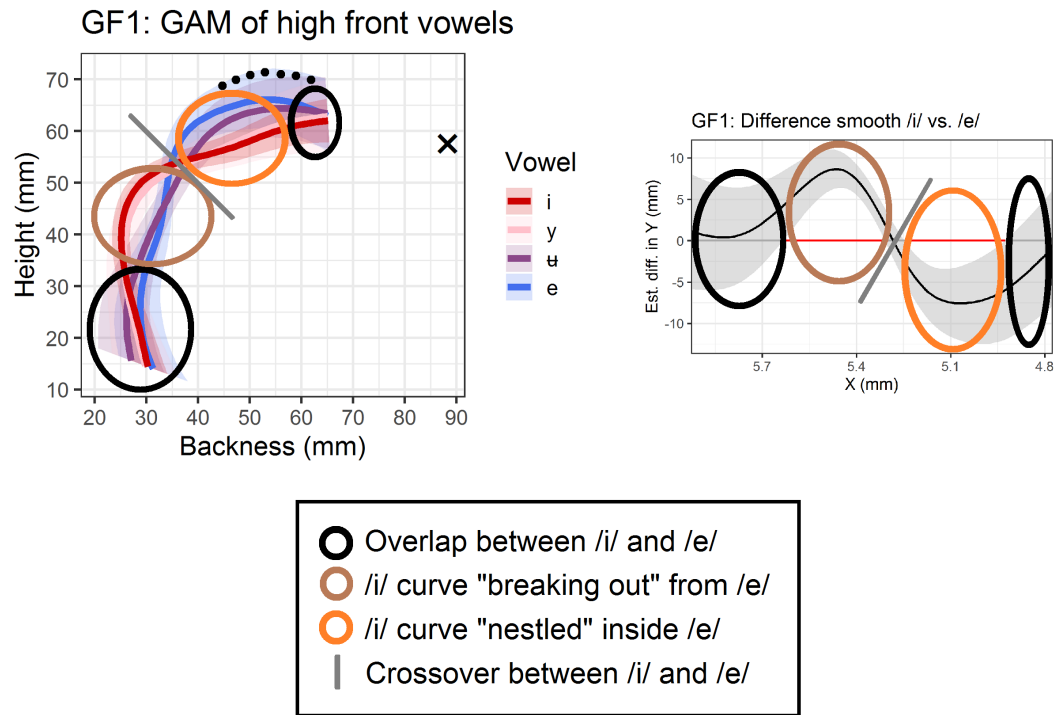


Figure 5.22: Example of relationship between GAMMs and difference smooths, comparing /i:/ and /e:/ splines.

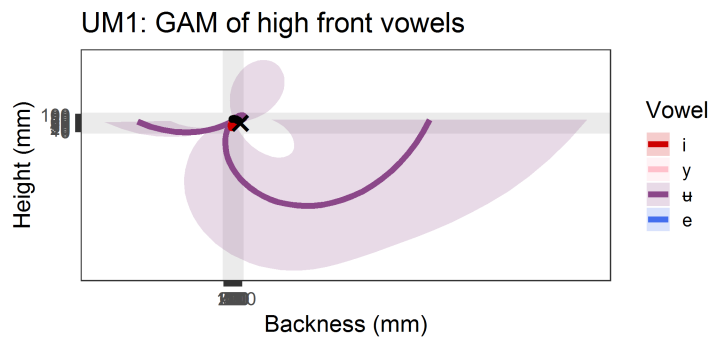


Figure 5.23: Example of erroneous GAMM.

# Chapter 6

## Acoustic results

### 6.1 Chapter overview

This chapter presents the results of the acoustic analysis of Viby-i. The first part reports the results of the static analysis, using data from the first 10% of the vowel duration. Section 6.2 briefly discusses the auditory impressions of Viby-i from the segmentation process. Section 6.3 shows the position of Viby-i in the acoustic F1/F2 vowel space, comparing it to other vowels in the system, and discussing the signs of a possible vowel shift. Section 6.4 provides a more detailed analysis of individual speakers' formant values for /i:, y:, e:/.

These sections are followed by a number of statistical tests: Section 6.5 tests the overlap between Viby-i and other vowels in the system, focusing particularly on /y:/ . Section 6.6 explores the relationship between F1 and F2 in the production of Viby-i. Section 6.7 investigates whether variation in Viby-i production can be attributed to differences in linguistic context. Section 6.8 performs a similar analysis of sociolinguistic factors.

Finally, Section 6.9 investigates the dynamic aspects of Viby-i, using data sampled at every 10% interval across the vowel duration. Subsection 6.9.1 visually examines formant values, formant intensity, and frication over the course of the vowel. Subsection 6.9.2 explores the occurrence of frication during Viby-i based on auditory transcriptions. Subsection 6.9.3 uses bandpass-filtered zero-crossing rate to quantify the level of frication across Viby-i, and investigates its relationship with linguistic context and individual speaker patterns.

### 6.2 Auditory impressions and acoustic parameters

Auditorily, all 34 speakers in the sample were classified as using some form of Viby-i, but, as suggested by the pilot study (Westerberg, 2016), there seemed to be a gradient difference in the kind of Viby-i they were producing. Some instances had a 'darker' or 'thicker' quality, while others were perceptually closer to a standard [i:]. As a full perceptual study is beyond the scope of this thesis, this analysis will rely on the previous finding that the 'damped' quality of Viby-i



appears to be primarily driven by a low F2, rather than a high F1 or high F3 (Westerberg, 2016). In my impression, variation in F1 could have an effect on Viby-i, contributing to the vowel’s perceived height, but this phenomenon appears to be distinct from variation in F2, which seems to contribute to the specific impression of Viby-colouring.

In terms of ‘buzziness’, most of the participants’ Viby-i productions were not perceived as being particularly fricated during the vocalic portion, but most had fricated offglides, which will be described in more detail in Section 6.9. An interesting phenomenon, which could contribute to the impression of intra-vowel frication, however, was the occurrence of end-frication before an epenthetic schwa (see Fig. 6.1). The fact that frication in these cases occurred in the middle of (what listeners might consider) the vowel phoneme, could be one explanation why Viby-i has been described as ‘buzzy’. However, it is also possible that intra-vowel frication is (or was) more common in varieties other than the ones studied here.

As seen in Fig. 6.1, there was also a tendency for the formants to weaken during part of Viby-i, which often coincided with the offglide. This formant weakening was usually perceived as either /j/ (if simultaneously fricated) or /j/ (if not fricated). This formant weakening could also be related to the impression of ‘dampedness’ described by Björsten and Engstrand (1999).

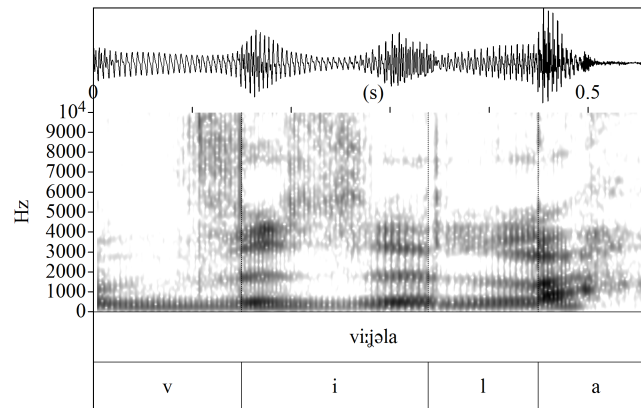


Figure 6.1: Example of frication and formant weakening in the middle of a vowel segment, followed by schwa.

### 6.3 Position of /i:/ in the Swedish vowel space

Acoustic data from all 34 speakers at the 10% timepoint is displayed in Fig. 6.2, showing the distribution of all nine long Swedish vowels in the F1/F2 plane in Lobanov normalised format (Lobanov, 1971). Mean formant values (Hz) for each vowel are also provided in Table 6.1.

As the figure shows, most vowels are located in the acoustic region where we expect them to be, with the major exception of /i:/, and possibly /y:/. In this set, /i:/ has a visibly lower F2 than /e:/, but is similar to /e:/ in F1. In raw Hertz values, F1 for /i:/ and /e:/ are nearly identical, at

around 400 Hz for women, and 340 Hz for men, but there is a large difference in F2, with /i:/ being lower than /e:/ by around 425 Hz for women, and 285 Hz for men. As reported in previous literature (e.g. Björsten & Engstrand, 1999), Viby-i is acoustically centralised, particularly in F2, and it no longer appears to be the high front vowel of the system. Since these results meet the criteria of how Viby-i is defined in this thesis, it can be concluded that most of the speakers in this sample use Viby-i.

The figure also shows considerable overlap between /i:/ and the other high vowels, particularly /y:/, and to some extent /ɥ:/ . This suggests that /y:/ could also be Viby-coloured, as previous literature suggests (e.g. Gross & Forsberg, 2019). Both /i:/ and /y:/ produce similar formant values, with /y:/ having slightly lower formants across F1 to F4, presumably as a result of lip-rounding. Auditorily, /i:/ and /y:/ also tended to be very similar. Although the formant values for /i:/ overlap somewhat with /ɥ:/ in F1 and F2, the higher formant values are different, suggesting that these vowels would not overlap perceptually. Indeed, I did not perceive /ɥ:/ to be Viby-coloured, but believe that its low F2 is primarily caused by lip-rounding. The acoustic overlap between different vowels will be investigated further in Section 6.5.

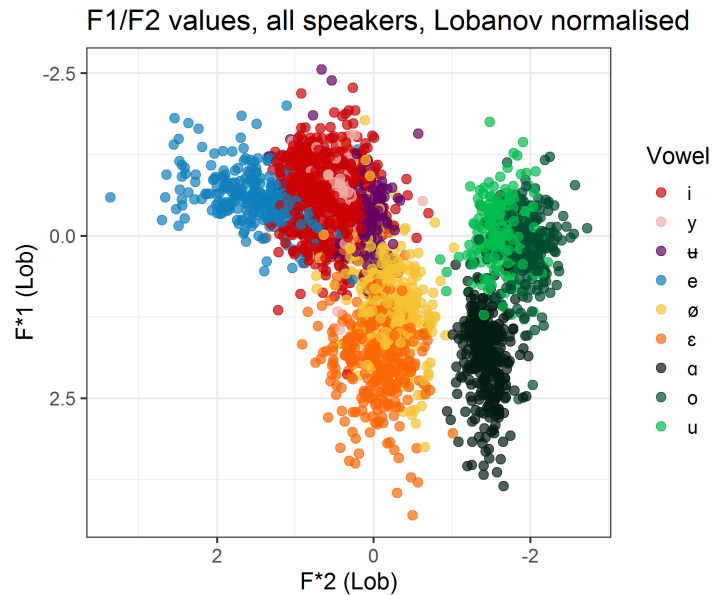


Figure 6.2: All speakers' vowels on the F1/F2 plane (Lobanov normalised).

The centralised position of /i:/ in the acoustic vowel space, and the overlap with /e:/ in F1, raises the question of whether /i:/ has shifted down and back, or whether /e:/ has shifted up and forward. To investigate this, historical formant values for Standard Swedish /i:/ and /e:/ were collated from Eklund and Traunmüller (1997); Engwall and Badin (1999); Ericsson (2005); Fant (1959, 1972); Fant, Henningsson, and Stålhammar (1969); Kuronen (2000); Malmberg (1971); Nord (1986); Stålhammar et al. (1973); Traunmüller and Öhrström (2007). An F1/F2 plot of these values (divided by gender) is presented in Fig. 6.3, with the studies listed according to year of publication. A full table of these formant values is provided in Appendix E.

Vowel	Female				Male			
	F1	F2	F3	F4	F1	F2	F3	F4
/i:/	398	1,944	3,209	4,387	337	1,709	2,742	3,664
/y:/	393	1,890	3,075	3,977	332	1,662	2,576	3,365
/ʊ:/	403	1,854	2,772	3,939	359	1,598	2,338	3,266
/e:/	397	2,369	2,945	4,125	341	1,994	2,512	3,409
/ø:/	513	1,657	2,627	3,944	453	1,385	2,278	3,297
/ɛ:/	591	1,694	2,762	4,121	517	1,470	2,373	3,419
/ɑ:/	586	1,059	2,815	3,908	520	913	2,513	3,311
/o:/	437	811	2,973	3,994	397	748	2,588	3,307
/u:/	426	932	2,910	3,984	383	856	2,474	3,289

Table 6.1: Mean F1 to F4 values (Hz) for all nine vowels across the sample, divided by gender.

Mean F1/F2 values for Swedish /i:/ and /e:/ in previous studies

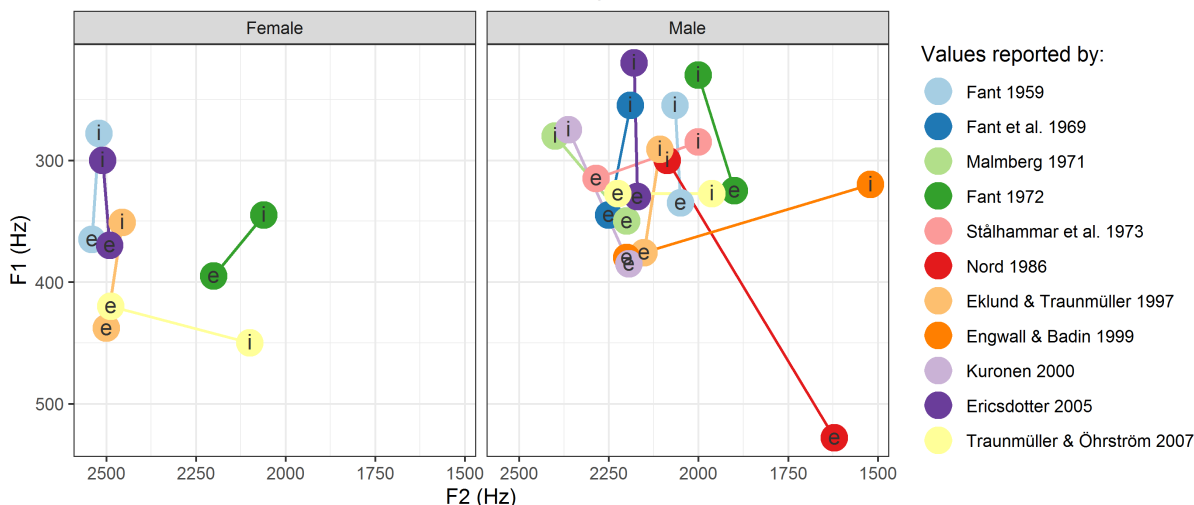


Figure 6.3: Historical F1/F2 values (Hz) for standard /i:/ and /e:/ from the literature.

As the figure shows, there is a great deal of variation in the relationship between /i:/ and /e:/ in these studies: In most cases, /i:/ is higher than /e:/, but it is not always fronter (e.g. Ericsdotter, 2005; Fant, 1959). In fact, relatively few of these tokens display a relationship similar to that on the Swedish vowel quadrilateral (Fig. 6.4). In addition, one study appears to show signs of Viby-colouring (Traunmüller & Öhrström, 2007). As a general rule, however, F1 for /e:/ in the historical data ranges roughly between 350-450 Hz for women, and 300-400 Hz for men. F2 ranges between 2,200-2,600 Hz for women, and 1,900-2,300 for men. These /e:/ values are comparable to those in the present study. Thus, the acoustic position of /e:/ does not appear to have changed in real time. On the other hand, there may be a real-time change in the acoustic position of /i:/ based on the historical data, whereby more recent studies seem to be more likely to acoustically lower and/or retract this vowel. However, this pattern is not fully consistent, and further research is needed to investigate whether Swedish /i:/ has shifted down in recent decades.

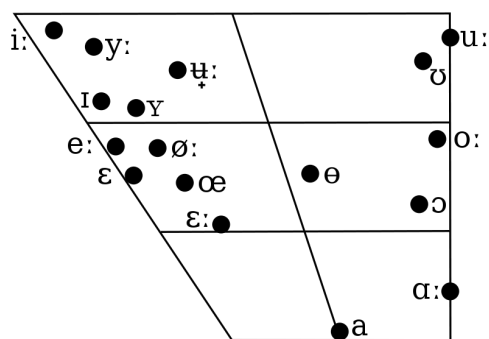


Figure 6.4: Position of Central Standard Swedish vowels on the vowel quadrilateral (Engstrand, 1999: 140).

Although most vowels in the historical data do not meet the criteria of a Viby-i as defined in this thesis (an /i:/ vowel with a lower F2 than /e:/, and a similar F1 to /e:/), an acoustically high front /i:/ is still relatively rare, despite the data representing Central Standard Swedish. It appears that some degree of F2-lowering has been occurring since at least the 1960s (Fant et al., 1969). There is also an interesting gender difference in the Fant (1972) set, where the female speakers have a lowered F2, while the males do not. Following sociolinguistic theories about change in progress (e.g. Eckert, 1989), this gender divide could indicate that women were leading a vowel change for /i:/ during this period.

To summarise this section, /i:/ is acoustically centralised in the present data, with a low F2, and a similar F1 to /e:/. This acoustic profile meets my working definition of Viby-i. Although the lowering of F2 is visually striking in the acoustic vowel space, we can also infer based on historical data that /i:/ also has a relatively high F1 compared to a standard [i:].

## 6.4 Individual speakers' use of Viby-i

Although Viby-i appears to be prevalent in the sample overall, the question remains of whether the individual speakers all use this vowel, or whether some of them use standard [i:]. To investigate this, the mean formant values for /i:/ and /e:/ were compared for each individual speaker (see Table 6.2, full vowel set available in Appendix F).

As the table shows, /i:/ and /e:/ have similar F1 values in most speakers, but in roughly half the speakers, /i:/ has a higher F1 than /e:/, while in the other half, /i:/ is lower. The mean difference in either direction is around 18 Hz, with a maximum of around 50 Hz. In F2, there is more uniformity. In all speakers but one (UM1), F2 is lower for /i:/ than for /e:/. The mean difference across the sample is around 360 Hz, but in one speaker (GF5), the difference is as great as 730 Hz. Thus, we can conclude that nearly all speakers in the sample use Viby-i in one form or another.

For UM1, /i:/ and /e:/ are virtually the same in F2, with /i:/ being only 2 Hz higher than /e:/. However, this speaker still has a relatively low F1, meaning that the vowel is unlikely

to have the auditory vowel quality of a standard [i:]. Although this speaker does not meet the working definition of Viby-i used in this thesis, there is nevertheless a considerable displacement of his /i:/ vowel from an acoustically high front position, and his production can therefore be considered a borderline case.

Speaker	F1 /i:/	F1 /e:/	F1 /i:/ diff.	F2 /i:/	F2 /e:/	F2 /i:/ diff.
GF1	427	407	20	1,756	2,323	-568
GF2	460	410	50	1,825	2,415	-590
GF3	392	365	27	1,899	2,488	-588
GF4	393	388	5	1,783	2,279	-496
GF5	399	398	1	1,932	2,667	-735
GF6	369	372	-3	1,997	2,268	-271
GM1	321	332	-11	1,683	2,023	-339
GM2	365	400	-35	1,841	2,064	-224
GM3	396	377	19	1,592	2,045	-453
GM4	380	351	29	1,562	2,007	-445
GM5	319	304	15	1,793	1,925	-132
GM6	313	311	1	1,682	2,055	-373
SF1	431	392	39	1,892	2,552	-661
SF2	397	410	-12	1,965	2,358	-393
SF3	373	383	-10	2,115	2,621	-506
SF4	368	407	-38	1,975	2,517	-542
SF5	406	399	8	2,076	2,587	-512
SF6	369	403	-34	2,050	2,140	-90
SM1	333	349	-16	1,660	1,917	-256
SM2	296	321	-25	1,670	1,910	-240
SM3	324	347	-23	1,644	1,784	-141
SM4	359	341	19	1,536	1,863	-327
SM5	356	375	-20	1,753	1,877	-124
SM6	324	305	19	1,758	2,224	-466
UF1	408	412	-4	1,896	1,929	-33
UF2	413	414	-1	1,732	2,316	-584
UF3	442	438	4	2,081	2,628	-548
UF4	401	425	-24	1,952	2,106	-154
UF5	363	374	-11	2,022	2,258	-236
UF6	356	359	-3	2,059	2,179	-120
UM1	312	340	-28	1,915	1,913	2
UM2	318	297	21	1,531	2,117	-586
UM3	362	357	5	1,708	1,985	-277
UM4	320	341	-20	2,018	2,201	-183

Table 6.2: Individual mean F1 and F2 values for /i:/ and /e:/ (Hz), with calculated difference.

## 6.5 Overlap between /i:/ and other vowels

As the previous two sections have shown, there is some overlap in the F1/F2 plane between /i:/ and the other vowels in the acoustic high front space /e:/, y:/, ʉ:/. Since Viby-coloured /i:/ and /y:/ in particular have been reported to be affected by perceptual ambiguity (Gross & Forsberg, 2019), this section investigates whether perceptual overlap is likely to occur between /i:/ and other vowels, using statistical modelling of the raw formant distributions (F1 to F4), as well as perceptually scaled acoustic data.

### 6.5.1 Formant distributions

Although F1 and F2 are often said to be the most important factors for the perception of different vowel sounds (Ladefoged, 2001: 39), higher formants also contribute to overall vowel quality. In Swedish, higher formants may be particularly informative, as Swedish has several degrees of lip-rounding, which is associated with a lowering of all formants, but particularly F3 (Ladefoged, 2001: 46). Thus, an overlap in the F1/F2 space does not necessarily equate to an overlap in perception, as higher formants can provide disambiguating cues. This section tests whether each vowel in the sample can be said to have a unique formant ‘profile’ with regard to the first four formants, and uses the results to infer whether perceptual overlap is likely to occur.

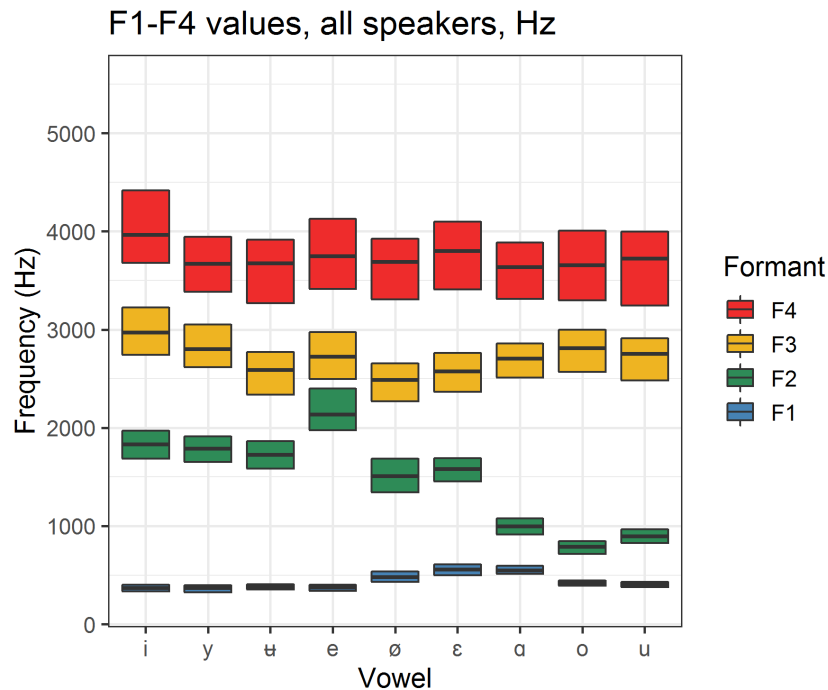


Figure 6.5: F1 to F4 distributions (Hz) for each vowel, based on all speakers.

The F1 to F4 distributions for all vowels and speakers are shown in Fig. 6.5. The values are not normalised, as it is not customary to normalise F3 and F4 (Watt, Fabricius, & Kendall, 2011). Visually, the F1 values for /i:/, e:/, y:/, ʉ:/ are very similar, and there is also overlap between /i:/,

	<i>Dependent variable:</i>			
	f1_hz	f2_hz	f3_hz	f4_hz
	(1)	(2)	(3)	(4)
vowel-y	−5.017 (7.859)	−51.183 (36.009)	−148.716*** (34.529)	−357.424*** (38.429)
vowel-u	12.310 (7.859)	−101.707** (36.009)	−422.245*** (34.529)	−426.691*** (38.429)
vowel-e	1.001 (7.865)	358.406*** (36.022)	−247.796*** (34.550)	−258.637*** (38.460)
vowel-ø	115.220*** (7.864)	−305.431*** (36.020)	−527.554*** (34.547)	−409.878*** (38.455)
vowel-ɛ	182.408*** (7.755)	−209.988*** (35.087)	−404.226*** (33.980)	−256.055*** (37.962)
vowel-a	185.437*** (7.866)	−843.469*** (36.025)	−317.598*** (34.554)	−420.266*** (38.465)
vowel-o	52.198*** (7.576)	−1,059.330*** (33.636)	−198.791*** (33.055)	−373.706*** (37.148)
vowel-u	36.771*** (7.698)	−965.000*** (34.505)	−282.523*** (33.660)	−385.405*** (37.716)
Constant	369.568*** (7.765)	1,833.805*** (26.699)	2,988.929*** (46.131)	4,046.776*** (66.682)
Observations	4,264	4,264	4,264	4,264

*Note:*

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

Table 6.3: LMER output showing the difference in distribution between different vowels (F1 to F4).

y:, ʉ:/ in F2. The F2 for /e:/ is higher than for the other high front vowels, indicating that it has a greater difference in vowel quality. For /i:, y:, ʉ:/, there is a ‘stepladder’ effect in F2, F3, and F4, which is likely to be the result of increasing lip-rounding from /i:/ to /ʉ:/. Viewed as a whole, it is thus clear that some differences are likely to exist in the overall formant profile of each vowel.

To test whether these formant distributions were statistically different from each other, four mixed-effects models were run (one for each formant), specifying F1 to F4 as dependent variables, ‘vowel’ as the independent variable, and ‘speaker’ and ‘word’ as random intercepts. A summary of these models is provided in Table 6.3, with /i:/ as the baseline variable.

The models show that the overall formant distributions for /i:, e:, y:, ʉ:/ are indeed different: There is no significant difference between /i:/ and the other high front vowels in F1, nor between /i:/ and /y:/ in F2, but in all other instances, there is a significant difference in the distributions of F2, F3 and F4 ( $p < 0.01$ ). The acoustic difference between /i:/ and /e:, y:, ʉ:/ thus appears to be robust: The effect size ranges between 100–430 Hz (in the same directions as in Fig. 6.5), and the significance levels are high. This result shows that each vowel has a distinct acoustic formant

profile, and that they are likely to be auditorily distinct from /i:/.

With regard to the documented perceptual overlap between /i:/ and /y:/ (Gross & Forsberg, 2019), the statistical models show that these two vowels do overlap to a greater extent than any of the others in the sample, as they are the only pair that have equivalent F1 *and* F2 values. Since the present study does not provide perceptual data, it cannot determine to what extent these formant patterns affect the perceived similarity between /i:/ and /y:/. However, based on the first two formants, these two vowels are likely to have greater similarities in vowel quality than the others in the set.

### 6.5.2 Bark scaling

A second perspective that might be useful when evaluating how these vowels may be perceived by listeners is to use a perceptual scaling method. This section uses the Bark Difference Metric (Syrdal & Gopal, 1986) to provide a (normalised) image of the acoustic vowel space that more closely resembles human hearing.

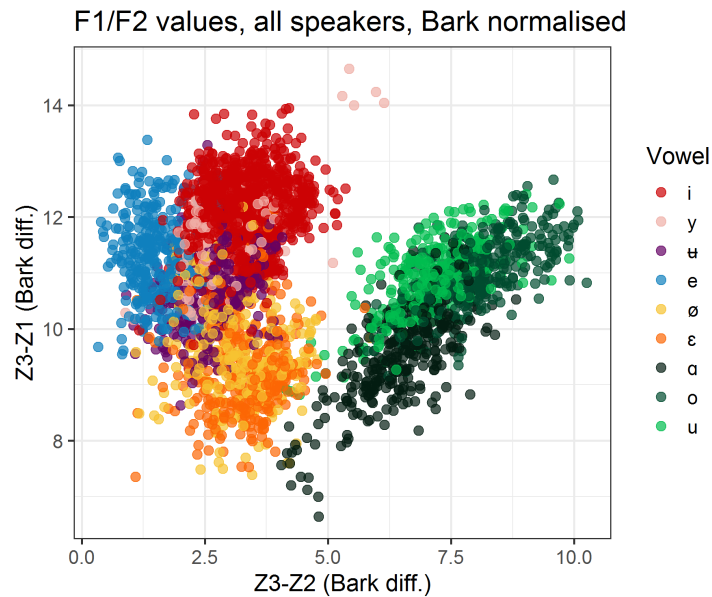


Figure 6.6: All speakers' vowels on the F1/F2 plane (Bark Difference Metric).

The full vowel set for all speakers in the Bark-normalised F1/F2 space is shown in Fig. 6.6. Even after the Bark transformation, there is still considerable overlap between many vowels in the system, particularly /i:, y:, ʏ:/. Nevertheless, the vowels seem to cluster in somewhat separate positions, which could indicate that their distributions are sufficiently different for them to be perceptually distinct. To test this, two mixed effects models, similar to the ones in the previous section, were fitted to the Bark-scaled data, as shown in Table 6.4.

These models show a similar picture to the Lobanov normalised data, i.e. that /e:/ and /ʌ:/ have significantly different distributions from /i:/ in both the 'height' and 'backness' dimension ( $p < 0.01$ ). The difference between /i:/ and /y:/ is more robust in the Bark-scaled data, in



	<i>Dependent variable:</i>	
	Z3.Z2	Z3.Z1
	(1)	(2)
vowel-y	-0.165 (0.096)	-0.298** (0.107)
vowel-u	-0.605*** (0.096)	-1.112*** (0.107)
vowel-e	-1.733*** (0.096)	-0.580*** (0.107)
vowel-ø	-0.028 (0.096)	-2.309*** (0.107)
vowel-ɛ	-0.032 (0.095)	-2.565*** (0.105)
vowel-a	3.220*** (0.096)	-2.378*** (0.107)
vowel-o	4.864*** (0.094)	-0.939*** (0.102)
vowel-u	3.937*** (0.095)	-0.996*** (0.104)
Constant	3.226*** (0.066)	11.907*** (0.094)
Observations	4,264	4,264
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001		

Table 6.4: LMER output showing the difference in distribution of different vowels (Bark).

that these vowels now only overlap in ‘height’, while their ‘backness’ is statistically different ( $p < 0.01$ ). However, the effect size is only around -0.3 units, which is very small in the scale of the entire vowel space. In order to reliably determine if listeners can perceive these differences, further perceptual work would need to be carried out.

In summary, this section has shown that Swedish vowels experience a great degree of crowding in the high front part of the acoustic vowel space, but that listeners are unlikely to experience perceptual ambiguity as a result of this, except possibly in the case of /i:/ and /y:/. Nevertheless, the effort required to maintain these contrasts may encourage shifts within the vowel system, or the use of additional cues to help listeners separate these vowels.

## 6.6 Relationship between F1 and F2 within Viby-i

Having established that virtually all speakers in the sample used Viby-i, this section investigates whether speakers who have a low F2 are also more likely to have a high F1, as both of these acoustic characteristics are typically associated with Viby-i in the literature (e.g. Björsten & Engstrand, 1999).

The distribution of Lobanov normalised F1 against F2 across the whole speaker sample is shown in Fig. 6.7. Based on this figure, there is no clear relationship between F1 and F2; instead, as previous results suggest, F2 is fairly variable, while F1 has a narrower range of values. Looking at the area where the values overlap the most, however, there could be a downward trend, whereby higher F1 values would be associated with slightly lower F2 values. To test this relationship statistically, a linear mixed-effects model (LMER) was run, using normalised F2 as the dependent variable, and normalised F1 as the fixed independent variable. ‘Speaker’ and ‘word’ were entered as random intercepts. The model output is shown in Table 6.5.

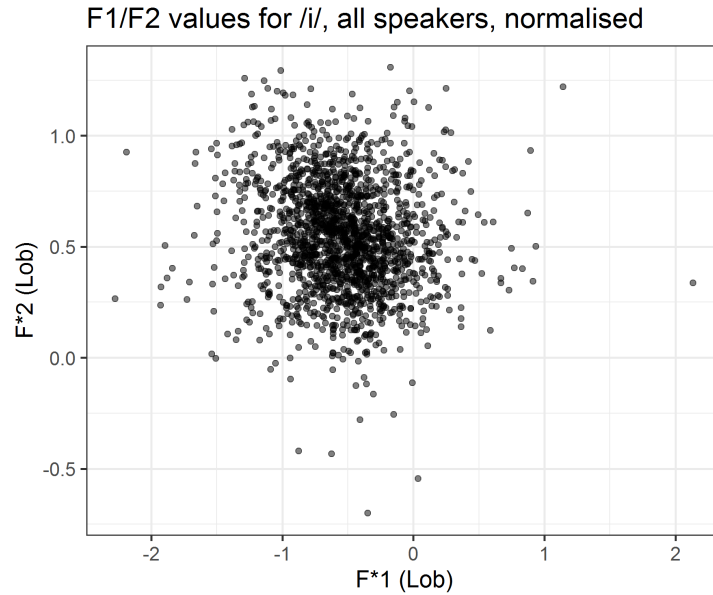


Figure 6.7: Distribution of normalised F1 against normalised F2 for /i:/ for all speakers.

The model shows that, within this data set, there is a tendency for F2 to decrease as F1 increases, as suggested above. However, this trend is very small (0.016 units on the normalised scale, which, plotted on the scale of Fig. 6.2, would barely be visible), and the effect is not statistically significant ( $p = 0.15$ ). We can thus conclude that, within this sample, a lower F2 for /i:/ is not usually accompanied by a higher F1. This result supports the finding in my previous work (Westerberg, 2016) that Viby-i mainly appears to be characterised by a low F2, while F1 contributes less to the acoustic profile of this vowel. In other words, Viby-i can be produced with either a low or a high F1, but it seems to be the low F2 that defines this vowel acoustically.

	<i>Dependent variable:</i>
	f2_lob
f1_lob	−0.016 (0.011)
Constant	0.539*** (0.038)
Observations	1,766
<i>Note:</i>	*p<0.05; **p<0.01; ***p<0.001

Table 6.5: LMER output showing the relationship between normalised F1 and normalised F2.

## 6.7 Linguistic effects on Viby-i formants

This section investigates the effects of linguistic factors on /i:/ in terms of normalised F1 and F2. As before, the formant values were taken at the 10% timepoint. The factors included in the analysis are word frequency, vowel duration, preceding consonant frication, preceding consonant voicing, following consonant backness, and following consonant complexity, as outlined in Section 3.5. I will begin by testing these factors for multicollinearity and removing those that are correlated. I will then go through the results of the final LMER models, examining whether any of the linguistic factors have a significant relationship with F1 or F2.

### 6.7.1 Variable selection

For context, Table 6.6 provides a summary of the word list items, and how they were coded with regard to the factors used in the analysis. The table is roughly divided into two separate lists (plosive and fricative), which cannot be entered into the same model, since the fricative set is biased towards front, complex, and voiced consonant environments. The fricative set also lacks word-final and non-lingual tokens. These two parts of the word list will therefore be tested separately.

To test whether any of the independent variables were correlated with each other, a correlation matrix was generated (using the Pearson method), as illustrated in Fig. 6.8. Note that the matrix uses binary dummy variables for categorical variables; thus, if a factor is correlated to e.g. ‘following consonant backness’, it means that it is also correlated to ‘following consonant frontness’, since these two variables represent two sides of the same coin.

The figure confirms that ‘preceding consonant frication’ is correlated with both ‘following consonant complexity’ and ‘following consonant frontness’ ( $p < 0.05$ ). It also shows the correlation between the first two of these factors and ‘word frequency’ ( $p < 0.05$ ). Thus, in addition to testing the two parts of the word list separately, ‘word frequency’ also had to be excluded from the analysis.

Following consonant		Preceding consonant			
		Plosive		Fricative	
Complexity	Backness	Voiced	Voiceless	Voiced	Voiceless
Simple	Null	BI	PI		
	Neutral	BIBEL	PIPA		
	Front	BITA	PITA	VITA	
	Back	BIGA	PIGA	VIKA	FIKA
Complex	Front			VILA	FILA
	Front			VIRA	FIRA
	Front			VISA	
	Front			VINA	

Table 6.6: Summary of word list items and their assigned categories.

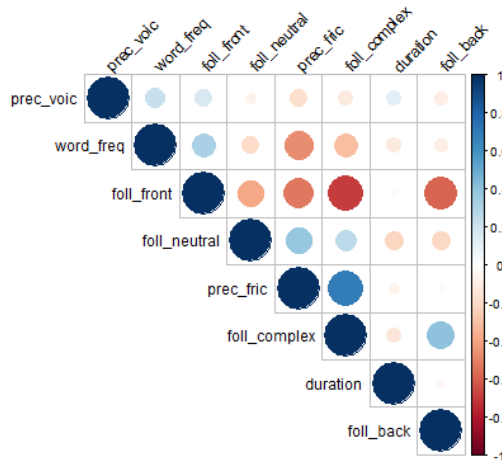


Figure 6.8: Correlation matrix of linguistic factors.

Three sets of models were thus constructed for the analysis: One testing the effect of following consonant backness in the plosive set (e.g. BITA vs. BIGA), one testing for the effect of following complexity in the fricative set (e.g. VITA vs. VIRA), and one comparing the plosive set to the fricative set (e.g. BITA vs. VITA). Vowel duration and preceding consonant voicing (e.g. PITA vs. BITA; FIKA vs. VIKA) were also entered into the first two models. All models used ‘speaker’ as a random intercept.

### 6.7.2 Plosive set: Duration, preceding voicing, following backness

The first set of models used the plosive word list set to test the effects of duration, preceding consonant voicing, and following consonant backness on normalised F1 and F2. Since following consonant backness had four levels (front, back, neutral, and null), Table 6.7 presents the results for the baseline level (front) first, with the different combinations of re-levelled results at the bottom of the table. Visualisations of the significant results are provided in Figs. 6.9 and 6.10.

	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
vowel_dur_ms	0.001 (0.0003)	0.0004* (0.0002)
prec_voic-yes	-0.008 (0.028)	-0.143*** (0.013)
foll_place-front/back	-0.033 (0.039)	0.062*** (0.018)
foll_place-front/neutral	0.008 (0.037)	0.031 (0.017)
foll_place-front/null	0.114* (0.049)	0.075** (0.023)
Constant	-0.714*** (0.072)	0.600*** (0.041)
foll_place-back/ntrl	0.041 (0.039)	-0.031 (0.018)
foll_place-back/null	0.147** (0.043)	0.013 (0.020)
foll_place-null/ntrl	-0.106* (0.050)	-0.044 (0.024)
Observations	829	829
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001		

Table 6.7: LMER output showing the effect of vowel duration, preceding consonant voicing, and following consonant backness on normalised F1 and F2 within the plosive set.

The results show that, in the plosive word list set:

- F2 increases with vowel duration ( $\beta = 0.0004$ ,  $p = 0.02$ ). Although this effect is statistically significant, the change in F2 is extremely small.
- F2 is lower when the preceding consonant is voiced ( $\beta = -0.14$ ,  $p < 0.001$ ), meaning that words like BITA have a ‘stronger’ Viby-i than words like PITA.
- F1 is higher in word-final contexts than in words like BITA, BIGA, or BIBEL ( $\beta = 0.11$ ,  $p = 0.02$ ;  $\beta = 0.15$ ,  $p = 0.001$ ;  $\beta = 0.11$ ,  $p = 0.04$ ).

- F2 is higher in front contexts like BITA compared to back or null contexts like BIGA and BI ( $\beta = 0.06$ ,  $p < 0.001$ ;  $\beta = 0.08$ ,  $p = 0.002$ ). Back and null contexts group together. Neutral contexts like BIBEL are intermediate, and not significantly different from any of the other groups.

If we interpret F2-lowering as an indication of Viby-i strength, these results show that Viby-i is stronger when vowel duration is short, when the preceding consonant is voiced, and when the following consonant is velar, or when the vowel occurs word-finally. To return to the predictions made in Section 3.5, Viby-i does not seem to need a long vowel duration to reach its target; rather, the reducing effects of a shorter duration may work in this vowel's favour. Preceding voicing (lack of aspiration) favours F2-lowering; if this context disfavors frication, there may be a complementary relationship between the two. The fact that Viby-i is stronger before a back consonant indicates that this sound may use a backed articulation. If this is the case, the vowel is able to achieve this articulation equally well in word-final positions.

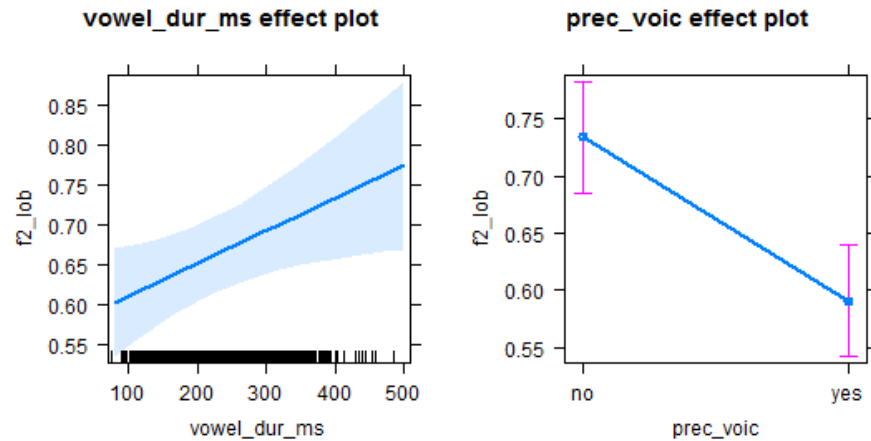


Figure 6.9: Significant effects of duration and preceding consonant voicing on normalised F2 in the plosive set.

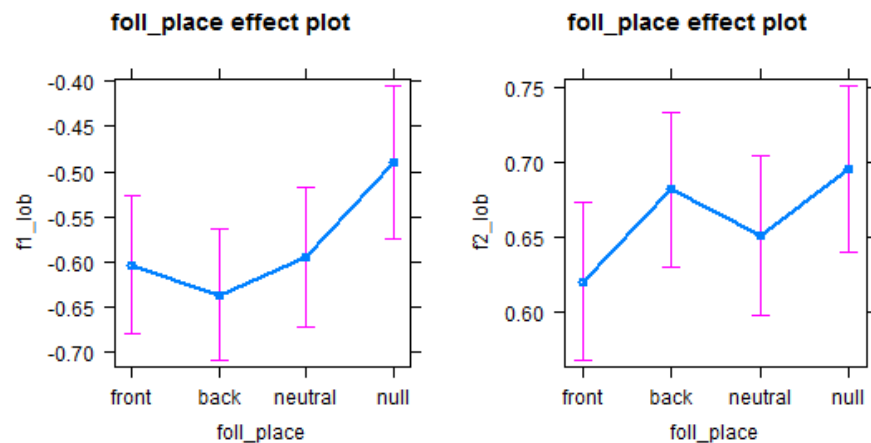


Figure 6.10: Significant effects of following consonant place on normalised F1 and F2 in the plosive set.

### 6.7.3 Fricative set: Duration, preceding voicing, following complexity

The second set of models used the fricative word list set to test the effects of duration, preceding consonant voicing, and following consonant complexity on normalised F1 and F2. The models are provided in Table 6.8, and significant results are shown in Figs. 6.11 and 6.12.

	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
vowel_dur_ms	0.001* (0.0003)	0.0001 (0.0001)
prec_voic-yes	-0.215*** (0.025)	-0.067*** (0.012)
foll_manner-complex	0.00004 (0.025)	-0.035** (0.013)
Constant	-0.500*** (0.063)	0.501*** (0.040)
Observations	937	937
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001		

Table 6.8: LMER output showing the effect of vowel duration, preceding consonant voicing, and following consonant manner on normalised F1 and F2 within the fricative set.

The results show that, in the /f, v/ word list set:

- F1 increases with vowel duration, but again, the coefficient is very small ( $\beta = 0.001$ ,  $p = 0.03$ ).
- Both F1 and F2 are lower when the preceding consonant is voiced ( $\beta = -0.215$ ,  $p < 0.001$ ;  $\beta = -0.067$ ,  $p < 0.001$ ), meaning words like VIKa have a ‘stronger’ Viby-i than words like FIKA. However, this lowering effect is relatively small in F2 compared to F1.
- F2 is lower when the following consonant is complex, meaning that Viby-i is ‘stronger’ in words like VILa, VIRa, VISa, VINa than in words like VITa, VIKa ( $\beta = -0.035$ ,  $p = 0.007$ ).

In terms of F2, Viby-i in the fricative set behaves the same way as in the plosive set, in that it is stronger when the previous consonant is voiced. The fact that F1 also follows this pattern in the fricative set could indicate that voicing affects the formants more generally, rather than applying specifically to Viby-i. The finding that Viby-i is stronger before complex consonants suggests that this sound may use a complex secondary articulation which is similar to one or several of these sounds.

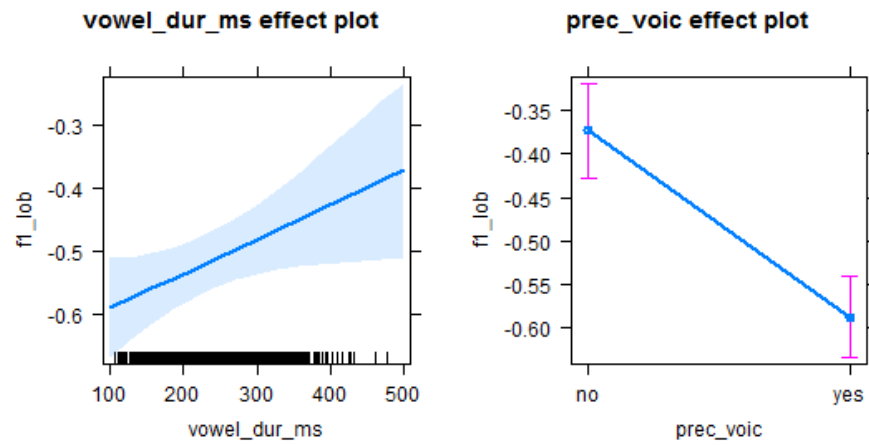


Figure 6.11: Significant effects of vowel duration and preceding consonant voicing on normalised F1 in the fricative set.

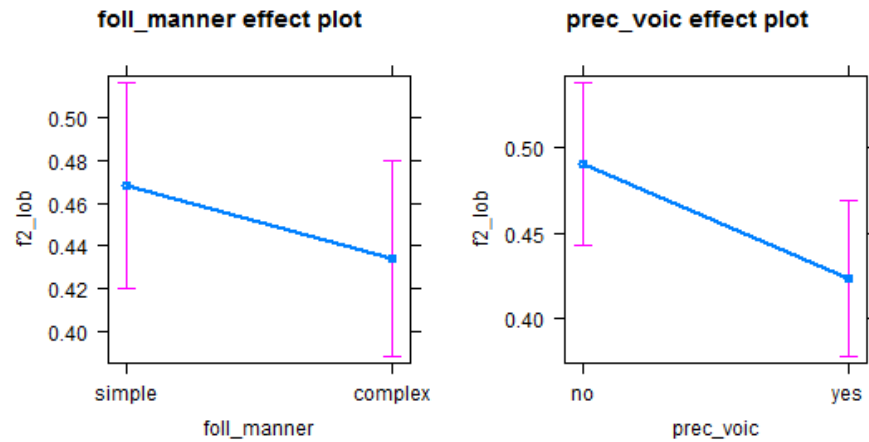


Figure 6.12: Significant effects of preceding consonant voicing and following consonant manner on normalised F2 in the fricative set.

#### 6.7.4 Full word list set: Preceding manner

The third and final model compares the fricative and plosive word list sets, to investigate the effect of preceding manner on normalised F1 and F2. The models are provided in Table 6.9, and the significant results are shown in Fig. 6.13.

The results show that the two word list sets differ significantly in both F1 and F2, with the fricative set having a higher F1, and a lower F2, than the plosive set ( $\beta = -0.064$ ,  $p < 0.001$ ;  $\beta = -0.217$ ,  $p < 0.001$ ). The difference is greater in F2 than in F1. Based on these results, words like VITA have a stronger Viby-i than words like BITA. Thus, the consonant context that was predicted to have more frication also has more F2-lowering. If it is the case that vowels preceded by /f, v/ are more fricated than vowels preceded by /p, b/, this result would contradict the suggestion that F2-lowering and vowel frication are complementary for Viby-i.



	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
prec_manner-plos	−0.064*** (0.018)	0.217*** (0.009)
Constant	−0.516*** (0.024)	0.446*** (0.022)
Observations	1,766	1,766

*Note:* \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Table 6.9: LMER output showing the effect of preceding consonant manner on normalised F1 and F2.

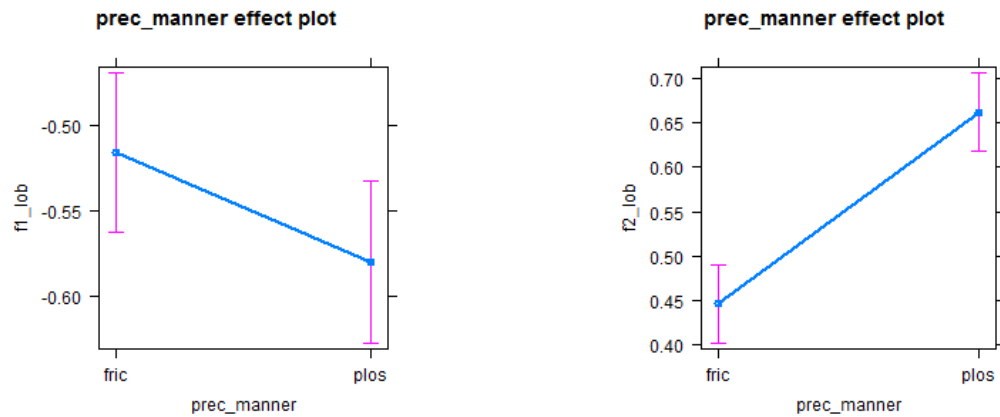


Figure 6.13: Significant effects of preceding consonant manner on normalised F1 and F2.

### 6.7.5 Summary of linguistic results

Based on all three models, we can see that duration contributed to a small but reliable increase in normalised F1 in the fricative word list set, and a similar increase in normalised F2 in the plosive set. To some extent, this refutes the hypothesis that a longer Viby-i would have more exaggerated characteristics because the tongue has more time to reach its target. Instead, it seems like the reduction that results from a short vowel duration may actually favour Viby-i, since it renders it more centralised.

Preceding consonant voicing had a consistent decreasing effect on F2 in both word list sets. It also increased F1 in the fricative set. No hypothesis was formed for this variable, as it was primarily intended to measure the relationship between pre-aspiration and intra-vowel frication, which will be discussed in Section 6.9. However, since voicing affected both F1 and F2 in the same way, it does not appear to specifically condition Viby-i, but may have a more general effect on the vowel formants.

Within the plosive set, Viby-i was stronger in velar, non-lingual, and word-final contexts compared to alveolar/dental contexts, indicating that Viby-i may use a backed articulation, which

can primarily be achieved when the following tongue gesture is ‘not front’. Within the fricative set, *Viby-i* was stronger in complex (liquid, nasal, sibilant) contexts compared to simple (plosive) contexts. This result could indicate that *Viby-i* uses a similar tongue gesture or secondary articulation to one or several of these sounds.

As a visual summary, Table 6.10 ranks the word list items according to their predicted degree of F2-lowering, and Fig. 6.14 similarly ranks F2 for the word list items from highest to lowest. Note that as F2 decreases, F1 continues to fluctuate from word to word. This pattern again demonstrates that a *Viby-i* with a low F2 does not necessarily also have a high F1. Finally, it should be noted that there may be interactions between the factors, which could not be investigated here. For example, in the fricative set, most vowels were front and followed by a complex consonant. In these cases, it is difficult to judge to what extent each factor contributes to F2 lowering.

← Weaker <i>Viby-i</i> (higher F2)				Stronger <i>Viby-i</i> (lower F2) →			
Plosive				Fricative			
Voiceless		Voiced		Voiceless		Voiced	
Null/back	Fr./Neut.	Null/back	Fr./Neut.	Simple	Complex	Simple	Complex
PI, PIGA	PITA, PIPA	BI, BIGA	BITA, BIBEL	FIKA	FILA, FIRA	VITA, VIKA	VILA, VIRA, VISA, VINA

Table 6.10: Summary of linguistic effects on normalised F2, categorised by word list item.

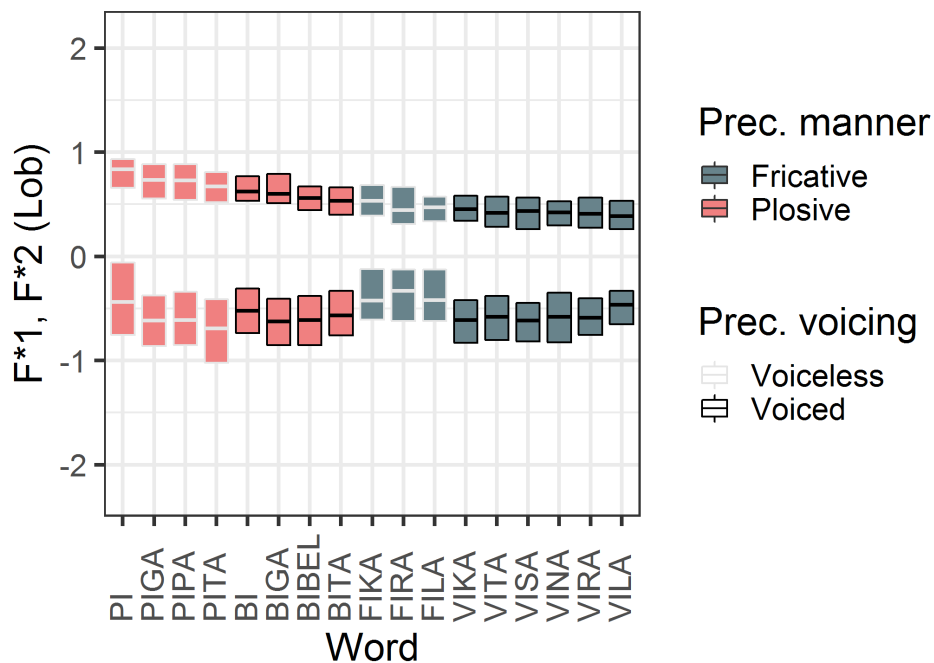


Figure 6.14: Word list items in order of F2, from high to low.

## 6.8 Social effects on Viby-i formants

This section investigates the effects of social factors on /i:/ in terms of normalised F1 and F2. The data is again taken from the first 10% of the vowel duration. The factors included in the analysis are age, gender, city of origin, distance from city centre, social interaction, local rootedness, dialect contact, and prescriptivism. Note that education (a proxy for socioeconomic class) is not included, as this variable was too homogeneous across the sample.

The first part of the analysis will test for multicollinearity and exclude any correlated factors. The second part of the analysis will present the suggested LMER models and their results. The third and final section will carry out an exploratory step analysis to investigate which of the social factors contribute to explaining variation in the data.

### 6.8.1 Variable selection

A correlation matrix was generated, as illustrated in Fig. 6.15. As the figure shows, there was a correlation between ‘local rootedness’ and ‘social interaction’ ( $p < 0.05$ ), and a separate correlation between ‘dialect contact’ and ‘gender’ ( $p < 0.05$ ). Since gender has previously been described as affecting Viby-i (e.g. Kotsinas, 2007), ‘dialect contact’ was dropped from the model. ‘Local rootedness’ was also dropped, as this data was less normally distributed than ‘social interaction’ (see Figs. 6.16).

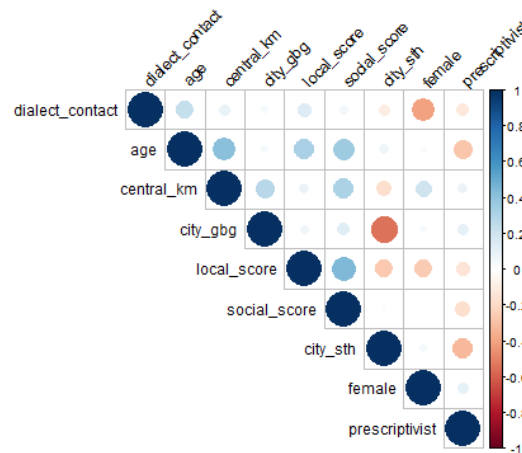


Figure 6.15: Correlation plot of social factors.

The set of models for the sociolinguistic analysis thus tested for the effects of age (years), gender (male, female), city (Gothenburg, Stockholm, Uppsala), distance from city centre (km), social interaction (score from 2-16), and prescriptivism (score from 2-8). The models also included a three-way interaction between age, gender, and city, and an interaction between city and distance from city centre. ‘Speaker’ and ‘word’ were used as random intercepts.

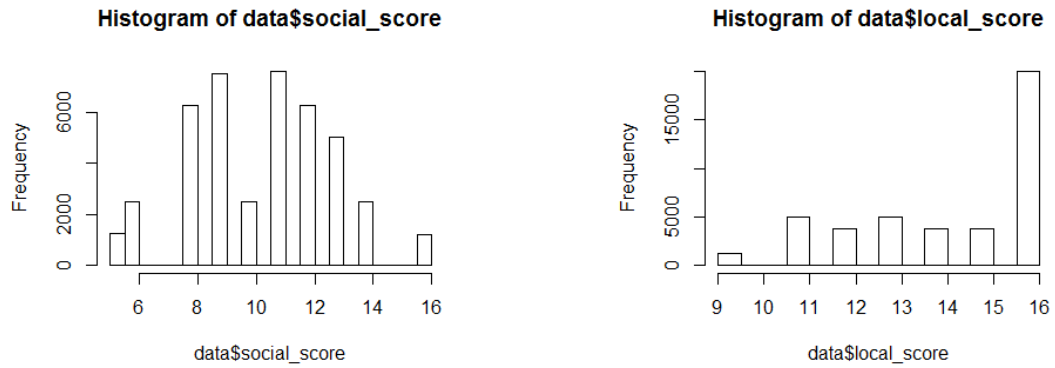


Figure 6.16: Distribution of scores for ‘social interaction’ and ‘local rootedness’ respectively, with an uneven distribution in the latter.

### 6.8.2 Initial model

The output of the social LMER model is presented Tables 6.11, 6.12, and 6.13 below. The three tables represent the three different city levels (Gothenburg, Stockholm, Uppsala), and are listed with the baseline city at the top of each table.

Based on these models, there is a significant effect of age in Gothenburg, whereby older speakers have a slightly higher F1 than younger speakers ( $\beta = 0.007$ ,  $p = 0.04$ ). However, this effect is very small, and not particularly meaningful in relation to Viby-i. The lack of significant results for the other factors could be attributed to unstructured variability in the data, but it could also be due to over-fitting. Given that the sample, from a statistical viewpoint, is quite small, it is possible that entering too many factors that do not improve the model could potentially obscure meaningful patterns. To explore which factors actually contribute to explaining the data, the following section uses step modelling to exclude factors which did not improve the model fit.

### 6.8.3 Step model

The models described above were entered into a ‘step’ function, which tests the explanatory power of each factor one by one, and removes factors that do not contribute to improving the model. The step model for normalised F1 kept the fixed factors ‘city’ and ‘gender’ ( $F = 10.7$ ,  $p < 0.001$ ;  $F = 7.4408$ ,  $p = 0.01$ ), and the random intercepts ‘speaker’ and ‘word’ ( $AIC = 1,778.8$ ,  $p < 0.001$ ;  $AIC = 1,831.2$ ,  $p < 0.001$ ). The step model for normalised F2 kept the fixed factor ‘city’ ( $F = 7.6$ ,  $p < 0.001$ ) and the random intercepts ‘speaker’ and ‘word’ ( $AIC = -475.9$ ,  $p < 0.001$ ;  $AIC = -163.3$ ,  $p < 0.001$ ). The results for these models are shown in Table 6.14, and the significant results are presented in Figs. 6.17 and 6.18. The baseline value was Gothenburg, but re-levelled values are provided at the bottom of the table.

Baseline: Gothenburg	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
city-Stockholm	0.031 (0.213)	0.006 (0.254)
city-Uppsala	0.190 (0.181)	0.253 (0.216)
age	0.007* (0.003)	0.002 (0.004)
gender-Male	0.027 (0.223)	-0.166 (0.267)
central_km	-0.001 (0.001)	-0.001 (0.001)
social_score	0.005 (0.009)	-0.020 (0.011)
prescriptivist	0.003 (0.027)	0.028 (0.032)
city-Stockholm:age	-0.006 (0.004)	0.003 (0.005)
city-Uppsala:age	-0.008 (0.004)	-0.001 (0.005)
city-Stockholm:gender-Male	-0.186 (0.288)	0.378 (0.345)
city-Uppsala:gender-Male	-0.282 (0.390)	0.362 (0.466)
age:gender-Male	-0.003 (0.007)	0.010 (0.008)
city-Stockholm:central_km	0.001 (0.006)	0.006 (0.007)
city-Uppsala:central_km	-0.001 (0.002)	0.001 (0.003)
city-Stockholm:age:gender-Male	0.004 (0.009)	-0.014 (0.010)
city-Uppsala:age:gender-Male	0.008 (0.015)	-0.019 (0.018)
Constant	-0.708*** (0.239)	0.413 (0.286)
Observations	1,766	1,766

*Note:* \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Table 6.11: LMER output showing the effect of age, gender, city, distance from city centre, social interaction, and prescriptivism on normalised F1 and F2 (Gothenburg).

Baseline: Stockholm	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
cityGothenburg	−0.031 (0.213)	−0.006 (0.254)
cityUppsala	0.159 (0.228)	0.247 (0.273)
age	0.001 (0.004)	0.005 (0.004)
genderMale	−0.159 (0.159)	0.212 (0.189)
central_km	0.001 (0.006)	0.004 (0.007)
social_score	0.005 (0.009)	−0.020 (0.011)
prescriptivist	0.003 (0.027)	0.028 (0.032)
cityGothenburg:age	0.006 (0.004)	−0.003 (0.005)
cityUppsala:age	−0.002 (0.005)	−0.004 (0.006)
cityGothenburg:genderMale	0.186 (0.288)	−0.378 (0.345)
cityUppsala:genderMale	−0.096 (0.329)	−0.017 (0.393)
age:genderMale	0.002 (0.004)	−0.004 (0.004)
cityGothenburg:central_km	−0.001 (0.006)	−0.006 (0.007)
cityUppsala:central_km	−0.002 (0.006)	−0.005 (0.008)
cityGothenburg:age:genderMale	−0.004 (0.009)	0.014 (0.010)
cityUppsala:age:genderMale	0.004 (0.012)	−0.006 (0.014)
Constant	−0.677* (0.263)	0.420 (0.314)
Observations	1,766	1,766

*Note:* \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Table 6.12: LMER output showing the effect of age, gender, city, distance from city centre, social interaction, and prescriptivism on normalised F1 and F2 (Stockholm).

Baseline: Uppsala	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
cityStockholm	−0.159 (0.228)	−0.247 (0.273)
cityGothenburg	−0.190 (0.181)	−0.253 (0.216)
age	−0.001 (0.002)	0.001 (0.003)
genderMale	−0.255 (0.294)	0.195 (0.351)
central_km	−0.001 (0.002)	−0.0003 (0.002)
social_score	0.005 (0.009)	−0.020 (0.011)
prescriptivist	0.003 (0.027)	0.028 (0.032)
cityStockholm:age	0.002 (0.005)	0.004 (0.006)
cityGothenburg:age	0.008 (0.004)	0.001 (0.005)
cityStockholm:genderMale	0.096 (0.329)	0.017 (0.393)
cityGothenburg:genderMale	0.282 (0.390)	−0.362 (0.466)
age:genderMale	0.005 (0.011)	−0.010 (0.014)
cityStockholm:central_km	0.002 (0.006)	0.005 (0.008)
cityGothenburg:central_km	0.001 (0.002)	−0.001 (0.003)
cityStockholm:age:genderMale	−0.004 (0.012)	0.006 (0.014)
cityGothenburg:age:genderMale	−0.008 (0.015)	0.019 (0.018)
Constant	−0.518*** (0.147)	0.667*** (0.176)
Observations	1,766	1,766
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001		

Table 6.13: LMER output showing the effect of age, gender, city, distance from city centre, social interaction, and prescriptivism on normalised F1 and F2 (Uppsala).

	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
city-Stockholm	−0.175*** (0.039)	0.134** (0.044)
city-Uppsala	−0.135** (0.042)	0.168*** (0.046)
gender-Male	−0.091** (0.033)	
Constant	−0.402*** (0.040)	0.451*** (0.044)
city-STH/UPP	−0.041 (0.042)	−0.033 (0.046)
Observations	1,766	1,766
<i>Note:</i>	*p<0.05; **p<0.01; ***p<0.001	

Table 6.14: LMER output (from step modelling) showing the effect of city on normalised F1 and F2, and the effect of gender on normalised F1.

In summary, the social analysis shows that:

- ‘City’ explains most of the variation in F1 and F2 for /i:/. Gender also contributes to explaining variation in F1.
- F1 is significantly higher in women than in men ( $\beta=-0.091$ ,  $p=0.007$ ). Note that this difference persists even though the data is normalised.
- F1 is significantly higher in Gothenburg compared to Stockholm and Uppsala ( $\beta=-0.175$ ,  $p<0.001$ ;  $\beta=-0.135$ ,  $p=0.002$ ). Stockholm and Uppsala pattern together.
- F2 is significantly lower in Gothenburg compared to Stockholm and Uppsala ( $\beta=0.134$ ,  $p<0.003$ ;  $\beta=0.168$ ,  $p<0.001$ ). Stockholm and Uppsala pattern together.

If we interpret F2-lowering (and accompanying F1-raising) as an indication of Viby-i strength, these results show that Viby-i has a stronger realisation in Gothenburg than in Stockholm and Uppsala, and that Stockholm and Uppsala produce equivalent versions of this vowel with regard to normalised F1 and F2. Women also have a higher F1 than men, indicating that their Viby-i could be stronger. However, since men and women are not significantly different in F2, this difference is likely to be less audible in terms of Viby-colouring.



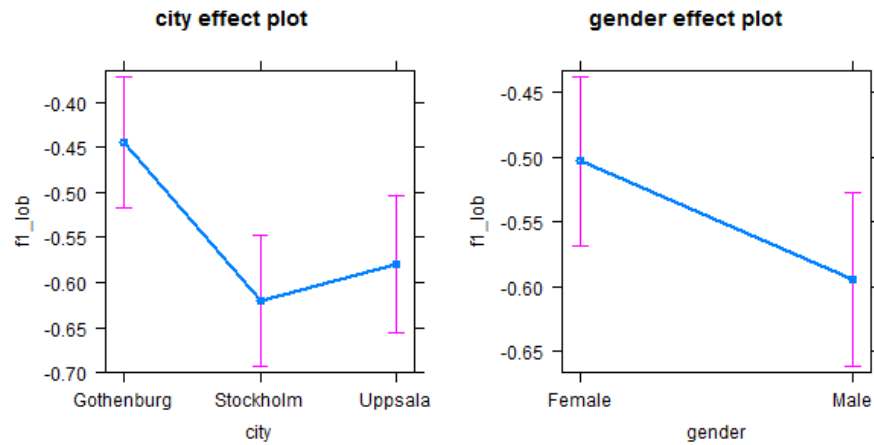


Figure 6.17: Statistical effects on F1 by city and gender.

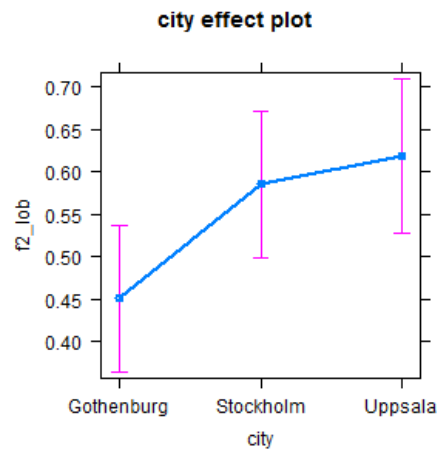


Figure 6.18: Statistical effects on F2 by city.

#### 6.8.4 Summary of social results

To return to the hypotheses made about social factors in Section 5.6, the lack of difference in acoustics between the different social groupings is somewhat surprising, given that there were theoretical justifications to include these factors in the analysis (e.g. Björsten & Engstrand, 1999; Frid et al., 2015; Grönberg, 2004; Gross & Forsberg, 2019; Kotsinas, 2007). The results indicate that *all* speakers in this sample used Vi-by-i, including older age groups, and speakers from Uppsala, who were less expected to use this vowel.

As predicted, Gothenburg has a ‘stronger’ Vi-by-i than Stockholm and Uppsala. This difference could be caused by dialectal differences, or by a more advanced progression towards Vi-by-i in Gothenburg. The fact that Stockholm and Uppsala pattern together indicates that Vi-by-i could have spread to Uppsala from Stockholm.

The difference in F1 by gender could also suggest that women have a stronger Vi-by-i, as reported in the literature (e.g. Björseth, 1958; Kotsinas, 2007), however, this result should be treated as tentative, as previous parts of the analysis found that F2 seems to contribute more to

Viby-colouring than F1 does.

Overall, there seems to be a great deal of unstructured variation in the sample, but it is worth noting that even the systematic variation (e.g. by city) is happening on quite a small scale. One reason why few systematic differences appeared in the sociolinguistic analysis could be that the sample is already fairly uniform in terms of normalised F1 and F2. In other words, it could be the case that the social factors are not informative because speakers are already producing Viby-i more or less consistently across different social groups (see Fig. 6.19). Indeed, the findings of this thesis point towards Viby-i being relatively well-established in all three cities.

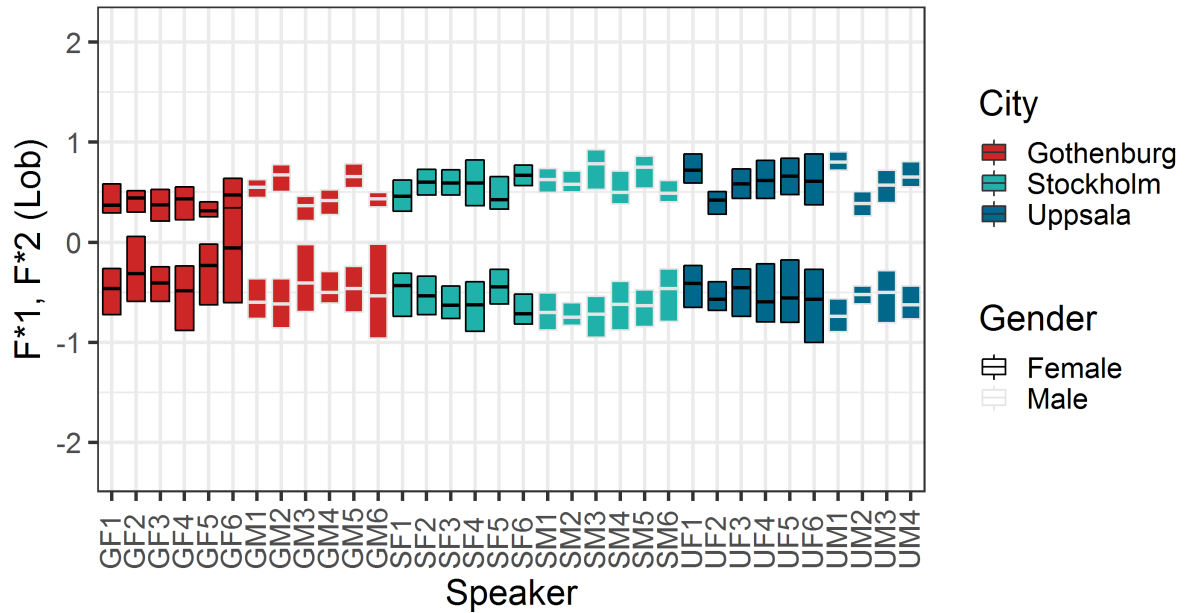


Figure 6.19: Distribution of normalised F1 and F2 for /i:/ by individual speakers.

## 6.9 Viby-i dynamics

Since Swedish vowels have been described as having dynamic properties with relation to both formant movement and fricative noise (Engstrand, 1999: 141), this section investigates whether these dynamic properties also affect Viby-i. The dynamic analysis uses F1, F2, and bpZCR data sampled at every 10% of the vowel duration. Section 6.9.1 begins by providing a visual introduction to some of the patterns observed for Viby-i in the spectrograms, and inspects the dynamic formant measurements. Section 6.9.2 explores the auditory transcriptions of frication during Viby-i, and provides a qualitative analysis of the different frication patterns found in the data. Section 6.9.3 then investigates the relationship between bpZCR and different kinds of frication, focusing on linguistic environments that condition higher frication rates.

### 6.9.1 Spectrograms and formant measurements

During the segmentation process, it became evident that most of the Viby-i tokens were relatively monophthongal, with the exception of the coarticulatory pattern known as a “velar pinch” (e.g. Baker, Mielke, & Archangeli, 2008: 61), whereby F2 increases and F3 decreases before a velar consonant (Fig. 6.20). However, a number of other dynamic patterns were visible in the spectrograms, including many different kinds of non-diphthongal offglides. Truly modal vowels (Fig. 6.21) were relatively rare in the sample. Instead, most speakers showed patterns of formant weakening (Figs. 6.22 and 6.23), high- and low- frequency end-frication (Figs. 6.24 and 6.25), or both occurring together (Figs. 6.26 and 6.27). These phenomena could occur vowel-finally or vowel-medially, as shown below. Overall, formant weakening appeared to coincide with the impression of a /j/-like production.

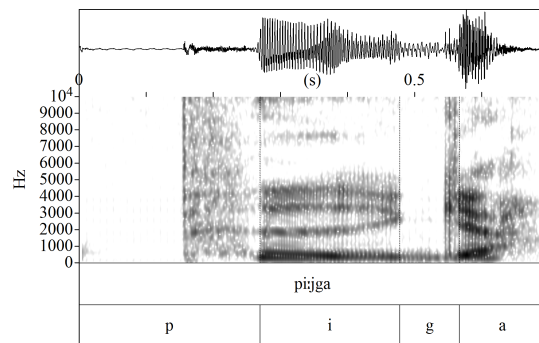


Figure 6.20: Velar pinch during Viby-i in speaker GF1.

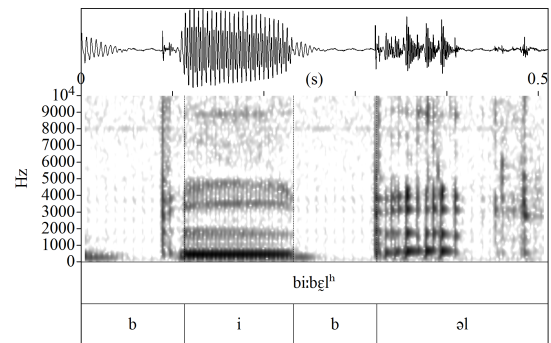


Figure 6.21: Modal Viby-i in speaker UF2.

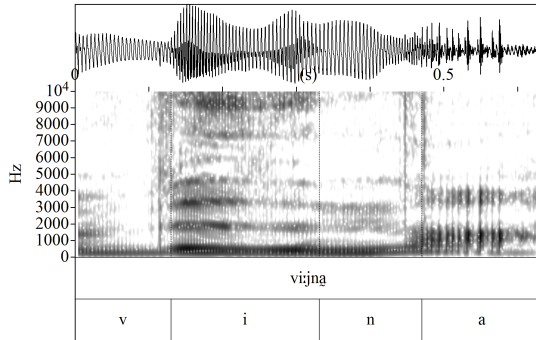


Figure 6.22: Final formant weakening during Viby-i in speaker SF1.

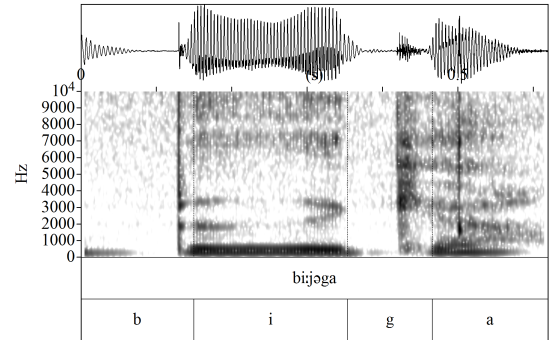


Figure 6.23: Medial formant weakening during Viby-i in speaker GF5.

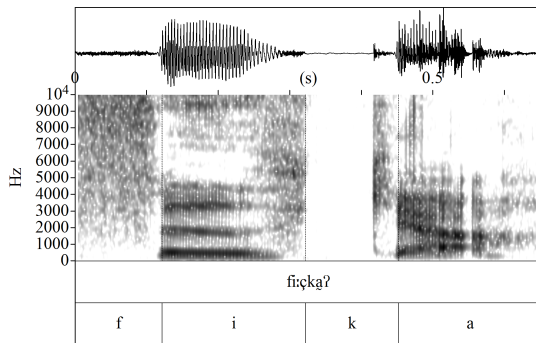


Figure 6.24: Low frequency (velar) end-frication during Viby-i in speaker SF1.

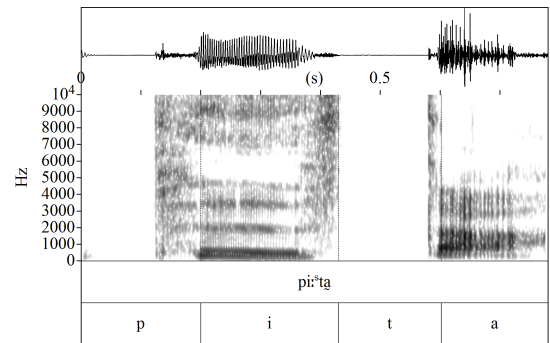


Figure 6.25: High frequency (alveolar) end-frication during Viby-i in speaker SF1.

To investigate whether F1 and F2 are as dynamically stable as they appear from the spectrograms, the mean values from each 10% timepoint across /i:/ are presented in Fig. 6.28. On this figure, the points represent individual formant measurements (F1 dark, F2 light), while the blue lines represent the mean trajectories of these values. In order to capture variation in the original data, the values are presented in Hertz. For this reason, the graph divides the data by gender.

As Fig. 6.28 shows, the mean formant trajectories across Viby-i are relatively stable, with a slight increase in both F1 and F2 towards the end of the vowel, which could be attributed to coarticulation with the following consonant. However, another interesting pattern in this data is the increasing variability of the measurement points around the mean as the vowel progresses. This variability appears to begin around 40-50% of the duration, and continues to increase until the end of the vowel. Thus, while the formant data is clustered relatively close to the mean values in the first 20-30% of the vowel, the range of values in the last 20-30% is roughly twice as large.

Some of this variability can likely be attributed to the fact that this data is taken from a number of different speakers and linguistic contexts. However, based on the spectrograms, this pattern could also indicate that the formant tracker was having increasing difficulties identifying

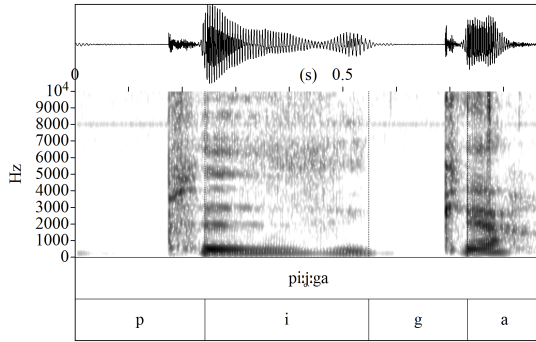


Figure 6.26: Frication and formant weakening during Viby-i in speaker GF4.

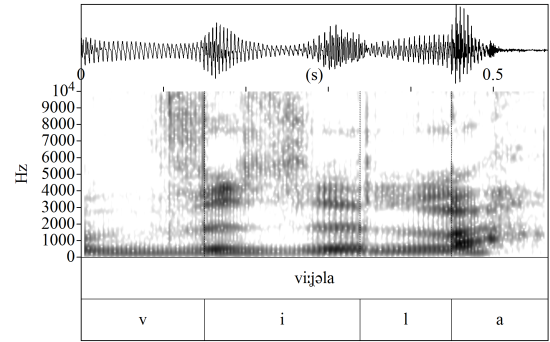


Figure 6.27: Frication and formant weakening during Viby-i in speaker GF1.

### F1 and F2 trajectories across Viby-i

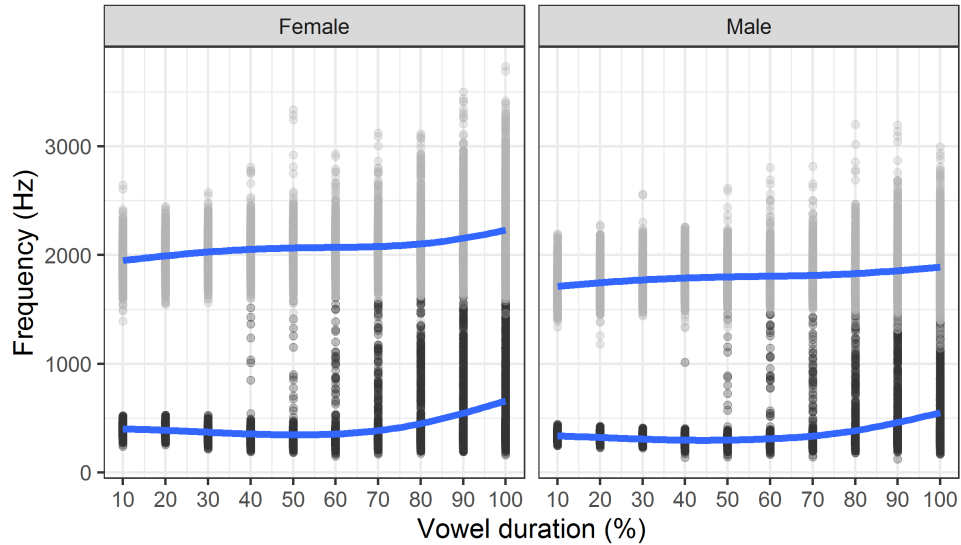


Figure 6.28: Time-normalised formant trajectories (Hz) across Viby-i by gender. F1 dark grey, F2 light grey. Blue lines represent overall means for each formant.

F1 and F2, possibly as a result of formant weakening and/or frication. As previously shown in Figs. 6.23, 6.26 and 6.27, these phenomena could affect the vowel from around 40% onwards. Further support for this theory is provided by the fact that words which tended to exhibit formant weakening and/or frication vowel-medially (e.g. VIRA, Fig. 6.29) had more variable formant measurements near the middle of the vowel, while words that tended to exhibit these patterns word-finally (e.g. BITA, Fig. 6.30) had more variable measurements near the end of the vowel. The use of different offglide patterns for different word contexts will be discussed further in Section 6.9.2.

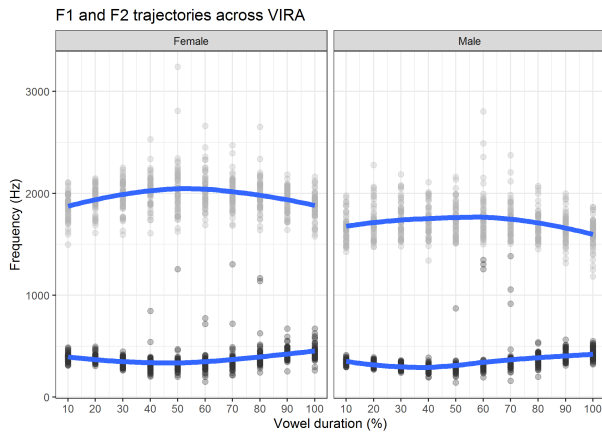


Figure 6.29: Mean F1 and F2 trajectories (Hz) for VIRA.

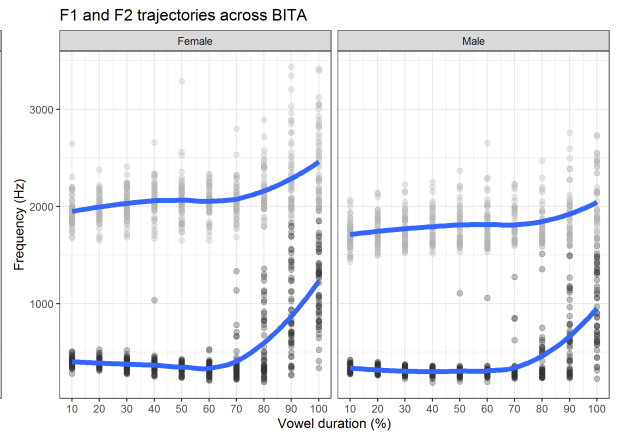


Figure 6.30: Mean F1 and F2 trajectories (Hz) for BITA.

## 6.9.2 Auditory transcriptions

Swedish high vowels are known for having fricated offglides, which in the case of standard [i:] usually consist of a closing gesture towards [j], with an optional epenthetic [ə] (Elert, 1995: 40). In the literature on Viby-i, it has not been clear whether the ‘buzziness’ usually attributed to this vowel is caused by frication during the vocalic portion, or whether it is simply the case that Viby-i also takes fricated offglides.

During transcription, around 5% of the /i:/ tokens were auditorily coded as ‘buzzy’ during the vocalic portion. This kind of frication was highly variable between individuals (Fig. 6.31), but seemed to favour word contexts where the preceding consonant was voiced (Fig. 6.32). Impressionistically, intra-vowel frication seemed to arise from a dental or alveolar constriction, and often produced ‘hairy’ waveforms, where the frication co-occurred with periodicity (Fig. 6.33).

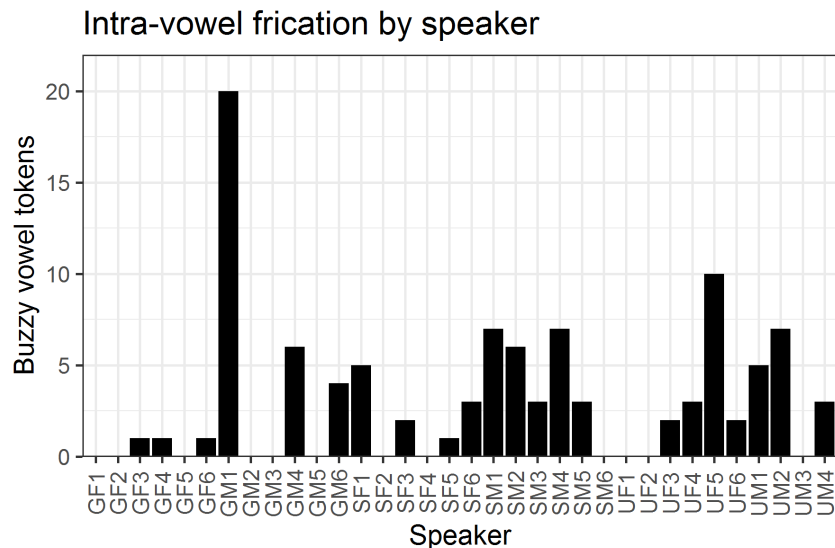


Figure 6.31: Zero-crossing rate across Viby-i by speaker.

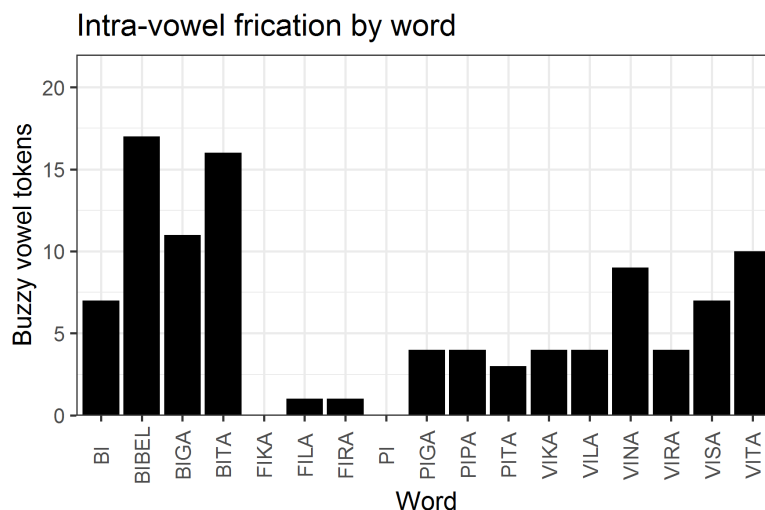


Figure 6.32: Zero-crossing rate across Viby-i by word.

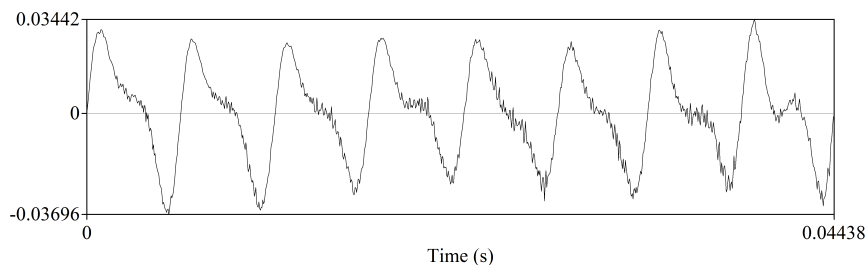


Figure 6.33: Frication superimposed on waveform for /i:/ during VILA.

End-frication was more common in the sample than intra-vowel frication, occurring in nearly every token. These fricated offglides were mainly conditioned by the following consonant, although they were also used in word-final position, as shown by Helgason (e.g. 2002). A summary of the different types of offglide possible for Viby-i is shown in Table 6.15. The proportional use of each of these types, organised by word, is shown in Fig. 6.34.

The most common type was the palatal glide [i:j], followed by its voiceless and fricated counterpart [i:ç]. The third most common was the voiced and fricated [i:j̥], which can be considered a more emphatic version of [i:j]. Together, these three categories constitute the majority of the offglides in the sample, although their proportions varied depending on the word context.

As Fig. 6.34 shows, [i:j] was prevalent when the following consonant was voiced, or when the vowel was word-final. Oppositely, [i:ç] was more common when the following consonant was voiceless. This pattern thus appears to be a coarticulatory effect, caused by speakers either maintaining or switching off voicing in order to accommodate the following consonant (Ohala & Solé, 2010). Some notable exceptions to this rule were PIPA and VISA, which still favoured [i:j]. It is not clear why this might be the case; PIPA still has a fairly large proportion of voiceless [i:ç] tokens, but the voiced version [i:j] was more than twice as common. VISA, on the other

	Simple offglide	With schwa
No closure	[i:]	[i:ə]
Palatal glide (voiced)	[i:j]	[i:jə]
Palatal frication (voiced)	[i:ɟ]	[i:ɟə]
Palatal frication (voiceless)	[i:]	[i:ç]
Dental frication (voiceless)	[i:s]	[i:sə]
Dental frication (voiced)	[i:z]	[i:zə]

Table 6.15: Viby-i offglide types. Entries in grey are theoretically possible, but were not observed in this sample.

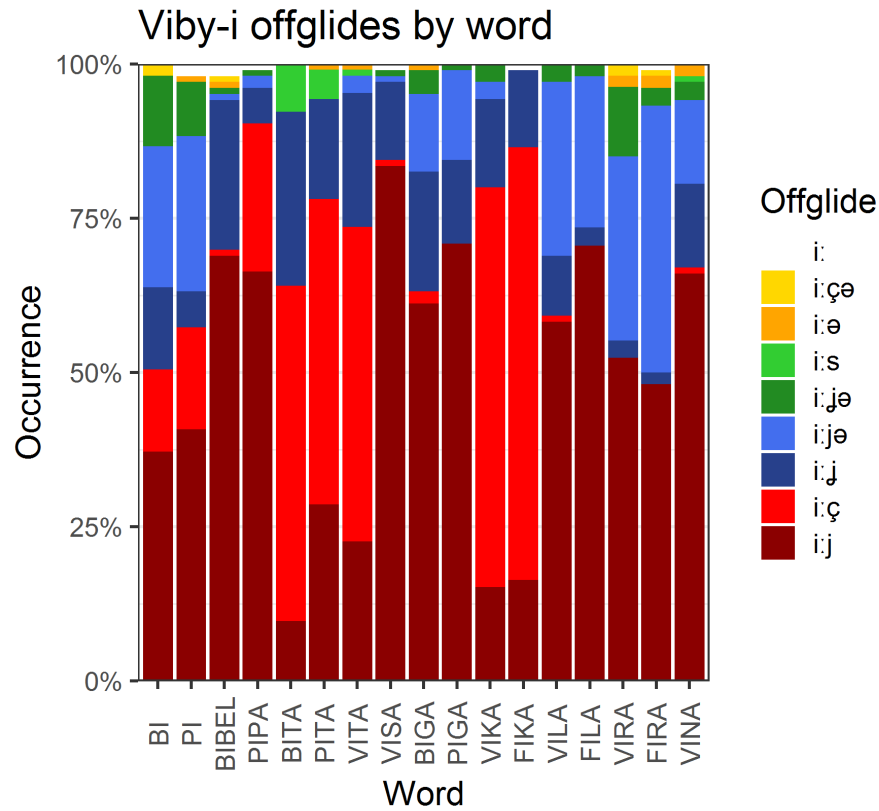


Figure 6.34: Viby-i offglides by word across the sample.

Word	/i:/ duration (ms)	Word	/i:/ duration (ms)
BI	306	PIGA	211
VIRA	289	VITA	209
PI	270	BIBEL	207
FIRA	268	FIKA	206
VILA	241	BITA	206
BIGA	240	VISA	199
VIKA	232	PITA	178
VINA	223	PIPA	171
FILA	222		

Table 6.16: Mean vowel durations across word list items.



hand, barely uses any voiceless offglides at all. This might have to do with the fact that Swedish does not have a voiced sibilant phoneme /z/, meaning that even if voicing persisted into /s/, there would be no risk of lexical confusion. It is also possible that [s] blocks the articulatory gesture for [ç], as two types of turbulence would need to be generated at different locations within a relatively short timeframe.

The offglides which used epenthetic schwa, either with or without preceding frication, were more common in monosyllables, and in words where the vowel was followed by /l, r/, and to some extent, /n, g/. The use of epenthetic schwa seemed to coincide with a longer vowel duration, but it is not evident whether epenthetic offglides were used to ‘fill’ the vowel duration, or if the duration became longer as a result of using them. The mean durations for /i:/ in different words is shown in Table 6.16 for comparison with Fig. 6.34. The words with the longest durations, i.e. BI, PI, VIRI, FIRA, VILA, were more likely than other words in the sample to use epenthetic offglides, particularly [i:jə].

The occurrence of different offglides also varied by speaker. The offglide patterns for all 34 speakers is shown in Fig. 6.35. Although it may not be informative to examine these patterns in detail, it is worth noting that [i:ç] and [i:ç] occurred at similar rates in most speakers. However, a small number of speakers – GM1, GM5, GM6, SM2, SM5, and UM1 – seemed to prefer [i:j] over the other variants. Another small group – UF2, GF3, GF5, and GM2 – used more non-fricated offglides than the rest of the sample, but the occurrence of these variants is still low. Thus, although there are some indications of individual preference, most speakers used a closing gesture for /i:/, supporting reports in the literature that these offglides are common for high vowels in Central Swedish (Elert, 1995; Engstrand, 1999).

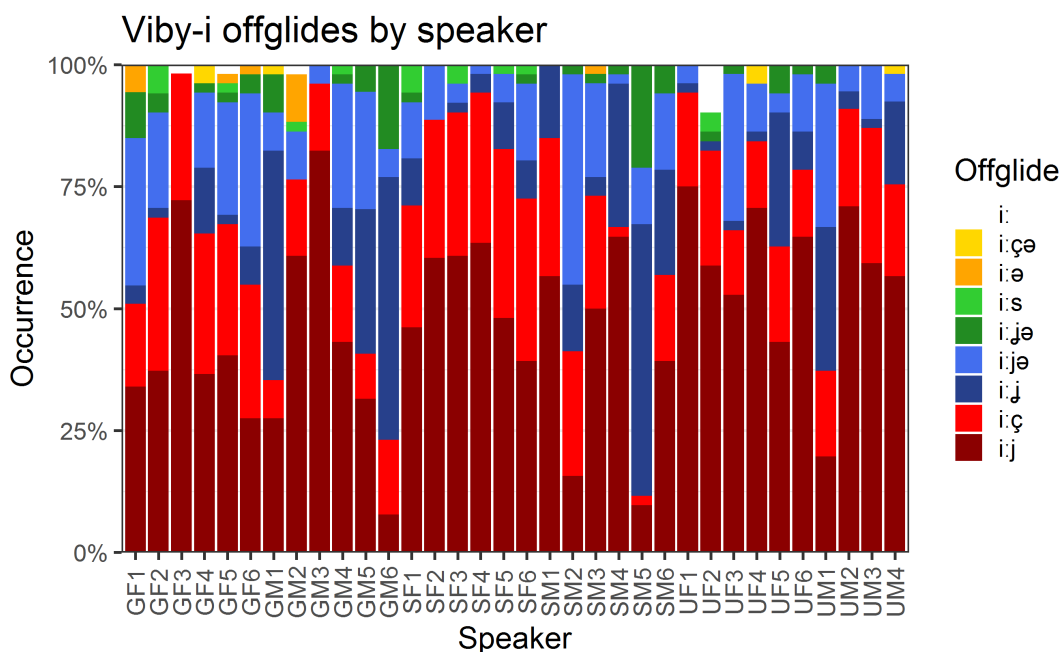


Figure 6.35: Viby-i offglides by speaker across the sample.

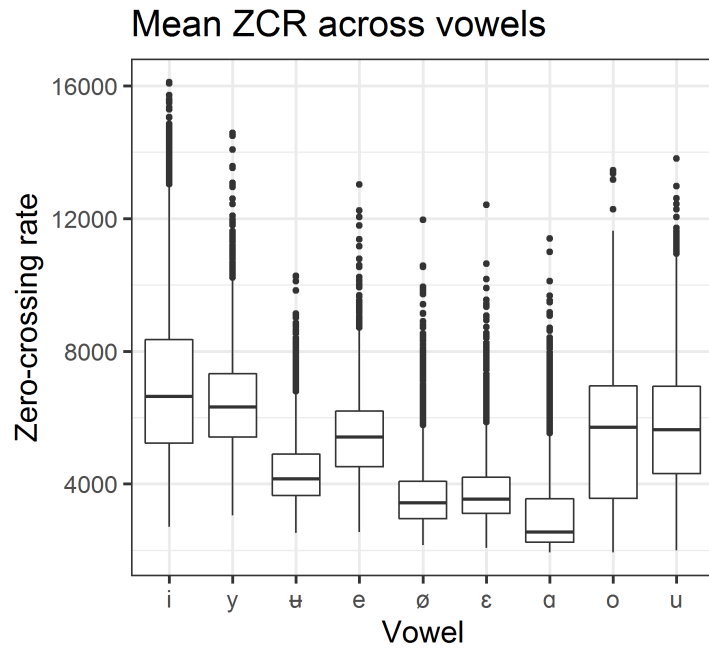


Figure 6.36: Overall difference in bpZCR between all vowels in the sample.

### 6.9.3 Zero-crossing rate

To quantify some of the frication patterns seen in the previous sections, bandpass-filtered zero-crossing rate (bpZCR) was measured for each 10% interval of the vowel duration. Taken as a whole, the mean bpZCR for /i:/ was higher than for most of the other vowels in the sample, with the exception of /y:/ (Fig. 6.36). This could indicate that both /i:/ and /y:/ had more frication than the other vowels, however this finding should be regarded as tentative, as bpZCR can also be sensitive to differences in formant structure (see Section 5.8.4). In the case of Vi-by-i, it is also worth noting that this mean value is likely to include both intra-vowel frication and end-frication.

#### Intra-vowel frication

To investigate the effect of intra-vowel frication on bpZCR, a comparison was made between the auditorily ‘buzzy’ and ‘non-buzzy’ /i:/ tokens over the course of the vowel duration. Fig. 6.37 provides a visualisation of how these vowel types behaved in normalised time. Note that even though the trajectories are continuous, the raw data was sampled in 10% blocks.

As the figure shows, the ‘buzzy’ and ‘non-buzzy’ trajectories have virtually the same shape, but the ‘buzzy’ tokens have a consistently higher bpZCR. There thus appears to be a quantifiable difference in frication during the vocalic portion of some Vi-by-i tokens. In addition to a gradient difference in F2-lowering, Vi-by-i may thus use different degrees of intra-vowel frication.

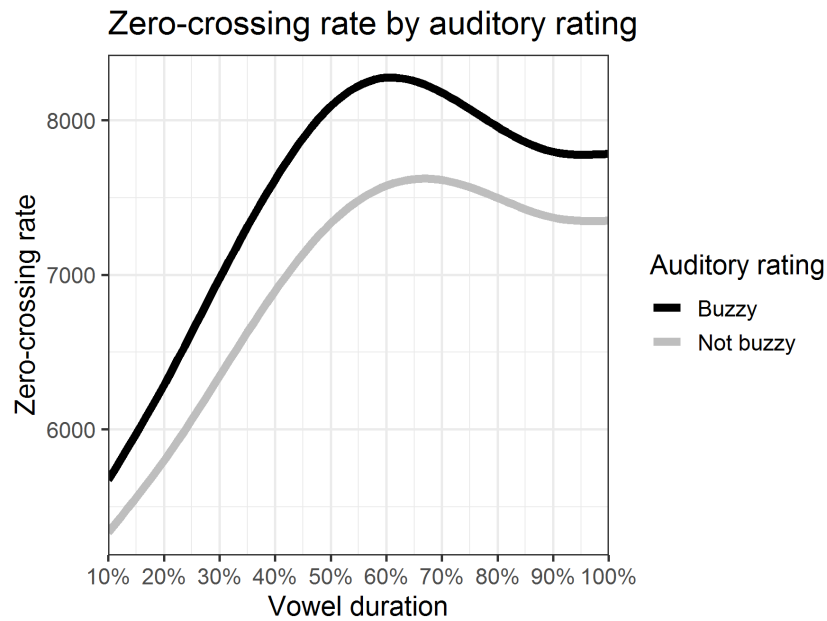


Figure 6.37: Overall difference in bpZCR between auditorily ‘buzzy’ vs. ‘not buzzy’ /i:/.

### End-frication

In addition to intra-vowel frication, the data also showed patterns of end-frication. Since the analysis in Section 6.9.2 revealed that word environment seems to condition offglide type, Fig. 6.38 displays the bpZCR in normalised time across Vi-by-i by word. The figure shows that the starting values for all /i:/ contexts are relatively similar, at 5,000–6,000 zero-crossings per second. However, beginning from the outset of the vowel, the bpZCR gradually increases, with most word contexts peaking at around 50–70% of the vowel duration. This corresponds roughly to the timepoint when the formant tracker started having difficulties identifying F1 and F2. For most word contexts, the bpZCR falls again towards the end of the vowel, ending at a similar, but usually slightly higher, value than where it started. In a few contexts, the bpZCR continues to increase throughout the vowel duration, with some reaching a maximum of around 10,000 zero-crossings per second at the very end of the vowel. A small subset finishes at a lower value of around 8,500.

Investigating the correspondence between bpZCR and word, we can see that there is some correspondence in the occurrence of a high bpZCR and a loss of accuracy in the formant tracker. For example, in the specific cases of VIRA and BITA, previously presented in Figs. 6.29 and 6.30, the bpZCR peaks near the middle of the vowel for VIRA, while it peaks near the end of the vowel for BITA.

There is also a fairly intuitive correspondence between the bpZCR and the auditory transcriptions of the offglides for /i:/. Words that favour a voiceless fricative offglide [i:ç], e.g. BITA, PITA, VITA, VIKA, FIKA, have continuous upward trajectories, with a high ending bpZCR. Since voiceless frication is more likely to be aperiodic than voiced frication, it is indeed ex-

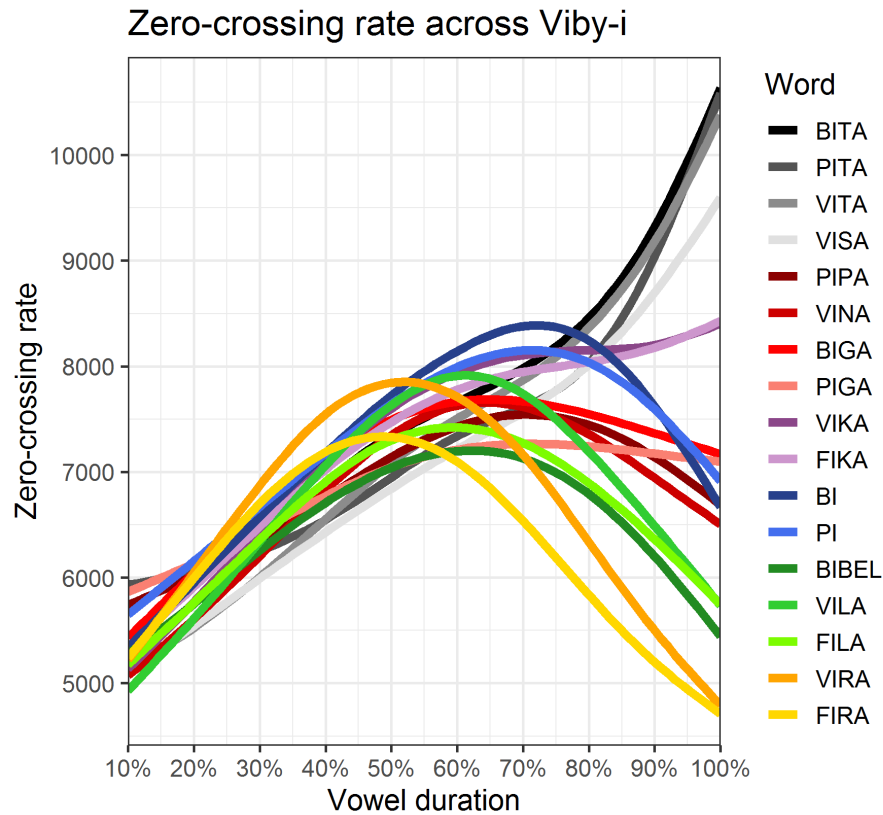


Figure 6.38: Zero-crossing rate across Viby-i by word.

pected to find a higher bpZCR in these environments. In addition, front consonant environments, e.g. BITA, VISA, have a higher ending bpZCR than back consonant environments, e.g. VIKA, FIKA. This pattern is also expected, since frication produced at the front of the vocal tract usually has a higher frequency than frication produced at the back, due to the smaller resonating space.

Words that commonly produce Viby-i with end-frication followed by an epenthetic schwa, e.g. BI, BIGA, VILA, VINA, have bpZCR peaks near the middle of the vowel, followed by a drop, which could indicate a return towards a more vowel-like production, i.e. schwa. This pattern is particularly pronounced in VIRA and FIRA, which also tend to have an earlier frication peak than the other word contexts. Conversely, BI and PI have a later frication peak, and end with a relatively high bpZCR. This high ending value could be the result of persisting voiced frication towards the end of the vowel, e.g. [i:j] or [i:j̥].

### Effect of preceding voicing and aspiration

To return to the hypotheses made in Section 5.4, this final section examines whether the voicing or aspiration of the preceding consonant encourages frication during Viby-i. Fig. 6.39 presents the dynamic bpZCR data, organised by preceding consonant /p, b, f, v/. Based on this figure, all environments had a similar starting bpZCR, with the exception of /p/, which was higher. Thus,

although the vowel segment did not begin until the onset of F2, when there was no longer any visible aspiration from /p/, it appears that this environment nevertheless encourages more frication at the start of Viby-i. This could be because some aspiration was nevertheless carried over into the vowel, or it could indicate that Viby-i uses similar strategies to consonant aspiration (e.g. breathiness, devoicing) to achieve intra-vowel frication. The fact that /f, v/ had lower starting points suggests that preceding labiodental frication did not contribute to increasing buzziness at the start of Viby-i.

For large parts of the vowel duration, however, all four consonant environments have similar bpZCR values, until around 60%, where most of them peak. The difference between the values at this point is relatively small; around 250 zero-crossings per second. Thus, for the overall impression of buzziness during the vocalic portion, it appears that the voicing and aspiration of the preceding consonant does not have a great effect.

The overall shapes and end-points of these trajectories also vary, but this is likely to be an effect of coarticulation with the following consonant. For example, the low end-point of the curve for /f/ can be attributed to the fact that this context consisted of the words FIRA, FILA, FIKA. Since /r, l/ conditioned an early bpZCR peak followed by a sharp dip, this explains why the end-point for /f/ is lower than for the other environments.

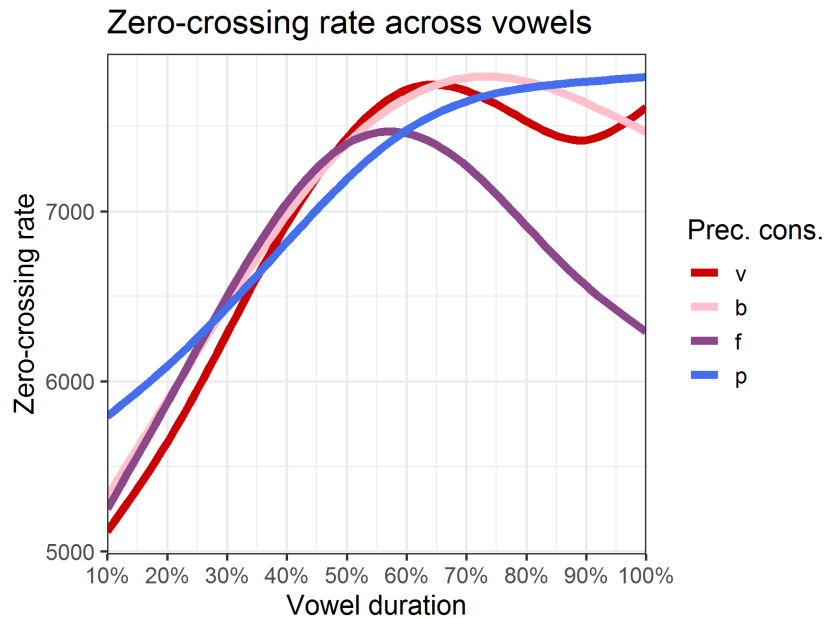


Figure 6.39: Zero-crossing rate across Viby-i by preceding consonant /v, b, f, p/

#### 6.9.4 Summary of Viby-i dynamics

In summary, the dynamic profile of Viby-i is characterised by relative stability in the trajectories of F1 and F2, with the exception of some coarticulatory movement towards the end of the vowel.

However, Viby-i displays unexpected dynamic qualities in the shape of formant weakening and frication, which sometimes occur together, and sometimes separately. Impressionistically, the two appear to serve similar functions, as formant weakening appears to give the impression of a /j/-like offglide, and frication is usually produced with a similar closing gesture. Both of these dynamic phenomena could thus be argued to contribute to greater contrasts between different vowel phonemes in a crowded acoustic space, as suggested by e.g. Bruce (2010) and Elert (1995).

The offglides observed for /i:/ in this sample begin relatively early, sometimes as early as 40% into the vowel duration. In many cases, frication occurs near the middle of the vowel, followed by an epenthetic schwa. Periods of intense frication or formant weakening appear to increase the risk of error, or at least result in greater variability, during formant measurement. This finding supports the methodological decision to use formant values from the first 10% of the vowel for the static analysis, as opposed to the customary 50% timepoint.

The realisation of vowel offglides for Viby-i appears to be conditioned by word context. High-frequency end-frication is usually produced before voiceless alveolar or dental consonants. Lower frequency frication, or glides towards schwa, are more common in word-final contexts, and contexts where the following consonant is voiced. However, individual speakers also seem to have some personal preferences in what offglide type they use.

Frication during the vocalic portion of Viby-i was relatively rare, but was found to contribute to a higher overall zero-crossing rate. Intra-vowel frication was also found to be variable across speakers and linguistic contexts, but since the tokens were relatively few, it was not possible to investigate this phenomenon in detail. However, it did appear that intra-vowel frication, like F2-lowering, could contribute to Viby-i in a gradient way.

## 6.10 Summary of acoustic results

The acoustic analysis has shown that /i:/ in the current sample is characterised by a low F2 and a high F1 with relation to /e:/. This realisation of /i:/ corresponds to the acoustic criteria used to define Viby-i in this thesis, and resembles the formant values reported in previous literature for this sound (e.g. Björsten & Engstrand, 1999; Schötz et al., 2011). Surprisingly, none of the 34 speakers used a standard high front [i:], which mirrors the findings of previous studies carried out as part of this project (Westerberg, 2013, 2016). These results could indicate that Viby-i is more widespread in Central Sweden than previously assumed. They also suggest that /i:/ has shifted down and back, leaving the high front part of the acoustic vowel space empty.

The degree to which speakers displayed F2-lowering and F1-raising was variable across the sample, supporting the suggestion that Viby-colouring is a gradient phenomenon (Westerberg, 2016). However, as other work has reported (Björsten & Engstrand, 1999; Gross & Forsberg, 2019; Schötz et al., 2011), Viby-colouring mainly seems to be associated with F2-lowering,

which was relatively great across the sample, while F1-raising was usually smaller and more variable. In addition to these phenomena occurring in /i:/, this vowel showed some overlap with /y:/ in the acoustic space, which could potentially result in perceptual ambiguity (Gross & Forsberg, 2019). This finding implies that speakers who use Viby-i also apply Viby-colouring to /y:/, as previous literature has suggested (e.g. Björsten & Engstrand, 1999; Bruce, 2010; Elert, 1995).

The production of Viby-i was influenced to some extent by vowel duration, preceding consonant voicing and frication, and following consonant place and manner. In particular, Viby-i seemed to be ‘stronger’ in contexts where vowel duration was short, where the preceding consonant was voiced or a fricative, and where the following consonant was complex or back, non-lingual, or word-final. These results could be interpreted as Viby-i being likely to have a centralised, backed articulation. The relationship between F2-lowering and preceding voicing and frication is difficult to explain, but it can be noted that had the environment that had the highest predicted F2, i.e. /p/, appeared to be slightly more fricated than the other contexts. Thus, it is possible that F2-lowering and vowel frication have a complementary relationship, but this phenomenon would need to be investigated further.

Social factors had a surprisingly small effect on Viby-i production overall, but there was a relatively strong effect of city, whereby Gothenburg speakers produced Viby-i with a lower F2 than Stockholm and Uppsala speakers. This result aligns with previous literature, in that the shift towards Viby-i might have progressed further in Gothenburg, where this sound is believed to be older (Björseth, 1958; Elert, 1995). In addition, it supports the prediction of this thesis that Viby-i might have spread from Stockholm to Uppsala, since Viby-i was realised similarly in both cities. The lack of difference between age groups, genders, and speakers with different lifestyles and attitudes, is in itself an interesting result, as it could point towards Viby-i being produced relatively consistently in all three cities, at least by speakers from a middle-class background.

The dynamic analysis revealed that, while Viby-i is produced with relatively stable formant values, the vowel is affected by a number of dynamic offglides, which can feature formant weakening, frication, and epenthetic vowels. The offglides mainly appear to be conditioned by the following consonant (Ohala & Solé, 2010). However, they do not seem to be greatly affected by the preceding consonant, with the exception of /p/, which slightly increased the initial frication rate. Audible frication during the vocalic portion itself was relatively rare, but Viby-i did show signs of containing more frication than most of the other vowels, by virtue of having a higher zero-crossing rate overall. Thus, there is some support for the description of Viby-i as “buzzy” (Borgström, 1913), although most of the frication found in this vowel seems to come from fricated offglides, similar to those used for high vowels in Central Standard Swedish (Engstrand, 1999: 141).

# Chapter 7

## Articulatory results

### 7.1 Chapter overview

This chapter presents the results of the articulatory analysis of Viby-i. The static analysis provides data of the tongue and lips sampled at the 10% timepoint, corresponding to the static part of the acoustic analysis, while the dynamic analysis uses data from the 10%, 50%, and 90% timepoints.

The first part of the chapter describes the overall tongue gesture used for Viby-i. Section 7.2 provides a visual inspection of the mean tongue splines for all nine long vowels, describing the position of /i:/ in the articulatory vowel space with respect to the nearest high front vowel /e:/. This section also discusses the different tongue shapes used for Viby-i. Section 7.3 presents GAMM data for the high front vowels, and tests whether any parts of the tongue curve are significantly different between /i:, y:, e:/.

Section 7.4 then presents a set of normalised articulatory measures taken at different points along the tongue curve, and evaluates how well each of these measures expresses the relationship between Viby-i acoustics and articulation. Since a mismatch is found, Section 7.5 additionally provides a description of the lip posture used for Viby-i. Section 7.6 then statistically tests whether any of the tongue point measures are correlated to differences in F1/F2, and whether the tongue gesture used for Viby-i is socially stratified.

Finally, Section 7.7 uses dynamic data taken to investigate the effects of tongue movement on formants and vowel frication. These effects are first contextualised using mean splines and GAMMS, which are qualitatively investigated in relation to the formant patterns observed in the acoustic analysis. The correlation between bandpass-filtered zero-crossing rate and articulatory narrowing at the front and back of the mouth is then tested to investigate whether either of these movements significantly contribute to frication during Viby-i.



## 7.2 Qualitative description of /i:/ tongue gesture

Mean tongue splines were generated for all nine long vowels as produced by each speaker, sampled at 10% of the vowel duration. The full set of mean tongue splines for each speaker (with annotations) is presented in Appendix G. This data excludes speaker SM3, whose ultrasound data was corrupted. The mean tongue plots represent the vocal tract in midsagittal section, facing right (Fig. 7.1). This orientation is conventional in most UTI literature (e.g. Preston et al., 2017; Stone, 2009), despite depicting the tongue in the opposite direction to the vowel quadrilateral. In addition to the tongue splines, each graph also contains a palate spline (dark grey), and a bite plate spline (light grey). The end of the bite plate represents the approximate position of the front teeth.

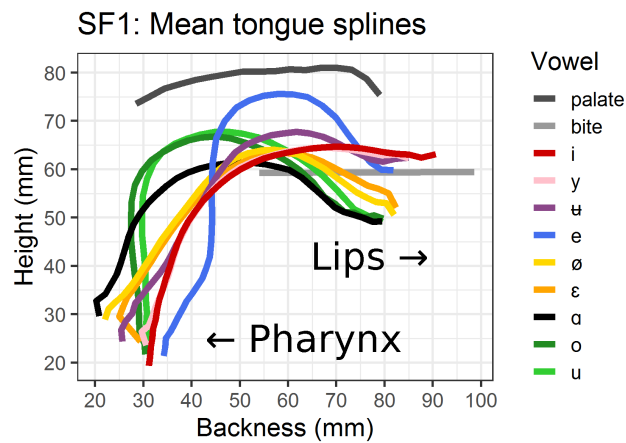


Figure 7.1: Illustration of how the UTI data is orientated.

For this part of the analysis, the mean tongue splines for /i:/ will be visually compared to the nearest high front vowel /e:/ with regard to the following parameters:

- Tongue body height and backness, referring to the highest point of the tongue, with some notes about the overall mass of the tongue.
- Tongue tip height, and whether the tongue tip is pointing up or down. Note that the term ‘tip’ here refers to the observed tip in the ultrasound images, which may not correspond to the actual tongue tip, since this part of the ultrasound image is sometimes obscured.
- Post-dorsal retraction, which collectively refers to either backing, constriction, or both, at the post-dorsal region of the tongue, i.e. the part of the tongue that faces the pharynx.

The shape of the tongue will also be discussed. The tongue shapes in the dataset have been divided into four categories based on visual criteria: Arched, front-bunched, back-bunched, and double-bunched. The arched category represents the canonical, convex shape described in the vowel literature (e.g. Ogden, 2009: 56) (Fig. 7.2). For arched tongue shapes, the highest point of the tongue tends to coincide with the point of maximum narrowing in the vocal tract. ‘Bunching’

refers to visible constriction of the tongue, signalled by a protrusion of the tongue surface, which can occur either at the front, at the back, or both. Front-bunching thus shows a main constriction at the front of the tongue (Fig. 7.3), and back-bunching at the back (Fig. 7.4). Double-bunching uses a dual constriction, and is characterised by a dip in the tongue surface between the raised front and back (Fig. 7.5).

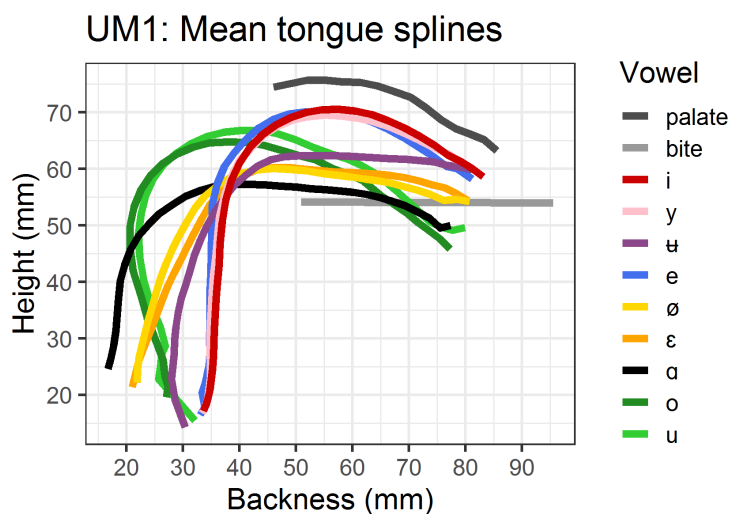


Figure 7.2: Example of an arched tongue shape for /i:/ (red).

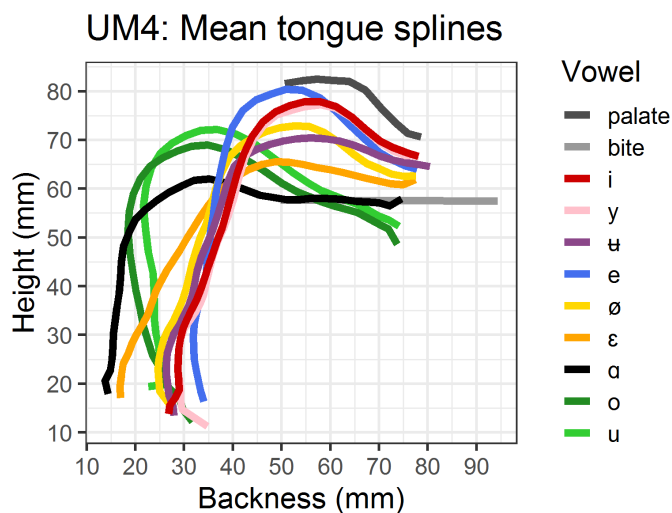


Figure 7.3: Example of a front-bunched tongue shape for /i:/ (red).

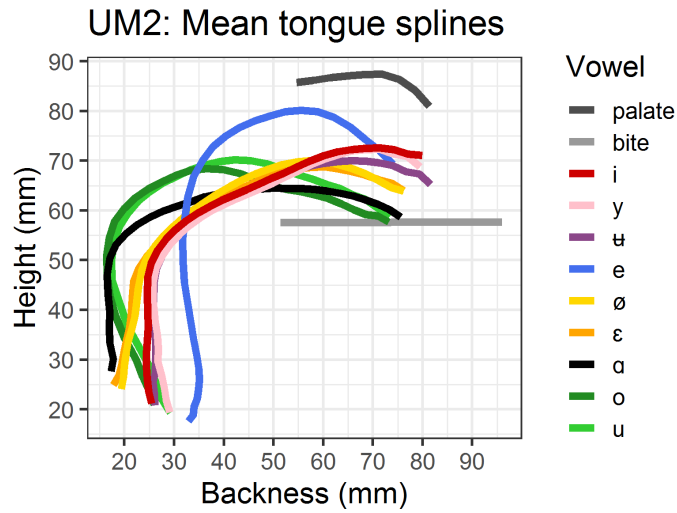


Figure 7.4: Example of a back-bunched tongue shape for /i:/ (red).

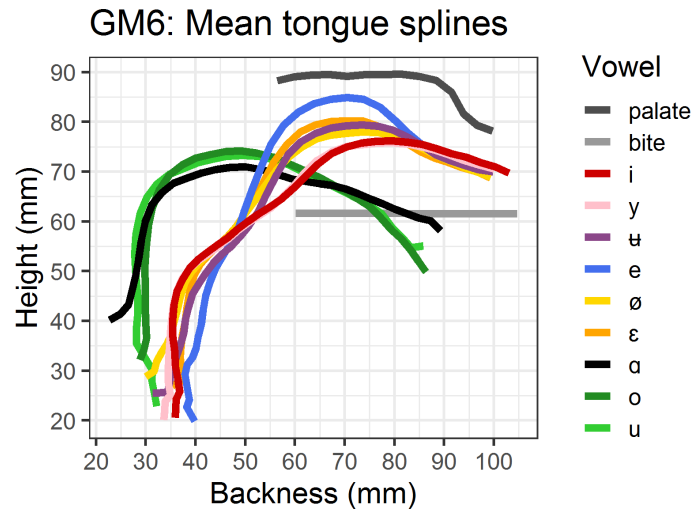


Figure 7.5: Example of a double-bunched tongue shape for /i:/ (red).

### 7.2.1 Tongue body position

The first part of this analysis concerns the overall position of the tongue for /i:/ in the articulatory space. A few qualitative patterns can be observed from the mean spline data: Firstly, the tongue gesture for /i:/ is lower than /e:/ in most speakers (22 of 33) (Figs. 7.6, 7.7). In the remaining speakers, the tongue body has a similar height for /i:/ and /e:/ (Fig. 7.8). Secondly, even when the tongue heights are similar, the tongue body for /i:/ tends to be fronter than /e:/, creating a characteristic gap between the two splines. In addition, a small number of speakers (SF6, UF6, UM1) have more or less overlapping tongue gestures for /i:/ and /e:/ (Fig. 7.9). The finding that /i:/ is mostly lowered and fronted in the articulatory space is surprising, as the acoustic data indicated a tongue gesture that was similar in height to /e:/, but more backed. Instead, we find

the opposite pattern. What both the acoustic and articulatory results have in common, however, is that /i:/ no longer seems to be the high front corner vowel of this system.

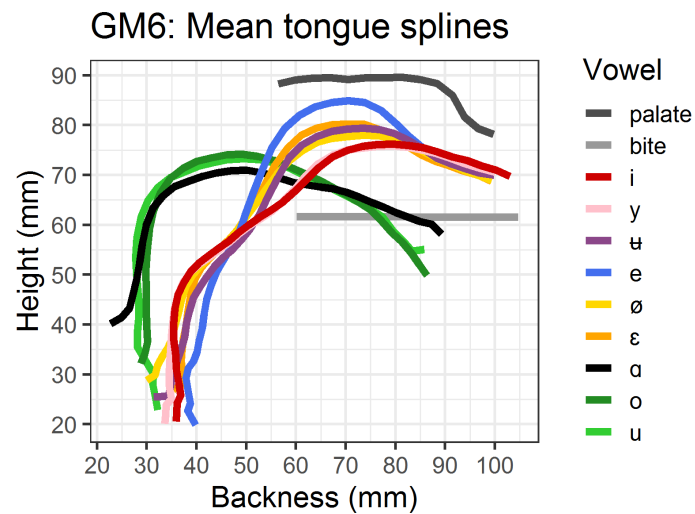


Figure 7.6: Low mean tongue height for /i:/ compared to /e:/.

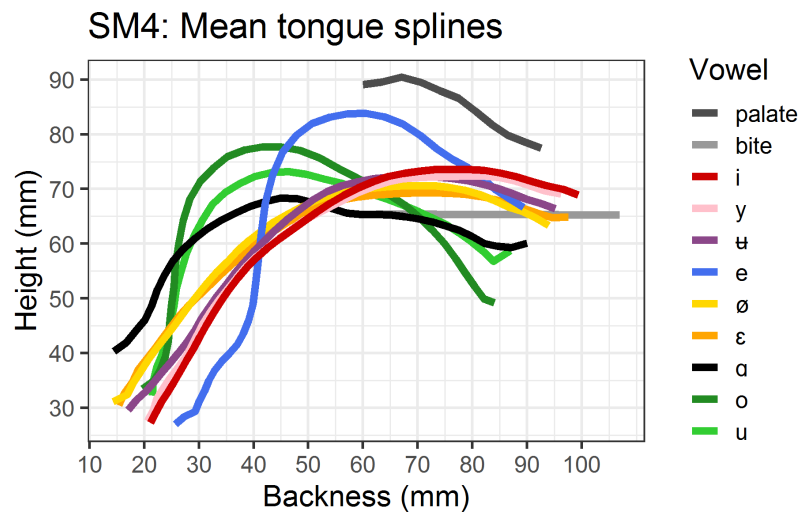


Figure 7.7: Very low mean tongue height for /i:/ compared to /e:/.

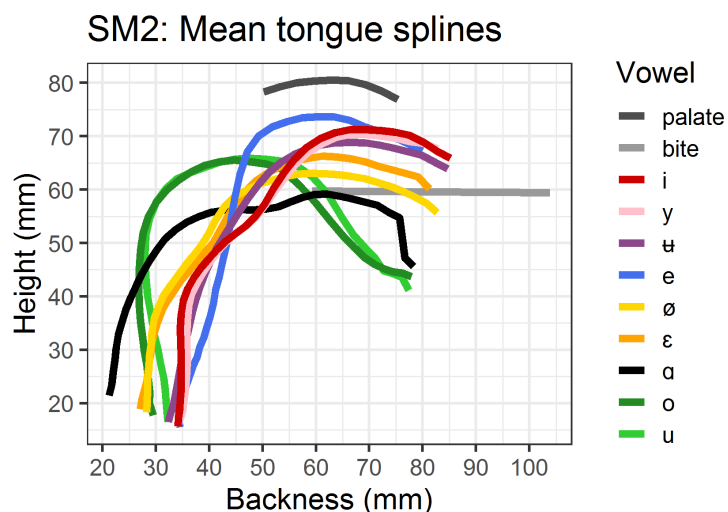


Figure 7.8: Similar mean tongue heights for /i:/ and /e:/.

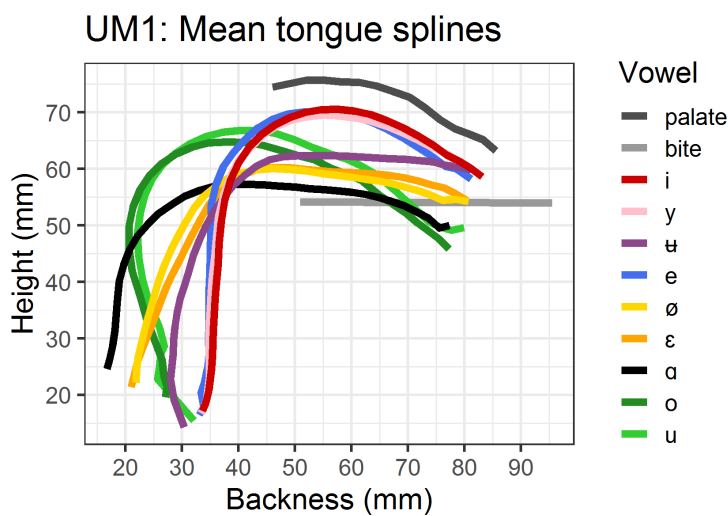


Figure 7.9: Overlapping mean tongue gestures for /i:/ and /e:/.

Regarding the tongue position of the other vowels in the system, /e:/, /y:/, /ɨ:/, /ø:/, /ɛ:/, /ɑ:/, /o:/, /u:/, they mostly behave in predictable ways:

- In most speakers, /e:/ has a high tongue body, situated near the other front vowels, but it is not usually the frontest. Since /i:/ is no longer high, /e:/ appears to take the place of the high front corner vowel in most speakers.
- In all speakers, /y:/ overlaps more or less perfectly with /i:/, indicating that /i:/ and /y:/ are either distinguished in other ways (e.g. lip-rounding), or that they may be subject to perceptual overlap (Gross & Forsberg, 2019).
- The tongue gesture for /ɨ:/ is usually lower than /e:/, but is variable across the sample. Sometimes, it patterns with /i:/, /y:/ (e.g. GF6, SM5), other times with /e:/ (e.g. GF5,

GM3), and other times with /ɛ:/, ø:/ (e.g. GM5, UM3). Some of this variation may be dialectal, or it could be the case that speakers are compensating for differences in tongue gesture through different degrees of lip-rounding (Lawson et al., 2019).

- The tongue gestures for /ø:/ and /ɛ:/ overlap in many speakers, but also seem to be subject to dialectal variation. Both these vowels usually have a similar tongue body height to /i:/, but in the Gothenburg group, they are slightly higher. When /ɛ:/ and /ø:/ differ in height, /ɛ:/ is usually lower (e.g. SF5, UF4, UM4), but the opposite pattern can also be found (e.g. GM1, GM3, SM2).
- /ɑ:/ is situated below /o:/ and /u:/ in most speakers, or, in a few speakers, these vowels have a similar height (e.g. GF6, GM6). The height of the back vowels compared to the front vowels is variable however, and does not appear to operate on the same scale. For example, /ɑ:/ has a similar tongue body height to /i:/, or even higher, in some speakers (e.g. GF6, GM2), but these vowels were still clearly distinct in F1 (see Section 6.5).
- The tongue gestures for /o:/ and /u:/ are backed and relatively high. The splines for these two vowels commonly overlap, suggesting that they are distinguished in other ways (e.g. lip-rounding, offglides). In some speakers, /u:/ is the higher of the two (e.g. GM1), while in others, /o:/ is higher (e.g. SM4).

My description of tongue position has primarily focused on tongue height, since backness was difficult to categorise visually. For example, the term ‘backness’ could describe the highest point of the tongue, the overall position of the tongue body, or retraction at the back or root of the tongue. In terms of Viby-i, determining backness is also problematised by the possibility of a double constriction; one at the highest point of the tongue body, and one at the post-dorsal region. This issue will be returned to in Section 7.4, when different measures for quantifying tongue backness are evaluated. For the purposes of this description, ‘backness’ was operationalised as the position of the highest point of the tongue.

Summaries of how the tongue body position for /i:/ compares to /e:/ across the sample are provided in Figs. 7.10 and 7.11. The plots are divided by city and gender. As the graphs show, fronting of the highest point of the tongue occurs at similar rates across all three cities, but males in Gothenburg and Stockholm are slightly more fronted than the other groups. In terms of tongue lowering, this is most common in Gothenburg, followed by Stockholm, and then Uppsala. This pattern repeats across both men and women. This finding corresponds to the acoustic result that Viby-i had a higher F1 in Gothenburg compared to the other two cities. However, the acoustic analysis also showed that Gothenburg had a significantly lower F2 than Stockholm and Uppsala, and this pattern is not matched by tongue backing, at least not going by the highest point of the tongue.

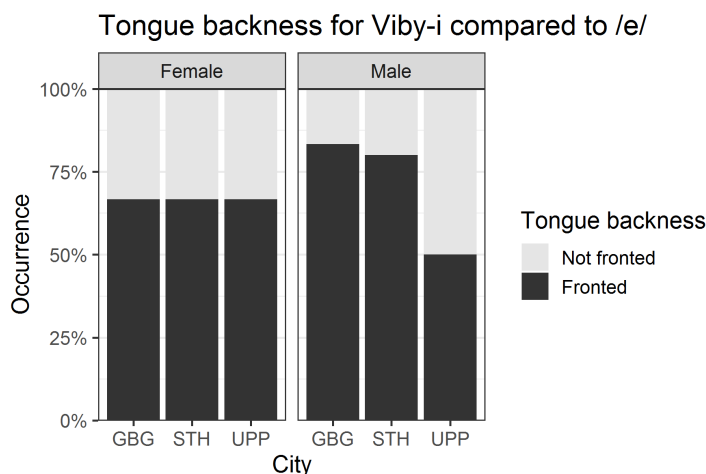


Figure 7.10: Distribution of tongue body fronting of /i:/ against /e:/ across the sample.

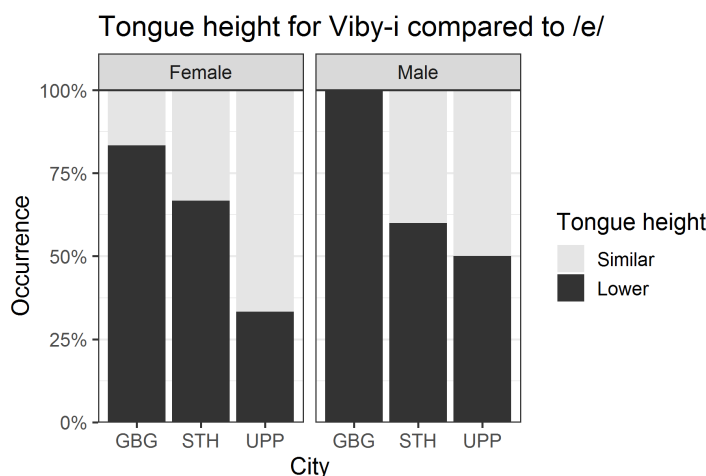


Figure 7.11: Distribution of tongue body lowering of /i:/ against /e:/ across the sample.

## 7.2.2 Tongue tip

The behaviour of the observed tongue tip for /i:/ is relatively variable. It is classified as ‘down’ in roughly half the speakers (17 of 33) (Fig. 7.12) and ‘up’ in the other half (16 of 33) (Fig. 7.13). However, the orientation of the tongue tip does not always predict its height relative to other vowels. In most speakers (18 of 33), the tongue tip for /i:/ is higher than /e:/, and the highest of all vowels in the system. This height is achieved through both tip-up and tip-down gestures. When the tongue tip is not higher than /e:/, the two vowels usually have a similar tip height (12 of 33 speakers) (Fig. 7.14). There are also a few speakers (GF1, GM1, UF3) who use a tip-up gesture, but nevertheless have a lower tongue tip for /i:/ than for /e:/ (Fig. 7.15). Overall, all speakers use tongue tip gestures that are relatively high and front, and thus close to either the alveolar ridge or the teeth. It thus seems possible that speakers use the tongue tip to generate frication during Viby-i.

Some of the variability in tongue tip behaviour may be due to partial data, since the tongue

tip is often obscured in ultrasound images. However, there are still some overall trends in tongue tip behaviour with regard to city and gender. Fig. 7.16 shows that women in this sample are slightly more likely than men to have a higher tongue tip for /i:/ than for /e:/, and to use a tip-up gesture. This gender difference is particularly pronounced in Stockholm, and to some extent in Uppsala, since the Uppsala males did not use any tip-up gestures at all. These differences could be related to patterns in the production of vowel frication during Viby-i, as examined in Section 7.7.3.

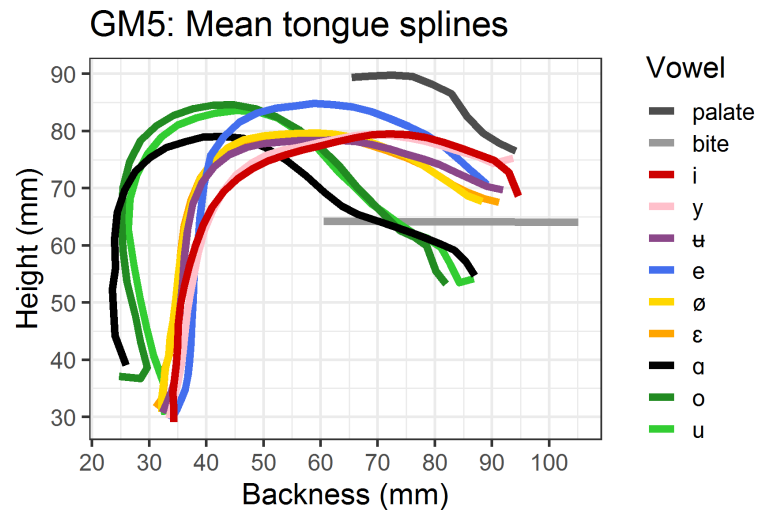


Figure 7.12: High tip-down gesture for /i:/.

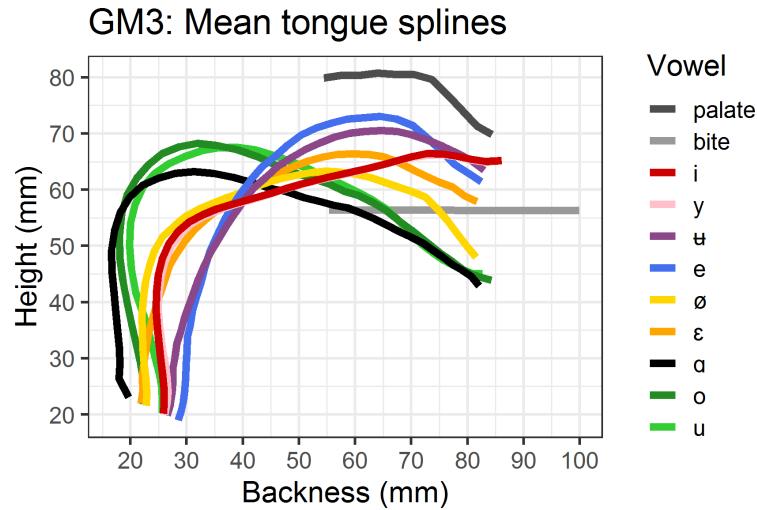


Figure 7.13: High tip-up gesture for /i:/.



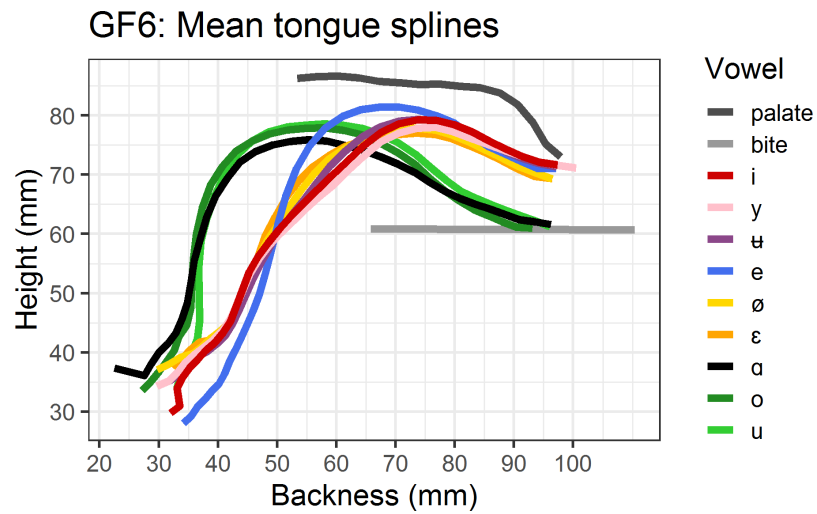


Figure 7.14: Similar tip gestures between /i:/ and /e:/.

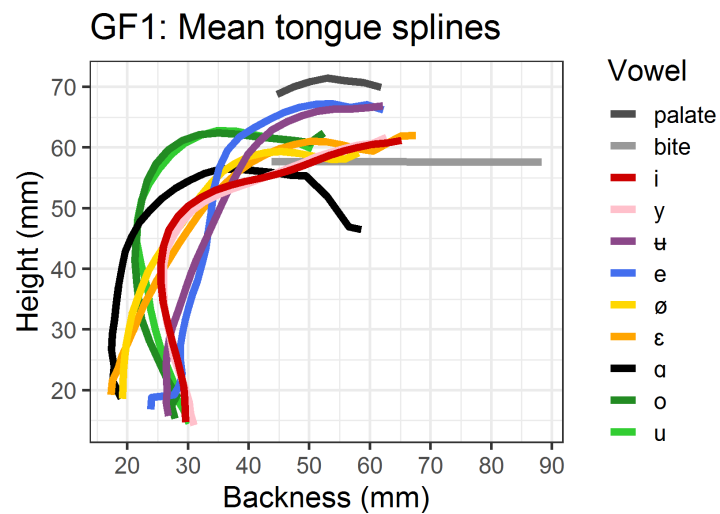


Figure 7.15: Lower tip gesture for /i:/ than for /e:/.

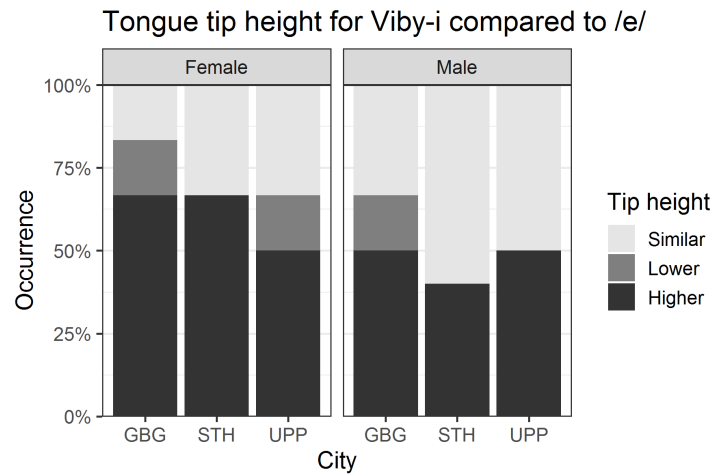


Figure 7.16: Distribution of tongue tip height for /i:/ compared to /e:/ across the sample.

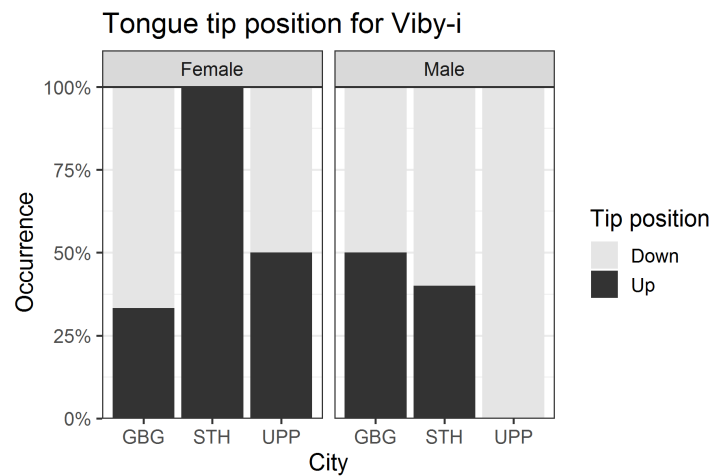


Figure 7.17: Distribution of tongue tip gesture for /i:/ across the sample.

### 7.2.3 Tongue retraction

As mentioned previously, the tongue position for Viby-i can be regarded as fronted with respect to the highest point of the tongue, but most speakers (24 of 33) have additional post-dorsal retraction, resulting in narrowing against the pharynx. This retraction can take the form of visible constriction (Fig. 7.18), tongue backing (Fig. 7.19), or both occurring together (Fig. 7.20). The combination of post-dorsal retraction and tongue body fronting frequently results in the tongue splines for /i:/ crossing over /e:/ at some point along the tongue body.

The degree of tongue retraction for /i:/ is equal to or greater than /e:/ in all speakers, but in terms of the entire vowel system, /i:/ is not particularly retracted. Compared to the other vowels, it is usually similar to or fronter than /ɛ:, ø:/, and never as backed as any of the back vowels.

Nevertheless, the occurrence of tongue retraction is unexpected for an /i:/ vowel, and is likely to have consequences for its acoustic output. This issue will be explored further in Section 7.4.

As Fig. 7.22 shows, backing of the post-dorsal region is relatively prevalent for /i:/ across the sample, and is particularly common in Gothenburg and Uppsala women. Post-dorsal constriction is less common (Fig. 7.21), and appears to be dis-preferred by Stockholm speakers of both genders. However, since many speakers combined these two strategies, there may be complex interactions between these social patterns that are not visible here. The relationship between tongue retraction and social factors will be investigated further in Section 7.6.4.

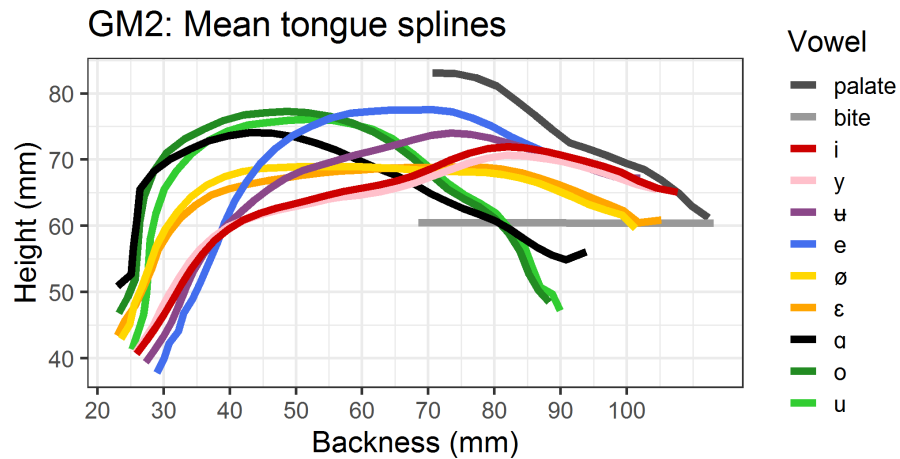


Figure 7.18: Post-dorsal constriction for /i:/.

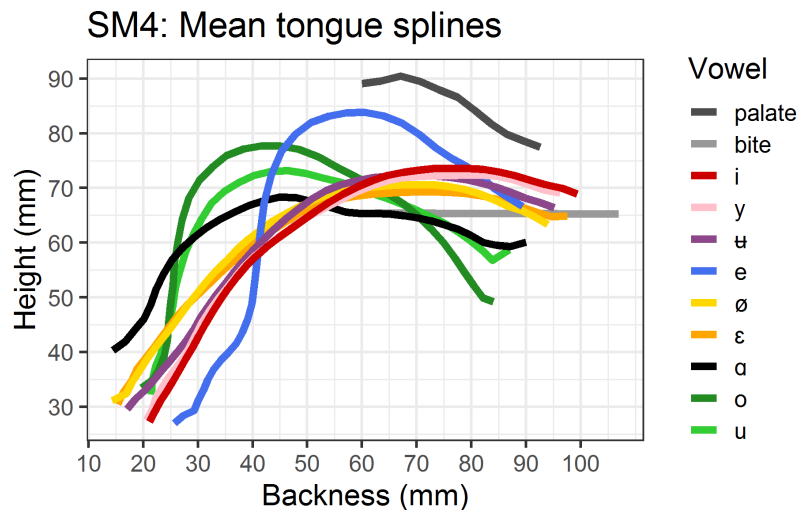


Figure 7.19: Post-dorsal backing for /i:/ compared to /e:/.

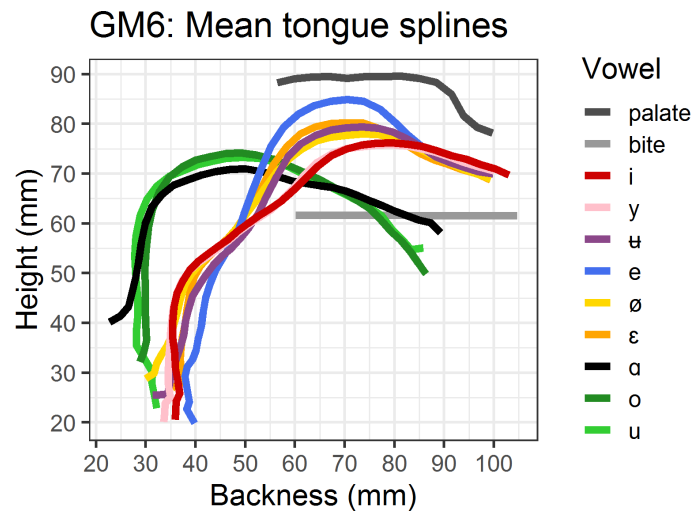


Figure 7.20: Post-dorsal constriction and backing for /i:/ compared to /e:/.

### 7.2.4 Tongue shape

The type of retraction used by speakers (constriction, backing, or both) also affects the overall shape of the tongue. Most speakers in the sample (13 of 33) use an arched tongue shape for /i:/. However, a similar number (12 of 33) use a double-bunched shape, with constrictions at both the front and the back. Double-bunched articulations of Viby-i have previously been reported in Westerberg (2016), and correspond well to the articulation of this sound as described by Borgström (1913), as will be addressed in Chapter 8. The effect of tongue shape on Viby-i acoustics will be tested indirectly through measures of tongue backness in Section 7.4. In addition to arched and double-bunched tongue shapes, front-bunching and back-bunching also occurs in the sample, but are less common (4 of 33 speakers respectively).

Fig. 7.23 shows the distribution of different tongue shapes for /i:/ across the city and gender groups. Overall, the shapes occur at similar rates across the sample, but front-bunched tongue shapes are slightly less common in males. Stockholm also appears to favour an arched tongue shape over the other two cities. Although the acoustic analysis showed that Gothenburg had a lower F2, and thus a stronger Viby-i, double-bunching does not appear to be more common in Gothenburg. This finding, coupled with the prevalence of arched tongue shapes in the sample, indicates that the production of Viby-i may not be as reliant on tongue shape as on the overall position of the tongue.

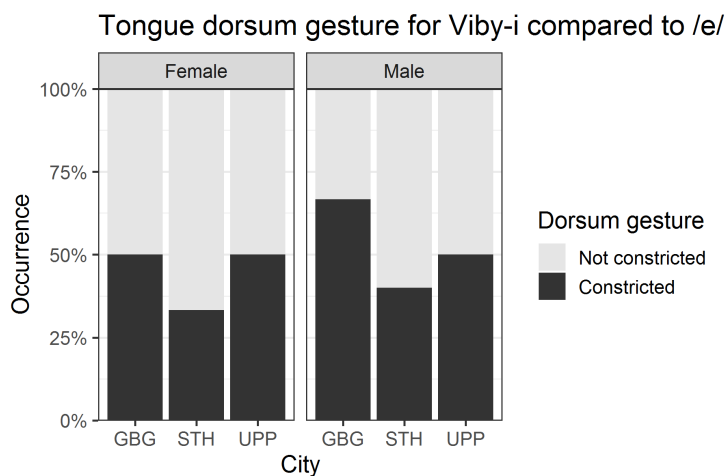


Figure 7.21: Distribution of post-dorsal constriction for /i:/ across the sample.

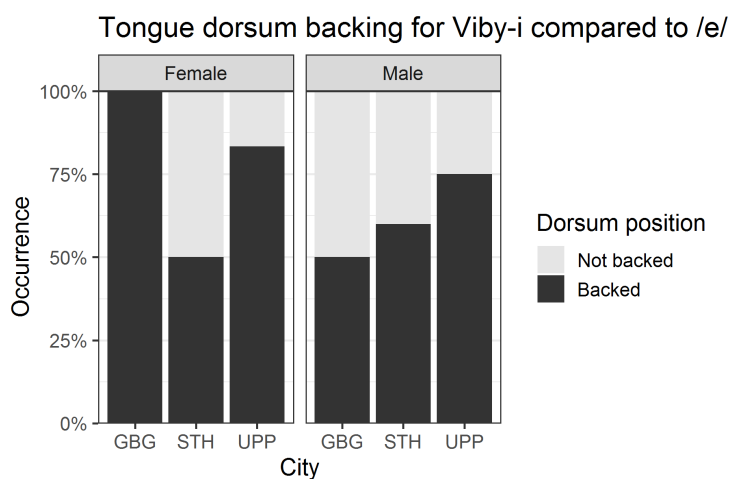


Figure 7.22: Distribution of post-dorsal backing for /i:/ compared to /e:/ across the sample.

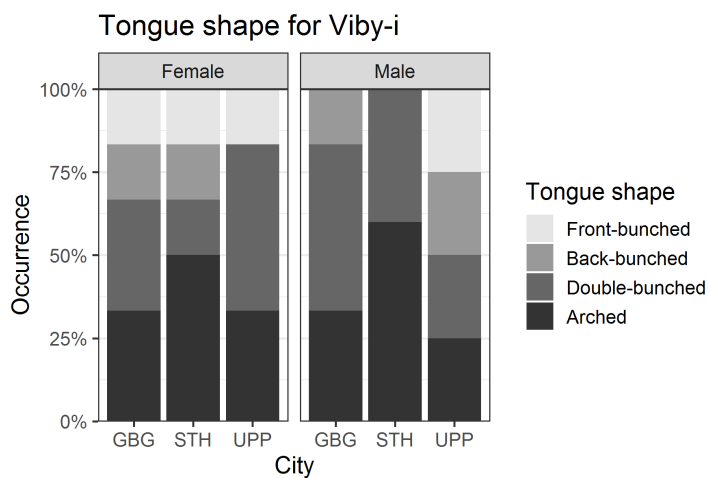


Figure 7.23: Distribution of tongue shapes for /i:/ across the sample.

### 7.2.5 Summary of mean tongue gestures

From this data, it appears that the characteristic tongue gesture for *Viby-i* includes a relatively fronted tongue body, which is lower than expected compared to the other vowels in the system. In particular, the tongue body for /i:/ tends to be lower than /e:/, and either similar to, or more fronted than /e:/. This lowering is especially common in the Gothenburg speakers. Although tongue lowering appears to correspond well to F1, the low F2 of *Viby-i* is not matched by backing of the highest point of the tongue. It may thus be the case that speakers are using different strategies to produce the low F2 of *Viby-i*.

Another prevalent pattern in the data is the backing and constriction of the post-dorsal part of the tongue towards the pharynx. In many speakers, this combination of tongue body fronting and post-dorsal retraction results in a distinct double-bunched shape, where the front and the back of the tongue are raised, but the middle is lowered. It is possible that this post-dorsal retraction helps *Viby-i* achieve the acoustically backed impression signalled by its low F2. However, a number of speakers also used arched tongue shapes to achieve an equivalent vowel quality. Thus, it seems possible to produce *Viby-i* through different articulatory strategies.

Despite *Viby-i* having a lower tongue body than /e:/ for most of the sample, the tongue tip for /i:/ is usually the highest in the system. This height can be achieved through either a tip-up or a tip-down gesture. The tongue tip is also frequently fronted, and close to the alveolar ridge or teeth, meaning that it is possible that speakers are using the tongue tip to generate frication during *Viby-i*. A summary of these results with regard to the individual speaker plots in Appendix G is provided in Table 7.1.

Speaker	Body		Tip		Post-dors.		Shape			
	Low	Front	Up	High	Constr.	Backed	Front bunch	Back bunch	Dbl. bunch	Arch
GF1	X		X		X	X		X		
GF2	X	X		X	X	X			X	
GF3	X	X		X		X				X
GF4	X	X		X	X	X			X	
GF5	X		X	X		X				X
GF6		X				X	X			
GM1	X	X	X							X
GM2	X	X			X				X	
GM3	X		X	X	X	X		X		
GM4	X	X	X	X	X	X			X	
GM5	X	X		X						X
GM6	X	X			X	X			X	
SF1	X		X	X		X				X
SF2		X	X	X	X				X	
SF3	X	X	X	X		X				X
SF4	X		X		X	X		X		
SF5	X	X	X							X
SF6		X	X	X			X			
SM1	X	X		X						X
SM2		X			X	X			X	
SM4	X	X		X		X				X
SM5		X	X		X	X			X	
SM6	X		X							X
UF1		X		X		X	X			
UF2	X			X		X				X
UF3	X		X			X				X
UF4		X		X	X				X	
UF5		X	X		X	X			X	
UF6		X	X		X	X			X	
UM1										X
UM2	X				X	X		X		
UM3	X	X		X	X	X			X	
UM4		X		X		X	X			
Total	65%	70%	48%	55%	48%	67%	12%	12%	36%	39%

Table 7.1: Summary of mean tongue spline patterns for /i:/ compared to /e:/ by individual speaker.

### 7.3 GAMM analysis of high front vowels

To investigate whether the visually observed patterns in the mean tongue splines were statistically significant, a set of GAMM models were generated of the high front vowels /i:/, y:/, e:/, ɛ:/, and tested for differences along the tongue curve. The full set of GAMM plots, and their respective difference smooths, are presented in Appendix H. Note that these plots still represent the static data at the 10% timepoint, in a similar way to the mean spline plots (see Fig. 7.24). In the GAMM plots, the palate is shown as a dotted line, and the approximate position of the front teeth is marked by a cross. The GAMM data omits eight speakers whose data was corrupted, bringing the sample size to 26.

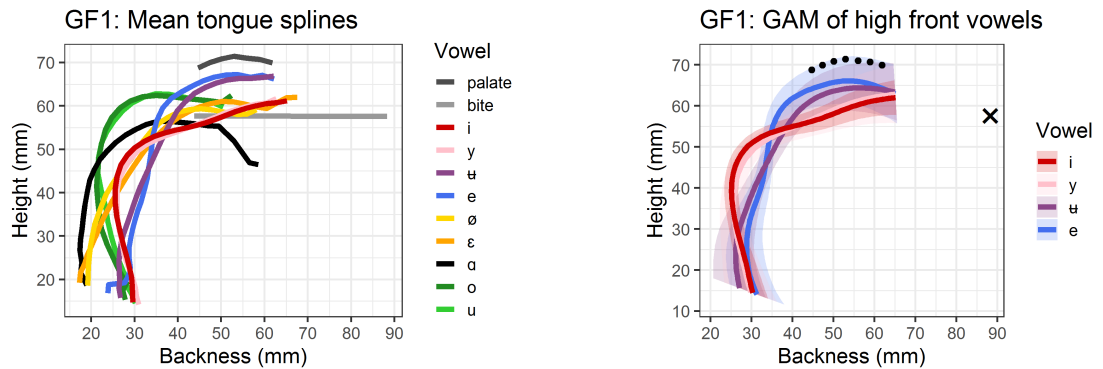


Figure 7.24: Comparison between mean spline plot and GAMM plot.

#### 7.3.1 /i:/ vs. /y:/

The GAMM plots show a similar relationship between the tongue curves for /i:/ and /y:/ as the mean spline data, in that the curves for the two vowels are usually superimposed on one another (Fig. 7.25). Similarly, the difference smooths comparing the /i:/ and /y:/ curves show that there is no significant difference between these two vowels at any point along the tongue curve, for any of the speakers, with the exception of SM5 (Fig. 7.26). For this speaker, there is a small but significant difference near the middle of the smooth, corresponding roughly to a small dip near the middle of the tongue body, where /i:/ is lower. Overall, however, the visual similarity between the tongue shapes for /i:/ and /y:/ is confirmed by the statistical analysis: Most of the time, the speakers in this sample use the same tongue gesture for /i:/ and /y:/. This result suggests that the main articulatory difference between /i:/ and /y:/ is likely to be lip posture, as described in the canonical literature on Swedish (e.g. Ladefoged & Maddieson, 1996: 295), or, if this is not the case, that the vowels overlap perceptually (Gross & Forsberg, 2019). The finding that /i:/ and /y:/ use the same tongue gesture also contrasts with Frid et al. (2015), whose speakers used separate tongue shapes for these vowels.



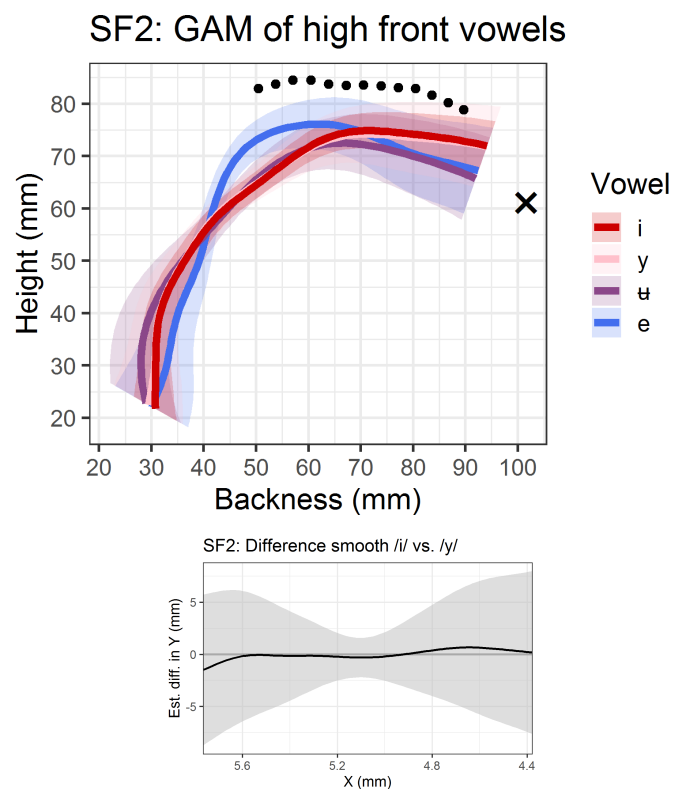


Figure 7.25: GAMM plot and difference smooth showing no significant difference between /i:/ and /y:/.

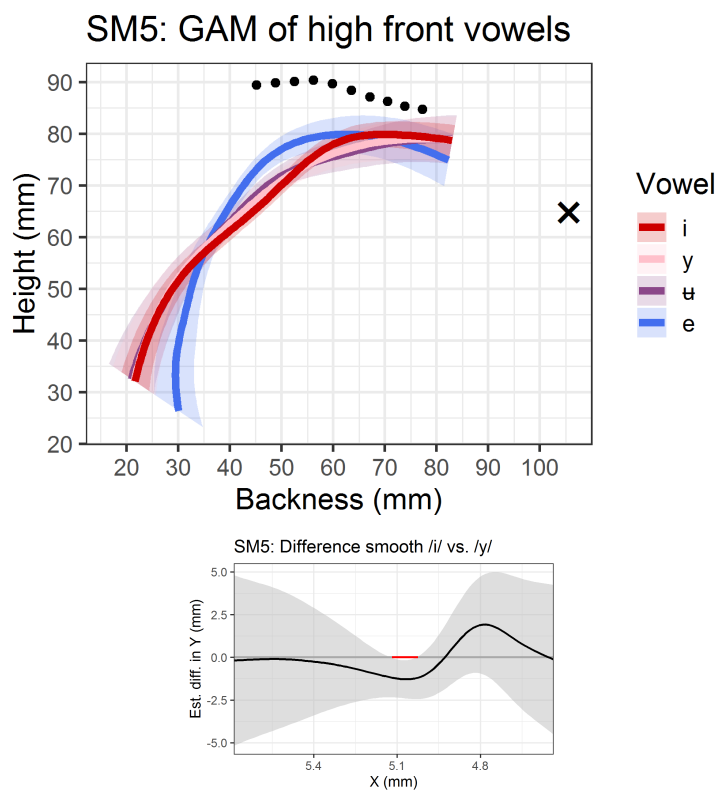


Figure 7.26: GAMM plot and difference smooth showing an interval of significant difference between /i:/ and /y:/.

### 7.3.2 /i/ vs. /e/

The characteristic gap between the tongue curves for /i:/ and /e:/ persists in the GAMM plots (Fig. 7.27, 7.28). The difference smooths show that the tongue curves for /i:/ and /e:/ are significantly different for at least one part of the tongue curve in all speakers except UF6, whose curves overlap entirely (Fig. 7.29). In most speakers (21 of 26), the difference smooths show a recurrent pattern where /i:/ is higher than /e:/ around the first half of the smooth, corresponding to post-dorsal retraction, and lower around the second half of the smooth, corresponding to tongue body lowering. In some speakers (7 of 25), there is an additional significant difference at the end of the smooth, signifying tongue tip raising for /i:/ (Fig. 7.28).

A small number of speakers (4 of 26) only had one interval of significant difference between /i:/ and /e:/, which usually occurred near the middle of the difference smooth. This difference manifested itself as a combination of tongue body lowering and fronting (Fig. 7.30). Overall, the GAMM plots for /i:/ and /e:/ correspond to the trends observed in the mean spline data, i.e. that /i:/ usually has a lower and fronter tongue body, a more retracted post-dorsal region, and a higher tip compared to /e:/.

### 7.3.3 Summary of GAMM analysis

The GAMM analysis has shown that the tongue lowering and post-dorsal retraction for Viby-i compared to /e:/ is statistically significant in most speakers, and that a high tongue tip is also common in the production of Viby-i. These results confirm the patterns observed in the mean splines. The GAMMs have also demonstrated that the tongue gestures for Viby-i and /y:/ are not significantly different from one another in this sample. These two vowels may nevertheless be distinguished by differences in lip posture, which will be addressed in Section 7.5.

A summary of the statistically significant results from this analysis is provided in Table 7.2.

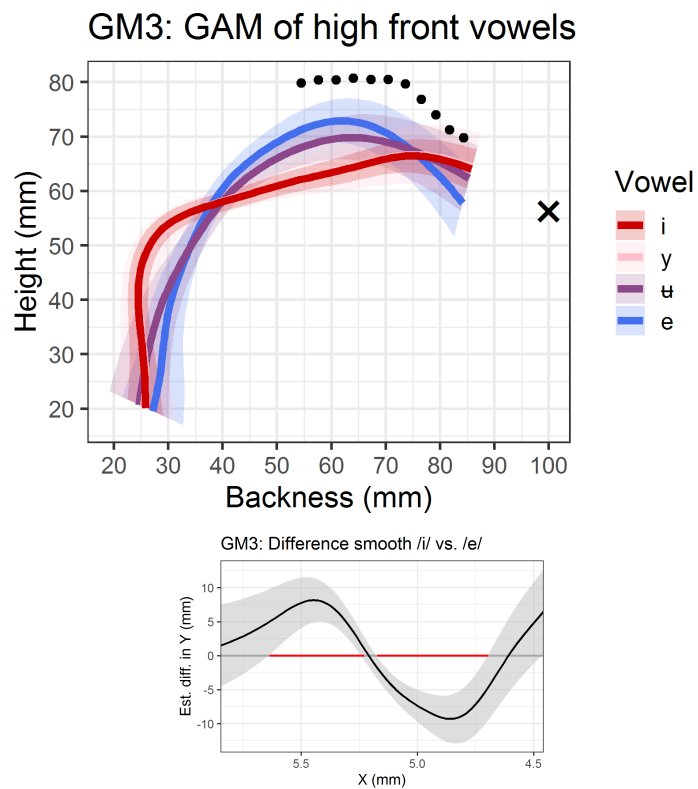


Figure 7.27: GAMM plot and difference smooth showing significant tongue retraction and lowering for /i:/ compared to /e:/.

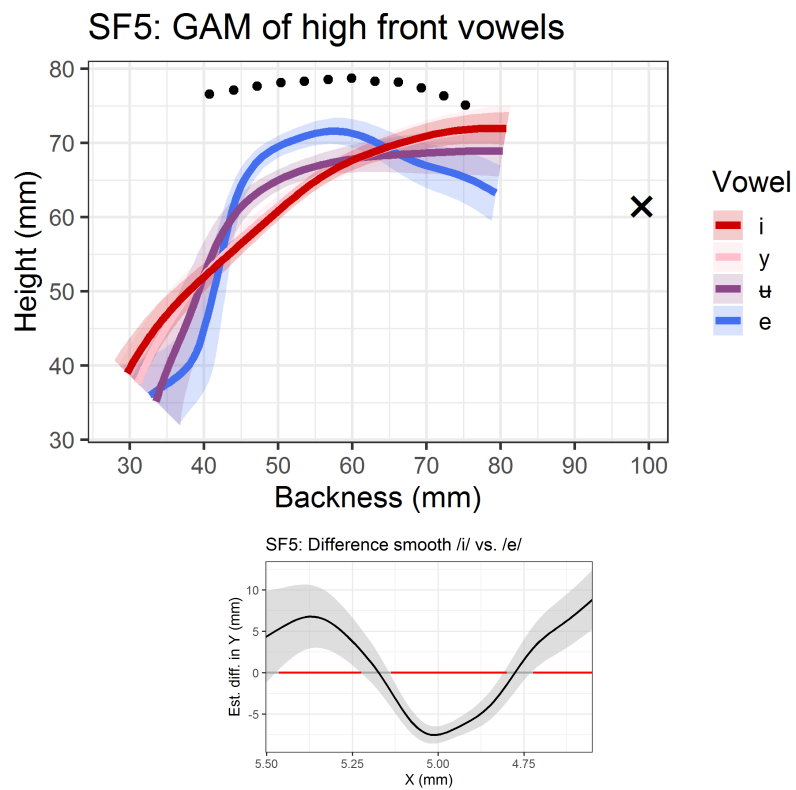


Figure 7.28: GAMM plot and difference smooth showing significant tongue retraction, lowering, and tip raising for /i:/ compared to /e:/.

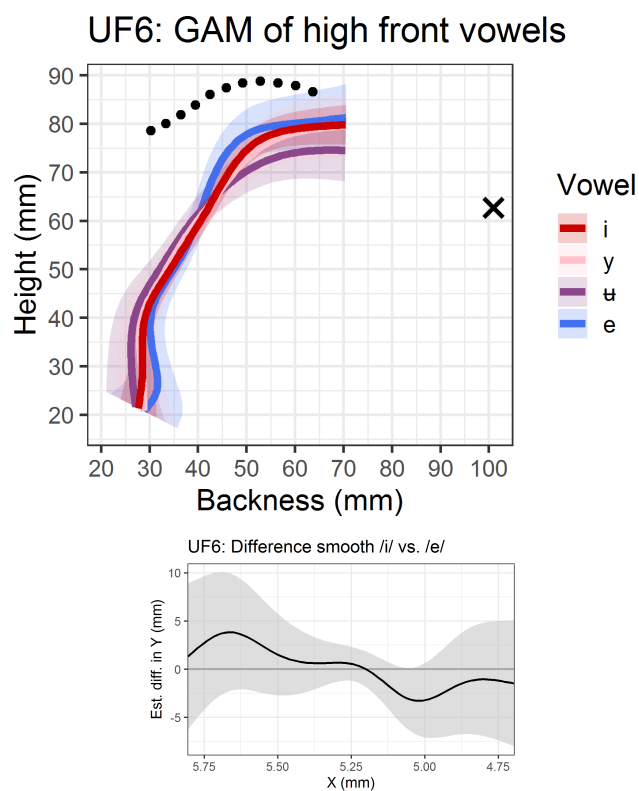


Figure 7.29: GAMM plot and difference smooth showing no significant difference between /i:/ and /e:/.

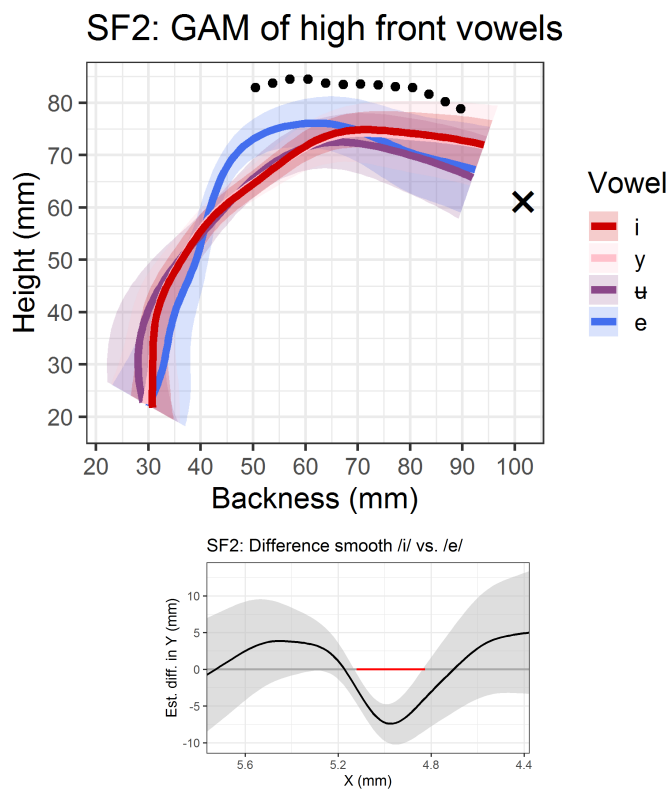


Figure 7.30: GAMM plot and difference smooth showing significant fronting/lowering of /i:/ compared to /e:/.

Speaker	<i>/i:/</i> vs. <i>/e:/</i>			<i>/i:/</i> vs. <i>/y:/</i>
	Significant retraction	Significant Lowering/fronting	Significant Tip-raising	Any sig. difference
GF1	X	X		
GF2	X	X		
GF3	X	X		
GF4	X	X		
GF6	X	X		
GM1		X		
GM2	X	X		
GM3	X	X	X	
GM4	X	X	X	
GM5	X	X	X	
GM6	X	X		
SF1	X	X		
SF2		X		
SF3	X	X	X	
SF5	X	X	X	
SF6		X		
SM2	X	X		
SM4	X	X	X	
SM5	X	X		X
SM6	X	X		
UF1	X	X	X	
UF2	X	X		
UF5	X	X		
UF6				
UM2	X	X	X	
UM3				
UM4		X		
Total	81%	96%	31%	4%

Table 7.2: Summary of GAMM tongue patterns for */i:/* compared to */y/, e:/*.

## 7.4 Normalised tongue point measures

As the previous sections have indicated, the highest point of the tongue for Viby-i does not seem to correspond directly to its low acoustic F2. To explore this issue, a number of articulatory single-point measures were used to investigate the correspondence between specific parts of the tongue and the resulting F1 and F2. As described in Section 5.9.3, the measures used in this analysis are:

- (A) Highest point of the tongue – ‘peak’
- (B) Frontest part of the tongue – ‘tip’
- (C) Backest part of the tongue – ‘backest’
- (D) Backest point at half tongue height – ‘midback’
- (E) Most constricted point (highest perpendicular point when a line is drawn between root and tip) – ‘max constriction’

An example of how these points correspond to the tongue curve is shown in Fig. 7.31. The measures consist of both x and y coordinates at each point. Following Lawson et al. (2019), the measures were Lobanov normalised (Lobanov, 1971) to enable comparisons between multiple speakers.

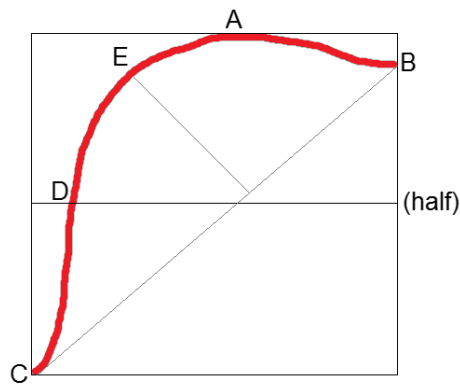


Figure 7.31: Measurement points used for the normalised articulatory analysis.

### 7.4.1 Highest point of the tongue

The highest point of the tongue has been a conventional measure of tongue body position since the earliest phonetic work on vowels (e.g. Bell, 1867; D. Jones, 1917). The analysis therefore begins by exploring how the highest point of the tongue in the current dataset compares to the acoustic F1 and F2 of Viby-i. Fig. 7.32 shows the normalised vowel space from the acoustic analysis, and Fig. 7.33 shows the corresponding values for the highest point of the tongue.

For ease of comparing the acoustic and articulatory data, the tongue point measurements are presented in the same orientation as the vowel quadrilateral (facing left).

The first noticeable difference between the acoustic and articulatory datasets is that the shape of the vowel space is different. The articulatory space is smaller, particularly in the vertical dimension, leading to more overlap between different vowel phonemes. It thus appears that small changes in the vocal tract can have relatively large acoustic consequences. Secondly, the overlap between different vowels in the articulatory space indicates that vowel contrasts are probably maintained using several factors, not just the highest point of the tongue. For example, vowels that have a similar highest point may still exhibit e.g. different gestures at other parts of the tongue curve, or complementary lip-rounding. Overall, the spatial relationships between the vowels in the articulatory data resemble those in the acoustic data, with the notable exception of /i:/ and /e:/.

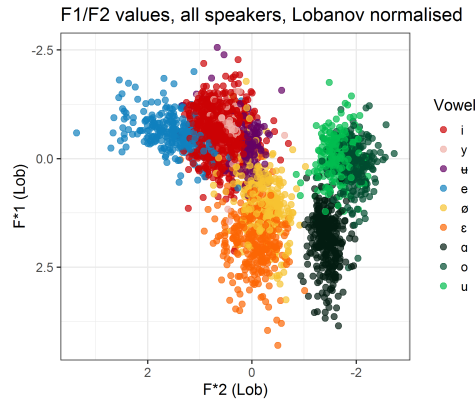


Figure 7.32: Normalised F1 and F2 values for all vowels and speakers.

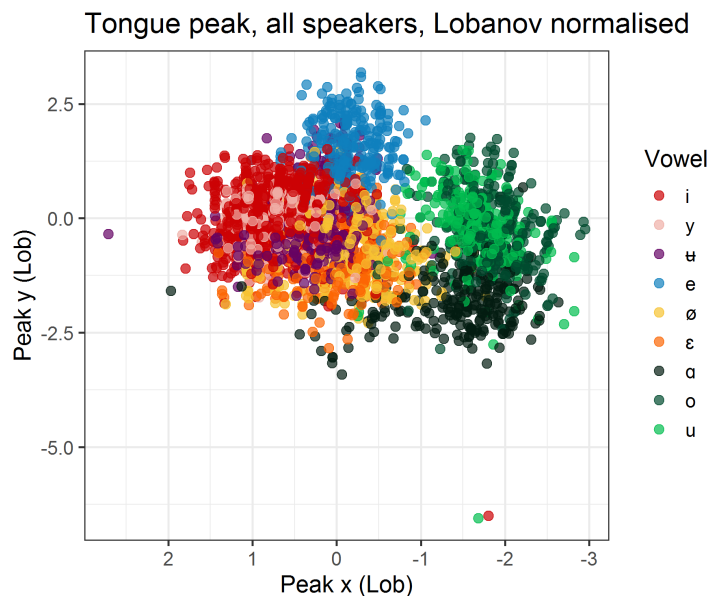


Figure 7.33: Normalised highest point of the tongue for all vowels and speakers.

In the acoustic data, /i:/ and /e:/ had a similar F1, but /i:/ had a lower F2. We would therefore expect these vowels to have a similar tongue height, but for /i:/ to have a backer tongue body. Instead, the articulatory data shows that the highest point of the tongue for /i:/ is both lower and fronter than /e:/. This mismatch between acoustics and articulation indicates that the highest point does not adequately capture the overall tongue gesture of Viby-i, and suggests the use of a more complex gesture than traditionally attributed to vowels in the general literature.

The highest point data also relates to the previous investigation of whether /e:/ has shifted into the space where a high front /i:/ would normally be (see Section 6.3). In line with previous findings, the articulatory results support that /e:/ has not shifted forward, but that it may have shifted up. However, as we do not have any articulatory data for standard [i:], it is difficult to estimate how high /e:/ actually is. To answer this question, more articulatory data would be required, preferably from other languages, or from other varieties of Swedish.

### 7.4.2 Most constricted point of the tongue

As an alternative to the highest point of the tongue, a measure of ‘maximum constriction’ is also investigated. This measure is similar to the highest point, but is not limited to the vertical dimension. Instead, it corresponds to the point where the tongue would be highest if the spline was rotated so that the start and end points of the spline were both level. In this way, the measure approximates the place of maximum constriction in the vocal tract. To some extent, this may resemble historical data of the highest point of the tongue, since X-ray images of the tongue are often rotated forward compared to UTI images (Fig. 7.34).

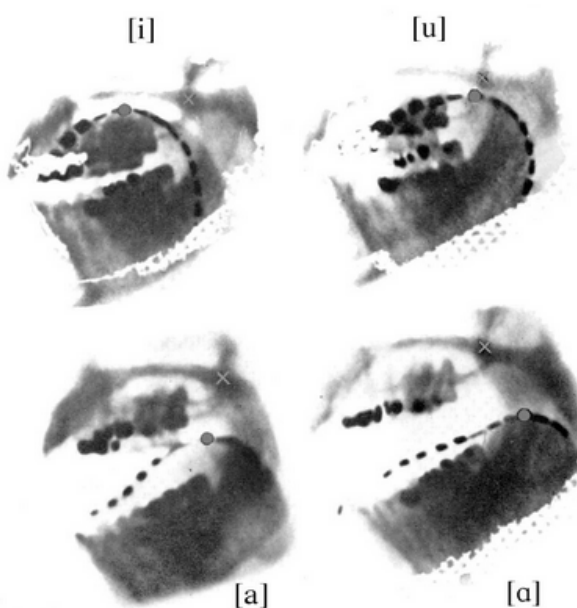


Figure 7.34: X-rays showing highest point of the tongue for Cardinal Vowels (after D. Jones, 1972: frontispiece). Note the slight forward tilt.



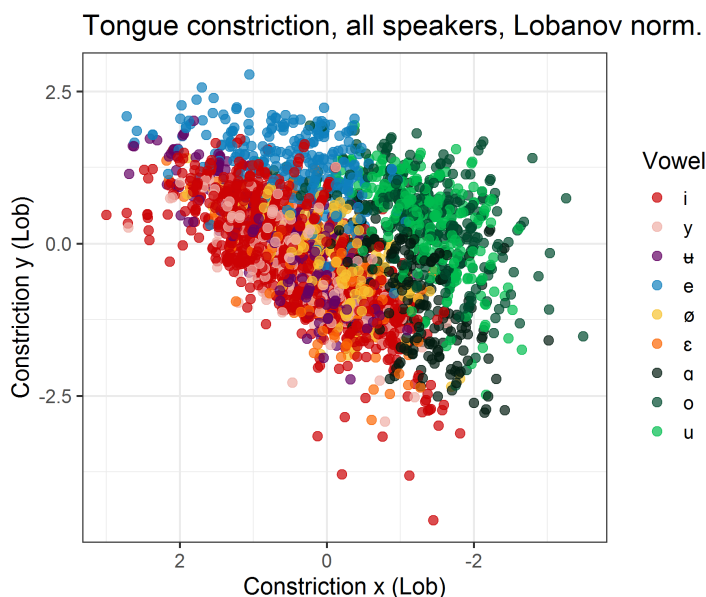


Figure 7.35: Normalised most constricted point of the tongue for all vowels and speakers.

The normalised maximum constriction data is presented in Fig. 7.35. Compared to the highest point, maximum constriction corresponds slightly worse to the acoustic data: The data points for each vowel are less clustered, and do not correspond particularly well to the F1/F2 plot, although the relationship between the vowels is roughly the same as for the highest point.

One noticeable pattern in the maximum constriction plot is that /i:/ is particularly variable in both the horizontal and vertical dimensions compared to the other vowels. In other words, the main constriction can be as high and front as /e:/, or as low and back as /ɑ:/. High constrictions tend to be fronter, while low constrictions tend to be backer. This variability could result from the fact that /i:/ was produced in a greater number of word contexts than the other vowels. However, it is also possible that the maximum constriction measure is capturing different parts of the tongue for different speakers, due to variations in tongue shape. For example, a speaker with an arched tongue shape would have a backer and lower main constriction than a speaker with a front-bunched tongue shape.

Although the presence of lower, backer tokens in the maximum constriction data shows that /i:/ is sometimes produced with tongue retraction – information that was not accessible from the highest point measure – maximum constriction is still not able to capture the parts of the tongue that are responsible for the observed F1 and F2 values. Indeed, it seems unlikely that any single point would be able to express the complex articulatory-acoustic relationship that results in Viby-i. However, it is possible that several measures could be combined to quantify tongue gesture in a more holistic way.

### 7.4.3 Retraction measures

Although the highest point and maximum constriction measures were able to capture some aspects of the tongue gesture for Viby-i, neither was particularly successful in demonstrating why the acoustic data is characterised by F2-lowering. To investigate this issue, two measures were taken to quantify post-dorsal retraction: The first, ‘backest’, simply measures the backest part of the tongue spline. This point could be located at the root, but also higher up on the tongue back, as captured by the y-value for this measure. The second, ‘midback’, samples the backness of the spline at half of the tongue’s total height.

Representations of the vowel space using the ‘backest’ and ‘midback’ measures are presented in Figs. 7.36 and 7.37. Both plots show a better correspondence with the acoustic data in terms of tongue retraction than the previous plots, although /i:/ for the most part overlaps with /e:/. Thus, the degree of tongue backing implied by the acoustic data is still not accounted for by the articulatory measures.

In terms of height, the backest point for /i:/ tends to be fairly low, indicating that the measurements were usually taken near the tongue root. Vertical variation in this measure also implies differences in tongue shape; higher tokens are more likely to correspond to an arched shape, while lower tokens signal constriction at the lower post-dorsal region, i.e. back-bunching or double-bunching.

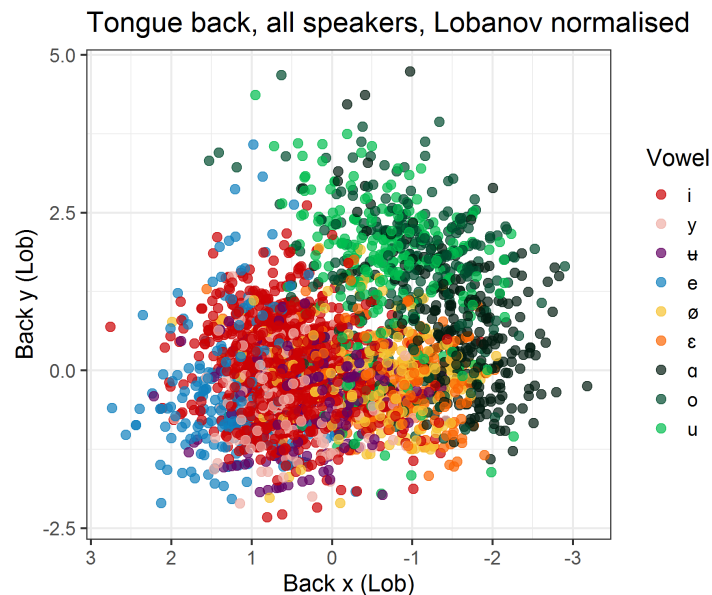


Figure 7.36: Normalised backest point of the tongue for all vowels and speakers.

The midback measures show a similar picture, but because all samples were taken at a similar height, the vertical dimension is more compressed, and potentially less informative in terms of tongue shape. Instead, the y-value for this measure reveals more about the overall height of the tongue, since this is the factor that determined where the midback measure was taken. For this reason, the vertical dimension of the midback plot resembles that of the highest point plot. The

combination of vertical data from the highest point, and horizontal data from the midback, has been shown by Lawson et al. (2019) to be more informative of the overall tongue gesture for vowels than the highest point by itself. However, in the current sample, the midback measure is only able to account for some of the acoustic characteristics of Viby-i.

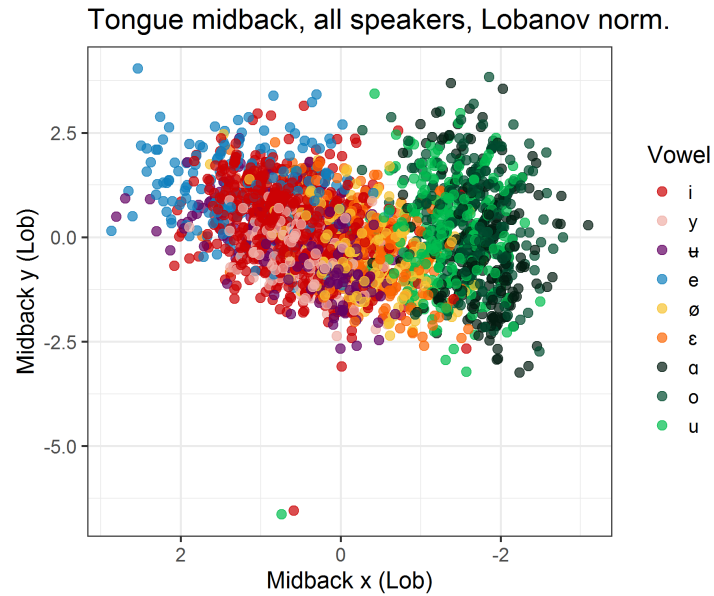


Figure 7.37: Normalised midback tongue point for all vowels and speakers.

### 7.4.4 Tongue tip measure

The final articulatory measure investigated here is the frontest part of the tongue, i.e. the observed tongue tip. This measure is primarily of interest to examine the source of fricative noise during Viby-i. However, it could also have an influence on the formants.

Fig. 7.38 shows the distribution of the tongue tip measures in the current sample. This data clearly differs from the other measures examined here, but the position of the tongue tip nevertheless seems to be influenced by overall tongue position, as variation in tongue tip height roughly follows the order of the cardinal vowels. This phenomenon could also be affected by tongue shape. For example, a vowel like /u:/ is relatively low and back, but its shape also tends to be more clearly arched, meaning that the tongue tip is pointing down. This arching could explain why the tongue tip for /ɑ:/, which usually has a flatter tongue shape, is higher than /o:/ and /u:/.

In terms of tongue tip backing, it is unsurprising to see that the back vowels have backer tongue tips, but there is also an unexpected amount of horizontal variation in /i:, y:, ʌ:, e:/ . This variation could be caused by differences in spline lengths; if the spline is short because of poor visibility on the ultrasound, the tongue tip measurement will appear backed. However, it is also possible that horizontal variation in tongue tip behaviour reflects overall tongue backing, which could have acoustic consequences. The correspondence between tongue tip position and acoustic output will be investigated further in Section 7.6. The effect of tongue tip gesture on frication will be studied in Section 7.7.3.

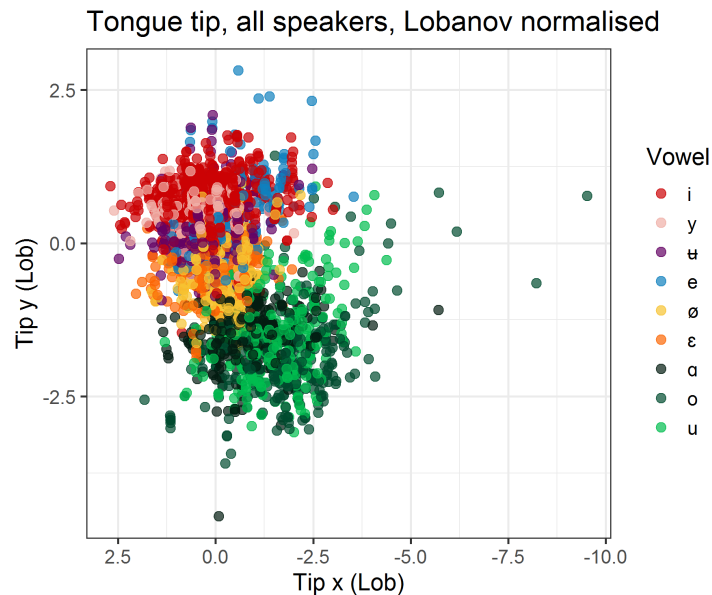


Figure 7.38: Normalised tongue tip measures for all vowels and speakers.

## 7.5 Lip posture

Since none of the tongue point measures were able to fully capture the correspondence between the articulation of Viby-i and its acoustic output, it is relevant to examine how lip posture contributes to this vowel. Due to time constraints, it was not possible to obtain quantitative measurements of the lip data, but a visual inspection will be carried out here.

Some representative examples of lip postures for /i:/, y:/, e:/ are provided in Figs. 7.39-7.42. The pictures were taken at 50% of the vowel duration, since most of the pictures at 10% were obscured by motion blur. The lip images represent two speakers of each gender who produced Viby-i with a low F2 (UF2, UM2), and two speakers who had a relatively high F2 (SF5, UM4). All speakers' lip images for these vowels are available in Appendix K.

As the figures show, there is no visible tendency for speakers to use lip-rounding when producing Viby-i. At the same time, the lips are not as spread as expected for a standard [i:] based on Engstrand (2004: 98). Overall, the lip posture for /i:/ is similar to /e:/, but with a slightly smaller mouth opening. As in Westerberg (2016) and Gross and Forsberg (2019), the lip posture for Viby-i can thus be described as spread, but with a tendency towards a more 'lax' or neutral posture. Since the lips for Viby-i are not rounded, it is unlikely that they contribute to the lowering of F2 observed in the acoustics. For the purposes of this analysis, it will therefore be assumed that the formant values for Viby-i are mainly influenced by tongue gesture, although future research would benefit from investigating the effect of other articulators in more detail.

To address the question of articulatory overlap between /i:/ and /y:/, the data shows that all speakers in the sample produce /y:/ with visible lip-rounding. Thus, the overlap in tongue gesture is somewhat compensated for by the lips. However, the tongue may still affect /y:/ in terms of Viby-colouring, which could possibly explain some of the perceptual overlap between /i:/ and /y:/ reported by Gross and Forsberg (2019). In person, listeners are nevertheless likely to have the visual cue of lip-rounding to help them disambiguate these two vowels. Lip-rounding can also account for the significantly lower F3 and F4 values observed for /y:/ compared to /i:/ in Section 6.5.

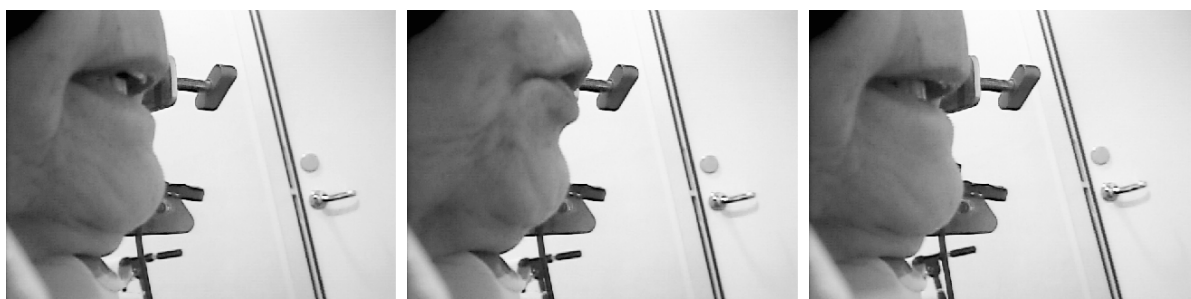


Figure 7.39: Lip postures for /i:/, y:/, e:/ in a female speaker with a high F2.



Figure 7.40: Lip postures for /i:, y:, e:/ in a male speaker with a high F2.



Figure 7.41: Lip postures for /i:, y:, e:/ in a female speaker with a low F2.



Figure 7.42: Lip postures for /i:, y:, e:/ in a male speaker with a low F2.

## 7.6 Effect of tongue point measures on F1 and F2

Since the tongue point measures and the lip data were both unable to explain the apparent mismatch between the articulatory gesture for Viby-i and its resulting acoustic output, this section statistically tests the correspondence between the normalised tongue point measures and the first two acoustic formants (also normalised). The goal is to identify possible articulatory parameters that could inform how the low F2 and high F1 of Viby-i are achieved.

### 7.6.1 Correlation between highest point and F1/F2 for all vowels

Before investigating Viby-i specifically, it is informative to examine how well the conventional measure of the highest point of the tongue corresponds to F1 and F2 overall. Two LMER models

were therefore constructed using data from all nine vowels /i:, y:, ʊ:, e:, ø:, ɛ:, a:, o:, u:/ at the 10% timepoint. The first model uses F1 as the dependent variable, and the highest point of the tongue (y) as the independent variable. The second model uses F2 as the dependent variable, and the highest point of the tongue (x) as the independent variable. Both models include ‘speaker’ as a random intercept. The results are provided in Table 7.3, and illustrated in Figs. 7.43 and 7.44.

	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
peak_y_norm	−0.607*** (0.003)	
peak_x_norm		0.769*** (0.002)
Constant	−0.109*** (0.013)	0.011 (0.012)
Observations	115,751	115,751

*Note:* \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Table 7.3: LMER output showing the effect of the highest point of the tongue on normalised F1 and F2 for the full set of vowels.

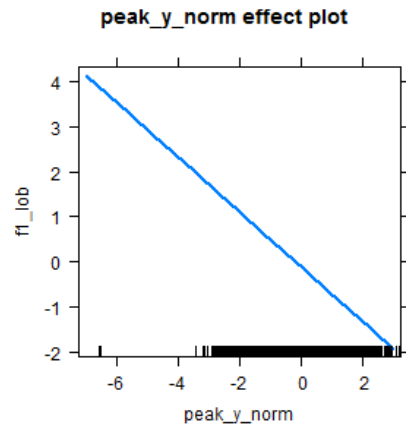


Figure 7.43: Significant negative correlation between tongue lowering (highest point) and F1 for the full vowel set.

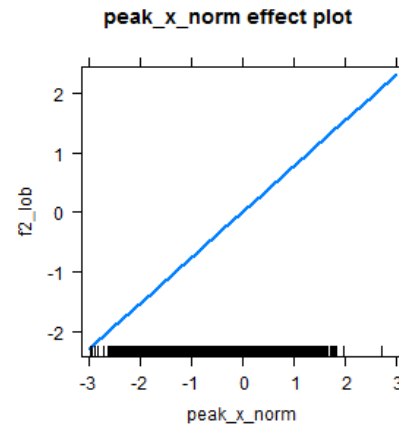


Figure 7.44: Significant positive correlation between tongue backing (highest point) and F2 for the full vowel set.

The models show that, for the dataset as a whole, there is a correlation between the highest point of the tongue and F1/F2. These relationships follow the pattern described in the literature (e.g. Delattre, 1951): Tongue height is negatively correlated with F1 ( $\beta=-0.61$ ,  $p<0.001$ ), and tongue frontness is positively correlated with F2 ( $\beta=0.77$ ,  $p<0.001$ ). Both relationships are statistically significant and have similar coefficients, indicating that the relationship between tongue height and F1 is roughly as strong as the relationship between F2 and tongue backness.

Thus, the highest point is able to capture the expected correspondence between articulation and acoustics in the overall vowel sample.

### 7.6.2 Correlation between highest point and F1/F2 for Viby-i

To examine whether the expected relationships still hold true for Viby-i, the same models were run on a subset of the data, consisting of only /i:/ at the 10% timepoint. The model summaries are shown in Table 7.4, with the significant effects shown in Figs. 7.45 and 7.46. In these models, the highest point of the tongue retains a similar coefficient and level of significance with F1 as before ( $\beta=-0.1$ ,  $p<0.001$ ). However, while the relationship between the highest point and F2 remains significant, this coefficient has become weaker, and changed direction ( $\beta=-0.01$ ,  $p<0.001$ ). In other words, in the case of Viby-i, F1 corresponds roughly to the highest point of the tongue, but F2 instead *increases* slightly with tongue backing, which is the opposite of the pattern described in the vowel literature (e.g. Delattre, 1951; Stevens & House, 1955).

	<i>Dependent variable:</i>	
	f1_lob	f2_lob
	(1)	(2)
peak_y_norm	-0.104*** (0.004)	
peak_x_norm		-0.010*** (0.003)
Constant	-0.537*** (0.022)	0.547*** (0.023)
Observations	48,319	48,319
<i>Note:</i> * $p<0.05$ ; ** $p<0.01$ ; *** $p<0.001$		

Table 7.4: LMER output showing the effect of the highest point of the tongue on normalised F1/F2 for /i:/.



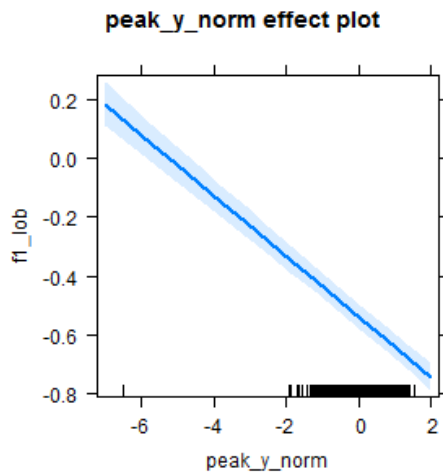


Figure 7.45: Significant negative correlation between tongue lowering (highest point) and F1 for /i:/.

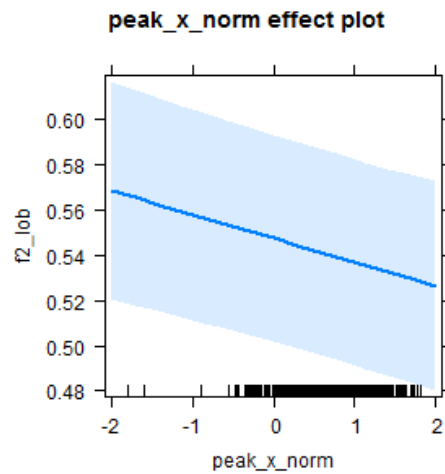


Figure 7.46: Significant negative correlation between tongue backing (highest point) and F2 for /i:/.

### 7.6.3 Correlation between multiple tongue measures and F1/F2 for Viby-i

The unexpected correlation between F2 and *fronting* of the highest point of the tongue for Viby-i supports the suggestion made in Section 7.4.1 that the highest point does not adequately capture the articulatory parameters that are important for this vowel. Specifically, the highest point measure cannot account for the low F2 of Viby-i, since this vowel is usually produced with a fronted tongue body. In addition, the fact that tongue fronting contributes to a *lower* F2 seems to indicate that speakers are using compensatory strategies to work against the raising effect on F2 that normally occurs when the tongue is fronted. To examine whether any of these compensatory gestures can be identified, this section carries out an exploratory statistical analysis of the normalised tongue measures presented in Section 7.4, to investigate whether they affect the F2 of Viby-i specifically.

#### Variable selection

Before constructing the model, the variables were tested for multicollinearity using the Pearson method. The following variables were entered into the correlation test:

- Peak (x, y)
- Tip (x, y)
- Backest (x, y)
- Midback (x, y)
- Max constriction (x, y)
- Tongue shape dummies (bbunch, fbunch)

The tongue shape dummy variables were based on the tongue shape classifications made in Section 7.2. Each tongue spline was assigned a binary value for front-bunching and back-bunching. Arched tongue shapes had a zero in both categories, while double-bunched shapes had a one in both categories.

The results of the correlation test are illustrated in Fig. 7.47. Based on significant correlations ( $p < 0.05$ ) within this set, ‘peak\_y’, ‘midback\_x’, ‘constr\_y’, and ‘bbunch’ had to be excluded from the analysis. Thus the final model included: the backness of the highest point of the tongue, both tip measures, both backest measures, midback height (corresponding to overall tongue height), backness of maximum constriction, and front-bunching. As before, ‘speaker’ was also used as a random intercept.

Since this analysis has an exploratory focus, and the variables entered into the model are not predicted to have a specific outcome on F2, the variables were entered into a ‘step’ model to eliminate those that did not contribute to explaining the data. The model eliminated ‘fbunch’, but kept all the remaining factors.

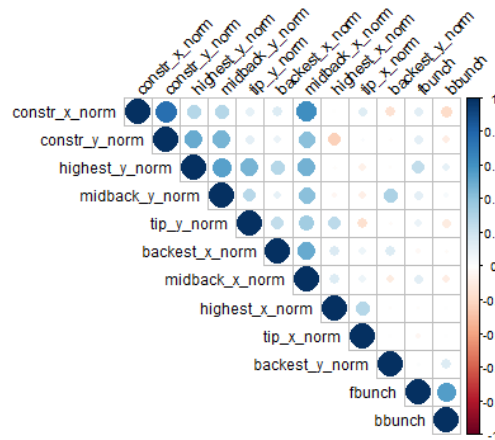


Figure 7.47: Correlation plot of tongue point measures.

### Final model

The output of the final model, comparing the F2 of Viby-i to the articulatory measures, is shown in Table 7.5. The significant effects are shown in Figs. 7.48 and 7.49. In these plots, higher x-values correspond to either fronting or raising in the articulatory space.

The model shows that, firstly, a large number of tongue parameters appear to contribute to the production of F2 for Viby-i. Although all factors entered into the final model predicted F2 with a high level of significance ( $p < 0.001$ ), their coefficients vary in both direction and size.

For ease of reading, the following factors (in order of size) are associated with F2-*lowering* for Viby-i:

- Retracting the backest part of the tongue ( $\beta=0.14$ ,  $p<0.001$ )
- Raising the backest part of the tongue ( $\beta=-0.06$ ,  $p<0.001$ )
- Fronting the highest point of the tongue ( $\beta=-0.05$ ,  $p<0.001$ )
- Retracting the main constriction of the tongue ( $\beta=-0.04$ ,  $p<0.001$ )
- Lowering the tongue tip ( $\beta=-0.04$ ,  $p<0.001$ )
- Fronting the tongue tip ( $\beta=-0.02$ ,  $p<0.001$ )
- Lowering the midback, i.e. lowering the tongue ( $\beta=-0.01$ ,  $p<0.001$ )

The results of the model thus appear to correspond well with the observations made in previous sections about the defining characteristics of the Viby-i tongue shape: The tongue body is low and fronted, with a fronted tongue tip, and the post-dorsal region is retracted and often constricted. This constriction usually takes place at a higher point than the tongue root. These factors are all correlated with a lower F2, although this analysis cannot determine their causality; for example, fronting the tongue body is unlikely to lower F2 in itself, but it may correlate with other gestures that do. Similarly, raising of the tongue tip was found by Björsten and Engstrand (1999) to lower F2, which is the opposite of the pattern found here.

Based on these findings, however, there is evidence to support that the low F2 of Viby-i is the result of multiple parts of the vocal tract working together. Even though the highest point of the tongue is relatively fronted, which should produce a *high* F2, the cumulative effect of additional tongue lowering and post-dorsal backing/constriction appears to cancel out the effects of overall tongue fronting. Further discussion of how these negative gestures may affect the formants will be provided in Chapter 8.

	<i>Dependent variable:</i>
	f2_lob
peak_x_norm	−0.050*** (0.003)
tip_x_norm	−0.023*** (0.002)
tip_y_norm	0.038*** (0.004)
backest_x_norm	0.139*** (0.002)
backest_y_norm	−0.057*** (0.002)
midback_y_norm	0.011*** (0.001)
constr_x_norm	0.036*** (0.002)
Constant	0.480*** (0.020)
Observations	48,319
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001	

Table 7.5: LMER output showing the effect of normalised tongue point measures on F2 for Viby-i.

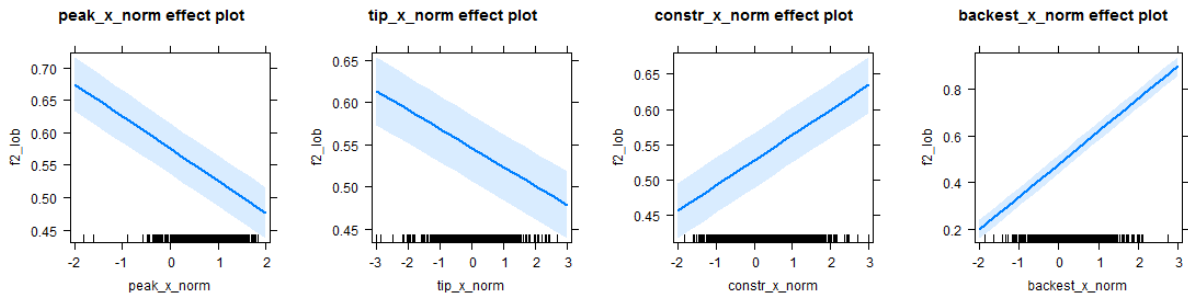


Figure 7.48: Horizontal tongue point measures significantly correlated with F2 for Viby-i.

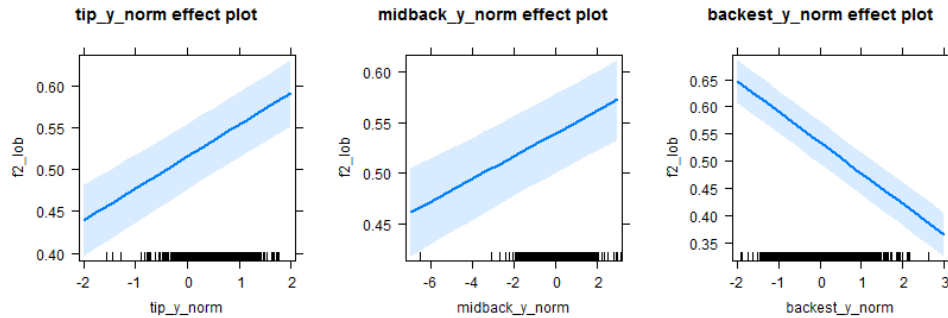


Figure 7.49: Vertical tongue point measures significantly correlated with F2 for Viby-i.

### 7.6.4 Sociolinguistic variation in tongue gesture

Although very few of the predicted social factors were found to affect Viby-i in the acoustic analysis, it is still possible that socially conditioned variation could exist covertly, i.e. in the form of ‘hidden’ tongue gesture variation (Lawson et al., 2014). To investigate this, three LMER models were constructed, using the dependent variables ‘peak\_y’, ‘tip\_y’, and ‘constr\_x’. These measures were selected because they had been found to correlate significantly with F2, and also to contribute visually to the characteristic tongue shape of Viby-i, as observed in Section 7.2. The independent variables entered into these models were ‘age’, ‘gender’, ‘city’, and interactions between the three. These were the social factors that, based on the literature (e.g. Bruce, 2010; Elert, 1995; Kotsinas, 2007), were most likely to affect production of Viby-i in general. Random intercepts were used for ‘speaker’ and ‘word’. Again, the models were ‘stepped’ to eliminate any variables that were not meaningful.

#### Tongue lowering

The first model, investigating the correlation between social factors and the height of the tongue peak, eliminated ‘age’ and ‘gender’ from the model, but kept ‘city’ ( $F = 7.4$ ,  $p = 0.002$ ). The model output is shown in Table 7.6, with the re-levelled data at the bottom of the table. The results are also visualised in Fig. 7.50. The model shows that, in terms of tongue peak height, Gothenburg is significantly lower than Stockholm ( $\beta = 0.44$ ,  $p < 0.006$ ) and Uppsala ( $\beta = 0.53$ ,  $p < 0.002$ ), while Stockholm and Uppsala pattern together. This result is similar to the acoustic finding that F1 is higher in Gothenburg than in Stockholm and Uppsala. It is also possible that,

since the previous analysis found a correlation between F2 and tongue lowering, this result also corresponds to the significantly lower F2 in Gothenburg compared to the other two cities.

Baseline: Gothenburg	<i>Dependent variable:</i>
	highest_y_norm
city-STH	0.440** (0.146)
city-UPP	0.535** (0.156)
Constant	−0.287** (0.105)
city-STH/UPP	0.095 (0.162)
Observations	48,319

*Note:* \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

Table 7.6: LMER output showing the effect of city on the height of the tongue peak.

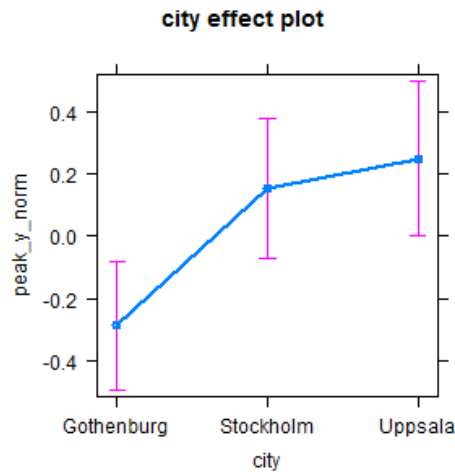


Figure 7.50: Significant correlation between city of origin and tongue peak height.

### Tongue tip raising

The second model investigates the correlation between social factors and the height of the tongue tip. Although tongue tip raising was found to contribute to a *higher* F2 for Viby-i, this factor could nevertheless be connected to different levels of frication in this vowel, which may be socially conditioned. However, after ‘stepping’ this model, only the random factors remained. In other words, none of the social categories were able to predict tongue tip behaviour in any meaningful way. Thus, we can conclude that there is no correlation between tongue tip raising

and age, gender, or city, and that the sample contained unstructured variation conditioned by individual speakers, and by the linguistic context in which the vowel appeared.

### Post-dorsal constriction

The third and final model compares the backness of the maximum constriction point across the social factors. The step model kept two variables: ‘age’ and ‘city’. The model output is presented in Table 7.7, with the significant factors illustrated in Fig. 7.51. The model shows that the maximum tongue constriction was significantly backer in younger speakers than in older speakers ( $\beta=0.007$ ,  $p<0.04$ ), regardless of their city or gender. Gothenburg also had a more backed tongue constriction than Stockholm ( $\beta=0.36$ ,  $p<0.02$ ) and Uppsala ( $\beta=0.40$ ,  $p<0.02$ ). Again, Uppsala and Stockholm were not significantly different from each other.

The fact that the age result has a much smaller coefficient than the city result could account for the fact that there was no apparent difference in F2 between young and old speakers in the acoustic analysis. However, similarly to tongue lowering, post-dorsal constriction also appears to contribute to the low F2 achieved by Gothenburg speakers, while being used to a lesser degree by speakers from Stockholm and Uppsala.

Baseline: Gothenburg	<i>Dependent variable:</i>
	constr_x_norm
age	0.007* (0.003)
city-STH	0.360* (0.139)
city-UPP	0.395** (0.149)
Constant	−0.158 (0.155)
city-STH/UPP	0.035 (0.154)
Observations	48,319
<i>Note:</i> * $p<0.05$ ; ** $p<0.01$ ; *** $p<0.001$	

Table 7.7: LMER output showing the effect of age and city on the backness of the main tongue constriction.

### Summary of social effects on tongue gesture

In summary, sociolinguistic factors appear to have some influence on the tongue gesture used for *Viby-i*. The Gothenburg speakers, who were also found to have a higher F1 and lower F2 in the acoustic analysis (and thus, a stronger *Viby-i*), had significantly greater degrees of tongue lowering, and a backer main constriction than Stockholm and Uppsala. Younger speakers also had a more backed constriction than older speakers, but the size of this effect was relatively small. However, the difference between younger and older speakers could potentially be interpreted as a sign that the articulation of *Viby-i* has changed in apparent time. Based on the mean spline data in Appendix G, it indeed appears that younger speakers are more likely to use a back-bunched or double-bunched tongue shape, compared to older speakers, who seem to prefer front-bunched or arched tongue shapes.

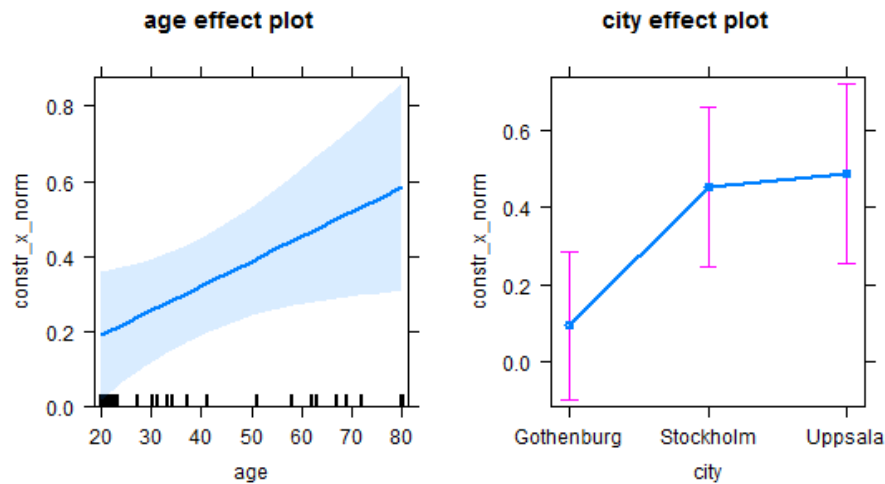


Figure 7.51: Significant correlation between city of origin, age, and post-dorsal constriction.



## 7.7 Tongue dynamics of Viby-i

This section carries out an abridged version of the analysis provided for the static data in the previous sections, using data from the 10%, 50%, and 90% timepoints to provide a dynamic perspective of how Viby-i is realised over time. Since Section 6.9 established that the formant trajectories for Viby-i are more or less stable, this analysis will primarily focus on factors that could identify sources of frication during Viby-i.

### 7.7.1 Visual inspection of dynamic plots

The full set of mean dynamic tongue splines is available in Appendix I, with corresponding GAMM trajectories and difference smooths available in Appendix J. A visual inspection of the mean splines shows that the dynamic tongue gestures can roughly be divided into two categories, based on the following consonant: Vowel tokens followed by a non-lingual /p, b/ or front /t, l, r, s, n show very little tongue movement over the course of the vowel, while tokens followed by a back /k, g/ show a raising or backing of the tongue body towards the palate at the 90% timepoint. For most of the back consonant contexts, the splines at 10% and 50% are virtually the same, indicating that the coarticulatory gesture towards the velum occurs quite late in the vowel.

Examples of the two dynamic categories (front and back), as seen in the mean spline data, are provided in Figs. 7.52 and 7.53. As the figures show, over the course of the vowel, there may also be some pivoting movement of the tongue root, and slight variations in the behaviour of the tongue body and tip. However, it is difficult to determine visually whether these movements are large or consistent enough to be considered meaningful. In order to assess this, a GAMM analysis was carried out to statistically compare the different timepoints to each other using difference smooths.

The GAMMs show a similar picture to the mean plots: Most of the time, the splines at the 50% timepoint are not statistically different from the 10% timepoint, regardless of consonant context. However, in a few speakers (GF6, SM5, SF5, SM5, UF5, UM4), and for the front context only, the 50% timepoint includes a significant raising of the tongue blade, and/or significant fronting of part of the tongue body. These differences tend to be very small, but when they do occur, they usually persist until the 90% timepoint.

For the back context, none of the speakers show significant tongue movement at the 50% timepoint. At 90%, however, nearly all speakers (24 of 26) display significant raising of the tongue body towards the velum, and significant shifting forward of the tongue root or post-dorsal region. This coarticulatory movement is highly consistent across the sample, and is relatively large in terms of effect size.

In summary, the dynamic movement of /i:/ over the course of the vowel appears to be mainly coarticulatory, in that speakers raise the tongue body towards the velum at the very end of the vowel if there is a velar consonant following. In the few speakers that display other dynamic

movement, it mainly consists of slight raising or lowering of the tongue body, and slight fronting of the tongue root or post-dorsal region. There is no statistically significant pattern of tongue tip raising.

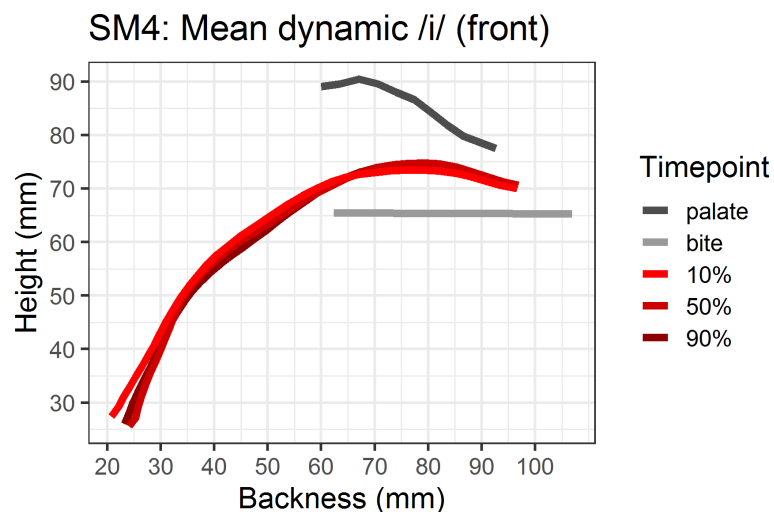


Figure 7.52: Mean dynamic tongue splines for /i:/ before a front consonant.

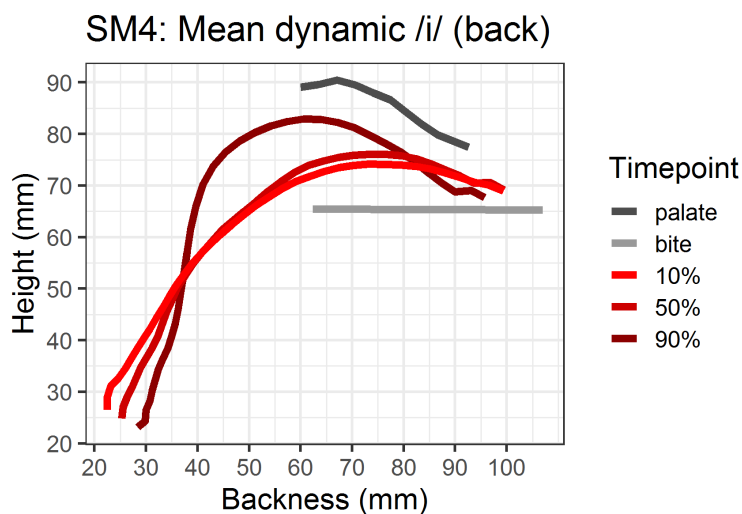


Figure 7.53: Mean dynamic tongue splines for /i:/ before a back consonant.

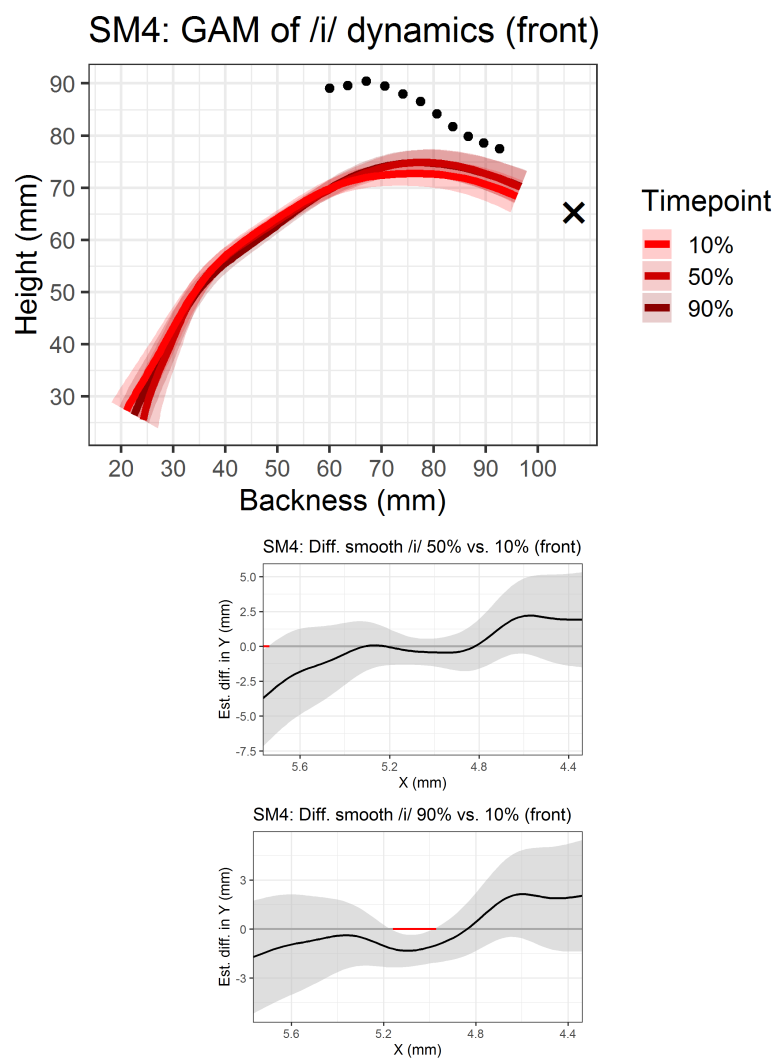


Figure 7.54: GAMM plot and difference smooths showing gesture dynamics for /i:/ before a front consonant.

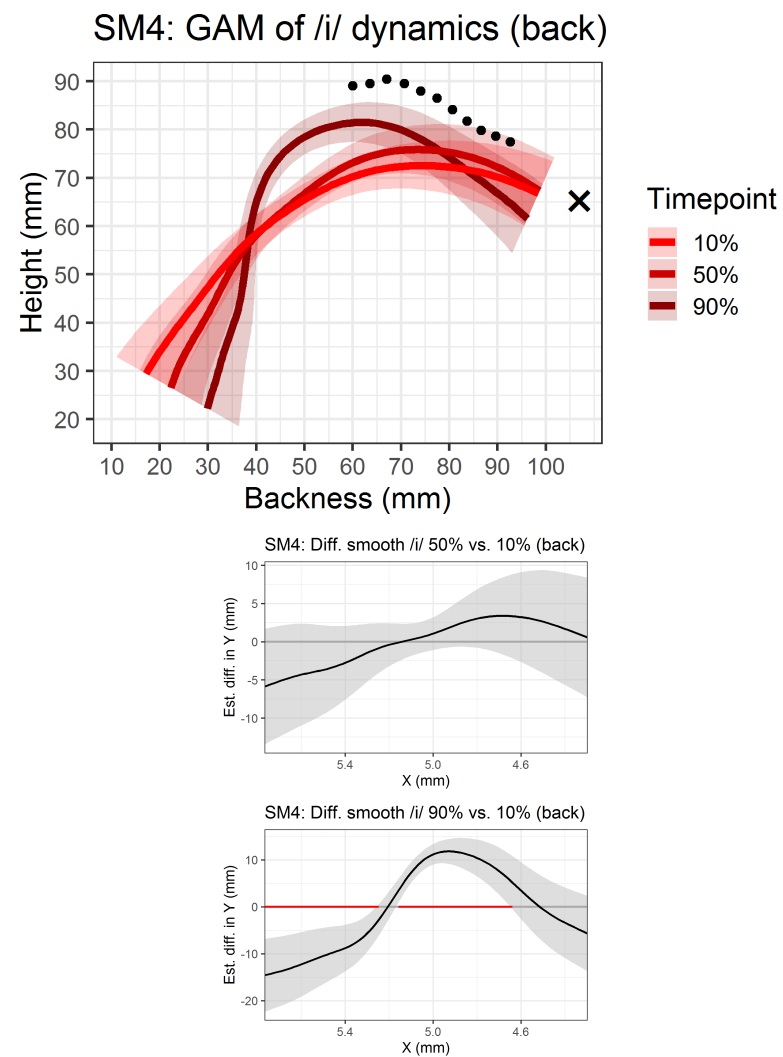


Figure 7.55: GAMM plot and difference smooths showing gesture dynamics for /i:/ before a back consonant.

### 7.7.2 Dynamic influences of articulation on formants

As mentioned above, the pattern of tongue body raising before a velar consonant appears to be a coarticulatory gesture, which corresponds well to the observation in Section 6.9.1 that the main form of visible formant movement during *Viby-i* was ‘velar pinching’ (Fig. 7.56). The static acoustic analysis also showed that velar contexts seemed to disfavour the F2-lowering associated with *Viby-i*. However, as shown in Fig. 7.57, the velar context did affect the *Viby-i* offglide. For voiceless velar contexts in particular, e.g. FIKA, VIKa, the raising of the tongue body tended to result in low-frequency frication, which was audibly velar in quality. Thus, since the dynamic tongue movement of *Viby-i* does not seem to have a strong effect on *Viby-colouring*, but a potentially greater effect on fricative offglides, the final section of this chapter will investigate the relationship between frication and tongue dynamics.

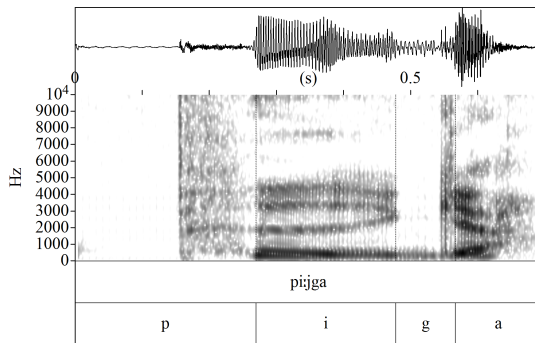


Figure 7.56: Velar pinch during *Viby-i* in speaker GF1.

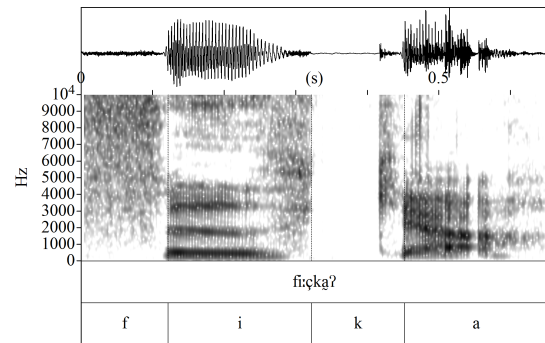


Figure 7.57: Low frequency (velar) end-frication during *Viby-i* in speaker SF1.

### 7.7.3 Dynamic influences of articulation on vowel frication

Since previous studies have suggested that the tongue tip may be important for generating ‘buzz’ or frication during the production of *Viby-i*, this part of the analysis statistically tests the relationship between vowel frication, tongue tip position, and main constriction height. For this purpose, an LMER model was constructed, using ‘bpZCR’ as the dependent variable, and ‘tip\_x’, ‘tip\_y’, and ‘max\_constr\_y’ as independent variables. These measures are intended to capture narrow constrictions at the front and back of the tongue respectively. Since the model included data with different baselines for buzziness, ‘timepoint’ and ‘word category’ were also included as independent variables. The latter denotes whether the vowel occurred before a front or a back consonant. A random intercept was included for ‘speaker’.

A summary of the model output is provided in Table 7.8, with plotted estimates in Fig. 7.58. The model shows that, disregarding the effects of timepoint and word category, all three articulatory measurements contribute significantly to an increase in zero-crossing rate, but some have a

stronger influence than others. The variable with the greatest influence on bpZCR was tongue tip height ( $\beta=797$ ,  $p<0.001$ ). The second strongest was post-dorsal constriction ( $\beta=108$ ,  $p<0.001$ ), and the least strong was tongue tip frontness ( $\beta=35$ ,  $p<0.001$ ). To some extent, the difference in the coefficients reflects the frequency of the fricative noise; for example, dental frication is likely to have a higher frequency than velar frication, and thus a higher bpZCR. Therefore, the tongue tip measure is associated with a higher increase in bpZCR than the maximum constriction measure. However, the difference between the two tongue tip measures indicates that tongue tip height was more important than tongue tip frontness for generating frication.

<i>Dependent variable:</i>	
zcr	
tip_x_norm	35.127*** (6.012)
tip_y_norm	797.442*** (11.228)
constr_y_norm	107.535*** (4.769)
timepoint_50%	1,907.654*** (9.682)
timepoint_90%	1,944.455*** (9.554)
word_cat_front	-24.268* (9.830)
Constant	4,939.206*** (231.079)
Observations	148,025

*Note:* \* $p<0.05$ ; \*\* $p<0.01$ ; \*\*\* $p<0.001$

Table 7.8: LMER output showing the effect of tongue tip position and maximum tongue constriction height on bpZCR over the course of /i:/.

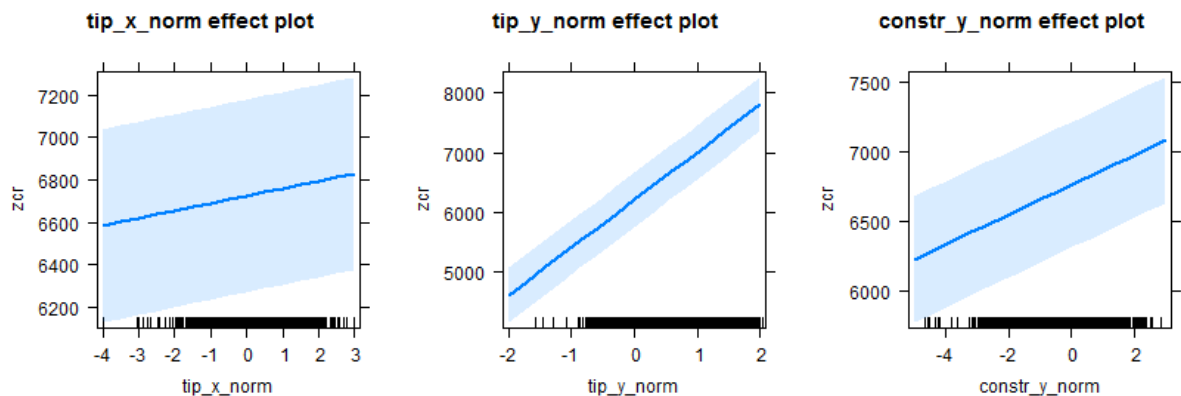


Figure 7.58: Significant effects of tip height, tip frontness, and max constriction height on bpZCR.

In summary, vowel frication during Viby-i is influenced most greatly by tongue tip raising (and, to a lesser extent, tongue tip fronting), as well as raising of the tongue body. However, since the GAMM analysis in Section 7.7.1 showed that the tongue tip for /i:/ does not appear to move over the course of the vowel, there are two potential explanations for how the frication is generated: Either, small degrees of tongue tip raising have a proportionally strong effect on bpZCR, since tongue tip raising can be observed in the GAMMS, even if it is not statistically significant. Else, the tongue tip is constantly in place to create frication, but requires a change in e.g. air velocity, to do so. The latter strategy has been proposed by Faytak and Lin (2015) to explain the source of turbulence in Mandarin apical vowels, where there was also little to no dynamic change in tongue gesture.

## 7.8 Chapter summary

The articulatory analysis has established that Viby-i in this sample is usually produced with a low, fronted tongue body, high tongue tip, and some degree of post-dorsal retraction. The dual ‘pull’ of the fronted tongue body and the retracted post-dorsal region results, in many speakers, in a distinct double-bunched tongue shape, which does not correspond to the tongue shape ascribed to vowels in the general literature (e.g. Catford, 1994; Ogden, 2009). However, many speakers also produced Viby-i with a canonical arched tongue shape, indicating that this vowel can be achieved in different ways. Post-dorsal retraction was found to be more common in younger speakers, and in speakers from Gothenburg.

To some extent, the articulatory findings correspond to those of the acoustic analysis: For example, tongue lowering was found to significantly correlate with a higher F1. Similarly, Gothenburg speakers, who had a higher F1 in the acoustic analysis, were also found to use a lower tongue body for Viby-i. However, the low F2 of Viby-i did not have the expected relationship with tongue backing, at least not when the highest point of the tongue was investigated. Instead, a number of different factors were found to correlate with F2, including tongue body *fronting*, tongue body lowering, tongue tip raising, and post-dorsal retraction. The lips did not appear to contribute to the low F2 of Viby-i.

The dynamic analysis found that the tongue gesture for Viby-i was mostly stable over the course of the vowel, corresponding to the formant stability observed in the acoustic analysis. However, velar consonant contexts encouraged coarticulatory raising of the tongue body, which resulted in ‘velar pinching’ in the spectrograms. Tongue gesture also had an impact on Viby-i frication: Tongue tip raising (and, to a lesser degree, fronting) contributed to a particularly high zero-crossing rate, indicating high-frequency dental or alveolar frication. Raising of the tongue body similarly contributed to velar frication, which had a lower frequency. The dynamic analysis thus showed a relatively good correspondence between the articulatory and acoustic data.

# Chapter 8

## Discussion

### 8.1 Chapter overview

This chapter returns to the research questions posed in Chapter 1, summarises the findings of the acoustic and articulatory analyses, links these findings to previous literature, and evaluates the implications of these results for Viby-i, within the wider contexts of vowel phonetics. To recap, the research questions which will be addressed are:

- What acoustic properties characterise Viby-i?
- How is Viby-i articulated by the tongue and lips?
- Is Viby-i subject to articulatory trade-off?
- Is there regional, social, or linguistic variation in the use of Viby-i?
- What methods are best used to describe and quantify the articulatory behaviour of Viby-i?

Section 8.2 begins by discussing the acoustic characteristics of Viby-i in terms of its formant values, formant dynamics, and use of frication. The findings of the articulatory analysis are examined, with particular focus on the unusual tongue gestures used for Viby-i, and the apparent mismatch between acoustics and articulation. Finally, the effects of linguistic environment and sociolinguistic factors on Viby-i are discussed, and some interpretations of these results are put forward.

Section 8.3 provides a general discussion of how the findings of the study relate to the wider literature. The section begins by discussing how Viby-i should be defined, and whether it conforms to the definition of a vowel sound. The occurrence of Viby-i in different languages is addressed, as well as the claim that Viby-i is becoming increasingly prevalent in Sweden. This section also discusses the contribution of the current study to phonetic ultrasound research.

Finally, Section 8.4 provides directions for future study, both on the topic of Viby-i, and in the area of vowel articulation in general.



## 8.2 Discussion of findings

### 8.2.1 Viby-i formant structure

The acoustic analysis investigated the formant structure of /i:/ compared to the other long vowels in the Swedish system: /y:/, ʉ:/, e:/, ø:/, ɛ:/, a:/, o:/, u:/ . The results showed that all speakers in the sample (with the exception of one borderline speaker) produced /i:/ with an F1 that was similar to /e:/, but with an F2 that was lower than /e:/ . Thus, all speakers in the sample were found to use some form of Viby-i, following the acoustic definition used in this thesis.

The analysis showed that Viby-i was mainly characterised by a low F2 compared to /e:/ . The mean F2 for Viby-i was around 1,710 Hz for men and 1,940 Hz for women, which was 200-500 Hz lower than /e:/ in most speakers. These F2 values were also lower than most reports of standard [i:] in Swedish (e.g. Eklund & Traunmüller, 1997; Ericsson, 2005; Fant, 1959, 1972), but not as low as previous reports of Viby-i (Björsten & Engstrand, 1999; Schötz et al., 2011). The finding that a low F2 was a prominent acoustic characteristic of Viby-i nevertheless agrees with previous descriptions of this vowel from the literature (e.g. Björsten & Engstrand, 1999; Ladefoged & Lindau, 1989). As suggested in Westerberg (2016), the unusual vowel quality of Viby-i seems to mainly be associated with a low F2, while F1 is more variable.

Previous literature states that Viby-i is also characterised by a high F1 (e.g. Björsten & Engstrand, 1999; Ladefoged & Lindau, 1989), although it does not explicitly state what the F1 should be compared to. In this study, the F1 of Viby-i was relatively high, in that it was similar to /e:/, with values of around 340 Hz for men and 400 Hz for women. These values resemble those found by Björsten and Engstrand (1999) and Schötz et al. (2011) for Viby-i. The F1 values for Viby-i were also high compared to standard [i:], whose values were 50-100 Hz lower than in the current sample (e.g. Eklund & Traunmüller, 1997; Ericsson, 2005; Fant, 1959, 1972). Based on these findings, Viby-i can be said to have shifted ‘down’ from the position where [i:] would normally be in the acoustic vowel space, leaving the ‘high front’ corner empty.

Ladefoged and Lindau (1989) also describe Viby-i as having a high F3, but again, it is not clear what the F3 should be compared to. In this study, both F3 and F4 were higher for Viby-i than for any other vowel in the system. The mean F3 was around 2,740 Hz for men and 3,210 Hz for women. These values are similar to Björsten and Engstrand (1999), but higher than those of Schötz et al. (2011). However, compared to studies of standard [i:], the reported F3 values for Viby-i are relatively low (c.f. Eklund & Traunmüller, 1997; Ericsson, 2005; Fant, 1959, 1972).

The mean F4 of Viby-i in this study was around 3,660 Hz for men and 4,390 Hz for women. These values are similar to previous reports of both Viby-i and Standard Swedish [i:]. There are no indications in the literature as to how the high F4 might relate to the overall quality of Viby-i. Measures of F4 across the Swedish vowel system in this study suggest that F4 can be lowered by lip-rounding, but since Viby-i was unrounded, F4 does not seem to play a great role in the

quality of this vowel.

Overall, Viby-i retains some ‘height’ in the acoustic vowel system, but its position is acoustically ‘backed’, particularly compared to the nearest high front vowel /e:/ . Based on these acoustic values, it is understandable why Viby-i has been described as a high central [i:] (Björsten & Engstrand, 1999). In addition, the acoustic profile of Viby-i overlaps significantly with /y:/ in both F1 and F2, although the two vowels are distinct in F3 and F4, probably due to lip rounding. To some extent, this overlap supports claims from previous literature (Björsten & Engstrand, 1999; Gross & Forsberg, 2019) that there may be perceptual overlap between these two vowels, and that /y:/ in these speakers was Viby-coloured.

In summary, this study supports claims from previous research that Viby-i is characterised acoustically by a low F2 and a high F1, but the degree of F2-lowering observed here is less extreme than in previous studies. For F1 and F3, there is a recurring issue of comparison; within the Swedish vowel system, F1 and F3 for Viby-i are both relatively high, but compared to a standard [i:], they are low. The study also reports F4 values, which can provide a useful baseline for future acoustic work on Viby-i.

## 8.2.2 Viby-i dynamics and frication

Since previous literature has described Viby-i as distinctly ‘buzzy’ or fricated (e.g. Björseth, 1958; Grönberg, 2004; Kotsinas, 2007), the current study also investigated aspects of frication during Viby-i production, which was measured through bandpass-filtered zero-crossing rate (bpZCR). The analysis found that Viby-i was sometimes produced with frication during the vocalic portion, with fricative noise superimposed on the periodic waveform, but this occurrence was relatively rare, occurring in around 5% of the vowel tokens. Much more common was the occurrence of end-frication, which resembled the offglide patterns described for high vowels in Standard Swedish (e.g. Elert, 1995; Engstrand, 1999).

The fricated offglides for Viby-i all consisted of a closing gesture in the vocal tract, e.g. [i:j], which could optionally be followed by an epenthetic schwa, e.g. [i:jə]. The exact realisation of the offglide was usually conditioned by the following consonant, although word-final contexts also used end-frication, as reported by Helgason (2002). The closing gesture of the vocal tract was also observed in the dynamic articulatory data, but only for velar contexts. Although the articulatory analysis showed that tongue tip raising and fronting were associated with greater levels of frication, these movements were *not* observed in the dynamic tongue contours. It may thus be the case that the end-frication was either generated by very small tongue tip movements, or that the tongue tip was already in place to produce the frication, with turbulence arising from a change in air velocity, as suggested by Faytak and Lin (2015) for fricated vowels in Mandarin.

As well as fricated offglides, the dynamic data also showed patterns of formant weakening, which sometimes co-occurred with end-frication, and appeared to serve a similar function. These patterns of frication and formant weakening often began around 40-50% of the vowel duration,

and appeared to coincide with increased variability in the formant tracker. The increased risk of formant measurement errors towards the end of the vowel thus justifies the decision to perform the static vowel analysis based on the 10% timepoint, rather than at 50%, which is more common in phonetic vowel studies. The exact cause of the formant weakening could not be examined in this thesis, but would be an interesting topic for future study, along with formant bandwidth.

### 8.2.3 Viby-i articulation

Based on what we know about vowel production – based on e.g. Jones’ vowel quadrilateral (1917), Delattre’s comparisons of acoustic output to tongue X-rays (1951), and Fant’s acoustic theory of speech production (1960) – the high F1 and low F2 of Viby-i leads us to expect an articulation with a low and backed tongue gesture. However, the ultrasound data revealed that, while the tongue gesture for Viby-i was indeed low, the highest point of the tongue was fronted rather than backed with relation to /e:/. At the same time, most speakers additionally retracted or constricted the post-dorsal region of the tongue, and raised the tongue tip, creating a complex articulation that looks more consonantal than vocalic. There was thus a mismatch between the acoustic output and our articulatory expectations for this sound.

The statistical analysis of these results, using GAMMs as well as mixed-effects modelling of measures representing different parts of the tongue, showed that the highest point of the tongue, while corresponding well to F1, did not have the expected relationship with F2. Instead of the low F2 being achieved through overall tongue backing, it seemed to result from a combination of tongue body lowering and post-dorsal retraction. A visual comparison of the lips for /i:, y:, e:/ showed that lip rounding was not present in Viby-i; thus, the lips did not appear to contribute to the low F2. Tongue tip raising was also found to *lower* F2, which opposed the findings of (Björsten & Engstrand, 1999: 1,959). In summary, the effects of tongue lowering and back retraction were seemingly strong enough to ‘override’ the high F2 that should have resulted from the fronted tongue body, raised tongue tip, and spread lips.

The ultrasound data also showed that Viby-i was associated with a number of different tongue shapes, which were roughly categorised as arched, front-bunched, back-bunched, and double-bunched, with arched and double-bunched shapes being the most common. Regardless of shape, the post-dorsal region of the tongue was usually relatively retracted compared to /e:/, and the tongue tip (pointing up or down) was close to the palate. The variation observed between individuals in the articulation of Viby-i is interesting given the many different proposed articulations of this vowel in the literature. Most theories are at least partially correct: Noreen (1903) states that Viby-i is produced apically (i.e. with a raised tongue tip), and that it can use either a tip-up or a tip-down gesture. Although Noreen does not describe tongue lowering or post-dorsal retraction, his description of the front part of the tongue is accurate. Björsten and Engstrand (1999) posit that the tongue shape for Viby-i is close to schwa, i.e. both backed and lowered compared to a standard [i:]. Although they do not predict the simultaneous tongue body

fronting and post-dorsal retraction, they too suggest that the tongue tip can be either up or down (Björsten & Engstrand, 1999: 1,959). Ladefoged and Lindau (1989) match the current findings better, suggesting a dual articulation where the tongue body is lowered and the blade is raised. This description is very close to the tongue gestures found in the present dataset. However, the hypothesis that best matches the findings of this thesis is that of Borgström (1913), who describes Viby-i as having a fronted tongue tip, and a tongue body where the front and back parts are raised, while the middle part is lowered. This is a very close approximation of the double-bunched tongue shapes observed in this sample, and it is impressive that Borgström could have arrived at this articulation introspectively over a hundred years ago. Perhaps the least accurate description of Viby-i is provided by Lundell (1878), who suggests that Viby-i is simply a high back unrounded [u:]. In this sample, a high back tongue gesture on par with [u:] was never used for Viby-i; however, that is not to say that Viby-i could not be produced this way, or that it might not have been produced this way in the past.

One of the most notable aspects of Viby-i production, which the preceding literature anticipates, is the multiple strategies available to speakers in producing this sound. It may be possible to characterise this phenomenon as articulatory trade-off in two senses: Firstly, in that speakers are able to achieve similar acoustic outputs using different articulatory strategies; and secondly, in that the strategies used by the speakers covertly vary from our expectations based on the acoustic data. Viby-i thus defies traditional models of vowel description by using multiple articulatory gestures, and by uncoupling F2 from the highest point of the tongue. The complexity of this articulation may be why researchers have found Viby-i difficult to parse.

### 8.2.4 Effect of linguistic environment on Viby-i

No previous studies have investigated the effect of linguistic environment on the production of Viby-i. The aim of investigating linguistically conditioned variation in this thesis was to examine whether certain coarticulatory processes would encourage a Viby-i with exaggerated acoustic characteristics, i.e. a lower F2, and possibly a higher F1. These coarticulatory environments could then be taken as indicators of how Viby-i was articulated. For example, if Viby-i had a lower F2 before a front consonant, it could be assumed that a prototypical Viby-i uses a relatively fronted tongue gesture.

The results showed that Viby-i had a lower F2 when the vowel duration was short, when it was preceded by a voiced rather than a voiceless consonant, and when it was preceded by a fricative rather than a plosive. Viby-i also had a lower F2 when the following consonant had a front place of articulation (bilabial, dental, or alveolar) compared to a back place of articulation (velar). Finally, Viby-i had a lower F2 when the following consonant was complex (liquid, nasal, or sibilant) rather than simple (plosive). These effects were only investigated directly in acoustics, but the articulatory data confirmed that F2-lowering was indeed associated with fronting of the tongue body, as well as articulatory complexity.

The fact that F2 was lower in short vowel tokens went against initial expectations; since *Viby-i* prototypically has a low F2, tokens with longer durations, and thus more time to reach their target, were hypothesised to have a lower F2. Instead, it appeared that shorter instances of *Viby-i* were more reduced, with a lower F2 as a result (Lindblom, 1963). For future work, it would be interesting to examine how this acoustic difference between short and long vowels manifests itself in tongue shape, i.e. whether shorter vowels use a more centralised tongue gesture (Browman & Goldstein, 1992).

The effect of preceding voicing and frication on F2 is somewhat difficult to explain. These linguistic environments were mainly included to investigate vowel frication, but were also found to affect formant values. It is possible that the presence of voicing or frication affects the timing of the articulatory gestures, or that the air stream is affected, but it seems unlikely that these mechanisms would affect the formant values. Furthermore, the presence of preceding /f, v/ did not encourage any visible labiodental element during *Viby-i*. Thus, to fully understand the effects of consonant environment on *Viby-i*, further research is needed.

The coarticulatory effects, whereby *Viby-i* had a lower F2 before front consonants and complex consonants, are also somewhat elaborate. The correlation between front consonant environments and a low F2 is matched in the articulatory results, where fronting the highest point of the tongue caused F2 to lower. This finding is counter-intuitive, but could perhaps be explained by the counteracting effect of simultaneous post-dorsal retraction and tongue body lowering. In addition, the dynamic analysis of the tongue curves showed that the tongue body tended to raise towards the velum before back consonants. Due to the tongue's hydrostatic properties, it may be difficult for speakers to achieve this raising at the same time as constricting a lower part of the tongue (Alwabari, 2017). Thus, coarticulatory velar raising seems to counteract the tongue lowering and post-dorsal retraction used to create the low F2 of *Viby-i*.

The effect of complex consonant environments is somewhat confounded with that of front consonants; all complex consonants were front, while the simple consonants were a mixture of back and front. It is therefore possible that the low F2 of complex consonant environments are another way of showing the same result as above. However, it is also possible that the articulatorily complex consonant environments encouraged the more complex articulatory settings found for *Viby-i*. Unfortunately, it was beyond the scope of this thesis to investigate the similarity between the vocal tract configurations for *Viby-i* and the lingual consonants, but this would be a valuable addition to future work, as similar bunched tongue gestures have also been found to occur in English /r/, where they have been shown to promote consonant-vowel coalescence (Lawson et al., 2013).

In summary, the analysis showed that *Viby-i* does vary, in both acoustics and articulation, based on linguistic context. However, the exact articulatory processes involved, and their relationship to surrounding consonants, still require further work.

### 8.2.5 Sociolinguistic variation of Viby-i

The sociolinguistic analysis investigated whether variation in Viby-i was correlated to a number of demographic factors, most notably city, age, and gender, but also lifestyle choices such as where speakers preferred to live, how socially interactive they were, and their attitudes towards their home city, local dialect, and non-standard language in general. Notably, no speakers in the sample used standard [i:]. Instead, gradient realisations of Viby-i, with different amounts of F2-lowering, were tested in the analysis.

The main factor found to influence the acoustic output of Viby-i was city of origin. Speakers from Gothenburg had a significantly lower F2 and higher F1 than Stockholm and Uppsala, which patterned together. This finding was also matched in the articulatory analysis, where Gothenburg speakers used a significantly lower tongue gesture with more post-dorsal retraction compared to Stockholm and Uppsala speakers. This finding largely corresponded to the study's hypothesis: Since Viby-i is believed to be older in Gothenburg than in Stockholm (e.g. Elert, 1995: 45); it was predicted that Gothenburg Viby-i would be acoustically and articulatorily further away from standard [i:] than the other two cities. The fact that Stockholm and Uppsala produced equivalent vowels (in both acoustics and articulation) was also anticipated, given that the two cities are geographically close. At the same time, it was unexpected for Viby-i to be so prevalent in Uppsala, since this vowel had not previously been documented there. This finding could indicate that Viby-i is more widespread in Central Sweden than previously assumed.

Although the sample was not balanced enough to investigate the effect of education (used here as a proxy for socioeconomic class), the finding that Viby-i was consistently used over standard [i:], regardless of age, gender, lifestyle, or attitudes, was somewhat surprising, as previous literature has found evidence of social stratification in this vowel (Grönberg, 2004; Gross, 2018; Kotsinas, 2007). The fact that everyone in the sample used Viby-i seems to confirm the prediction that it is becoming the new standard variant in Central Sweden (Bruce, 2010: 216), or even that this change has already taken place. A further discussion of this topic will be provided in Section 8.3.4.

It is difficult to tell whether the shift towards Viby-i constitutes a change from above or a change from below. In both cases, we expect the change to be led by young, female, upwardly mobile speakers (Labov, 2001: 272-280), which historically seems to be the case (e.g. Kotsinas, 2007). The fact that Viby-i is primarily associated with speakers from higher socioeconomic backgrounds, and that it is a well-known prestige variable, strongly suggests a change from above. Similarly, it appears to have become part of the prescriptive standard, since Uppsala speakers, who represent a variety close to Standard Swedish, were also using Viby-i. On the other hand, Viby-i is not necessarily perceived as a positive feature, as it is simultaneously described as both 'posh' and 'ugly' (Elert, 1995; Norstedt, 2019). Furthermore, speakers do not seem to be aware of their own use of this sound, and style-shifting is rare (Norstedt, 2019). Thus, it appears to have spread across Central Sweden relatively unnoticed, which would indicate a

change from below (Labov, 1994: 78). A change from below could perhaps also explain how Viby-i might have travelled from rural dialects into the cities. Regardless of which process led to this change, the results of this study indicate that the social stratification previously observed for Viby-i has now levelled out and stabilised, at least in terms of acoustics.

In the articulatory analysis, younger speakers produced Viby-i with significantly greater levels of tongue retraction than older speakers. Despite these differences in tongue gesture, the acoustic output between young and old speakers was not significantly different. This finding could imply the use of articulatory trade-off, in that speakers appear to be able to achieve similar acoustic results using multiple articulatory strategies (e.g. Guenther et al., 1999). The occurrence of covert, age-based variation is reminiscent of the finding that social class can covertly govern the tongue gesture used for /r/ in Central Scotland (Lawson et al., 2011, 2014).

In the context of age differences, it is tempting to describe the current phenomenon as a demonstration of Viby-i changing in apparent time. Older speakers, who may or may not have used standard [i:] when they were younger, seem to favour a tongue gesture that is more fronted, and thus more articulatorily similar to a standard [i:]. Younger speakers, who may have been less exposed to standard [i:], may be more flexible in their choice of articulatory strategies, and it is possible that post-dorsal retraction is a more articulatorily efficient way of producing their desired output. However, even though younger and older speakers were not significantly different in F1/F2, there could be a perceptible difference in their productions of Viby-i, which could be used as an age marker. Since the current study does not have a perceptual component, future work on the perception of Viby-i is encouraged to investigate this issue further.

Overall, the sociolinguistic analysis found fewer differences between demographic groups than expected, and instead found that Viby-i was surprisingly uniform across the sample, albeit a sample of fairly homogeneous social backgrounds. The current sample indicates that, at least in highly-educated, inner-city, culturally Swedish speakers producing word list speech, Viby-i is used consistently across age and gender groups. There are, however, some dialectal differences between the Eastern and Western parts of Central Sweden, which could be the result of geographic and temporal separation (if Viby-i did indeed spread from West to East); or it could be the case that Viby-i has arisen in different parts of the country at a similar time. Given that the sample was not large or balanced enough to represent the wider Swedish population, future work should continue to inform the social use of Viby-i by including data from a larger number of speakers with different educational and cultural backgrounds, as well as a wider range of speech styles.

## 8.3 General discussion

### 8.3.1 What counts as a Viby-i?

Before addressing the implications of this study's findings, it is worth noting that the working definition of Viby-i used in this thesis may be contested by other scholars. Here, Viby-i is acoustically defined as a vowel that functions as /i:/, but has a lower F2 than /e:/, and, secondarily, a similar or higher F1 than /e:/. This definition is based on patterns observed in previous literature (Björsten & Engstrand, 1999; Gross & Forsberg, 2019; Schötz et al., 2011), including my own previous work (Westerberg, 2013, 2016). However, other studies have only defined Viby-i auditorily, and provided data where F2 is much lower than in the current dataset. In addition, previous studies claiming to present acoustic data for standard [i:] (particularly Traunmüller & Öhrström, 2007) report formant values which, by the definition of this thesis, would qualify as examples of Viby-i. This lack of correspondence raises questions about what qualifies as a Viby-i, and on what grounds.

Part of the justification for how Viby-i is defined in this work comes from the observation that this vowel appears to exist on a gradient scale. In Westerberg (2016), a perceptual experiment showed that listeners were able to rank Viby-i productions based on perceptual similarity to, or difference from, standard [i:]. Tokens that were perceived as less [i:] -like, and more 'thick' or 'dark', were significantly more likely to have a low F2, while no such correlation was found for F1. Having worked closely with the audio recordings for this project, I have similarly observed that some speakers use a Viby-i that is more perceptually salient than others. What all these sounds have in common is that, in terms of their *phonetic* properties, they can more or less consistently be recognised as 'not [i:]'.

One of the issues when studying Viby-i, however, is that Swedish lay listeners in particular do not necessarily perceive Viby-i as 'not [i:]'. This may be because many Swedish speakers categorically use this sound to represent the /i:/ phoneme, or it may be because Viby-i is sufficiently stigmatised that speakers do not want to admit that they use it (Norstedt, 2019). Furthermore, if Viby-colouring is gradient, 'weaker' or less salient versions of this vowel may not meet listeners' criteria for what qualifies as a Viby-i. Researchers may similarly wish to study examples of this vowel that can indisputably be classified as Viby-i, since the boundary becomes increasingly fuzzy the closer this vowel comes to standard [i:].

In an attempt to show how widespread Viby-colouring of /i:/ (and, to some extent /y:/) is in Central Sweden, this study chooses to draw the line between Viby-i and standard [i:] at the point where the vowel is no longer the 'highest' and 'frontest' in the acoustic vowel space. Since /i:/ represents a corner vowel in most of the world's vowel systems (Lindblom, 1986: 15), a centralising shift in this part of the vowel space seems quite strange. Furthermore, most languages that have a high central unrounded [ɨ] phoneme, e.g. Turkish, Russian, or Romanian, tend to also have a separate phoneme for standard [i:]. It is this distribution of the vowel space,



where the high front corner is ‘empty’, that makes Viby-i unusual.

Thus, even though this study may present examples of Viby-i that are not particularly ‘strong’ from an auditory or acoustic standpoint, they nevertheless represent a pattern of vowel production that challenges traditional notions of what an /i:/ vowel ‘should’ be like. Rather than seeking out especially salient examples of this vowel for study, this thesis demonstrates that weaker instances of Viby-colouring can be found in randomly sampled speakers from various parts of Central Sweden, suggesting that Viby-i may be more common in this region than previously believed. Establishing a definition for Viby-i based on acoustic rather than auditory criteria also adds some level of objectivity and reproducibility to this work, since it is not always clear from previous studies whether the same sound is being described. Hopefully, the working definition used in this thesis, as well as the data presented on Viby-i, can act as a starting point from which a more detailed typology of this vowel can be developed.

### 8.3.2 Vowel or consonant?

A question that has recurred over the course of this project is whether Viby-i should, in phonetic terms, be regarded as a vowel or a consonant. This question arises from the fact that the sound has some ‘consonantal’ qualities in its use of frication, and its unusual range of tongue shapes. Since consonants are defined as sounds which manipulate the airflow with some kind of constriction (Ladefoged, 2001: 47), there are indeed aspects of Viby-i that could be regarded as consonantal.

One of the specific criteria for vowel sounds given by Ogden is that there is “free passage of air ... without friction” (Ogden, 2009: 56). The fact that Viby-i is often produced with some degree of frication starting around the midpoint of the vowel, means that there is probably some form of constriction causing turbulence in the airflow. In addition, Viby-i defies traditional vowel description in the sense that the highest point of the tongue – a reference point that has been used to describe vowels for over a century (Catford, 1994: 132) – is not adequately able to capture the tongue shape used for Viby-i. The use of multiple constrictions to produce a vowel sound seems unusual, and this kind of strategy is more reminiscent of the tongue gestures observed for approximant /r/ (e.g. Delattre & Freeman, 1968; Guenther et al., 1999; Zhou et al., 2008) or *schwar* (Lawson et al., 2013).

On the other hand, as with *schwar*, the boundary between vowel and consonant is not always clear. This is why, for example, the phonetic definition of a vowel borrows the phonological criterion of functioning as the nucleus of a syllable (Laver, 1994). In purely phonetic terms, *schwar* and approximant consonants such as [ɹ] or [w] are not technically different from the classical definition of a vowel sound, except in phonological function. Thus, although [ɹ] and *schwar* in particular may have more complex articulations than most vowels, their main articulatory characteristics are still similar.

In addition, Viby-i is not unique in the Swedish vowel system for using frication. Fricated offglides are common to all high vowels in Central Swedish, including standard [i:] (Elert, 1995:

40). Fricated vowels have also been described in other languages, such as Mandarin (Faytak & Lin, 2015), and a number of smaller languages from parts of Asia, Africa, and North America (Connell, 2000: 233). Even in languages like English, where frication is not canonical, vowels may still be fricated in some speech styles, particularly in emphatic or hyper-articulated speech. The presence of frication could, at least for some vowels, be regarded as a secondary feature superimposed on the intended vowel quality, in a similar way to how nasalisation or creak may provide phonemic contrast in some languages, but not in others.

My impression of Viby-i is that the distinct ‘damped’ vowel quality appears to be separate from frication, since frication does not always co-occur with Viby-colouring. Although this issue was not specifically researched in this thesis, it is hypothetically possible to produce a Viby-i with no frication at all. Further investigation into the fricative aspects of Viby-i, and particularly intra-vowel frication, would thus be an interesting topic for future study.

In summary, even though Viby-i has some qualities that could be described as consonantal, the standpoint of this thesis is that Viby-i is still functionally a vowel, and should still be categorised phonetically as a vowel sound, albeit one which requires a more detailed articulatory description than most vowels. In terms of its complex tongue shape, it is, if anything, similar to an approximant consonant, but since the line between vowels and approximants is drawn using phonotactic criteria (Ogden, 2009: 78), Viby-i fits better into the vowel category. The presence of frication complicates this classification, since it implies an obstruction in the vocal tract, defying the criteria for both vowels and approximants. However, since frication does not appear to be obligatory for the production of Viby-i, it can be argued that Viby-i is better described as a vowel with consonantal features, rather than a consonant with a vocalic function.

### 8.3.3 How exotic is Viby-i?

There has been some disagreement in the previous literature about whether Viby-i should be regarded as an “exotic” sound (Schötz et al., 2011), or whether it is equivalent to more common vowel categories in other languages, such as Turkish (Björsten & Engstrand, 1999) or Mandarin (Faytak & Lin, 2015). This is a difficult question to answer, because some aspects of this sound seem to overlap with vowels in other languages, while others diverge.

For example, Björsten and Engstrand (1999) found that Swedish and Turkish listeners treated Viby-i as more or less equivalent to Turkish [i]. However, the degree of perceived similarity between these vowels varied slightly depending on which dialect the Viby-i came from (Fig. 8.1). Swedish listeners also rated Turkish [i] as a better example of a ‘damped’ vowel than the Viby-i tokens from Gothenburg and Kräklinge, indicating that Swedish speakers did not necessarily make a distinction between the two languages. The Turkish listeners consistently preferred their own variant to the Swedish vowels, indicating that they could hear a difference. As Björsten and Engstrand do not provide acoustic values for the stimuli, we do not know how acoustically similar these vowels were, but the authors report that Swedish listeners were

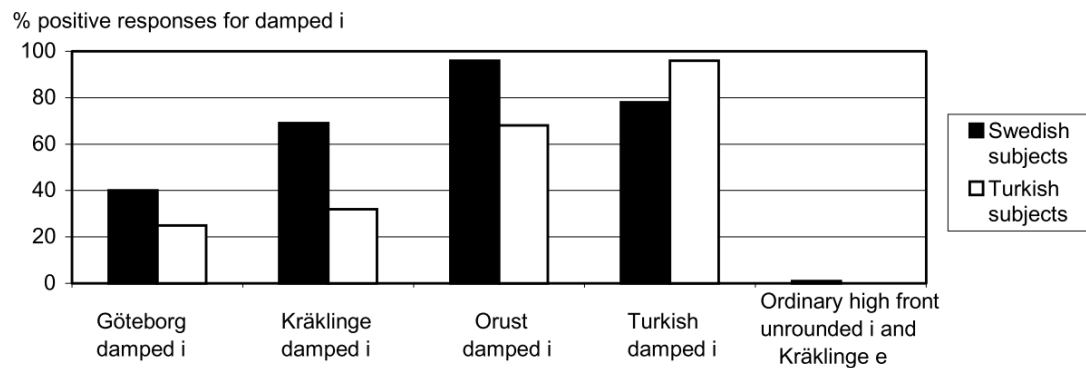


Figure 8.1: Acceptability ratings of Turkish [i] and Swedish Viby-i as ‘damped /i:/’ (Björsten & Engstrand, 1999).

probably using F2 as a perceptual cue (Björsten & Engstrand, 1999: 1,958-1,959). However, it is possible that, while F2 is the most important cue for Swedish listeners, Turkish listeners might be able to identify additional cues (e.g. frication), that do not agree fully with their phonological category. Similarly, I have met Mandarin speakers who accept Viby-i as an example of the apical vowels [ɿ] or [ɨ], while I perceive the Mandarin vowels as different from Viby-i. Thus, from a perceptual perspective, there are similarities between Viby-i and sounds in other languages, but they may not overlap perfectly. Since there also appears to be variation within Sweden, similarities with other languages could depend on which accent of Viby-i is used for comparison.

In terms of acoustics, no cross-linguistic comparison of Viby-like vowels yet exists. The closest would be Björsten and Engstrand (1999), who used acoustic data to classify Viby-i as an example of [i], and then used the UCLA Phonological Segment Inventory Database (UPSID) to evaluate how common this sound is in the world’s languages (Björsten & Engstrand, 1999: 1,959). The UPSID inventory showed that around 14% of the world’s languages use [i], and that it is widespread over different geographical regions, meaning that it is relatively common. However, as noted in the previous section, it is not particularly common for a language to have a phoneme for [i] but not for [ɨ]. In addition, Viby-i may have other acoustic properties that distinguish it from sounds in other languages, e.g. frication, or variation in formant bandwidths. Further work is therefore needed to establish more clearly what the acoustic properties of Viby-i are, so that it can be compared to other languages.

Finally, as this thesis has shown, Viby-i may involve some degree of articulatory trade-off to create a low F2 while maintaining a high F1. Instead of the tongue being low and backed, as the acoustic values suggest, the tongue body is low and fronted, while the post-dorsal part of the tongue is retracted, and the tongue tip is raised. We do not know whether these articulatory strategies are also common in other languages, as articulatory studies of vowel sounds are still rare. However, Faytak and Lin (2015) provide some illuminating ultrasound data of Mandarin apical vowels, which have been said to resemble Viby-i in both vowel quality and acoustics. Fig. 8.2 shows the tongue gestures used for [i, ɿ, ɨ] (dotted lines) compared to [ɕ, s, ʃ] (thin solid lines). For [ɿ] in particular, the tongue appears to have a similar gesture to Viby-i: The

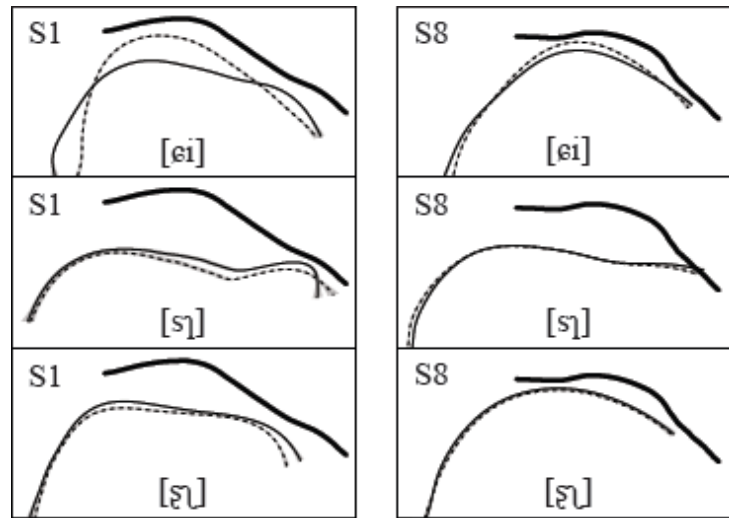


Figure 8.2: Tongue contours of two Mandarin speakers, comparing [i, ɿ, ɥ] to preceding [ɛ, s, ʂ].

tongue body is low, the blade/tip is fronted and raised, and the post-dorsum is relatively retracted compared to the other vowels. Both speakers also display a small dip in the tongue surface, reminiscent of the double-bunched tongue shape observed in the current dataset. Although it is difficult to compare individual speakers' tongue shapes in this way, there are clear similarities in tongue gesture between Viby-i and Mandarin [ɿ, ɥ].

In summary, there is probably not sufficient evidence to establish whether Viby-i is 'exotic' in terms of its occurrence in the world's languages. In many ways, Viby-i bears similarities to sounds in other languages, and it is thus unlikely that it would be entirely unique. However, in the context of Germanic languages, bearing in mind its phonological function and its apparent use of covert articulatory variation, it is quite unexpected to find this sound in Swedish. An interesting direction for future work would therefore be to investigate the acoustic and articulatory similarity between Viby-i and similar sounds in other languages, as well as the historical appearance of Viby-i in Swedish.

### 8.3.4 Change in progress?

Contrary to this study's predictions, no apparent-time variation was found for Viby-i acoustics, but an age-based difference was found in the tongue gesture used for this vowel. While older speakers tended to use a more fronted articulation, younger speakers were more likely to use post-dorsal retraction, either on its own, or through double-bunching. Since there was no statistically significant difference in the acoustic output based on age, it appears that speakers of different ages were producing equivalent vowel qualities using different articulatory strategies.

While age-related differences in tongue gesture could be an indication of Viby-i changing over time, the study found no age difference in the use of Viby-i over standard [i:]. The hypothesis was that older speakers would use a 'weaker' Viby-i (i.e. a lower F1 and higher F2) or no Viby-i at all, and that age differences would be greater in locations where this vowel was

more recent. In other words, older Uppsala speakers were expected to use *Viby-i* the least, since this sound had not been documented in this population, while older Gothenburg speakers were expected to use it the most, since *Viby-i* has been present in Gothenburg since at least the 1950s (Björseth, 1958). Based on Kotsinas (2007) showing that Stockholm *Viby-i* was stratified by age, gender, and class in the 1990s, the current study also expected to find continued patterns of social stratification, which can indicate change in progress.

Surprisingly, the analysis showed that the use of *Viby-i* was relatively consistent across the sample, regardless of age, gender, or city. Since the study could not compare speakers according to socioeconomic class, it is unclear if these results would also hold true for working-class speakers; but for the middle-class speakers in this sample, it seems that *Viby-i* was more or less ubiquitous. To some extent, this finding supports the association of *Viby-i* with social prestige, but it also raises the possibility that *Viby-i* could be equally pervasive in other socioeconomic groups or speech styles. Whether middle-class speakers have adopted this feature as a way of showing social aspiration (Labov, 1966), or whether *Viby-i* has always been present in middle-class speech, cannot be demonstrated from the available data. What can be observed, however, is that *Viby-i* is possibly older and more widespread in Central Sweden than the literature suggests.

It is possible that *Viby-i* in this sample was age-graded, since older speakers, who may have less incentive to conform to language norms, sometimes return to non-standard features that they used when they were younger (Eckert, 1997), or adopt new ones as they age (Harrington, 2006; Sankoff & Blondeau, 2007). However, if *Viby-i* was an innovation in their city of origin, it seems surprising that none of the older speakers would have resisted it. Although there seems to be an age-based difference in tongue gesture, the vowel qualities produced by young and old speakers were equivalent in acoustics. An apparent-time explanation, where *Viby-i* would have been present in the older speakers since their youth, therefore seems more likely, but further work is required to confirm this.

From an apparent-time perspective, it seems likely that *Viby-i* has been present in Gothenburg, Stockholm, and Uppsala long enough for the shift from standard [i:] to *Viby-i* to be complete. Using the example of the sociolinguistic s-curve (Chambers & Trudgill, 1988), the change may have started at different times, but all three cities now appear to have reached the plateau stage of language change, where *Viby-i* has become a standard feature used by more or less everyone. This may not be true for all social classes and speech styles, but the occurrence of *Viby-i* across the current sample appears to agree with the statement that *Viby-i* is rapidly taking over as the standard variant (Bruce, 2010: 216) and is “in the process of losing its social charge” (Riad, 2014: 21). Perhaps this change has simply occurred faster than expected.

From personal observation, it seems that *Viby-i* is currently more or less categorically used across Central Sweden, including in suburban and rural speech, although it varies in strength depending on the speaker. Perhaps a very slow, gradual increase in *Viby-colouring* (as part of a change from below) is the reason why many Swedish speakers seem unaware of their use of this

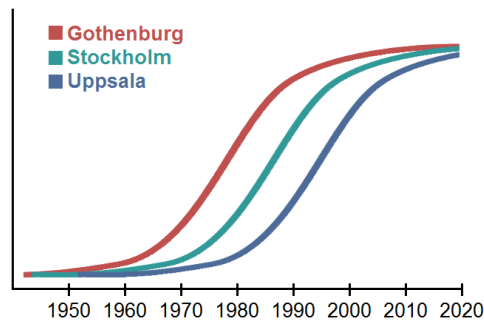


Figure 8.3: Schematic s-curves of how Viby-i might have changed over time.

vowel, and why it has also seemed to slip under the radar of many phonetic researchers.

How Viby-i has spread, how it is realised differently in different regions, and how it interacts with rural versions of this vowel, cannot be answered by this thesis, but provide interesting topics for future research. Based on the variation observed in the tongue gestures used for Viby-i, there may still be traces of change even when the vowel appears to be acoustically stable, and it is possible that there is more fine-grained variation in Viby-i production that has not been explored here, particularly with regards to social class (see Lawson et al., 2011, 2014).

Finally, to address the suggestion by Lindström (2014) that Viby-i has migrated from Bohuslän to Stockholm, and to resolve the apparent-time variation observed in the current study, it would be a valuable continuation to the current work to explore historical recordings of Viby-i from different parts of the country. Although these recordings are likely to be biased towards specific speech styles, some of which may not favour Viby-i, it would nevertheless be illuminating to see how far back in time this feature can be found, and who uses it. For now, although we cannot determine exactly how long Viby-i has existed in each city, it appears that the feature has caught up with itself in Gothenburg, Stockholm, and Uppsala.

### 8.3.5 Mismatch between acoustics and articulation

The acoustic and articulatory data in this study showed an unexpected relationship between the acoustic low F2 of Viby-i and the position of the highest point of the tongue, in a way that defies the traditional vowel literature (e.g. Delattre, 1951; Fant, 1960). Although the highest point of the tongue retained its expected relationship with F1 (the lower the tongue, the higher the F1), the advancement of the highest point was still correlated with F2, but in the ‘wrong’ direction. In the vowel literature, low F2 values are associated with tongue backing, which is often quantified through the highest point of the tongue (e.g. Blackwood Ximenes et al., 2017). However, for Viby-i, F2-lowering was instead associated with tongue body *fronting*. Having observed that Viby-i seemed prone to non-canonical tongue shapes, however, the analysis was extended to several other points on the tongue curve, to investigate their possible impact on F2.

The analysis showed that, although the low F2 of Viby-i was significantly correlated with a fronted tongue body and tongue tip, it was also correlated with a lowered tongue body, and

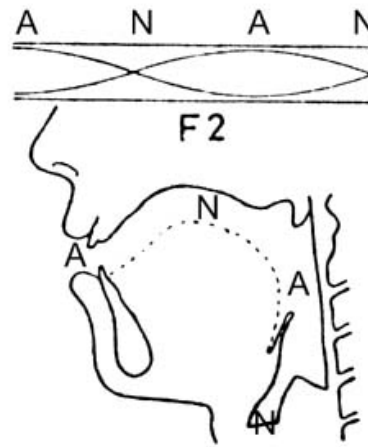


Figure 8.4: Velocity nodes (N) and antinodes (A) for F2 according to perturbation theory (Chiba & Kajiyama, 1941).

backing and/or constriction of the post-dorsal region. Since tongue body fronting and post-dorsal retraction tended to occur together, it is possible that tongue body fronting did not in fact lower F2, but that it was instead ‘cancelled out’ by simultaneous lowering and retraction.

The complex tongue shape for Viby-i and the resulting acoustic output could be explained through perturbation theory (Chiba & Kajiyama, 1941). In Fig. 8.4, the points labelled ‘A’ (antinodes) represent locations where narrowing of the vocal tract would lower F2, while the point labelled ‘N’ (nodes) represent locations where narrowing would raise F2. A lowered tongue body, retracted post-dorsum, and fronted tongue tip could thus all contribute to lowering F2 by compressing the antinodes by the lips and pharynx, while widening near the node at the palate.

Based on these results, the relationship between Viby-i and standard [i:] is somewhat reminiscent of the relationship between clear and dark /ɪ/. Both can be produced with the tongue tip touching the alveolar ridge, but dark /ɪ/ can additionally involve backing of the entire tongue body, or just the post-dorsal region (e.g. Turton, 2014). This retraction creates a lower F2 and produces its characteristic ‘dark’ quality (Sproat & Fujimura, 1993). Viby-i similarly appears to be a ‘dark’ version of [i:]: The tongue body is still relatively fronted, but simultaneous tongue lowering and post-dorsal retraction can change the quality of the vowel, while allowing it to still retain some similarities to [i:].

Based on the findings of this study, it appears that the articulation of Viby-i cannot be adequately described using only the highest point of the tongue; instead, additional parameters from the tongue, and possibly other parts of the vocal tract, are required. This finding supports Ladefoged’s observation that it is not always meaningful to compare the highest point of the tongue for two vowels that use different tongue shapes (Ladefoged, 1967: 1). The highest point of the tongue is still informative, and has been found by at least one recent study to correlate well with both F1 and F2 (Blackwood Ximenes et al., 2017). However, similarly to the current work, Lawson et al. (2019) find that it is relatively common for tongue gestures to share a highest point,

while differing in other aspects of the tongue curve. In particular, the highest point seems to be useful in explaining F1, but F2 has been shown, both by Lawson et al. (2019) and the present study, to be better explained by measures obtained from the post-dorsal region. To some extent, this issue might stem from the orientation in which the highest point of the tongue is presented in modern articulatory work, compared to historical data, such as X-rays (e.g. D. Jones, 1972; S. Jones, 1929; Russell, 1928).

Although this thesis has explained some of the acoustic and articulatory characteristics of Viby-i, there are still some aspects of this vowel that have not yet been studied. In acoustics, an analysis of the higher formants, as well as formant bandwidth or intensity, may be relevant. Articulatorily, it may also be valuable to further investigate coronal tongue behaviours (e.g. bracing, grooving, lateralisation), the impact of lip, jaw and larynx movement, and oral and nasal airflow, to mention some. The pilot study for this project (Westerberg, 2016) also revealed that the tongue for Viby-i is usually fronted enough for the tip to be visible in front-facing lip videos – a feature that listeners seem to have picked up on (Anonymous, 2007; Fagerström, 2019) – but this phenomenon could not be studied in the current work, as front-facing data was not available.

It is important to remember that the sample of this study, although large in terms of articulatory phonetic work, is small in terms of its ability to predict the behaviour of Central Swedish speakers overall. Although Viby-i was prevalent in this study, this vowel may be absent, or produced differently, in speakers from different social or dialectal backgrounds, or in different speech styles. Although there is still further work to be carried out in the description of Viby-i, the findings of the current study provide an interesting starting point, and represent a substantial step forward, not only in characterising Viby-i, but in understanding the relationship between the acoustics and articulation of vowel sounds.

### **8.3.6 Contribution to ultrasound methodology**

One of the goals of this study was to test and develop new ways of collecting and analysing ultrasound tongue imaging (UTI) data in the field. This section will discuss the innovative aspects of this study, and evaluate their contribution to UTI methodology for future research.

#### **Recording and processing UTI data**

Firstly, this study used a relatively new, portable ultrasound system (Telemed Echo-Blaster 128) to collect data on location in Sweden. While the ultrasound machine itself is very small, about the size of an external CD player, a lot of other equipment was needed to make the recordings. In particular, the additional equipment for recording simultaneous lip data added to the bulk of the set-up, as well as the time required to assemble the equipment. Recording additional types of articulatory data thus requires a compromise with portability and ease of set-up. In this case,



since recordings took place at academic institutions, it was possible to save time by keeping the assembled equipment locked in a sound studio. However, future researchers wishing to collect e.g. interview data in speakers' homes, would need to consider the time required to set up and troubleshoot the equipment before recording. Nevertheless, this study has shown that it is more than possible to collect ultrasound data in the field, to a much greater degree than was possible with earlier, larger, video-based ultrasound machines.

The study was also able to record a very large speaker sample (in the context of UTI research), although this sample is small compared to many acoustic studies. In total, 62 speakers were recorded, and 34 were selected for the analysis. Processing this amount of data was very challenging, as both the acoustic and articulatory data had to be segmented by hand. In total, around 14,700 acoustic segments and around 12,200 tongue splines were processed over the course of 36 months. Fortunately, for researchers of other languages than Swedish, automatic segmentation using forced alignment (e.g. Ludwig Maximilian University of Munich, 2019; Rosenfelder, Fruehwald, Evanini, & Yuan, 2011) may be an option for acoustic data, but transcription would still mostly have to be carried out manually. In addition, the AAA software (Articulate Instruments Ltd., 2019), used to create the tongue splines in the current work, has developed greatly over the few years that this project has run, and it is now possible to at least semi-automate the creation of tongue splines using predictive batch processing. This function is still under development, and requires relatively high-quality ultrasound images, as well as visual inspection and hand correction, but shows great potential in increasing the sample size of ultrasound studies, or for providing large quantities of dynamic tongue data. Large sample sizes of ultrasound data have also successfully been processed by Mielke and Roy (2009) using principal component analysis.

An innovative contribution to UTI methodology in the current study was the use of the bite plate, which is normally used to correct the rotation of the ultrasound image (e.g. Strycharczuk & Scobbie, 2017), to estimate the position of the front teeth (following Lawson et al., 2019). Since the length of the bite plate is known, and the end that has been inserted into the mouth can be seen when the tongue presses against it, it was possible to use a scaled fiducial line in AAA to estimate the position of the front teeth. This information is valuable, not only for establishing the position of the tongue in relation to the teeth, but also for estimating the size of the oral cavity, and evaluating how much of the tongue can be seen on the ultrasound image.

### **Analysing UTI data**

In addition to recording and processing UTI data, this study also presents some innovative methods of analysing articulatory vowel data. The most notable of these is the use of multiple normalised measurement points along the tongue curve to directly investigate the correspondence between different articulatory parameters and the vowel formants. While most articulatory work on vowels is either qualitative (e.g. Ladefoged et al., 1972) or focuses on the highest point of the

tongue (e.g. Blackwood Ximenes et al., 2017), this study attempts to quantify the tongue curve at multiple specific points of interest. In the case of Viby-i, some of these points were found to be more informative than the highest point of the tongue, and indeed necessary for describing this vowel in an adequate way. It would be interesting to see to what extent similar measurements would add to the articulatory description of other vowel sounds.

Following Lawson et al. (2019), the current study also uses z-score normalisation of articulatory measures, using the Lobanov (1971) method, which is already commonly applied to acoustic vowel data, and has been rated as one of the best-performing acoustic normalisation methods (Adank, Smits, & van Hout, 2004). The normalisation of single-point articulatory vowel measures is an important step towards comparing multiple speakers' data, as differences in vocal tract size and shape have so far meant that tongue gestures could only be compared within individual speakers. The pilot study to this project (Westerberg, 2016) explored whether Viby-i could be normalised by finding equivalent measurements relative to an 'anchor vowel' (in this case, /e:/), and creating a ratio between the two (for example, the highest point of /i:/ divided by the highest point of /e:/). However, this method presumes that the anchor vowel remains stable across speakers and timepoints, which may not be the case, particularly in Swedish. The use of z-scoring does not rely on a specific vowel, but uses all available data of e.g. the highest point of the tongue, and assigns each data point a score based on its distribution within this dataset. Since z-scoring is based on each value's position within a normal distribution, rather than the raw value in itself, relative differences in vocal tract size are normalised. The normalised articulatory data can then be directly compared to corresponding normalised acoustic values.

A few issues still arise in the use of normalised articulatory data points. For example, there is the issue of which points to select. In this study, the choice of measurement categories were based on visual inspection of the tongue splines, as well as information from previous literature about which points might be of interest, e.g. the highest point of the tongue (D. Jones, 1917), the tongue tip (Björsten & Engstrand, 1999), and the 'midback' measure (Lawson et al., 2019). However, some exploratory measures were also used, e.g. maximum constriction. Hypothetically, the more articulatory measures entered into a regression model, the more accurate the model should be in predicting the acoustic output, since it increasingly approximates the full tongue curve. However, most regression models also come with the limitation that too many co-varying factors will increase the risk of a type II error. However, because the tongue is a physical object, each point along its curve will to some extent be correlated with the previous and the following points. GAMM models are able to circumvent this problem (Sóskuthy, 2017), but at the moment, there is no methodology that combines GAMMs with normalised single-point articulatory measures. Thus, in normalising articulatory tongue data, phonetic researchers need to make careful, informed choices about which measurement points they wish to investigate.

It should also be noted that there is a choice between 'static' measurement points, e.g. where the tongue curve intersects a specific ultrasound fan line, and 'relative' measurement points, e.g.

the highest point of the tongue. The relative aspect comes from differences in tongue shape: For a sloped shape, the highest point of the tongue may be the same as the tongue tip. These cases will not correspond to traditional descriptions of the highest point, where an arched tongue shape is assumed, and where the highest point can be taken to represent the overall mass of the tongue body. Since this study found a large number of bunched tongue shapes with high tongue tips, the measurement points had to be manually inspected to ensure that the highest point was not biased towards the tip, which would have made the tongue body appear fronter than it actually was. The inspection showed that the highest point of the tongue largely corresponded to the position of the tongue body, rather than the tip, but researchers should be aware of this issue if automatically sampling single-point measures of the tongue surface.

In summary, this thesis contributes to UTI methodology in its use of fieldwork; its large sample size; and in methodological developments, e.g. using the bite plate to estimate the position of the teeth, using multiple single-point measurements to represent the tongue, and using z-score normalisation to compare articulatory data across multiple speakers. Although there are still issues to be resolved in articulatory vowel research, this study has hopefully brought the field one step closer to standardised, large-scale analysis methods for articulatory data, similar to those currently used for acoustic analysis.

## 8.4 Future directions

Although this study has provided a comprehensive description of Viby-i in acoustics and articulation, there are still a number of interesting directions for future research, some of which have been mentioned earlier in this chapter.

Firstly, more data is needed from a larger sample of speakers, to represent different geographical and social backgrounds, and to account for different speech styles. As this study only provides word list data of highly-educated, ethnically Swedish speakers from three Central Swedish cities, there are many sociolinguistic factors that have not yet been explored. Further sociolinguistic research could create a better understanding of how widespread Viby-i currently is in Central Sweden, and whether it has indeed become accepted as the standard variant. Such research could also tell us more about speakers' awareness of the social function in this sound, e.g. if they style-shift between different contexts, as indicated by Norstedt (2019). It would perhaps be possible to investigate these issues (at least acoustically) using large-scale crowd-sourced data, e.g. by allowing speakers to record themselves and log their location through a mobile phone application (see Leemann, Kolly, & Britain, 2018).

It would also be interesting to further investigate listeners' awareness of, and attitudes towards, Viby-i. Some work in this area has already been carried out by Norstedt (2019), but this work is limited to Stockholm. As a Viby-i user from the Gothenburg area, I would be interested in whether speakers from outside Stockholm are able to identify Viby-i in their own dialect, and

whether their attitudes towards this sound differs depending on which dialect it occurs in.

We also know very little about the rural link to *Viby-i*, and how it has developed over time. Since Lindström (2014) suggests that *Viby-i* has spread from Bohuslän to Stockholm, it would perhaps be relevant to start by investigating historical archives of data from these locations, to examine whether *Viby-i* is indeed ‘the same’ in both places, and when it first started to appear in urban environments. From a modern perspective, it would also be interesting to examine how rural *Viby-i*, which is usually regarded as having low prestige, behaves when it comes into contact with urban *Viby-i*, which has become high-prestige in standard language (Bruce, 2010).

There are also several acoustic aspects of *Viby-i* that would benefit from further study. In particular, its relationship to standard [i:] has not previously been investigated in the literature, and would provide a better understanding of the ways in which *Viby-i* differs from this sound. Furthermore, this study was not able to present bandwidth data, but it seems likely that properties beyond formant values may be relevant for the “damped” vowel quality associated with *Viby-i* (Björsten & Engstrand, 1999). Perceptual data of how the different acoustic aspects of *Viby-i* are perceived by listeners would also be extremely valuable.

In terms of articulation, there are some methodological limitations to the current UTI study, which could be addressed in future work. For example, UTI can only image the tongue surface, and to some extent the palate, but does not provide a full picture of the oral cavity. Furthermore, although midsagittal data has been said to be the most informative for vowel articulation (Russell, 1928: 119), three-dimensional behaviour could also be important for the production of *Viby-i*, and for vowels in general (Stone & Lundberg, 1996). The use of e.g. 3D ultrasound, MRI, EMA, or a combination of these methods (see Engwall, 2003), would help researchers obtain a more holistic picture of the tongue, and of the entire acoustic chamber. While such investigations rely heavily on the accessibility of advanced equipment, the rapid development in articulatory phonetic methods in recent years provides an encouraging context for these investigations to start.

# Chapter 9

## Conclusion

In conclusion, this thesis has provided a comprehensive phonetic description of Swedish Viby-i, filling a substantial gap in the literature about the acoustic, articulatory, and sociolinguistic qualities of this vowel.

The study has shown that Viby-i is acoustically characterised by a low F2, and to some extent a high F1, compared to /e:/. Viby-i is also frequently subject to end-frication, similar to other high vowels in Swedish (Engstrand, 1999). Frication of the vocalic portion of the vowel is also possible, but was less common in this sample. These acoustic parameters are likely to give rise to the “damped”, “buzzing” vowel quality associated with Viby-i (Engstrand et al., 1998: 1-2).

In articulation, Viby-i was found to be produced with a low and fronted tongue gesture, with simultaneous retraction of the post-dorsal region, as well as a high tongue tip, and spread lips. This complex articulation often resulted in a double-bunched tongue shape, but many speakers also used a canonical arched shape to achieve this sound. Overall, a great deal of variation was observed in the articulatory data, possibly due to anatomical differences between speakers, but relatively little variation was observed in acoustics. The articulatory analysis also found a surprising mismatch between acoustics and articulation, in that the low F2 of Viby-i did not seem to be produced by overall tongue backing, but by a combination of tongue lowering and post-dorsal retraction. This finding contradicts the traditional description of vowel articulation in the literature (e.g. Laver, 1994), and demonstrates the need to combine acoustic and articulatory data for a fuller understanding of vowel production.

The sociolinguistic analysis revealed that Viby-i could be found across age, gender, and lifestyle groups in Gothenburg, Stockholm, and Uppsala, suggesting that this vowel is widespread in Central Sweden. It is possible that Viby-i has already become the standard variant in the sampled populations, although further research is needed to confirm this. The three cities also varied somewhat in their production of Viby-i, with Gothenburg producing a higher F1 and lower F2 for this vowel, as well as using a lower tongue body. This result could be taken to support claims from the literature that Viby-i is older in Gothenburg (e.g. Elert, 1995), and that the shift towards this vowel has therefore progressed further there than in Stockholm and Uppsala. In

contrast with previous sociolinguistic research (Grönberg, 2004; Kotsinas, 2007), the current study did not find that *Viby-i* displayed the social stratification patterns associated with change in progress, or change in apparent time.

The production of *Viby-i* was also found to be affected by linguistic context, with duration and consonant environment both influencing the formant values of this vowel. The fricative offglides of *Viby-i* were also strongly conditioned by the following consonant. However, the presence of dynamic acoustic frication was not fully matched by dynamic movement in the articulatory data, prompting further investigation of this phenomenon.

The study provides several innovative contributions to current UTI methodology, notably the use of normalised articulatory measurement points, which can be compared to equivalent acoustic data. Although this study is not the first to provide articulatory data of *Viby-i* (see Frid et al., 2015), it is the first to provide an articulatory analysis of the correspondence between the tongue gesture and corresponding acoustic output for this vowel.

This thesis has thus contributed to an increased understanding of how *Viby-i* is produced, which may be applicable to the study of other vowel sounds. The work has also moved the field of articulatory vowel phonetics one step closer to a standard methodology that is able to describe the production of both conventional and unconventional vowels.

# References

- Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5), 3099–3107.
- af Klintberg, K., & Ripås, J. (2003). *Värsta språket (season 1, episode 3)*. Sveriges Television.
- Altairi, H., Brown, J., Watson, C., & Gick, B. (2017). Tongue retraction in Arabic: An ultrasound study. In *Proceedings of the Annual Meetings on Phonology*. New York.
- Alwabari, S. (2017). Articulatory constraints and sensitivity to coarticulation of Arabic pharyngealization. In *Proceedings of Ultrafest 8* (pp. 25–26). Potsdam.
- Anonymous. (2007). *Vem i helvete vill se era slappa tungor?* Retrieved 2019-09-13, from <http://www.vemihelvete.se/2007/12/vem-i-helvete-vill-se-era-slappa-tungor.html>
- Anonymous. (2015). *Swedish Pronunciation: Long i sound*. Retrieved 2019-11-19, from <https://forum.wordreference.com/threads/swedish-pronunciation-long-i-sound.3029643/>
- Articulate Instruments Ltd. (2008). *Ultrasound Stabilisation Headset User's Manual, Revision 1.4*. Edinburgh.
- Articulate Instruments Ltd. (2010). *SyncBrightUp User's Manual, Revision 1.10*. Edinburgh.
- Articulate Instruments Ltd. (2018). *Pulse-Stretch Unit*. Edinburgh.
- Articulate Instruments Ltd. (2019). *Articulate Assistant Advanced, version 2.17*. Retrieved 2019-03-26, from <http://www.articulateinstruments.com/downloads/>
- Ashby, M. G. (2016). *Experimental Phonetics in Britain, 1890-1940* (Doctoral thesis). University of Oxford.
- Atkinson, H. W. (1897). Tongue positions of vowel-sounds. *The Modern Language Quarterly*, 1(1), 13–20.
- Baer, T., Gore, J. C., Boyce, S. E., & Nye, P. W. (1987). Application of MRI to the analysis of speech production. *Magnetic Resonance Imaging*, 5, 1–7.
- Bailey, C.-J. N. (1973). *Variation and Linguistic Theory*. Washington D.C.: Center for Applied Linguistics.
- Baker, A., Mielke, J., & Archangeli, D. (2008). More velar than /g/: Consonant coarticulation

- as a cause of diphthongization. *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 60–68.
- Bakst, S. (2016). Differences in the relationship between palate shape, articulation, and acoustics of American English /r/ and /s/. *UC Berkeley Phonetics and Phonology Lab Annual Report*, 216–224.
- Barry, J. W., & Timmermann, G. (1985). Mispronunciations and compensatory movements of tongue-operated patients. *International Journal of Language & Communication Disorders*, 20(1), 81–90.
- Barth, E. (1907). Röntgenographische Beiträge zur Stimmphysiologie. *Archiv für Laryngologie und Rhinologie*, 19, 396–405.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Beck, J. M. (2010). Organic variation of the vocal apparatus. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 155–201). Oxford: Wiley-Blackwell.
- Bell, A. M. (1867). *Visible Speech: The Science of Universal Alphabetics, or Self-Interpreting Physiological Letters, for the Writing of All Languages in One Alphabet*. London: Simpkin, Marshall & Co.
- Berzelius, H. (2012). Bildspecial: Möt hemvändaren Hjalmar. *Nerikes Allehanda*, 20 April. Retrieved from <https://www.na.se/artikel/bildspecial-mot-hemvandaren-hjalmar>
- Björseth, B. (1958). *Göteborgsspråket*. Stockholm: Svenska bokförlaget.
- Björsten, S., & Engstrand, O. (1999). Swedish "damped" /i/ and /y/: Experimental and typological observations. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1957–1960). San Francisco.
- Bjurström, E. (1997). *Högt och lågt: Smak och stil i ungdomskulturen*. Umeå: Boréa.
- Blackwood Ximenes, A., Shaw, J. A., & Carignan, C. (2017). A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *The Journal of the Acoustical Society of America*, 142(1), 363–377.
- Block, J. R. (1968). The role of the speech clinician in determining indications for frenulotomy in cases of ankyloglossia. *The New York State Dental Journal*, 34(8), 479–481.
- Bodén, P. (2010). Pronunciation in Swedish multiethnolect. In P. Quist & B. A. Svendsen (Eds.), *Multilingual Urban Scandinavia: New Linguistic Practices* (pp. 65–78). Bristol: Multilingual Matters.
- Boersma, P., & Weenink, D. (1995). *Praat: Doing Phonetics by Computer*.
- Boersma, P., & Weenink, D. (2019). *Praat: Doing Phonetics by Computer, version 2.0.29*. Retrieved 2019-03-26, from <http://www.praat.org/>



- Borgström, M. (1913). Askermålets ljudlära. *Svenska landsmål och svenskt folkliv*, B(11).
- Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp: The corpus infrastructure of Språkbanken. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 474–478). Istanbul.
- Borneskog, I. (2011). *Deconstructing the brat phenomenon* (Masters dissertation). Stockholm University.
- Boyce, S. E., Tiede, M., Espy-Wilson, C. Y., & Groves-Wright, K. (2015). Diversity of tongue shape for the American English rhotic liquid. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Boyd, S. (2010). Sociolinguistics in the Nordic region. In M. J. Ball (Ed.), *Sociolinguistics around the World: A Handbook* (pp. 281–295). London: Routledge.
- Britain, D. (2009). One foot in the grave? Dialect death, dialect contact, and dialect birth in England. *International Journal of the Sociology of Language*(196-197), 121–155.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Haskins Laboratories Status Report on Speech Research*, 111-112, 23–42.
- Bruce, G. (2010). *Vår fonetiska geografi: Om svenskans accenter, melodi och uttal*. Lund: Studentlitteratur.
- Burris, C., Vorperian, H. K., Fourakis, M., Kent, R. D., & Bolt, D. M. (2014). Quantitative and descriptive comparison of four acoustic analysis systems: Vowel measurements. *Journal of Speech, Language, and Hearing Research*, 57(1), 26–45.
- Bye, P. (2004). *Evolutionary Typology and Scandinavian Pitch Accent*. Dordrecht: Kluwer Academic Publishers.
- Catford, J. C. (1981). Observations on the recent history of vowel classification. In R. Asher & E. Henderson (Eds.), *Towards a History of Phonetics* (pp. 19–34). Edinburgh: Edinburgh University Press.
- Catford, J. C. (1994). *A Practical Introduction to Phonetics*. Oxford: Clarendon Press.
- Chambers, J. K. (2003). *Sociolinguistic Theory: Linguistic Variation and its Social Significance* (2nd ed.). Oxford: Blackwell.
- Chambers, J. K., & Trudgill, P. (1988). *Dialectology*. Cambridge: Cambridge University Press.
- Chiba, T., & Kajiyama, M. (1941). *The Vowel: Its Nature and Structure*. Tokyo: Tokyo-Kaiseikan.
- Clark, J., Yallop, C., & Fletcher, J. (2007). *An Introduction to Phonetics and Phonology* (3rd ed.). Oxford: Blackwell.
- Cleland, J., Scobbie, J. M., Heyde, C., Roxburgh, Z., & Wrench, A. A. (2017). Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics & Phonetics*, 31(1), 35–55.
- Cohen, C. (2019). *Python script for converting AAA coordinates to CSV*.
- Connell, B. (2000). Mambila fricative vowels. In V. Carstens & F. Parkinson (Eds.), *Advances*

- in *African Linguistics* (pp. 233–250). Trenton: Africa World Press.
- Coretta, S. (2019). *rticulate: Ultrasound Tongue Imaging in R, version 1.5.0*. Retrieved 2019-08-11, from <https://github.com/stefanocoretta/rticulate>
- Darwin, E. (1804). *The Temple of Nature; or, the Origin of Society: A Poem, with Philosophical Notes*. Baltimore: M. and J. Conrad & Co.
- Delattre, P. (1951). The physiological interpretation of sound spectrograms. *Publications of the Modern Language Association of America*, 66(5), 864–875.
- Delattre, P., & Freeman, D. C. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics*, 6(44), 29–68.
- Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, 27(1-2), 47–55.
- Dickson, V., & Hall-Lew, L. (2017). Class, gender, and rhoticity: The social stratification of non-prevocalic /r/ in Edinburgh speech. *Journal of English Linguistics*, 45(3), 229–259.
- Donald, I., Macvicar, J., & Brown, T. G. (1958). Investigation of abdominal masses by pulsed ultrasound. *The Lancet*, 271(7032), 1188–1195.
- Eager, C., & Roy, J. (2017). *Mixed Effects Models are Sometimes Terrible*. Retrieved 2019-05-08, from <https://arxiv.org/abs/1701.04858>
- Eckert, P. (1989). The whole woman: Sex and gender difference in variation. *Language Variation and Change*, 1(3), 245–267.
- Eckert, P. (1997). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 151–167). Oxford: Blackwell.
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41, 87–100.
- Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, 54, 1–21.
- Ekstrand, L. (2016). Göteborgskan breder ut sig. *Göteborgs-Posten*, 27 Mar. Retrieved from <https://www.gp.se/nyheter/g\teborg/g\teborgskan-breder-ut-sig-1.6957>
- Elert, C.-C. (1964). *Phonological Studies of Quantity in Swedish*. Uppsala: Almqvist & Wiksell.
- Elert, C.-C. (1995). *Allmän och svensk fonetik*. Stockholm: Norstedts.
- Engstrand, O. (1999). Swedish. In F. Nolan et al. (Eds.), *Handbook of the International Phonetic Association* (pp. 140–142). Cambridge: Cambridge University Press.
- Engstrand, O. (2004). *Fonetikens grunder*. Lund: Studentlitteratur.
- Engstrand, O., Björnsten, S., Lindblom, B., Bruce, G., & Eriksson, A. (1998). Hur udda är Viby-i? Experimentella och typologiska observationer. *Folkmålsstudier*, 39, 83–95.
- Engstrand, O., Bruce, G., Elert, C.-C., Eriksson, A., & Strangert, E. (2000). *Yngre man, Orust* (SweDia database). Retrieved 2019-11-17, from <http://swedia.ling.gu.se/>

Gotaland/Bohuslan/Orust/ym.html

- Engwall, O. (2003). Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. *Speech Communication*, 41(2-3), 303–329.
- Engwall, O., & Badin, P. (1999). Collecting and analysing two- and three-dimensional MRI data for Swedish. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 3-4, 11–38.
- Ericsson, C. (2005). *Articulatory-Acoustic Relationships in Swedish Vowel Sounds* (Doctoral thesis). Stockholm University.
- Ericsson, J., & Nilsson, A. (2013). *SNN News*. TV4 Sweden.
- Erikson, Y., & Alstermark, M. (1972). Fundamental frequency correlates of the grave word accent in Swedish: The effect of vowel duration. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 13(2-3), 53–60.
- Eriksson, A. (2004). SweDia 2000: A Swedish dialect database. In *Proceedings of the Nordic Symposium on the Comparison of Spoken Languages* (pp. 33–48). Fredriksberg.
- Eriksson, S. (2008, jan). Stort hemlighetsmakeri kring varumärket Liidingö. *Dagens Media*, 15 Jan. Retrieved from <https://www.lidingosidan.se/lidingo/varumarket-liiidingo-registrerat/>
- Essner, C. (1947). Recherche sur la structure des voyelles orales. *Archives Néerlandaises de Phonétique Expérimentale*, 20, 40–77.
- Fagerström, J. (2019). *Jonas Fagerström (Instagram)*. Retrieved 2019-11-19, from <https://www.instagram.com/p/4FEolZvioM/>
- Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics*, 15, 3–108.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fant, G. (1972). Vocal tract wall effects, losses, and resonance bandwidths. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 13(2-3), 28–52.
- Fant, G. (1981). The source filter concept in voice production. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 22(1), 21–37.
- Fant, G., Henningsson, G., & Stålhammar, U. (1969). Formant frequencies of Swedish vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 10(4), 26–31.
- Faytak, M. (2014). Chain shifts, strident vowels, and expanded vowel spaces. In *Linguistic Society of America Annual Meeting, Extended Abstracts* (pp. 1–4). Minneapolis.
- Faytak, M., & Lin, S. (2015). Articulatory variability and fricative noise in apical vowels. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics using R*. London: SAGE Publications.
- Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, 32(1), 1–24.

- Frid, J., Schötz, S., Gustafsson, L., & Löfqvist, A. (2015). Tongue articulation of front close vowels in Stockholm, Gothenburg and Malmöhus Swedish. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Frid, J., Schötz, S., & Löfqvist, A. (2011). Age-related lip movement repetition variability in two phrase positions. In *Proceedings of FONETIK* (pp. 21–24). Stockholm, Sweden.
- Gårding, E. (1977). *The Scandinavian Word Accents*. Lund: Gleerups.
- Garlén, C. (1988). *Svenskans fonologi*. Lund: Studentlitteratur.
- Gick, B. (2019). How bodies talk. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 20–24). Melbourne.
- Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Towards universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, 34, 49–72.
- Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory Phonetics*. Chichester: Wiley-Blackwell.
- Google Maps. (2019). 'Measure Distance' Function. Retrieved 2019-03-26, from <http://www.maps.google.com/>
- Gordeeva, O. B., & Scobbie, J. M. (2010). Preaspiration as a correlate of word-final voice in Scottish English fricatives. In S. Fuchs, M. Toda, & M. Zygis (Eds.), *Turbulent sounds: An Interdisciplinary Guide* (pp. 167–208). New York: Walter de Gruyter.
- Gordeeva, O. B., & Scobbie, J. M. (2013). A phonetically versatile contrast: Pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality. *Journal of the International Phonetic Association*, 43(3), 249–271.
- Gósy, M. (2011). From stomatoscopy to BEA: The history of Hungarian experimental phonetics. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 172–175). Hong Kong.
- Götlind, J. (1940). *Västergötlands folkmål. Del I: Vokalerna*. Lund: Lundequistska bokhandeln.
- Grönberg, A. G. (2004). *Ungdomar och dialekt i Alingsås* (Doctoral thesis). Gothenburg University.
- Gross, J. (2018). Segregated vowels: Language variation and dialect features among Gothenburg youth. *Language Variation and Change*, 30, 315–336.
- Gross, J., Boyd, S., Leinonen, T., & Walker, J. A. (2016). A tale of two cities (and one vowel): Sociolinguistic variation in Swedish. *Language Variation and Change*, 28(2), 225–247.
- Gross, J., & Forsberg, J. (2019). Weak Lips? A possible merger of /i:/ and /y:/ in Gothenburg. *Phonetica*, 1–21.
- Grunmach, E. (1907). Die Röntgentechnik zur Untersuchung der Mund-, Schlund- und Nasenhöhle bei der Phonation. *Archiv für Laryngologie und Rhinologie*, 19, 405–407.
- Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English

- /r/ production. *The Journal of the Acoustical Society of America*, 105(5), 2854–2865.
- Gutzmann, H. (1930). Röntgenkinematographie der Sprechbewegungen. In *Verhandlungen: Kongress Internationaler Gesellschaft Logopaedie und Phonatrie 4* (pp. 94–100).
- Hårdstedt, V. (2009). Visst har upplänningen mål i mun. *Uppsala Nya Tidning*, 3 Feb. Retrieved from <https://www.unt.se/nyheter/uppsala/visst-har-%0Aupplanningen-mal-i-mun-307282.aspx>
- Harrell, F. E. . (2019). *Hmisc: Harrell Miscellaneous, version 4.2-0*. Retrieved 2019-11-28, from <https://cran.r-project.org/package=Hmisc>
- Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, 34(4), 439–457.
- Haugen, E. I. (1976). *The Scandinavian Languages: An Introduction to their History*. Cambridge, MA: Harvard University Press.
- Hay, J. (2011). Statistical analysis. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A Student's Guide* (pp. 198–214). London: Routledge.
- Heinz, J. (2011). *Acoustic and Auditory Phonetics* (Teaching materials). University of Delaware.
- Helgason, P. (2002). *Preaspiration in the nordic languages: Synchronic and diachronic aspects* (Doctoral thesis). University of Stockholm.
- Hellberg, M. (2012). *Svenska dialektmysterier: Bohuslänska (season 2, episode 2)*. Sveriges Television.
- Hlavac, M. (2018). *stargazer: Well-formatted regression and summary statistics tables, version 5.2.2*. Retrieved 2019-11-28, from <https://cran.r-project.org/package=stargazer>
- Hockett, C. F. (1950). Age-grading and linguistic continuity. *Language*, 26(4), 449–457.
- Holmberg, B. (1976). *Språket i Göteborg*. Stockholm: Esselte Studium.
- Holmes, J. (2013). *An Introduction to Sociolinguistics* (4th ed.). London: Routledge.
- Holmes-Elliott, S., & Levon, E. (2017). Reality check: Stance, style and the short vowel system of Southern British English. In *Paper presented at uk language variation and change 11*. Cardiff.
- Hoole, P., & Nguyen, N. (1997). Electromagnetic articulography in coarticulation research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 35, 177–184.
- House, A. S., & Fairbanks, G. (1953). The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25, 105–112.
- International Phonetic Association. (1949). *The Principles of the International Phonetic Association*. London: University College London, Department of Phonetics.
- International Phonetic Association. (2015). *The International Phonetic Alphabet (Revised to*

2015).

- Johansson, R. (2015). *A trip to Lyysekiil*. Retrieved 2019-11-17, from <https://nouw.com/rebeccaajhnsn/a-trip-to-lyyysekiil-19029821>
- Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359–383.
- Johnson, K. (1997). *Acoustic & Auditory Phonetics*. Oxford: Blackwell.
- Jones, D. (1917). *Everyman's English Pronouncing Dictionary*. London: J.M. Dent & Sons.
- Jones, D. (1972). *An Outline of English Phonetics* (9th ed.). Cambridge: W. Heffer & Sons Ltd.
- Jones, S. (1929). Proceedings of the Society of Radiographers: Radiography and pronunciation. *British Journal of Radiology*, 2, 149–152.
- Jönsson, A. (2014). *Tala om tungor: Dialekter i svensk film* (Undergraduate dissertation). Lund University.
- Joos, M. (1948). Acoustic phonetics. *Language*, 24, 1–136.
- Joy, J., Cooke, I., & Love, M. (2006). Is ultrasound safe? *The Obstetrician & Gynaecologist*, 8, 222–227.
- Källskog, M., Eklund, G., Danielsson, B., Hagren, K., Westerberg, A., Västerlund, R., & Reinhammar, M. (1993). *Uppländska: Språkprov med kommentar*. Uppsala: Almqvist & Wiksell.
- Karlgren, B. (1915). *Études sur la Phonologie Chinoise*. Uppsala: K.W. Appelberg.
- Kederstedt, D. (2007, feb). Kommunerna som har högst och lägst medelinkomst. *Svenska Dagbladet Näringsliv*, 22 Feb. Retrieved from <https://www.svd.se/kommunerna-som-har-hogst-och-lagst-medelinkomst>
- Kendall, T., & Thomas, E. R. (2018). *vowels: Vowel manipulation, normalization, and plotting, version 1.2-2*. Retrieved 2019-11-28, from <https://cran.r-project.org/package=vowels>
- Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In D. Britain & J. Cheshire (Eds.), *Social Dialectology: In Honour of Peter Trudgill* (pp. 223–244). Amsterdam: John Benjamins.
- Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., . . . Senda, M. (2005). Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner. *Acoustical Science and Technology*, 26(5), 465–468.
- Koenig, W., Dunn, H. K., & Lacey, L. Y. (1946). The sound spectrograph. *The Journal of the Acoustical Society of America*, 18, 19–49.
- Kotsinas, U.-B. (1988). Immigrant children's Swedish: A new variety? *Journal of Multilingual and Multicultural Development*, 9(1-2), 129–140.
- Kotsinas, U.-B. (2007). *Ungdomsspråk* (3rd ed.). Uppsala: Hallgren & Fallgren.
- Kuecker, K., Lockenvitz, S., & Müller, N. (2015). Amount of rhoticity in *schwar* and in vowel+/r/ in American English. *Clinical Linguistics and Phonetics*, 29(8-10), 623–629.

- Kuronen, M. (2000). *Vokaluttalets akustik i sverigesvenska, finlandssvenska och finska* (Doctoral thesis). University of Jyväskylä.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- Labov, W. (1966). The effect of social mobility on linguistic behavior. *Sociological Inquiry*, 36, 186–203.
- Labov, W. (1972). *Sociolinguistic Patterns*. Oxford: Blackwell.
- Labov, W. (1994). *Principles of Linguistic Change, Volume 1: Internal Factors*. Oxford: Blackwell.
- Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Oxford: Blackwell.
- Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonology, Phonetics, and Sound Change*. Berlin: Mouton De Gruyter.
- Ladefoged, P. (1962). *Elements of Acoustic Phonetics*. Chicago: University of Chicago Press.
- Ladefoged, P. (1967). *Three Areas of Experimental Phonetics: Stress and Respiratory Activity; The Nature of Vowel Quality; Units in the Perception and Production of Speech*. London: Oxford University Press.
- Ladefoged, P. (1976). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- Ladefoged, P. (1993). *A Course in Phonetics* (3rd ed.). Fort Worth: Harcourt Brace & Co.
- Ladefoged, P. (1996). *Elements of Acoustic Phonetics* (2nd ed.). Chicago: University of Chicago Press.
- Ladefoged, P. (2001). *Vowels and Consonants*. Oxford: Wiley-Blackwell.
- Ladefoged, P., DeClerk, J., Lindau, M., & Papçun, G. (1972). An auditory-motor theory of speech production. *Working Papers in Phonetics*, 22, 48–75.
- Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th ed.). Stamford: Cengage Learning.
- Ladefoged, P., & Lindau, M. (1989). Modeling articulatory-acoustics relations: A comment on Stevens' "On the quantal nature of speech". *Journal of Phonetics*, 17, 99–106.
- Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- Langenfelt, G. (1953). *Staden vid de ljusa vattnen*. Stockholm: KF:s bokförlag.
- Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples by employing nuclear magnetic resonance. *Nature*, 242, 190–191.
- Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, 15(2), 256–268.
- Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2013). Bunched /r/ promotes vowel merger to schwa: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal*

- of Phonetics*, 41(3-4), 198–210.
- Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2014). A socio-articulatory study of Scottish rhoticity. In R. Lawson (Ed.), *Sociolinguistics in Scotland* (pp. 53–78). London: Palgrave Macmillan.
- Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of articulatory and acoustic parameters for the GOOSE vowel across British Isles Englishes. *The Journal of the Acoustical Society of America*, 146.
- Lazić, D. (2012). *Hur talar uppländska ungdomar? En studie av uppländska dialektala särdrag* (Undergraduate dissertation). Uppsala University.
- Leemann, A., Kolly, M., & Britain, D. (2018). The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand*, 5, 1–17.
- Leinonen, T. (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects* (Doctoral dissertation). Groningen University.
- Leinonen, T. (2011). Aggregate Analysis of Vowel Pronunciation in Swedish Dialects. *Oslo Studies in Language*, 3(2), 75–95.
- Leplâtre, G., & Lawson, E. (2019). *R script for plotting and measuring tongue curves at multiple points*.
- Lindberg, M. (2016). Karta över dialektområden. *Ordbok över Finlands svenska folk-mål*. Retrieved from [http://www.sprakinstitutet.fi/sv/ordbocker/dialektordbocker/ordbok\\_over\\_finlands\\_svenska\\_folkmal/karta\\_over\\_dialektomraden](http://www.sprakinstitutet.fi/sv/ordbocker/dialektordbocker/ordbok_over_finlands_svenska_folkmal/karta_over_dialektomraden)
- Lindberg, S. (2005). *Hey Baberiba*. TV4 Sweden.
- Lindblad, P. (2010). *Fonetikens grunder* (Teaching materials). Lund University.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35, 1773–1781.
- Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 8(4), 1–29.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 13–44). Orlando: Academic Press.
- Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, 50(4), 1166–1179.
- Lindström, F. (2014). Därför surrar stekaren som fiskaren från Smögen. *Språktidningen*, Aug. Retrieved from <https://spraktidningen.se/artiklar/2014/08/darfor-surrar-stekaren-som-fiskaren-fran-smogen/>
- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different listeners. *The Journal of the Acoustical Society of America*, 49, 606–608.
- Ludwig Maximilian University of Munich. (2019). *Munich Automatic Segmentation System (MAUS)*. Retrieved 2019-05-08, from <http://www.bas.uni-muenchen.de/>



Bas/BasMAUS.html

- Lundell, J. A. (1878). Det svenska landsmålsalfabetet. *Svenska landsmål och svenskt folkliv*, 1(2), 11–158.
- Macaulay, R. K. S. (1977). *Language, social class and education: A Glasgow study*. Edinburgh: Edinburgh University Press.
- Macmillan, A. S., & Kelemen, G. (1952). Radiography of the Supraglottic Speech Organs: A Survey. *AMA archives of otolaryngology*, 55(6), 671–688.
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Malmberg, B. (1956). Distinctive features of Swedish vowels: Some instrumental and structural data. In *For Roman Jakobson: Essays on the Occasion of his Sixtieth Birthday* (pp. 316–321). The Hague: Mouton.
- Malmberg, B. (1971). *Svensk fonetik*. Lund: Gleerups.
- Mehner, H. (2017). *Om Göteborgs historia*. Retrieved 2017-03-27, from <http://goteborg.se/wps/portal/start/kommun-o-politik/kommunfakta/historia/om-historia/>
- Melin, L., & Melin, M. (2005). *Fiint språk: Språket som förhöjer, förför och förargar*. Stockholm: Schibsted Förlagen.
- Menzerath, P., & de Lacerda, A. (1933). *Koartikulation, Steuerung und Lautabgrenzung*. Berlin: Dümmler.
- Meyer, E. A. (1907). Röntgenologische Lautbilder. *Medizinisch-pädagogische Monatsschrift für die gesamte Sprachheilkunde*, 17, 225–243.
- Mielke, J., & Roy, J. (2009). Measuring articulatory similarity with algorithmically reweighted principal component analysis. *The Journal of the Acoustical Society of America*, 125(4), 2499–2499.
- Milroy, L. (1987). *Language and Social Networks*. New York: Blackwell.
- Moisik, S. R., & Dediu, D. (2017). Anatomical biasing and clicks: Evidence from biomechanical modeling. *Journal of Language Evolution*, 2(1), 37–51.
- Nair, A., Schellenberg, M., & Gick, B. (2015). A case study on the efficacy of ultrasound biofeedback in voice pedagogy. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Nearey, T. M. (1977). *Phonetic Feature Systems for Vowels*. Bloomington: Indiana University Linguistics Club.
- Network to Promote Linguistic Diversity. (2016). *Swedish*. Retrieved 2018-10-05, from <http://www.npld.eu/about-us/swedish/>
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85.
- Noiray, A., Iskarous, K., & Whalen, D. H. (2014). Variability in English Vowels is Comparable in Articulation and Acoustics. *Laboratory Phonology*, 5(2), 271–288.

- Nord, L. (1986). Acoustic studies of vowel reduction in Swedish. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 27(4), 19–36.
- Nordberg, B. (1975). Contemporary social variation as a stage in a long-term phonological change. In K.-H. Dahlstedt (Ed.), *The Nordic Languages and Modern Linguistics* (pp. 587–608). Stockholm: Almqvist & Wiksell.
- Noreen, A. (1903). *Vårt språk* (Vol. 1). Lund: Gleerups.
- Norrby, C., & Håkansson, G. (2015). *Introduktion till sociolingvistik* (2nd ed.). Lund: Studentlitteratur.
- Norstedt, L. (2019). *Investigating the Use of the Swedish Viby-i Across Topics* (Masters dissertation). University of Edinburgh.
- Ogden, R. (2009). *An Introduction to English Phonetics*. Edinburgh: Edinburgh University Press.
- Ohala, J. J., & Solé, M.-J. (2010). Turbulence and phonology. In S. Fuchs, M. Toda, & M. Żygis (Eds.), *Turbulent Sounds: An Interdisciplinary Guide* (pp. 37–101). The Hague: Mouton De Gruyter.
- Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(11), 151–168.
- Olsson, K. (1996). *Göteborgs historia: Näringsliv och samhällsutveckling, vol 3. Från industristad till tjänstestad, 1920-1995*. Stockholm: Nerenius & Santérus.
- Organisation for Economic Co-Operation and Development. (2020). *Income inequality (indicator)*. Retrieved 2020-10-19, from <https://data.oecd.org/inequality/income-inequality.htm>
- Pamp, B. (1978). *Svenska dialekter*. Stockholm: Natur och kultur.
- Panconcelli-Calzia, G. (1994). *Geschichtzahlen der Phonetik (1941) and Quellenatlas der Phonetik (1940)*. Philadelphia: John Benjamins.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184.
- Potter, R. K., Kopp, G. A., & Green, H. G. (1946). *Visible Speech*. New York: van Nostrand.
- Pratt, T., & D’Onofrio, A. (2017). Jaw setting and the California vowel shift in parodic performance. *Language in Society*, 46(3), 283–312.
- Preston, J. L., Byun McAllister, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., ... Whalen, D. H. (2017). Ultrasound images of the tongue: A tutorial for assessment and remediation of speech sound errors. *Journal of Visualized Experiments*, 119, 1–10.
- Python Software Foundation. (2019). *Python Language Reference, version 3.7*. Retrieved 2019-11-28, from [www.python.org](http://www.python.org)
- R Foundation for Statistical Computing. (2019). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved 2019-11-28, from <http://www.r-project.org/>

- Recasens, D. (1999). Lingual coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data, and Techniques in Speech Production* (pp. 80–104). Cambridge: Cambridge University Press.
- Riad, T. (2006). Scandinavian accent typology. *Language Typology and Universals*, 59(1), 36–55.
- Riad, T. (2014). *The Phonology of Swedish*. Oxford: Oxford University Press.
- Röntgen, W. C. (1895). *Über eine neue Art van Strahlen*. Würzburg: Stahel.
- Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). *FAVE (Forced Alignment and Vowel Extraction) Program Suite*. Retrieved 2019-11-28, from <http://fave.ling.upenn.edu>
- Rosner, B. S., & Pickering, J. B. (1994). *Vowel Perception and Production*. Oxford: Oxford University Press.
- Roy, J., & Levey, S. (2014). Mixed effects models and unbalanced sociolinguistic data: The need for caution. In *Proceedings of New Ways of Analyzing Variation 43*. Illinois.
- Russell, G. O. (1928). *The Vowel: Its Physiological Mechanism as Shown by X-ray*. New York: McGrath.
- Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, 83(3), 560–588.
- Schaeffler, F. (2005). *Phonological quantity in Swedish dialects: Typological aspects, phonetic variation and diachronic change* (Doctoral thesis). Umeå University.
- Schaeffler, F., Wretling, P., & Strangert, E. (2002). On the development of a quantity typology for Swedish dialects. In *Proceedings of Speech Prosody* (pp. 3–6). Aix-en-Provence.
- Scheier, M. (1897). Die Anwendung der X-strahlen für die Physiologie der Stimme und Sprache. In *Verhandlungen: Congress für inneren Medizin 15* (pp. 531–533).
- Schilling-Estes, N. (1998). Investigating "self-conscious" speech: The performance register in Ocracoke English. *Language in Society*, 17(1), 53–83.
- Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31(1), 26–35.
- Schötz, S., Frid, J., Gustafsson, L., & Löfqvist, A. (2013). Functional data analysis of tongue articulation in palatal vowels: Gothenburg and Malmöhus Swedish /i:, y:, u:/. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)* (pp. 1326–1330). Lyon.
- Schötz, S., Frid, J., Gustafsson, L., & Löfqvist, A. (2014). Tongue articulation dynamics of /i:, y:, u:/ in Stockholm, Gothenburg and Malmöhus Swedish. In *Proceedings of FONETIK* (pp. 17–22). Stockholm.
- Schötz, S., Frid, J., & Löfqvist, A. (2011). Exotic vowels in Swedish: An articulographic and acoustic pilot study of /i:/. In *Proceedings of the 17th International Congress of Phonetic*

- Sciences* (pp. 1766–1769). Hong Kong.
- Scobbie, J. M., Stuart-Smith, J., & Lawson, E. (2012). Back to front: A socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Italian Journal of Linguistics*, 24(1), 103–148.
- Scobbie, J. M., Wrench, A. A., & van der Linden, M. L. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In *Proceedings of the 8th International Seminar on Speech Production* (pp. 373–376). Strasbourg.
- Shao, B., & Ridouane, R. (2019). Apical vowel in Jixi-Hui Chinese: An articulatory study. In *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne.
- Shawker, T. H., Sonies, B., Stone, M., & Baum, B. J. (1983). Real-time ultrasound visualization of tongue movement during swallowing. *Journal of Clinical Ultrasound*, 11(9), 485–490.
- Sóskuthy, M. (2017). *Generalised Additive Mixed Models for Dynamic Analysis in Linguistics: A Practical Introduction* [Teaching materials]. Retrieved 2019-08-11, from [https://github.com/soskuthy/gamm\\_intro](https://github.com/soskuthy/gamm_intro)
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21, 291–311.
- Ståhle, C. I. (1981). *Stockholmsnamn och Stockholmsspråk*. Stockholm: Norstedts.
- Stålhammar, U., Karlsson, I., & Fant, G. (1973). Contextual effects on vowel nuclei. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 14(4), 1–18.
- Stark, J., Lindblom, B., & Sundberg, J. (1996). APEX: An articulatory synthesis model for experimental and computational studies of speech production. In *Proceedings of FONETIK*, pages = 45–48. Stockholm.
- Statistics Sweden. (2015). *Urbanisering: Från land till stad*. Retrieved 2019-11-15, from <https://www.scb.se/hitta-statistik/artiklar/2015/Urbanisering--fran-land-till-stad/>
- Statistics Sweden. (2019a). *Kommuner i siffror*. Retrieved 2019-07-05, from <https://www.scb.se/hitta-statistik/sverige-i-siffror/kommuner-i-siffror>
- Statistics Sweden. (2019b). *Studenter och examina i högskoleutbildning på grundnivå och avancerad nivå*. Retrieved 2019-07-05, from <https://www.scb.se/hitta-0Astatistik/statistik-efter-amne/utbildning-och-forskning/hogskolevasende/studenter-och-examina-i-hogskoleutbildning-pa-grundniva-och-0Aavancerad-niva>
- Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In P. Denes & E. Davis (Eds.), *Human Communication: A Unified View* (pp. 51–66). New York: McGraw-Hill.
- Stevens, K. N. (2000). *Acoustic Phonetics*. London: MIT Press.
- Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel

- articulation. *The Journal of the Acoustical Society of America*, 27(3), 484–493.
- Stone, M. (2009). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 455–501.
- Stone, M. (2010). Laboratory techniques for investigating speech articulation. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 9–38). Chichester: Wiley-Blackwell.
- Stone, M., & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*, 99(6), 3728–3737.
- Strange, W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2081–2087.
- Strycharczuk, P., & Scobbie, J. M. (2017). Whence the fuzziness? Morphological effects in interacting sound changes in Southern British English. *Laboratory Phonology*, 8(1), 1–21.
- Stuart-Smith, J., & Lawson, E. (2018). *Script for extracting formant and bandwidth info from multiple intervals and sound files and show where measurements were taken*.
- Sundberg, J., & Nordström, P. (1976). Raised and lowered larynx: The effect on vowel formant frequencies. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 17(2-3), 35–39.
- Sundgren, E. (2001). Men and women in language change: A Swedish case study. *Nordic Journal of Women's Studies*, 9(2), 113–123.
- Svahn, M., & Nilsson, J. (2014). *Dialektutjämning i Västsverige*. Gothenburg: Institutet för språk och folkminnen.
- Sweet, H. (1877). *A Handbook of Phonetics, Including a Popular Exposition of the Principles of Spelling Reform*. Oxford: Clarendon Press.
- Syrdal, A. K., & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, 79, 1086–1100.
- Trautmüller, H., & Öhrström, N. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, 35, 244–258.
- Trost, J. E. (1981). Articulatory additions to the classical description of the speech of persons with cleft palate. *Cleft Palate Journal*, 18(3), 193–203.
- Trudgill, P. (1974a). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, 2, 215–246.
- Trudgill, P. (1974b). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, P. (1986). *Dialects in Contact*. Oxford: Blackwell.
- Turk, A. E., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research*, 3, 1–28.

- Turton, D. (2014). Some /l/s are darker than others: Accounting for variation in English /l/ with ultrasound tongue imaging. *University of Pennsylvania Working Papers in Linguistics*, 20(2), 187–198.
- University of Duisburg-Essen. (2019). *Studying the history of English: Transmission and propagation of change* [Teaching materials]. Retrieved 2019-11-28, from [https://www.uni-due.de/SHE/SHE\\_Transmission.htm](https://www.uni-due.de/SHE/SHE_Transmission.htm)
- van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs, version 2.3*.
- van Santen, J. (1992). Contextual effects on vowel duration. *Speech Communication*, 11(6), 513–546.
- Vilhelmson, B. (1994). Att undersöka geografisk rörlighet some en aspekt på relationen livsstil - miljö. *Nordisk samhällsgeografisk tidskrift*, 16, 7–38.
- von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Vieweg.
- Watt, D., & Fabricius, A. (2002). Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1 ~ F2 plane. *Leeds Working Papers in Linguistics and Phonetics*, 9, 159–73.
- Watt, D., Fabricius, A., & Kendall, T. (2011). More on vowels: Plotting and normalisation. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A Student's Guide* (pp. 107–118). New York, NY, USA: Routledge.
- Wei, T., & Simko, V. (2017). *corrplot: Visualization of a correlation matrix, version 0.84*. Retrieved 2019-11-28, from <https://github.com/taiyun/corrplot>
- Wells, J. C. (1994). The Cockneyfication of RP? In G. Melchers & N. L. Johannesson (Eds.), *Nonstandard Varieties of Language* (pp. 198–205). Stockholm: Almqvist & Wiksell.
- Wenner, L. (2010). *När lögnare blir lugnare: En sociofonetisk studie av sammanfallet mellan kort ö och kort u i uppländskan* (Doctoral thesis). Uppsala University.
- Wessén, E. (1945). *Våra folkmål*. Stockholm: Fritzes.
- West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender and Society*, 1(2), 125–151.
- Westerberg, F. (2013). *I'm from Sweden: An Articulatory Study of Swedish /i/ using Ultrasound Tongue Imaging* (Undergraduate dissertation). University of Glasgow.
- Westerberg, F. (2016). *An Auditory, Acoustic, Articulatory and Sociophonetic Study of Swedish Viby-i* (Masters dissertation). University of Glasgow.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wikimedia Commons. (2019). *Image: Sverigekarta Landskap*. Retrieved 2019-11-15, from <https://commons.wikimedia.org/wiki/File:Sverigekarta-Landskap.svg>
- Wood, S. (1982). *X-ray and Model Studies of Vowel Articulation*. Lund: Lund University.
- Wood, S. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Boca Raton:

CRC Press.

- Wood, S. (2019). *Interpreting vowel articulation from formant frequencies*. Retrieved 2019-11-25, from <https://swphonetics.com/methods/vowel-articulation-from-formants/>
- Wrench, A. A., & Balch, P. (2015). Towards a 3D tongue model for parameterising ultrasound data. In *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). Glasgow.
- Wright, J. T., & Riordan, C. J. (1980). Articulatory compensation in the production of vowels with bite blocks. *The Journal of the Acoustical Society of America (Supplement 1)*, 68, S31.
- Zharkova, N. (2018). An ultrasound study of the development of lingual coarticulation during childhood. *Phonetica*, 75(3), 245–271.
- Zhou, X., Espy-Wilson, C. Y., Boyce, S. E., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *The Journal of the Acoustical Society of America*, 123(6), 4466–4481.
- Zimman, L. (2017). Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics*, 55(5), 993–1019.

# Appendix A

## Questionnaire (English)

### QUESTIONNAIRE

*Please answer all questions. If you don't wish to answer a particular question, draw a line through it.*

1. Year of birth:

2. Gender:

- ☐ Male
- ☐ Female
- ☐ Non-binary

3. City of birth:

4. Where in Sweden have you lived since you were born? (Continue overleaf if you need more space.)

City:

Part of city:

How old were you when you moved there?

How long did you live/have you lived there?

5. Where outside of Sweden have you lived since you were born?

- ☐ Never lived abroad

Countr(ies):

How old were you when you moved there?

How long did you live there?



6. Do you regularly visit any other cities in Sweden?

☐ No

Which one(s)?

How often?

For how long?

For what purpose? (e.g. work, friends, family)

7. Do you travel around [city] a lot, or do you tend to stay in the same area?

☐ Very mobile

☐ Quite mobile

☐ A little mobile

☐ Barely mobile

8. Which areas do you visit the most?

For what purpose? (e.g. work, friends, coffee, going out)

9. Are you considering moving somewhere else? (e.g. another country, city, part of the city)

☐ No

☐ Yes, to (fill in):

Because (fill in):

10. Where do you prefer to live?

☐ Centrally, in the middle of everything

☐ Outside the city centre, but close to everything

☐ In a suburb or smaller city

☐ Further away, e.g. in the countryside

11. How often do you usually travel abroad?

For how long?

12. Do most of your friends come from [city]?

☐ Yes

☐ No, most come from nearby cities

☐ No, most come from other parts of Sweden

☐ No, most come from other countries

13. How happy are you to be from [city], as opposed to any other place in Sweden? (If you're not from there, how happy are you to live there now?)

- ☐ Very happy
- ☐ Happy
- ☐ Not very happy
- ☐ Unhappy
- ☐ Don't care

14. Are you proud of [city] as a city?

- ☐ Yes, definitely
- ☐ Yes, a bit
- ☐ No, not really
- ☐ No, not at all
- ☐ Don't care

15. Do you feel at home in [city]?

- ☐ Yes, definitely
- ☐ Yes, a bit
- ☐ No, not really
- ☐ No, not at all
- ☐ Don't care

16. Do you like how people speak in [city]?

- ☐ Yes, definitely
- ☐ Yes, a bit
- ☐ No, not really
- ☐ No, not at all
- ☐ Don't care

17. What do you like/dislike the most about how people speak in [city]?

18. Are there any other Swedish dialects that you particularly like/dislike?

Why?

19. Do you think it's important to speak a certain way?

- ☐ Yes, definitely
- ☐ Yes, a bit
- ☐ No, not really
- ☐ No, not at all
- ☐ Don't care

Explain:

20. Do you get annoyed when people use "incorrect" language?

- ☐ Yes, definitely
- ☐ Yes, a bit
- ☐ No, not really
- ☐ No, not at all
- ☐ Don't care

21. Is Swedish your mother tongue?

- ☐ Yes, only Swedish
- ☐ Yes, Swedish and (fill in):
- ☐ No, my mother tongue is (fill in):

22. Where are your parents/guardians from? (Please only include those who were present when you grew up.)

23. Were there any other languages spoken in the home when you grew up?

- ☐ No
- ☐ Yes (fill in):

24. Which languages apart from Swedish do you speak now?

Language(s):

Level (fluent, good, or beginner):

25. Have you ever had any issues with your speech, e.g. caused by a speech impediment, or physical or neurological damage?

- ☐ No
- ☐ Yes (please describe):

If yes, when did you have this issue? For how long?

26. Have you ever been referred to a speech coach or speech therapist?

- ☐ No
- ☐ Yes, to a speech coach for (fill in):
- ☐ Yes, to a speech therapist for (fill in):

27. What level of education do you have? (Tick your highest level, finished or ongoing. For adult education, tick the equivalent level.)

- ☐ Mandatory school (years 1-9)
- ☐ Upper secondary school (years 10-12)
- ☐ Additional upper secondary education
- ☐ University
- ☐ Masters degree
- ☐ Doctorate degree
- ☐ Higher

28. If you attended upper secondary school, which programme did you study?

29. If you have a university degree, in what subject?

30. Are you a student at the moment?

- ☐ No
- ☐ Yes, in (fill in subject):

Level:

- ☐ Upper secondary school
- ☐ Additional upper secondary education
- ☐ University
- ☐ Masters degree
- ☐ Doctorate degree
- ☐ Higher

31. Do you have a profession? (This refers to a long-term career path, as opposed to a single job.)

- ☐ No
- ☐ Yes, (fill in):

If yes, how many years of work experience do you have?

32. Are you working at the moment?

- ☐ No
- ☐ Yes, full-time as (fill in):
- ☐ Yes, part-time as (fill in):

33. What do/did your parents/guardians do for a living?

34. What level of education do your parents/guardians have? (Tick for one, two, or more guardians)

- ☐ ☐ Mandatory school (years 1-9)
- ☐ ☐ Upper secondary school (years 10-12)
- ☐ ☐ Additional upper secondary education
- ☐ ☐ University
- ☐ ☐ Masters degree
- ☐ ☐ Doctorate degree
- ☐ ☐ Higher
- ☐ ☐ Don't know

35. What interests do you have in your spare time? (You can tick several options.)

- ☐ Cooking/baking
- ☐ Arts/crafts
- ☐ DIY/gardening
- ☐ Technology/vehicles
- ☐ Computer games/video games/role playing games
- ☐ Reading/writing/blogging/studying
- ☐ Fashion/shopping/design
- ☐ Hiking/excursions
- ☐ Travelling
- ☐ Sports/dancing
- ☐ Music/acting/film (making)
- ☐ Watching films/TV
- ☐ Theatre, museums, etc.
- ☐ Going out for coffee/food

- ☐ Going out dancing/clubbing
- ☐ Going to concerts/gigs
- ☐ Family activities
- ☐ Politics/volunteering

# Appendix B

## Questionnaire (Swedish)

### FRÅGEFORMULÄR

*Fyll i alla fält. Om det är någon fråga du inte vill svara på, dra ett streck genom fältet.*

1. Födelseår:

2. Kön:

- ☐ Man
- ☐ Kvinna
- ☐ Icke-binär

3. Födelseort:

4. Var i Sverige har du bott sedan du föddes? (Fortsätt på baksidan om du behöver mer plats.)

Ort:

Stadsdel:

Hur gammal var du när du flyttade dit?

Hur länge bodde/har du bott där?

5. Var utomlands har du bott sedan du föddes?

- ☐ Har inte bott utomlands

Land:

Hur gammal var du när du flyttade dit?

Hur länge bodde du där?

6. Besöker du regelbundet någon annan ort i Sverige?

- ☐ Nej

Vilken ort?

Hur ofta?

Hur länge?

I vilket syfte? (t.ex. jobb, vänner, familj)

7. Rör du dig mycket omkring [stad], eller stannar du oftast i samma område?

- ☐ Mycket rörlig
- ☐ Ganska rörlig
- ☐ Inte särskilt rörlig
- ☐ Knappt rörlig alls

8. Vilket/vilka områden besöker du mest?

I vilket syfte? (t.ex. jobb, vänner, fika, uteställen)

9. Funderar du på att flytta någon annanstans? (t.ex. annat land, stad, stadsdel)

- ☐ Nej
- ☐ Ja, till (fyll i):

För att (fyll i):

10. Var bor du helst?

- ☐ Centralt, mitt i smeten
- ☐ Utanför centrum men nära till allt
- ☐ I förort eller mindre stad
- ☐ Längre bort, t.ex. på landet

11. Hur ofta brukar du resa utomlands?

Hur länge?

12. Kommer de flesta av dina vänner från [stad]?

- ☐ Ja
- ☐ Nej, de flesta kommer från närliggande orter
- ☐ Nej, de flesta kommer från andra delar av Sverige
- ☐ Nej, de flesta kommer från andra länder



13. Hur nöjd är du med att komma just från [stad], snarare än någon annanstans i Sverige?  
(Om du inte kommer därifrån, hur nöjd är du med att bo där nu?)
- ☐ Jättenöjd
  - ☐ Nöjd
  - ☐ Inte så nöjd
  - ☐ Missnöjd
  - ☐ Bryr mig inte
14. Är du stolt över [stad] som stad?
- ☐ Ja, absolut
  - ☐ Ja, lite
  - ☐ Nej, inte direkt
  - ☐ Nej, inte alls
  - ☐ Bryr mig inte
15. Känner du dig hemma i [stad]?
- ☐ Ja, absolut
  - ☐ Ja, lite
  - ☐ Nej, inte direkt
  - ☐ Nej, inte alls
  - ☐ Bryr mig inte
16. Tycker du om hur folk pratar i [stad]?
- ☐ Ja, absolut
  - ☐ Ja, lite
  - ☐ Nej, inte direkt
  - ☐ Nej, inte alls
  - ☐ Bryr mig inte
17. Vad gillar/ogillar du mest med hur folk pratar i [stad]?
18. Finns det några andra svenska dialekter du tycker särskilt mycket/lite om?  
Varför?

19. Tycker du att det är viktigt att prata på ett visst sätt?

- ☐ Ja, absolut
- ☐ Ja, lite
- ☐ Nej, inte direkt
- ☐ Nej, inte alls
- ☐ Bryr mig inte

Utveckla:

20. Blir du irriterad på folk som använder "fel" språk?

- ☐ Ja, absolut
- ☐ Ja, lite
- ☐ Nej, inte direkt
- ☐ Nej, inte alls
- ☐ Bryr mig inte

21. Är svenska ditt modersmål?

- ☐ Ja, bara svenska
- ☐ Ja, svenska och (fyll i):
- ☐ Nej, mitt modersmål är (fyll i):

22. Varifrån kommer dina föräldrar/vårdnadshavare? (Inkludera bara de som var med när du växte upp.)

23. Talades det några andra språk i hemmet när du växte upp?

- ☐ Ja (fyll i):
- ☐ Nej

24. Vilka språk förutom svenska talar du nu?

Språk:

Nivå (flytande, bra, eller nybörjare):

25. Har du någon gång haft problem med talet, t.ex. på grund av ett talfel, eller en fysisk eller neurologisk skada?

- ☐ Nej
- ☐ Ja (beskriv):

Om ja, när hade du det här problemet? Hur länge?

26. Har du någon gång gått hos talpedagog eller talterapeut?

- ☐ Nej
- ☐ Ja, talpedagog för (fyll i):
- ☐ Ja, talterapeut för (fyll i):

27. Vad är din utbildningsnivå? (Kryssa högsta nivå, färdig eller pågående. För komvux, kryssa motsvarande nivå.)

- ☐ Grundskola
- ☐ Gymnasie
- ☐ Folkhögskola
- ☐ Högskola/Universitet
- ☐ Master
- ☐ Doktorand
- ☐ Högre

28. Om du gick gymnasiet, vilket program?

29. Om du har universitetsutbildning, i vilket ämne?

30. Är du student just nu?

- ☐ Nej
- ☐ Ja, i (fyll i ämne):

Nivå:

- ☐ Gymnasie
- ☐ Folkhögskola
- ☐ Högskola/Universitet
- ☐ Master
- ☐ Doktorand
- ☐ Högre

31. Har du ett yrke? (Syftar på långsiktig karriärsbana snarare än ett enskilt jobb.)

- ☐ Nej
- ☐ Ja, (fyll i):

Om ja, hur många års arbetserfarenhet?

32. Jobbar du just nu?

- ☐ Nej
- ☐ Ja, heltid som (fyll i):
- ☐ Ja, deltid som (fyll i):

33. Vad jobbar/jobbede dina föräldrar/vårdnadshavare som?

34. Vad har dina föräldrar/vårdnadshavare för utbildningsnivå? (Kryssa för en, två, eller flera vårdnadshavare.)

- ☐ ☐ Grundskola
- ☐ ☐ Gymnasie
- ☐ ☐ Folkhögskola
- ☐ ☐ Högskola/Universitet
- ☐ ☐ Master
- ☐ ☐ Doktorand
- ☐ ☐ Högre
- ☐ ☐ Vet inte

35. Vad har du för intressen på fritiden? (Du kan kryssa flera.)

- ☐ Matlagning/bakning
- ☐ Konst/hantverk
- ☐ Bygge/trädgård
- ☐ Teknologi/fordon
- ☐ Datorspel/tv-spel/rollspel
- ☐ Läsa/skriva/blogga/studera
- ☐ Mode/shopping/design
- ☐ Friluftsliv/utflykter
- ☐ Längre resor
- ☐ Sport/dans
- ☐ Musik/skådespel/film (göra)
- ☐ Titta på film/tv
- ☐ Gå på teater, museum, osv.
- ☐ Gå ut och fika/äta

- ☐ Gå ut och festa/klubba
- ☐ Gå på konserter/spelningar
- ☐ Familjeaktiviteter
- ☐ Politik/voluntärarbete

# Appendix C

## Information and consent forms (English)



### Information for participants

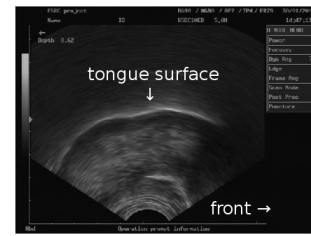
Hi! My name is Fabienne, and I'm a doctoral student in linguistics at the University of Glasgow and Queen Margaret University. In my research project, I investigate how the tongue and lips move when we speak, and I compare how people from different cities in Sweden use their tongue and lips in different ways. I use a technique called “ultrasound tongue imaging” (UTI), which is a safe and radiation-free method that uses sonar to show what is happening inside the mouth. I also film the lips moving (not the whole face), and make audio recordings of what is being said. Afterwards, I use a questionnaire to collect some information about you and your language background.

### How does it work?

UTI works in the same way as a medicinal ultrasound – using sonar. To see into your mouth, you hold an ultrasound sensor against the underside of your chin. The sensor sends out a soundwave that is good at travelling through soft tissue, but is reflected by air. The soundwave travels up through the chin until it reaches the surface of the tongue, where there is an air barrier. The signal then bounces back and is received by the sensor, which turns the echo into a picture.

There are a few things we have to do before we start recording:

1. Get you into a stabilisation headset. The headset holds the sensor in place, making sure that it doesn't move in relation to your head. You can still move your head normally. The headset is comfortable, but a little heavy, so just let me know if you want a break!



2. Adjust the picture. To make sure that the video is aligned with the screen, you will get to bite down on a small piece of plastic, which you press your tongue against. This creates a straight line, which is adjusted until it matches the computer screen. Don't panic – it's not like the dentist! You put the plastic plate into your mouth kind of like a spoon, and it only takes a few seconds.
3. Drink some water. When you swallow, the air in your mouth disappears, so that the signal can travel all the way up to the roof of your mouth. This gives me something to compare the tongue's position to.

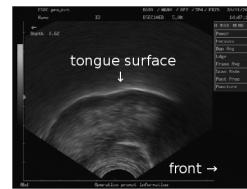
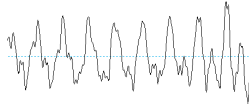
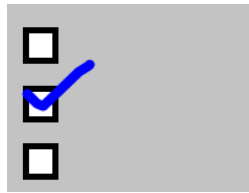
### **What do I have to do?**

You will be recorded either by yourself, or together with a friend. The recording itself consists of a number of sentences which you read from a computer screen when you hear a beep. This takes around 45 minutes. Afterwards it is optional to record a few short clips when either I or your friend asks you a few short interview questions. In total, the session takes approx. 1 hour per person. Afterwards, you have a chance to win cinema tickets or a gift card worth up to 150 SEK apiece.

### **What kind of materials are being collected?**

- Questionnaire: The answers are used to sort the recorded materials by different factors, e.g. age, gender, city.
- Audio recordings: The sound is investigated acoustically to compare e.g. how the sound-waves behave, and how the acoustic energy is distributed.
- Ultrasound recordings of the tongue: I compare different tongue shapes and tongue movements, and how they are used by different speakers.
- Video recordings of the lips: Give me extra information, since the tongue and lips often work together.

All materials are anonymous (you will be given a code name before we start), and your personal information will be treated confidentially. Only I and my supervisors have access to your information.



## How will the materials be used?

I will keep the materials for educational and academic purposes, e.g. research, teaching, conferences, publications, or databases. After the project is over, the materials may be stored with the UK Data Service (a public research database), University of Glasgow, and Queen Margaret University. General information from the questionnaire (e.g. age, gender, city, and different categories that you are part of) may be presented together with your recorded materials, since these factors can be relevant for how you speak. Your original answers will not be shown to anyone.

## Contact

For more information, visit [project website]. You can also contact me via email at [email address]. If you have any questions, or would like to talk to my supervisors, contact [first supervisor] [email address] or [second supervisor] [email address]. Thank you for your time!



## Consent to the use of data

### University of Glasgow, College of Arts Research Ethics Committee

I understand that Fabienne Westerberg is collecting data in the form of questionnaire answers, audio recordings, ultrasound tongue imaging video, and lip video for use in an academic research project at the University of Glasgow and Queen Margaret University Edinburgh.

The project investigates how the tongue and lips are used to create speech sounds, and compares the different articulation strategies used by speakers from different Swedish cities. Further information about the project can be found on the attached information sheet, or online at [project website].

I give my consent to the use of data for this purpose on the understanding that:

- Participation in the study is voluntary, and I can withdraw at any time before or during the recordings without stating a reason.
- The study is in no way related to my academic performance.
- All names and personal information will be anonymised for all types of use.
- All personal and sensitive information will be treated as confidential and kept in secure storage at all times.
- The materials will be kept by the researcher for future educational or academic use, e.g. research, teaching, conferences, publications, or speech databases.
- At the end of the project, anonymised materials may be deposited with the UK Data Service (a public research database), University of Glasgow and Queen Margaret University.
- Participation in the study may not benefit me directly.

Signed by the participant: \_\_\_\_\_ Date: \_\_\_\_\_

Signed by the researcher: \_\_\_\_\_ Date: \_\_\_\_\_

Optional: I wish to be contacted again about this project or future projects (the researcher will keep my contact details).

Signed by the participant: \_\_\_\_\_ Date: \_\_\_\_\_

[researcher's contact details]

[supervisors' contact details]

# Appendix D

## Information and consent forms (Swedish)



### Information till dig som deltar

#### INFORMATION TILL DIG SOM DELTAR

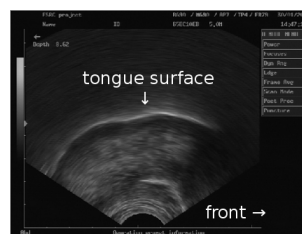
Hej! Jag heter Fabienne och är doktorand i lingvistik på University of Glasgow och Queen Margaret University. I mitt forskningsprojekt undersöker jag hur tungan och läpparna rör sig när vi pratar, och jämför hur folk i olika svenska städer använder tungan och läpparna på olika sätt. Jag använder en teknik som kallas “ultrasound tongue imaging” (UTI), en säker och strålningsfri metod som använder sonar för att visa vad som händer inuti munnen. Jag filmar även läpparnas rörelser (ej hela ansiktet) och gör ljudupptagningar av det som sägs. Efteråt samlar jag in lite information om dig och din språkbakgrund i ett frågeformulär.

### Hur funkar det?

UTI fungerar på samma sätt som ett medicinskt ultraljud – med hjälp av sonar. För att se in i munnen håller du en ultraljudssensor mot undersidan av hakan. Sensorn skickar ut en ljudvåg som är bra på att färdas genom mjuk vävnad men reflekteras när den möter luft. Ljudsignalen färdas upp genom hakan tills den når tungytan, där det finns en luftbarriär. Signalen studsar då tillbaka och tas återigen upp av sensorn, som förvandlar ekot till en bild.

Det finns några saker vi måste göra innan vi börjar spela in:

1. Sätta på dig ett stabiliseringsheadset. Headsetet håller sensorn på plats och ser till att den inte flyttar på sig i relation till huvudet. Du kan fortfarande röra huvudet normalt. Headsetet är bekvämt men lite tungt, säg bara till om du vill ta en paus!



2. Justera bilden. För att se till att videon är i linje med skärmen får du bita i en liten plastbit som du trycker tungan emot. Det skapar en rak linje som justeras tills den är i linje med datorskärmen. Ingen panik – det är inte som hos tandläkaren! Plastbiten stoppar du i munnen ungefär som en sked, och det tar bara ett par sekunder.
3. Dricka lite vatten. När du sväljer försvinner luften i munnen, så att signalen kan färdas hela vägen upp till gommen. Det ger mig något att jämföra tungans position med.

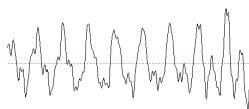
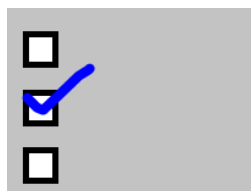
### Vad ska jag göra?

Du spelas in antingen själv eller tillsammans med en kompis. Själva inspelningen består av ett antal meningar som du läser upp från en datorskärm när du hör ett pip. Detta tar ca 45 minuter. Efteråt är det valfritt om du vill spela in ett par korta klipp när jag eller din kompis ställer några intervjufrågor till dig. Sammanlagt tar sessionen ca 1 timme per person. Efteråt har du chans att vinna biobiljetter eller presentkort med värde upp till 150 kr styck.

### Vad för material samlas in?

- Frågeformulär: Svaren används för att sortera det inspelade materialet efter olika faktorer, t.ex. ålder, kön, ort.
- Ljudinspelningar: Ljudet undersöks akustiskt för att jämföra t.ex. hur ljudvågorna beter sig och hur den akustiska energin är fördelad.
- Ultraljudsinspelningar av tungan: Jag jämför olika tungformer och tungrörelser och hur de används av olika personer.
- Videoinspelningar av läpparna: Ger mig extra information eftersom läpparna och tungan ofta jobbar tillsammans.

Allt material är anonymt (du får ett kodnamn innan vi börjar), och din personliga information behandlas konfidentiellt. Endast jag och mina handledare har tillgång till din information.



## Hur används materialet?

Jag behåller materialet för utbildnings- och akademiska syften, t.ex. forskning, undervisning, konferenser, publikationer eller databaser. Efter projektet kan materialet komma att lagras hos UK Data Service (en offentlig forskningsdatabas), University of Glasgow och Queen Margaret University. Allmän information från frågeformuläret (t.ex. ålder, kön, ort, och diverse kategorier som du passar in i) kan komma att presenteras tillsammans med ditt inspelade material, eftersom dessa faktorer kan vara relevanta för hur du pratar. Dina direkta svar kommer inte att visas för någon.

## Kontakt

För mer information, gå in på [projekthemsida]. Du kan även kontakta mig via mejl på [mejladress]. Om du har några frågor eller vill prata med mina handledare, kontakta [första handledare] [mejladress] eller [andra handledare] [mejladress]. Tack för din tid!

## Samtycke till användning av material

### University of Glasgow, College of Arts Research Ethics Committee

Jag förstår att Fabienne Westerberg samlar material i form av frågeformulär, ljudinspelningar, ultraljudsvideo av tungan, och video av läpparna för användning i ett akademiskt forskningsprojekt på University of Glasgow och Queen Margaret University Edinburgh.

Projektet undersöker hur tungan och läpparna används för att skapa språkljud, och jämför hur talare från olika svenska städer använder olika artikulationsstrategier. För mer information om projektet, se bifogat informationsblad eller projektets hemsida: [projekthemsida].

Jag samtycker till att materialet används för detta syfte, under förståelsen att:

- Mitt deltagande är frivilligt, och jag kan dra mig ur när som helst innan eller under inspelningarna utan att säga varför.
- Studien har ingenting att göra med mitt skolarbete eller mina betyg.
- Alla namn och all personlig information anonymiseras för alla sorters användning.
- All personlig och känslig information behandlas konfidentiellt och förvaras alltid på ett säkert ställe.
- Materialet behålls av forskaren för framtida användning i utbildnings- eller akademiska syften, t.ex. forskning, undervisning, konferenser, publikationer, eller databaser.
- Efter projektets slut kan anonymiserat material komma att lagras hos UK Data Service (en offentlig forskningsdatabas), University of Glasgow och Queen Margaret University.
- Jag kanske inte vinner någonting på att delta i studien.

Deltagarens underskrift: \_\_\_\_\_ Datum: \_\_\_\_\_

Forskarens underskrift: \_\_\_\_\_ Datum: \_\_\_\_\_

Valfritt: Jag vill bli kontaktad igen om det här projektet eller framtida projekt (forskaren behåller mina kontaktuppgifter).

Deltagarens underskrift: \_\_\_\_\_ Datum: \_\_\_\_\_

[forskarens kontaktuppgifter]

[handledarnas kontaktuppgifter]

Appendix E

Historical formant values for Swedish vowels

Study	Variety	N	Gender	Vowel	Mean F1	Mean F2	Mean F3	Mean F4
(Eklund & Traunmüller, 1997)	Stockholm	5	Male	i	291	2107	3135	NA
				y	285	1988	2745	NA
				ʉ	328	1679	2365	NA
				e	376	2152	2720	NA
(Engstrand, 2004)	Stockholm	5	Female	i	351	2455	3500	NA
				y	353	2319	3128	NA
				ʉ	386	1904	2686	NA
				e	438	2500	3203	NA
	Standard?	NA	Female	i	430	2700	3875	NA
				y	430	2475	3375	NA
				ʉ	NA	NA	NA	NA
				e	NA	NA	NA	NA
(Engwall & Badin, 1999)	Stockholm	1	Male	i	320	1520	2900	3460

(Ericsson, 2005)				y	360	1560	2880	3500
				ʏ	400	1720	2400	3260
				e	380	2200	2620	3520
	Standard?	1	Male	i	220	2180	3100	NA
				y	220	2040	2510	NA
				ʏ	280	1590	2310	NA
				e	330	2170	2480	NA
	Standard?	1	Female	i	300	2510	3370	NA
				y	310	2510	3080	NA
				ʏ	310	2060	2630	NA
				e	370	2490	3320	NA
(Fant, 1959)				i	255	2065	2960	3400
				y	260	1930	2420	3300
				ʏ	285	1635	2140	3310
				e	335	2050	2510	3400
	Stockholm	7	Male	i	278	2520	3460	NA
				y	270	2480	2930	NA
				ʏ	300	1910	2600	NA
				e	365	2540	2950	NA
	Stockholm	7	Female	i	230	2000	3000	NA
				y	245	1875	2075	NA
				ʏ	265	1480	2060	NA
				e	325	1900	2450	NA
(Fant, 1972)	Standard?	NA	Male	i	345	2060	3100	NA
				y	295	2000	2650	NA
	Standard?	NA	Female	i				
				y				

				ʈ		305	1770	2380	NA
				e		395	2200	2850	NA
(Fant et al., 1969)	Stockholm	24	Male	i		255	2190	3150	3730
				y		260	2060	2675	3310
				ʈ		285	1640	2250	3250
				e		345	2250	2850	3540
(Kuronen, 2000)	Nyköping	4	Male	i		275	2363	3304	3807
				y		285	2258	2994	3593
				ʈ		328	1733	2453	3310
				e		385	2194	2920	3310
(Malmberg, 1956)	Stockholm	NA	Male	i		300	2400	3500	NA
				y		300	2200	2600	NA
				ʈ		300	1800	NA	NA
				e		NA	NA	NA	NA
(Malmberg, 1971)	Standard?	NA	Male	i		280	2400	3400	NA
				y		280	2200	2600	NA
				ʈ		300	1800	2800	NA
				e		350	2200	3000	NA
(Nord, 1986)	Stockholm	4	Male	i		300	2087	NA	NA
				y		NA	NA	NA	NA
				ʈ		401.5	1189.5	NA	NA
				e		528	1620.5	NA	NA
(Stålhammar et al., 1973)	Stockholm	3	Male	i		285	2000	3015	3610
				y		310	1910	2675	3490
				ʈ		335	1550	2300	3325



	e		315	2285	2865	3465
(Traunmüller & Öhrström, 2007)						
	i	Male	327.5	1962.5	2625	3415
	y		307.5	1862.5	2692.5	3250
	ʉ		NA	NA	NA	NA
	e		327.5	2225	2532.5	3377.5
(Björsten & Engstrand, 1999)*						
	i	Female	450	2100	2845	4082.5
	y		415	2087.5	3220	3815
	ʉ		NA	NA	NA	NA
	e		420	2487.5	2717.5	2940
(Westerberg, 2013)*						
	i	Male	350	1590	2860	3590
	y		310	1560	2720	3330
	ʉ		320	1490	2320	3330
	e		340	2220	2730	3670
(Westerberg, 2016)*						
	i	Female	424	1965	3052	4177
	y		NA	NA	NA	NA
	ʉ		NA	NA	NA	NA
	e		491	2253	2918	4034
(Westerberg, 2016)*						
	i	Female	362	2110	3185	4144
	y		NA	NA	NA	NA
	ʉ		NA	NA	NA	NA
	e		395	2331	2933	3743
(Westerberg, 2016)*						
	i	Male	391	1754	2911	NA
	y		NA	NA	NA	NA
	ʉ		393	1997	2542	NA
	e		393	2406	2828	NA

Gothenburg	1	Female	i	337	1858	2854	NA
			y	NA	NA	NA	NA
			ʉ	331	1690	2236	NA
			e	344	2224	2594	NA
Varberg	2	Female	i	409	2120	3307	NA
			y	NA	NA	NA	NA
			ʉ	429	1919	2537	NA
			e	417	2492	3179	NA
Jönköping	1	Male	i	317	1865	2741	NA
			y	NA	NA	NA	NA
			ʉ	340	1613	2328	NA
			e	338	2188	2783	NA
Stockholm	1	Male	i	331	1696	2716	NA
			y	NA	NA	NA	NA
			ʉ	361	1508	2336	NA
			e	314	2071	2784	NA
Stockholm	6	Female	i	430	2085	3266	NA
			y	NA	NA	NA	NA
			ʉ	449	1933	2829	NA
			e	412	2584	3053	NA
Katrineholm	1	Female	i	415	2455	3628	NA
			y	NA	NA	NA	NA
			ʉ	510	2095	2845	NA
			e	405	2756	3430	NA

# Appendix F

## Individual formant values

Speaker	Vowel	Mean F1	Mean F2	Mean F3	Mean F4
GF1	i	427	1756	3229	4060
	y	428	1772	3009	3762
	ɯ	416	1937	2383	3765
	e	407	2323	2884	3961
	ø	585	1382	2440	3847
	ɛ	578	1689	2597	3913
	ɑ	603	1013	2826	3888
	o	451	785	3027	3856
	u	441	915	2903	3879
GF2	i	460	1825	3056	4294
	y	446	1715	2976	4027
	ɯ	425	1999	2692	3851
	e	410	2415	2892	4094
	ø	588	1508	2593	3828
	ɛ	618	1567	2666	3956
	ɑ	616	1147	2759	3809
	o	482	799	2892	4045
	u	477	973	2769	3982
GF3	i	392	1899	3137	4507
	y	380	1849	3077	4379
	ɯ	385	1934	2759	4091
	e	365	2488	2981	4206
	ø	524	1741	2646	4065
	ɛ	596	1861	2759	4426
	ɑ	631	1091	2941	3799
	o	406	776	3065	4000

	u	420	994	2863	4018
GF4	i	393	1783	2783	4159
	y	369	1833	2797	3998
	ʉ	387	1855	2797	4136
	e	388	2279	2733	4127
	ø	489	1661	2529	3946
	ɛ	518	1680	2576	3963
	ɑ	569	1084	2676	3816
	o	444	840	2847	3899
	u	426	911	2933	3874
GF5	i	399	1932	3429	5095
	y	403	1879	4020	4797
	ʉ	399	2197	2999	4185
	e	398	2667	3028	4624
	ø	429	1775	2712	4158
	ɛ	577	1993	3015	4589
	ɑ	552	1023	2925	4300
	o	409	775	2920	4247
	u	407	827	2979	4392
GF6	i	369	1997	3056	4375
	y	355	1918	2871	3854
	ʉ	316	2054	2859	3926
	e	372	2268	2869	4058
	ø	397	1863	2488	3820
	ɛ	438	1860	2691	3923
	ɑ	437	1011	2747	3780
	o	368	835	2769	3909
	u	346	892	2768	3939
GM1	i	321	1683	2659	3669
	y	309	1650	2564	3231
	ʉ	327	1696	2261	3118
	e	332	2023	2687	3359
	ø	488	1188	2268	3296
	ɛ	523	1572	2369	3442
	ɑ	545	809	2693	3495
	o	372	625	2789	3336
	u	344	755	2582	3327
GM2	i	365	1841	2836	3870

	y	345	1769	2710	3530
	ʏ	392	1718	2388	3336
	e	400	2064	2433	3560
	ø	483	1397	2318	3375
	ɛ	506	1529	2476	3522
	ɑ	568	1010	2271	3315
	o	418	834	2505	3465
	u	391	901	2430	3462
GM3	i	396	1592	2726	3439
	y	377	1576	2677	3506
	ʏ	363	1848	2412	3441
	e	377	2045	2589	3486
	ø	467	1301	2291	3483
	ɛ	499	1641	2411	3515
	ɑ	514	882	2557	3356
	o	414	767	2535	3239
	u	401	831	2499	3235
GM4	i	380	1562	2700	3791
	y	364	1539	2630	3520
	ʏ	373	1702	2252	3309
	e	351	2007	2496	3500
	ø	500	1380	2103	3315
	ɛ	510	1456	2156	3382
	ɑ	543	943	2442	3364
	o	405	792	2576	3308
	u	386	823	2452	3278
GM5	i	319	1793	2686	3657
	y	313	1713	2225	3161
	ʏ	329	1538	2091	3160
	e	304	1925	2320	3383
	ø	340	1430	2024	3140
	ɛ	422	1608	2155	3235
	ɑ	424	844	2601	3130
	o	361	706	2583	3243
	u	356	769	2524	3053
GM6	i	313	1682	2946	3716
	y	301	1621	2748	3283
	ʏ	309	1718	2277	3206

	e	311	2055	2569	3531
	ø	318	1636	2305	3268
	ɛ	380	1762	2468	3409
	ɑ	434	820	2615	3264
	o	381	702	2697	3441
	u	375	854	2652	3465
SF1	i	431	1892	3143	3997
	y	415	1904	3190	4044
	ʉ	459	1851	2845	3887
	e	392	2552	3027	3928
	ø	681	1519	2720	3732
	ɛ	696	1658	2820	3817
	ɑ	689	1091	2997	3788
	o	476	802	3034	3795
	u	470	1011	2904	3785
SF2	i	397	1965	3542	4447
	y	403	1912	3364	4075
	ʉ	401	1802	2899	3899
	e	410	2358	3054	4127
	ø	583	1556	2767	3750
	ɛ	674	1633	3000	3950
	ɑ	650	1108	3043	3806
	o	426	801	2943	4253
	u	437	837	2916	4277
SF3	i	373	2115	3258	4463
	y	385	2023	3164	3879
	ʉ	400	1969	2871	3845
	e	383	2621	3068	4199
	ø	523	1650	2662	3885
	ɛ	607	1799	2950	4226
	ɑ	584	1076	2747	3794
	o	418	771	3159	4017
	u	426	900	3070	3987
SF4	i	368	1975	3509	4738
	y	381	1839	3246	4029
	ʉ	413	1668	2751	4039
	e	407	2517	3094	4384
	ø	562	1459	2757	4074

	ɛ	582	1658	3020	4291
	ɑ	557	1067	2933	3996
	o	432	851	2870	4171
	u	413	988	2815	4170
SF5	i	406	2076	3430	4699
	y	393	1970	3162	4103
	ʉ	391	1920	2923	4011
	e	399	2587	3152	4259
	ø	417	2012	2877	4094
	ɛ	658	1913	2993	4336
	ɑ	615	1112	3038	4014
	o	435	830	3358	3981
	u	424	927	3256	3851
SF6	i	369	2050	3048	4466
	y	378	2032	2988	3895
	ʉ	408	1866	2864	4075
	e	403	2140	2923	4251
	ø	511	1723	2501	4408
	ɛ	583	1632	2429	4522
	ɑ	618	1047	2272	4485
	o	440	795	2888	4268
	u	442	945	2741	3874
SM1	i	333	1660	2497	3658
	y	326	1608	2398	3432
	ʉ	362	1537	2164	3089
	e	349	1917	2419	3307
	ø	527	1285	2162	3236
	ɛ	563	1311	2252	3326
	ɑ	544	951	2320	3148
	o	432	798	2361	3045
	u	410	862	2453	3307
SM2	i	296	1670	2712	3713
	y	294	1658	2582	3495
	ʉ	339	1581	2319	3346
	e	321	1910	2540	3490
	ø	473	1311	2479	3413
	ɛ	512	1448	2517	3501
	ɑ	542	969	2652	3258

	o	386	760	2742	3336
	u	344	838	2387	3133
SM3	i	324	1644	2722	3586
	y	329	1610	2604	3182
	ʉ	414	1348	2362	3144
	e	347	1784	2337	3348
	ø	474	1234	2193	3188
	ɛ	527	1266	2360	3402
	ɑ	525	919	2437	3312
	o	389	792	2481	3230
	u	388	873	2373	3162
SM4	i	359	1536	2425	3666
	y	369	1495	2336	3486
	ʉ	396	1453	2091	3286
	e	341	1863	2348	3361
	ø	458	1265	1993	3212
	ɛ	521	1359	2089	3472
	ɑ	527	993	1943	3174
	o	398	802	2263	3324
	u	412	975	2110	3231
SM5	i	356	1753	2691	3732
	y	351	1679	2341	3156
	ʉ	389	1535	2292	3253
	e	375	1877	2383	3403
	ø	439	1426	2261	3235
	ɛ	539	1393	2287	3318
	ɑ	526	892	2466	3184
	o	409	753	2556	3211
	u	405	823	2462	3236
SM6	i	324	1758	2819	3653
	y	342	1674	2583	3395
	ʉ	333	1701	2480	3394
	e	305	2224	2759	3509
	ø	403	1567	2403	3341
	ɛ	485	1557	2583	3519
	ɑ	507	881	2691	3435
	o	379	703	2755	3730
	u	362	820	2738	3597



UF1	i	408	1896	3176	4320
	y	388	1857	2772	3814
	ʉ	399	1602	2660	3921
	e	412	1929	2689	4000
	ø	553	1437	2568	3902
	ɛ	601	1499	2622	3977
	ɑ	566	1121	2627	3956
	o	432	857	2807	4129
	u	387	1028	2709	4118
UF2	i	413	1732	3423	4371
	y	392	1737	3191	3872
	ʉ	423	1693	2699	3862
	e	414	2316	2876	4139
	ø	590	1535	2547	3966
	ɛ	664	1505	2714	4105
	ɑ	610	1079	2746	3911
	o	456	788	3003	4024
	u	429	934	2948	4050
UF3	i	442	2081	3165	4418
	y	433	1978	2923	3839
	ʉ	473	1774	2680	3861
	e	438	2628	3076	4125
	ø	573	1723	2637	3911
	ɛ	748	1641	2779	4059
	ɑ	688	1095	2813	3778
	o	486	833	2954	3912
	u	474	846	2940	3934
UF4	i	401	1952	3218	4295
	y	414	1938	2927	3867
	ʉ	402	1699	2735	3911
	e	425	2106	2861	4038
	ø	455	1759	2619	3954
	ɛ	593	1535	2705	4168
	ɑ	594	982	2759	4006
	o	473	866	2898	4020
	u	438	964	2864	3993
UF5	i	363	2022	3034	4213
	y	361	1847	2800	3734

	ʉ	376	1724	2622	3804
	e	374	2258	2794	3766
	ø	394	1601	2394	3691
	ɛ	481	1604	2546	3850
	ɑ	508	943	2756	3718
	o	418	835	2985	3800
	u	412	933	2932	3803
UF6	i	356	2059	3118	4038
	y	360	2010	2916	3671
	ʉ	368	1847	2847	3815
	e	359	2179	3011	3963
	ø	384	1887	2823	3885
	ɛ	432	1747	2881	4151
	ɑ	465	974	3019	3711
	o	416	769	3088	3634
	u	390	936	3040	3835
UM1	i	312	1915	2761	3484
	y	314	1853	2600	3256
	ʉ	357	1498	2421	3247
	e	340	1913	2564	3139
	ø	506	1228	2407	3298
	ɛ	541	1289	2369	3375
	ɑ	508	888	2503	3218
	o	392	769	2498	3349
	u	381	881	2371	3281
UM2	i	318	1531	2888	3676
	y	301	1553	2825	3384
	ʉ	311	1472	2580	3225
	e	297	2117	2585	3305
	ø	518	1310	2383	3275
	ɛ	570	1369	2527	3489
	ɑ	554	909	2658	3376
	o	380	649	2774	3197
	u	359	871	2556	3150
UM3	i	362	1708	2750	3658
	y	370	1653	2684	3433
	ʉ	395	1545	2398	3200
	e	357	1985	2406	3351

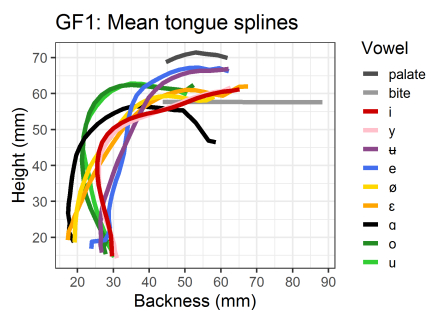
	ø	435	1509	2281	3156
	ɛ	570	1461	2352	3351
	ɑ	517	1016	2499	3418
	o	418	801	2396	3129
	u	408	914	2412	3212
UM4	i	320	2018	3044	3656
	y	306	1912	2711	3467
	ʉ	343	1689	2605	3453
	e	341	2201	2777	3537
	ø	416	1683	2576	3537
	ɛ	599	1534	2663	3461
	ɑ	548	895	2864	3550
	o	414	704	2945	3395
	u	406	899	2600	3493

# Appendix G

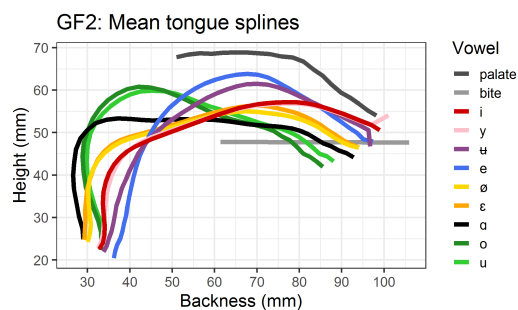
## Individual mean tongue splines

Mean ultrasound tongue contours for individual speakers' production of /i:, y:, ʊ:, e:, ø:, ɛ:, ɑ:, o:, u:/ at the 10% timepoint are provided on the following pages. The sample excludes SM3, whose ultrasound data was corrupted. The tongue is shown in midsagittal (profile) view, facing right. Annotations of tongue body height (backness), tongue tip height (gesture), tongue retraction (retraction type), and tongue shape are provided below each figure.

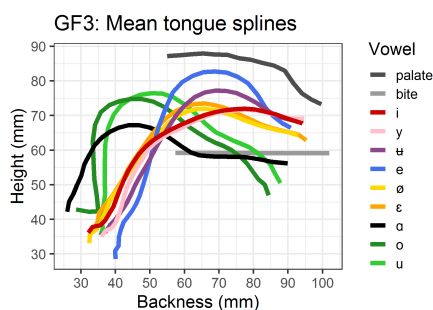
## Gothenburg females



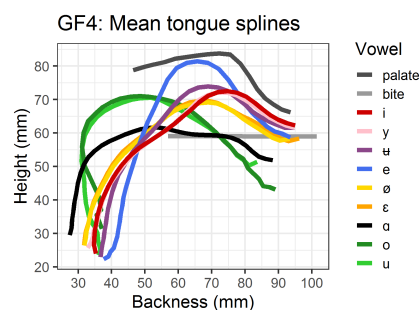
Body: Lower than /e:/ (not fronted)  
 Tip: Lower than /e:/ (up)  
 Back/root: Retracted (constricted + backed)  
 Shape: Back-bunched



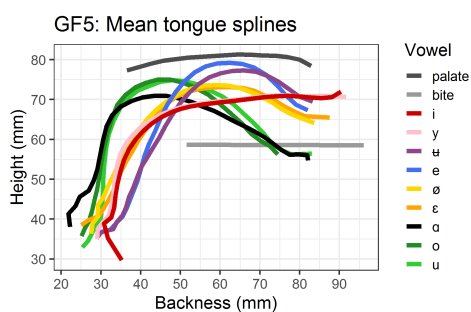
Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched



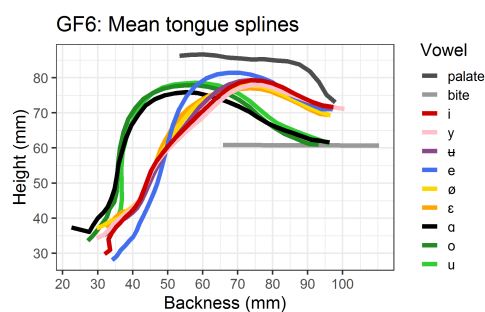
Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (backed)  
 Shape: Arched



Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched

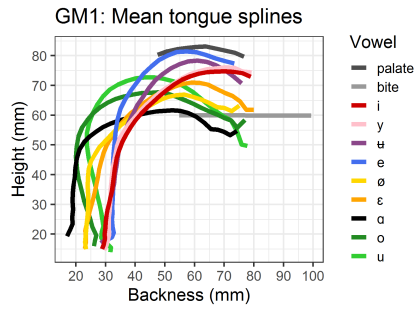


Body: Lower than /e:/ (not fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Retracted (backed)  
 Shape: Arched

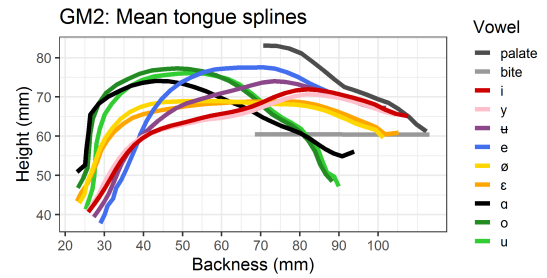


Body: Like /e:/ (fronted)  
 Tip: Like /e:/ (down)  
 Back/root: Retracted (backed)  
 Shape: Front-bunched

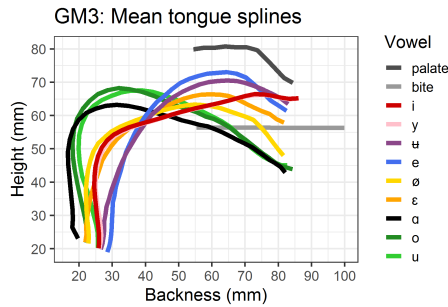
## Gothenburg males



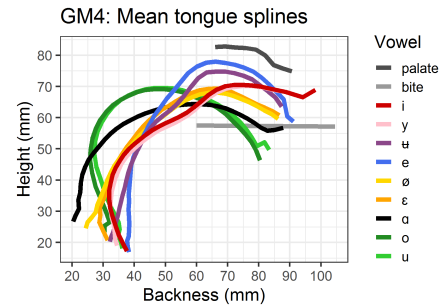
Body: Lower than /e:/ (fronted)  
 Tip: Lower than /e:/ (up)  
 Back/root: Not retracted  
 Shape: Arched



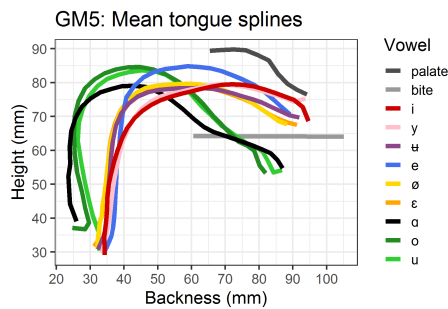
Body: Lower than /e:/ (fronted)  
 Tip: Like /e:/ (down)  
 Back/root: Retracted (constricted)  
 Shape: Double-bunched



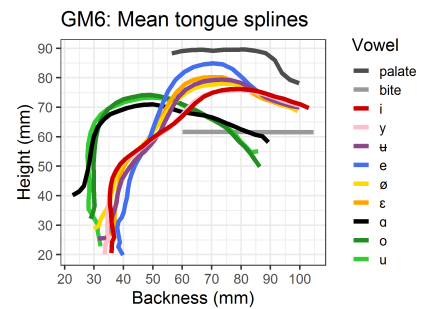
Body: Lower than /e:/ (not fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Retracted (constricted + backed)  
 Shape: Back-bunched



Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched

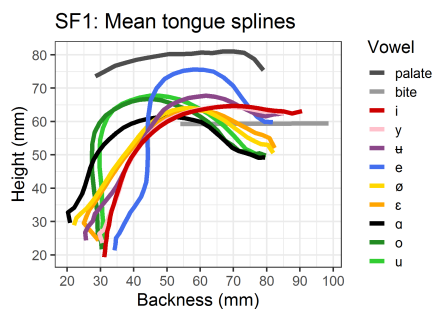


Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Not retracted  
 Shape: Arched

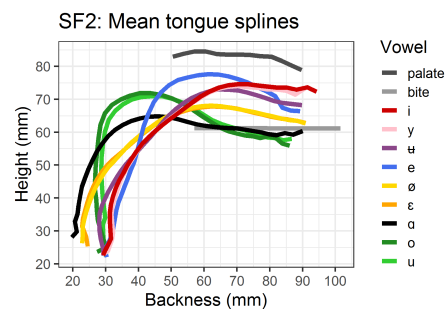


Body: Lower than /e:/ (fronted)  
 Tip: Like /e:/ (down)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched

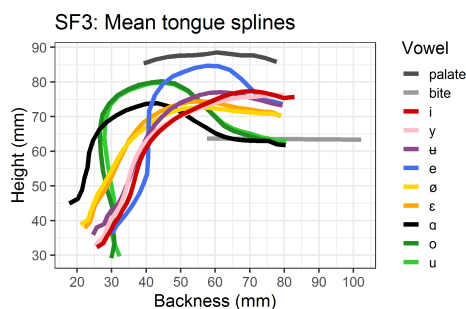
## Stockholm females



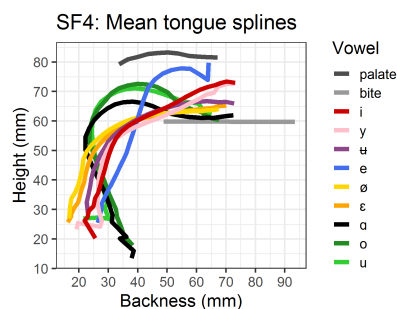
Body: Lower than /e:/ (not fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Retracted (backed)  
 Shape: Arched



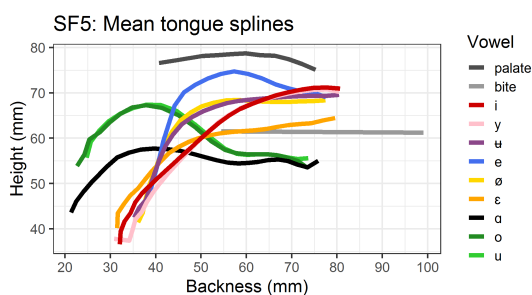
Body: Like /e:/ (fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Retracted (constricted)  
 Shape: Double-bunched



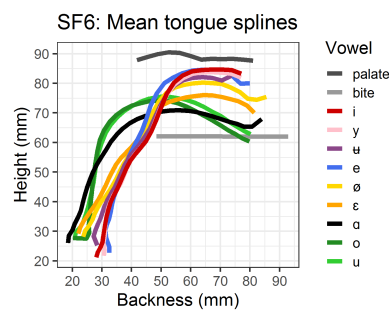
Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Retracted (backed)  
 Shape: Arched



Body: Lower than /e:/ (not fronted)  
 Tip: Like /e:/ (up)  
 Back/root: Retracted (constricted + backed)  
 Shape: Back-bunched

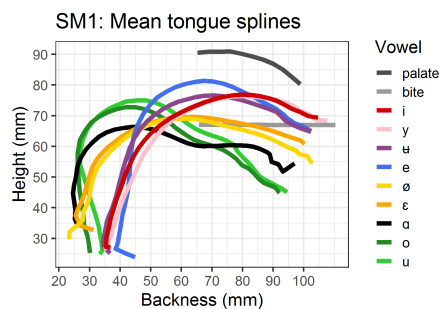


Body: Lower than /e:/ (fronted)  
 Tip: Like /e:/ (up)  
 Back/root: Not retracted  
 Shape: Arched

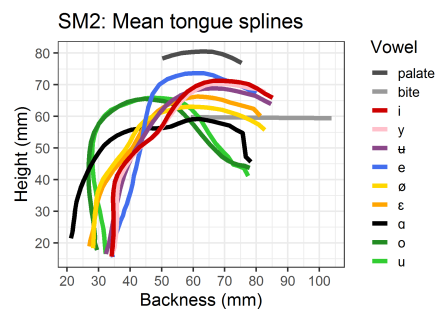


Body: Like /e:/ (fronted)  
 Tip: Higher than /e:/ (up)  
 Back/root: Not retracted  
 Shape: Front-bunched

## Stockholm males

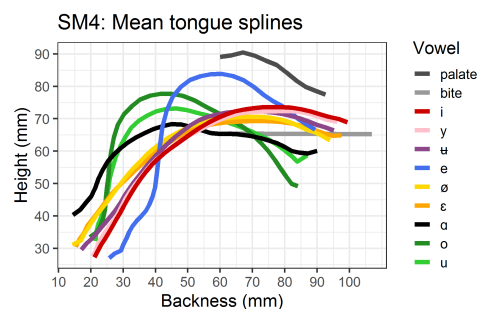


Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Not retracted  
 Shape: Arched

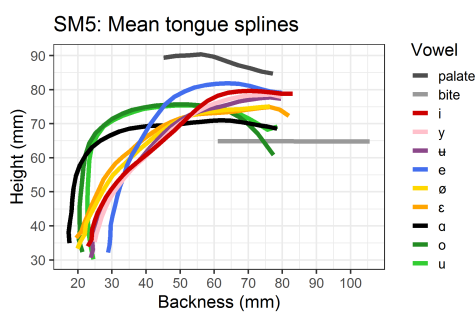


Body: Like /e:/ (fronted)  
 Tip: Like /e:/ (down)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched

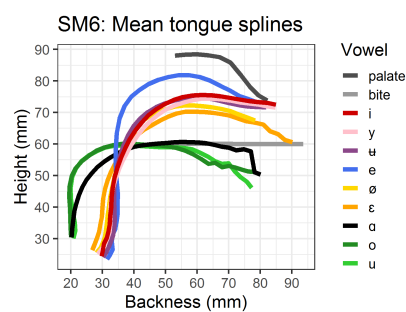
SM3: No ultrasound data



Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (backed)  
 Shape: Arched



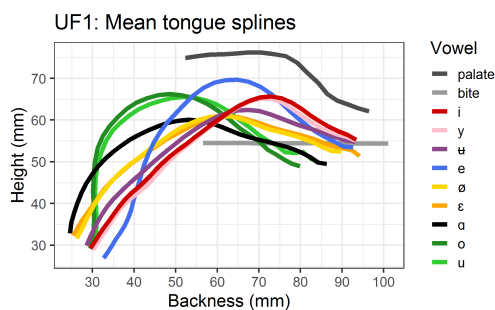
Body: Like /e:/ (fronted)  
 Tip: Like /e:/ (up)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched



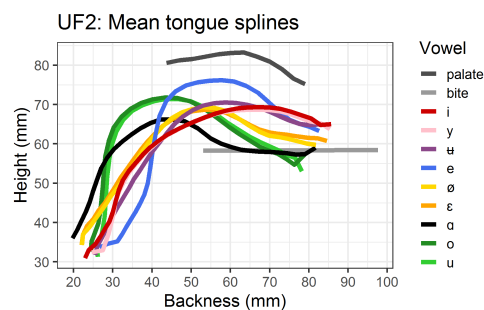
Body: Lower than /e:/ (not fronted)  
 Tip: Like /e:/ (up)  
 Back/root: Not retracted  
 Shape: Arched



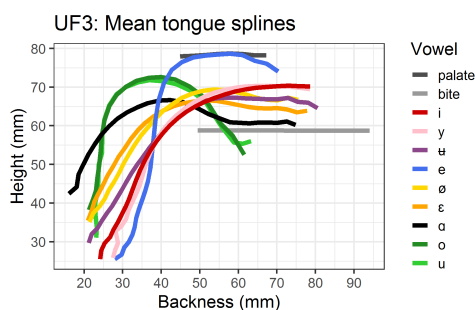
## Uppsala females



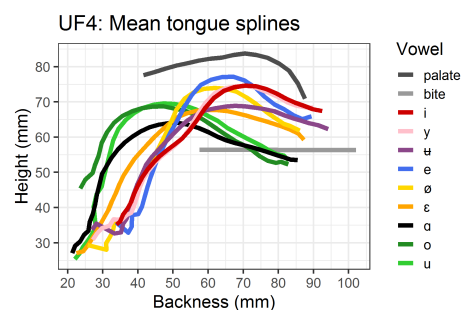
Body: Like /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (backed)  
 Shape: Front-bunched



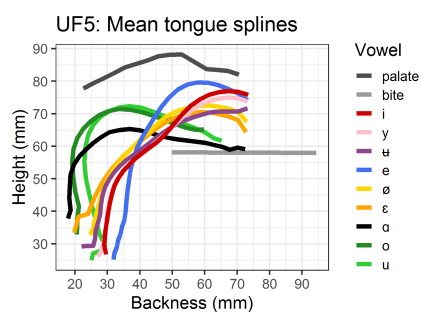
Body: Lower than /e:/ (not fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (backed)  
 Shape: Arched



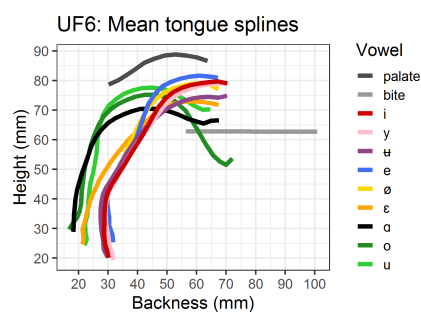
Body: Lower than /e:/ (not fronted)  
 Tip: Lower than /e:/ (up)  
 Back/root: Retracted (backed)  
 Shape: Arched



Body: Like /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (constricted)  
 Shape: Double-bunched

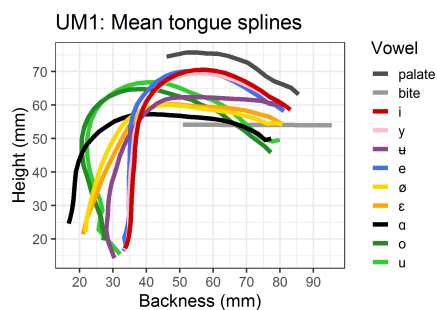


Body: Like /e:/ (fronted)  
 Tip: Like /e:/ (up)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched

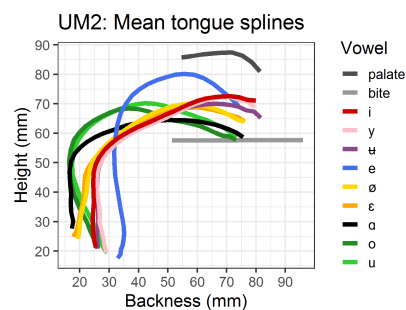


Body: Like /e:/ (fronted)  
 Tip: Like /e:/ (up)  
 Back/root: Retracted (constricted)  
 Shape: Double-bunched

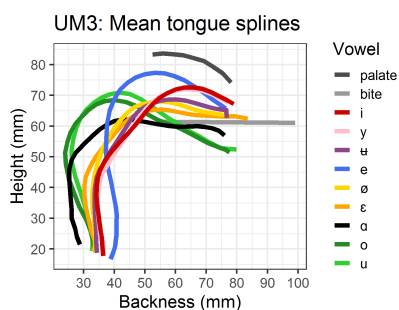
## Uppsala males



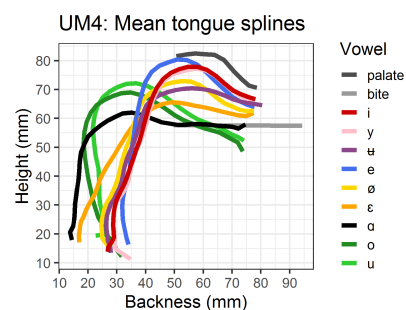
Body: Like /e:/ (not fronted)  
 Tip: Like /e:/ (down)  
 Back/root: Not retracted  
 Shape: Arched



Body: Lower than /e:/ (not fronted)  
 Tip: Like /e:/ (down)  
 Back/root: Retracted (constricted + backed)  
 Shape: Back-bunched



Body: Lower than /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (constricted + backed)  
 Shape: Double-bunched



Body: Like /e:/ (fronted)  
 Tip: Higher than /e:/ (down)  
 Back/root: Retracted (backed)  
 Shape: Front-bunched

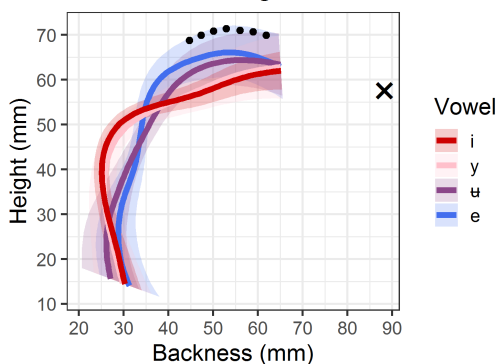
# Appendix H

## Individual tongue GAMMs

GAMM trajectories for individual speakers' tongue gestures for /i:, y:, ʊ:, e:/ at the 10% time-point are provided here. Difference smooths for /i:-y:/ and /i:-e:/ are positioned beside each figure. An interpretation of each difference smooth is annotated below the GAMM plots. The sample excludes GF5, SF4, SM1, SM3, UF3, UF4, UM1, and UM3 due to missing data.

## Gothenburg females

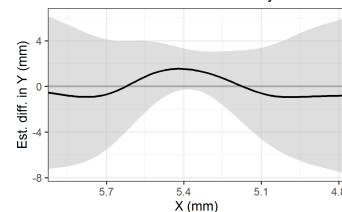
GF1: GAM of high front vowels



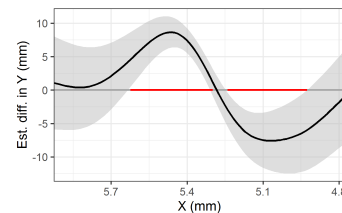
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower)

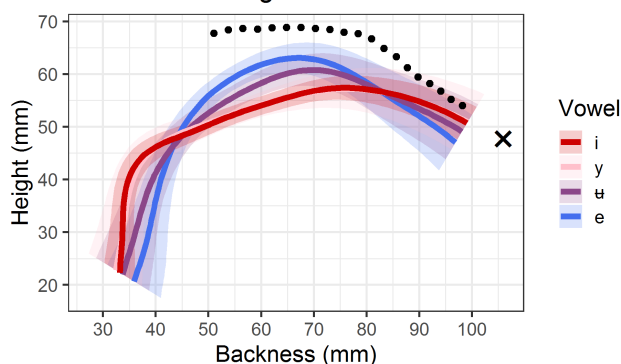
GF1: Difference smooth /i/ vs. /y/



GF1: Difference smooth /i/ vs. /e/



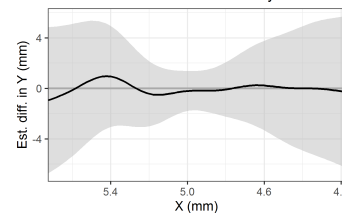
GF2: GAM of high front vowels



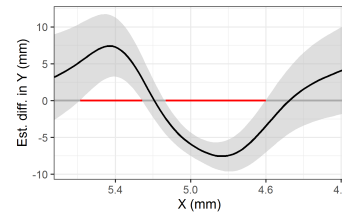
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower)

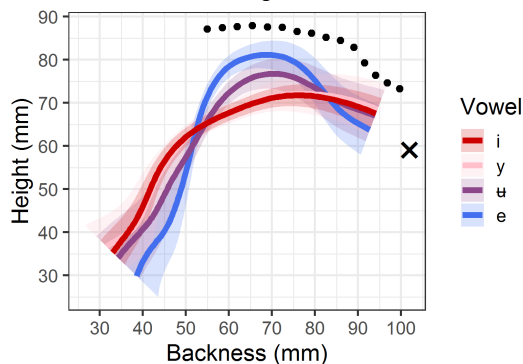
GF2: Difference smooth /i/ vs. /y/



GF2: Difference smooth /i/ vs. /e/



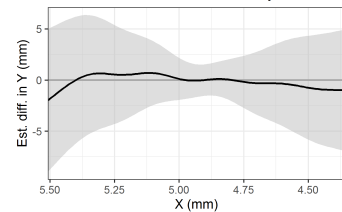
GF3: GAM of high front vowels



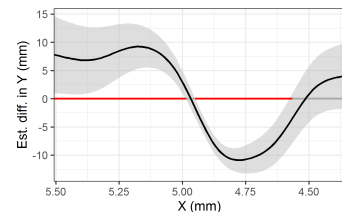
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower)

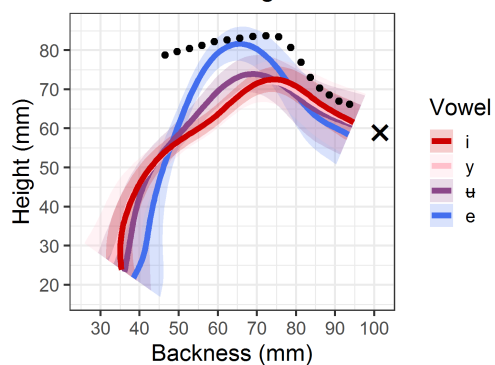
GF3: Difference smooth /i/ vs. /y/



GF3: Difference smooth /i/ vs. /e/

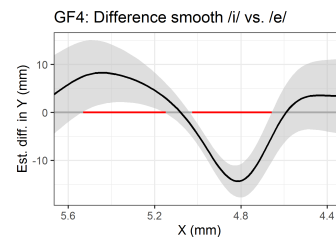
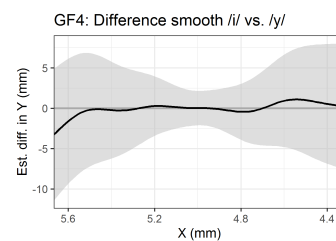


GF4: GAM of high front vowels



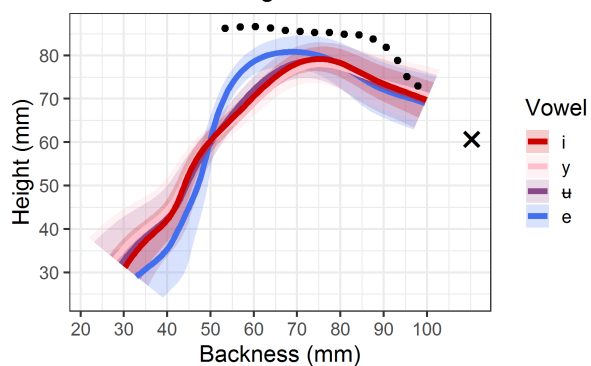
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter)



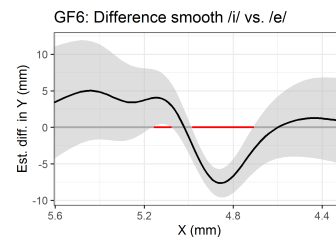
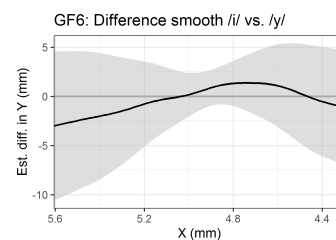
GF5: No GAMM data

GF6: GAM of high front vowels



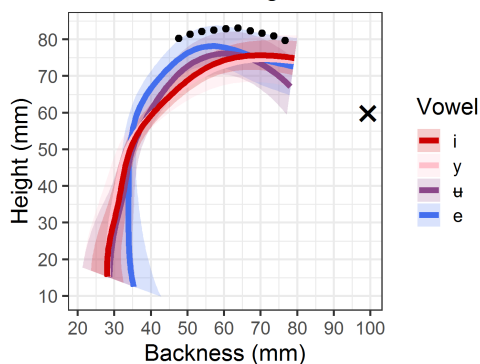
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer\*), tongue body (lower/fronter)

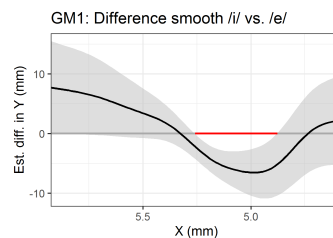
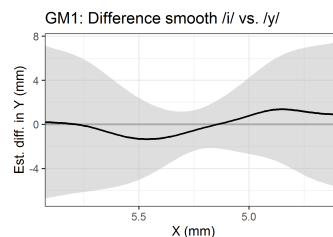


## Gothenburg males

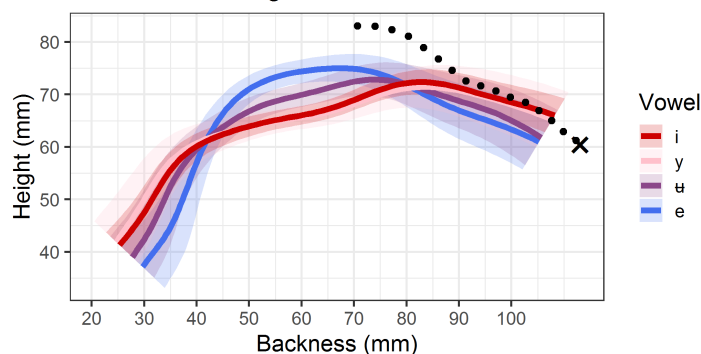
GM1: GAM of high front vowels



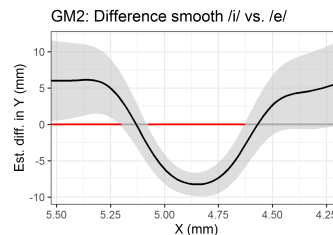
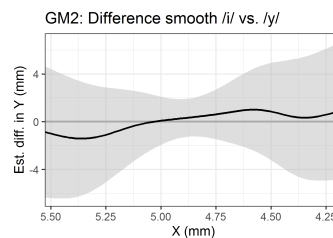
/i:/ vs. /y:/: No sig. diff.  
 /i:/ vs. /e:/: Tongue body (lower/fronter)



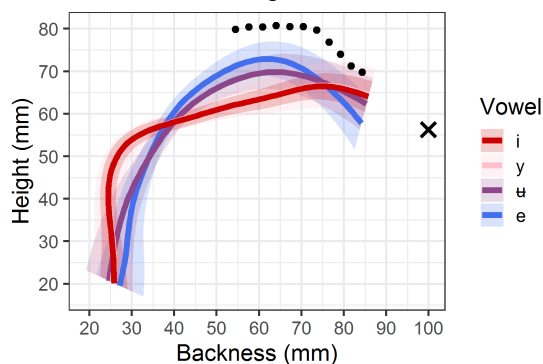
GM2: GAM of high front vowels



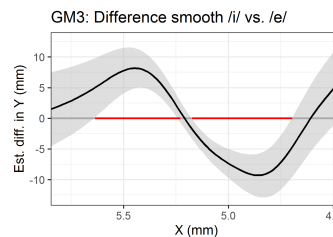
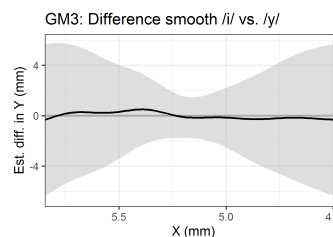
/i:/ vs. /y:/: No sig. diff.  
 /i:/ vs. /e:/: Tongue back (backer), tongue body (lower)



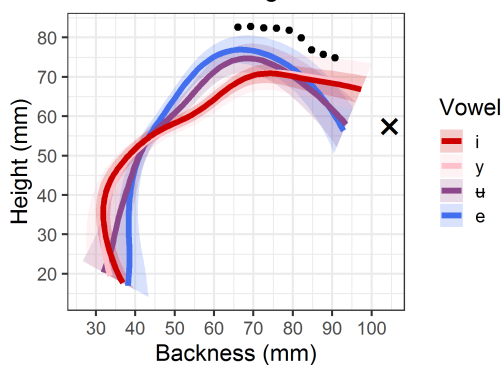
GM3: GAM of high front vowels



/i:/ vs. /y:/: No sig. diff.  
 /i:/ vs. /e:/: Tongue back (backer\*), tongue body (lower), tongue tip (higher\*)

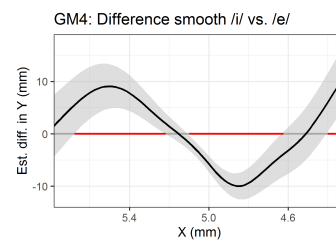
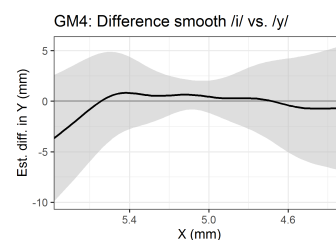


GM4: GAM of high front vowels

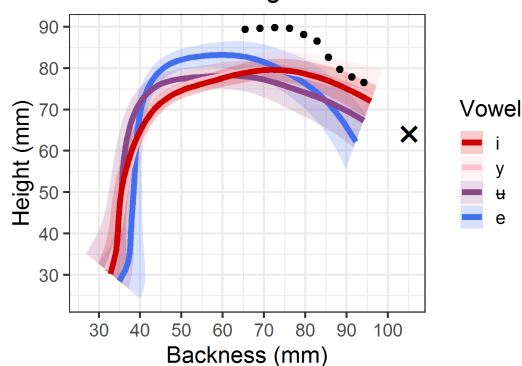


/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter), tongue tip (higher)

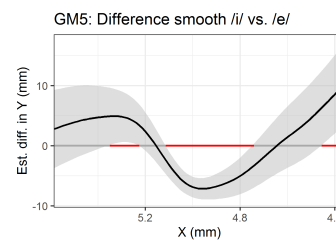
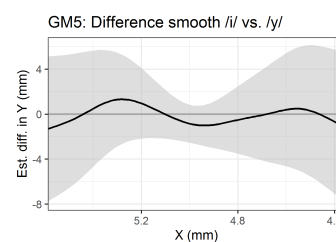


GM5: GAM of high front vowels

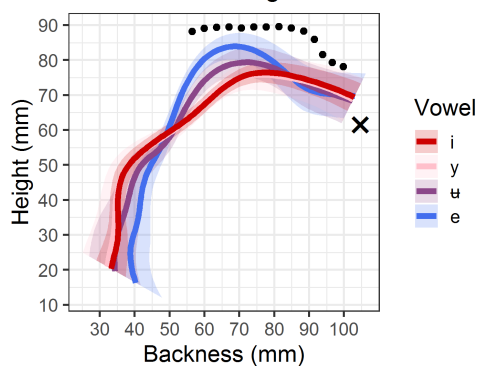


/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter), tongue tip (higher)

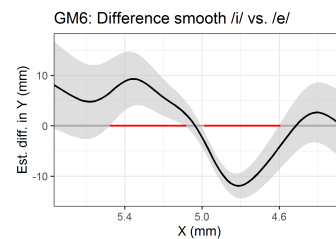
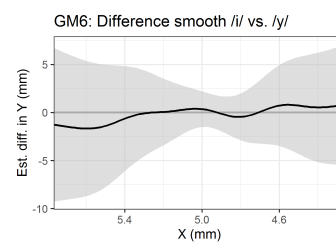


GM6: GAM of high front vowels



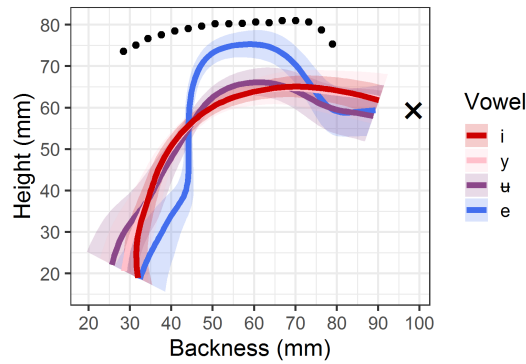
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter)



## Stockholm females

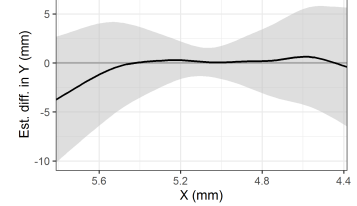
SF1: GAM of high front vowels



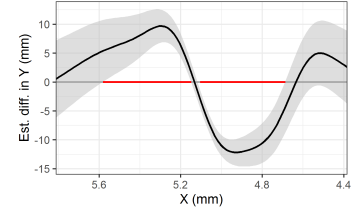
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower)

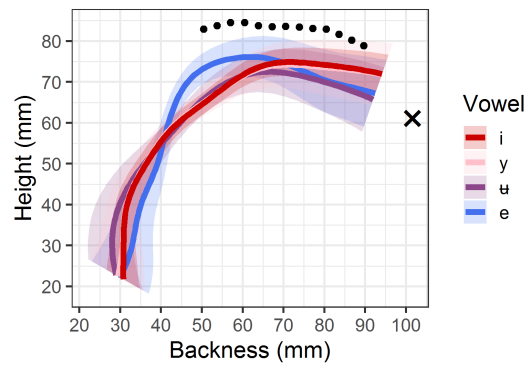
SF1: Difference smooth /i/ vs. /y/



SF1: Difference smooth /i/ vs. /e/



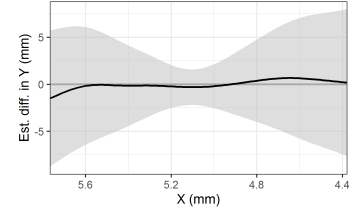
SF2: GAM of high front vowels



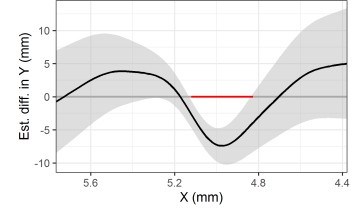
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue body (lower/fronter)

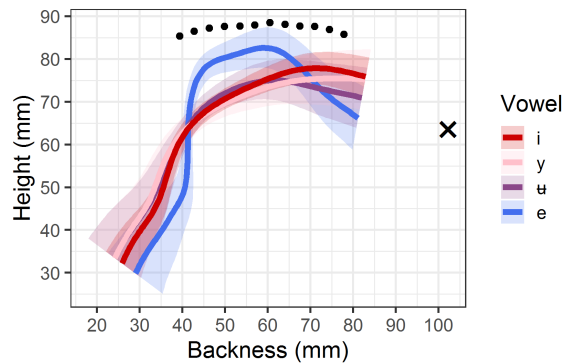
SF2: Difference smooth /i/ vs. /y/



SF2: Difference smooth /i/ vs. /e/



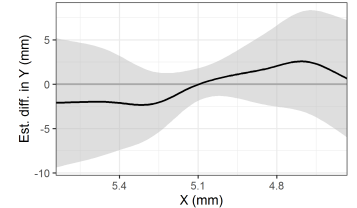
SF3: GAM of high front vowels



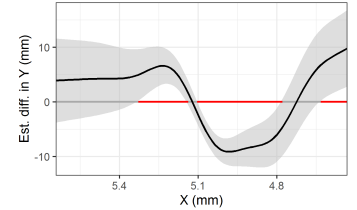
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower), tongue tip (higher)

SF3: Difference smooth /i/ vs. /y/

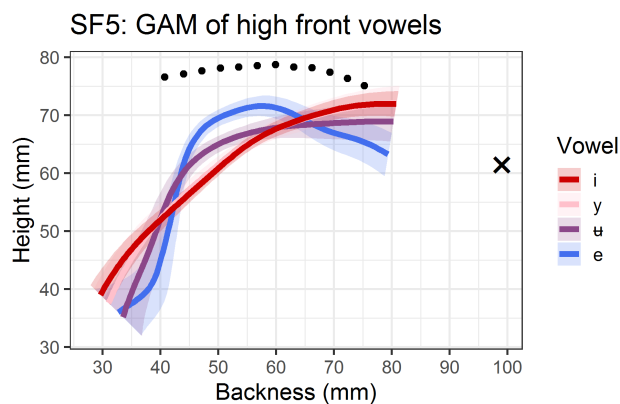


SF3: Difference smooth /i/ vs. /e/



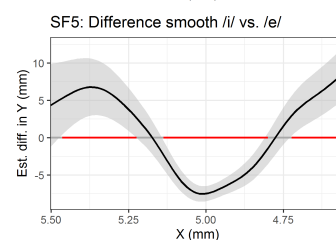
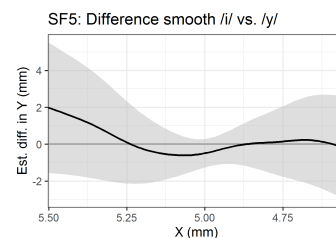


SF4: No GAMM data

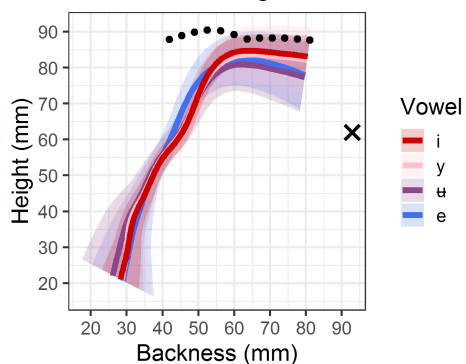


/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter), tongue tip (higher)

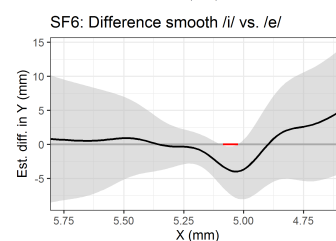
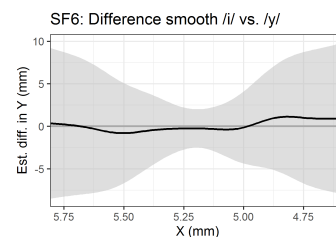


SF6: GAM of high front vowels



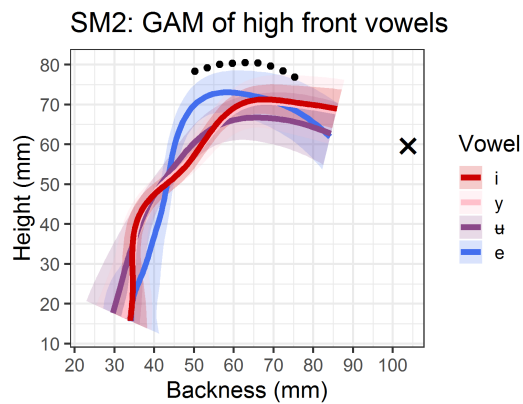
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue body (fronter\*)



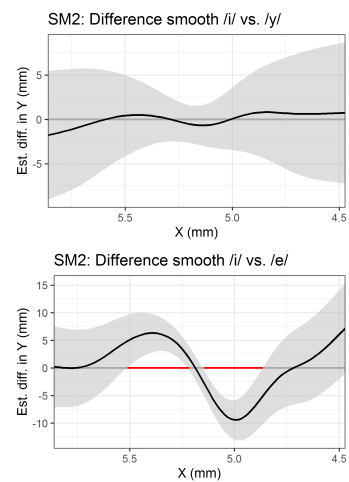
## Stockholm males

SM1: No GAMM data



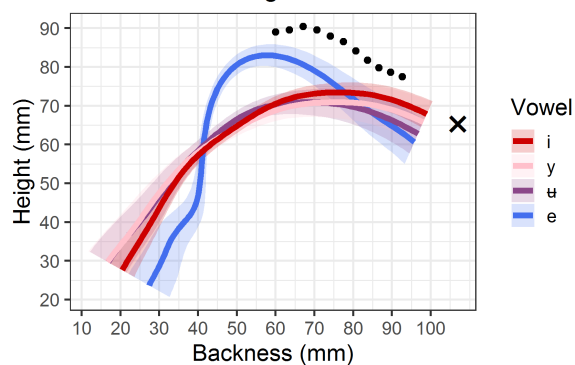
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter)



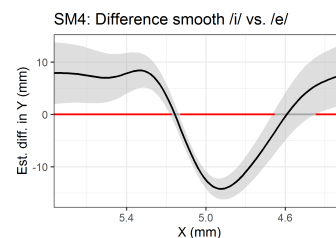
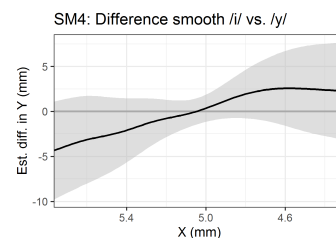
SM3: No GAMM data

SM4: GAM of high front vowels

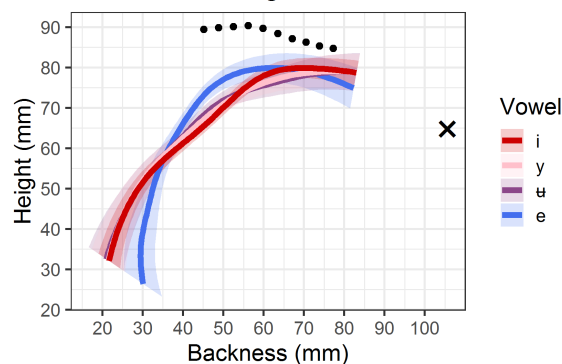


/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower), tongue tip (higher)

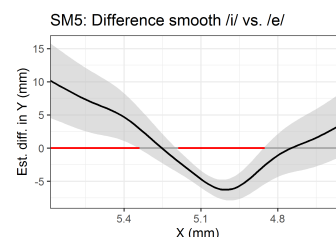
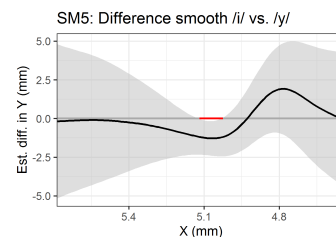


SM5: GAM of high front vowels

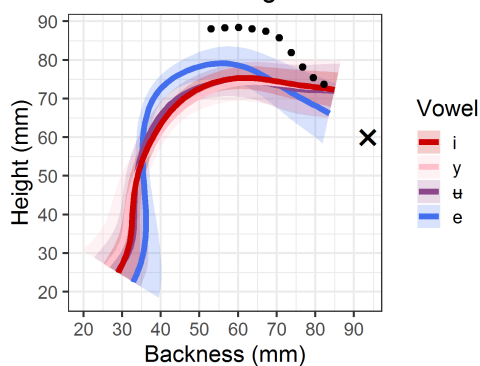


/i:/ vs. /y:/: Tongue body (lower\*)

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter)

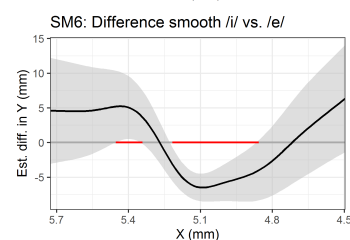
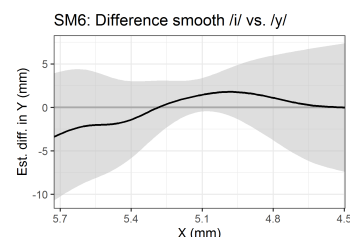


SM6: GAM of high front vowels



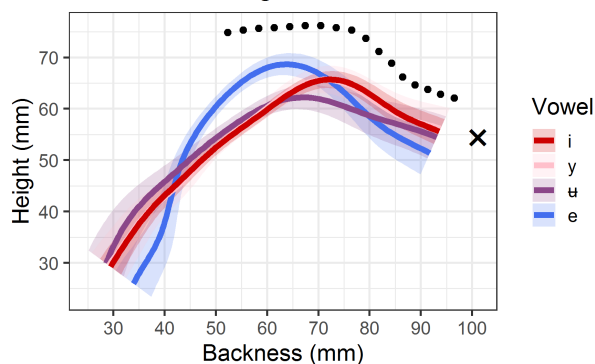
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower)



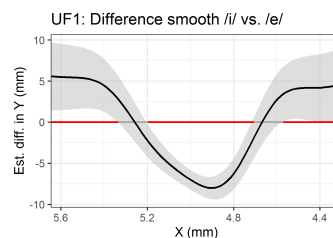
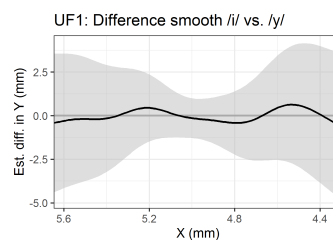
## Uppsala females

UF1: GAM of high front vowels

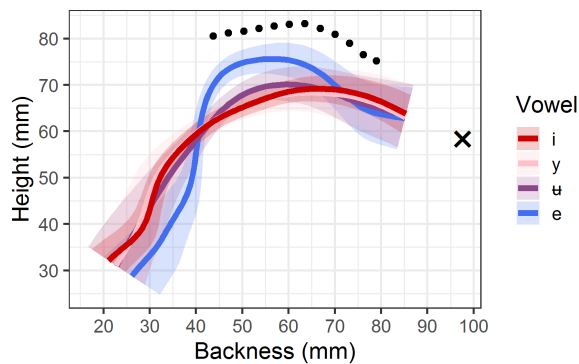


/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter), tongue tip (higher)

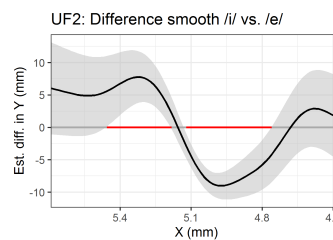
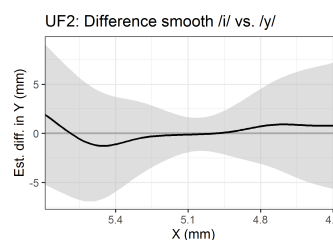


UF2: GAM of high front vowels



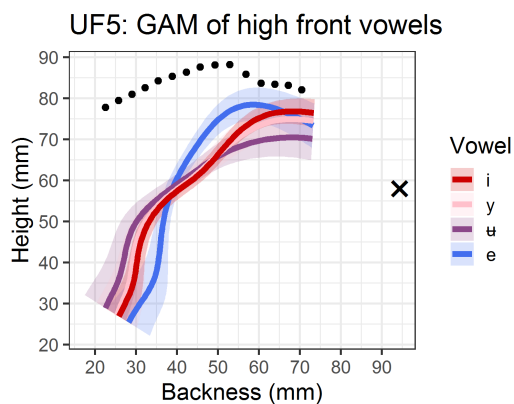
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower)



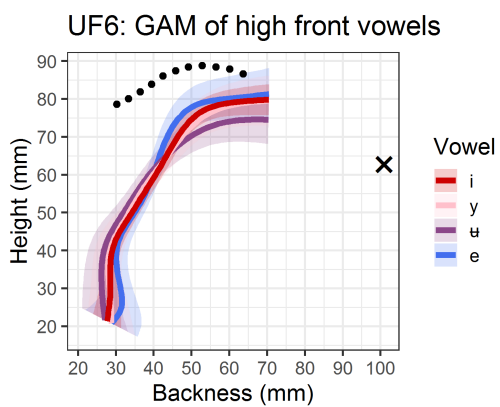
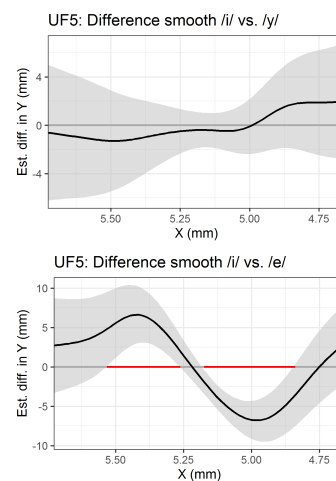
UF3: No GAMM data

UF4: No GAMM data



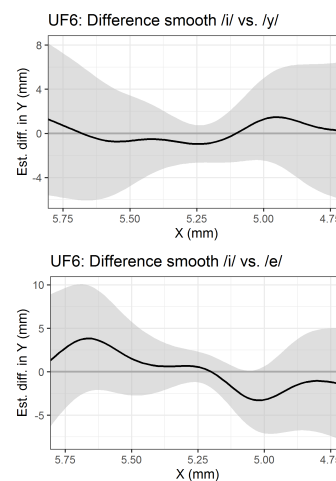
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower/fronter)



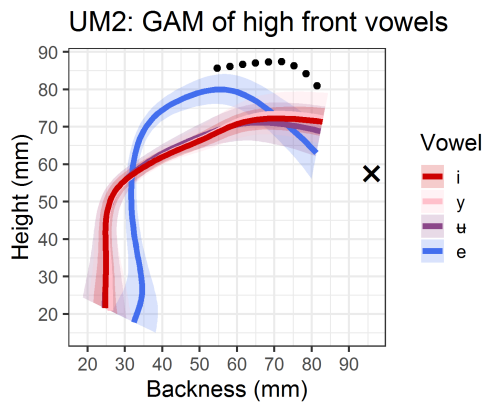
/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: No sig. diff.



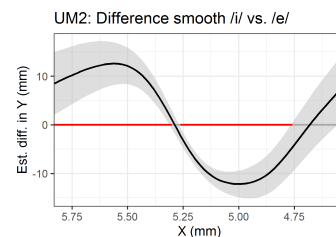
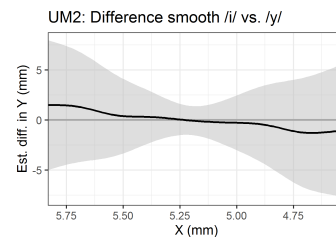
## Uppsala males

UM1: No GAMM data

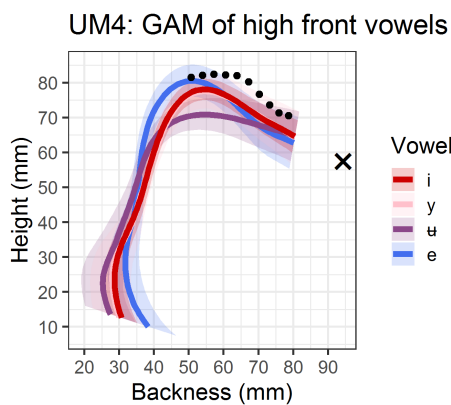


/i:/ vs. /y:/: No sig. diff.

/i:/ vs. /e:/: Tongue back (backer), tongue body (lower), tongue tip (higher)

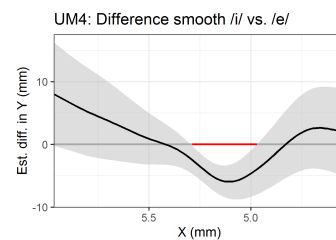
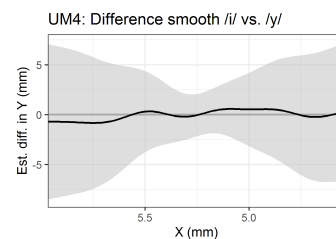


UM3: No GAMM data



/i:/ vs. /y:/: No sig. diff.

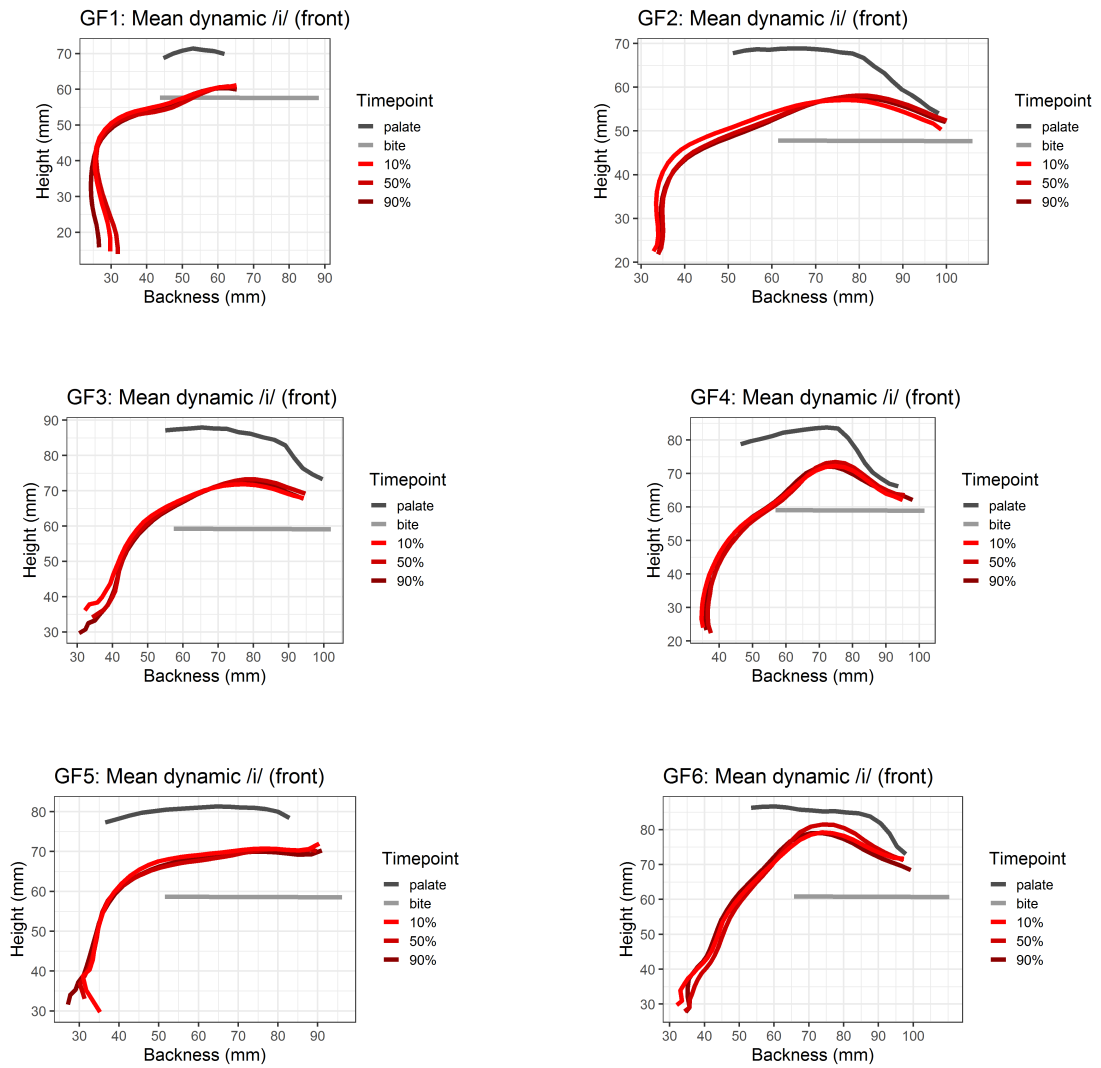
/i:/ vs. /e:/: Tongue body (lower/fronter)



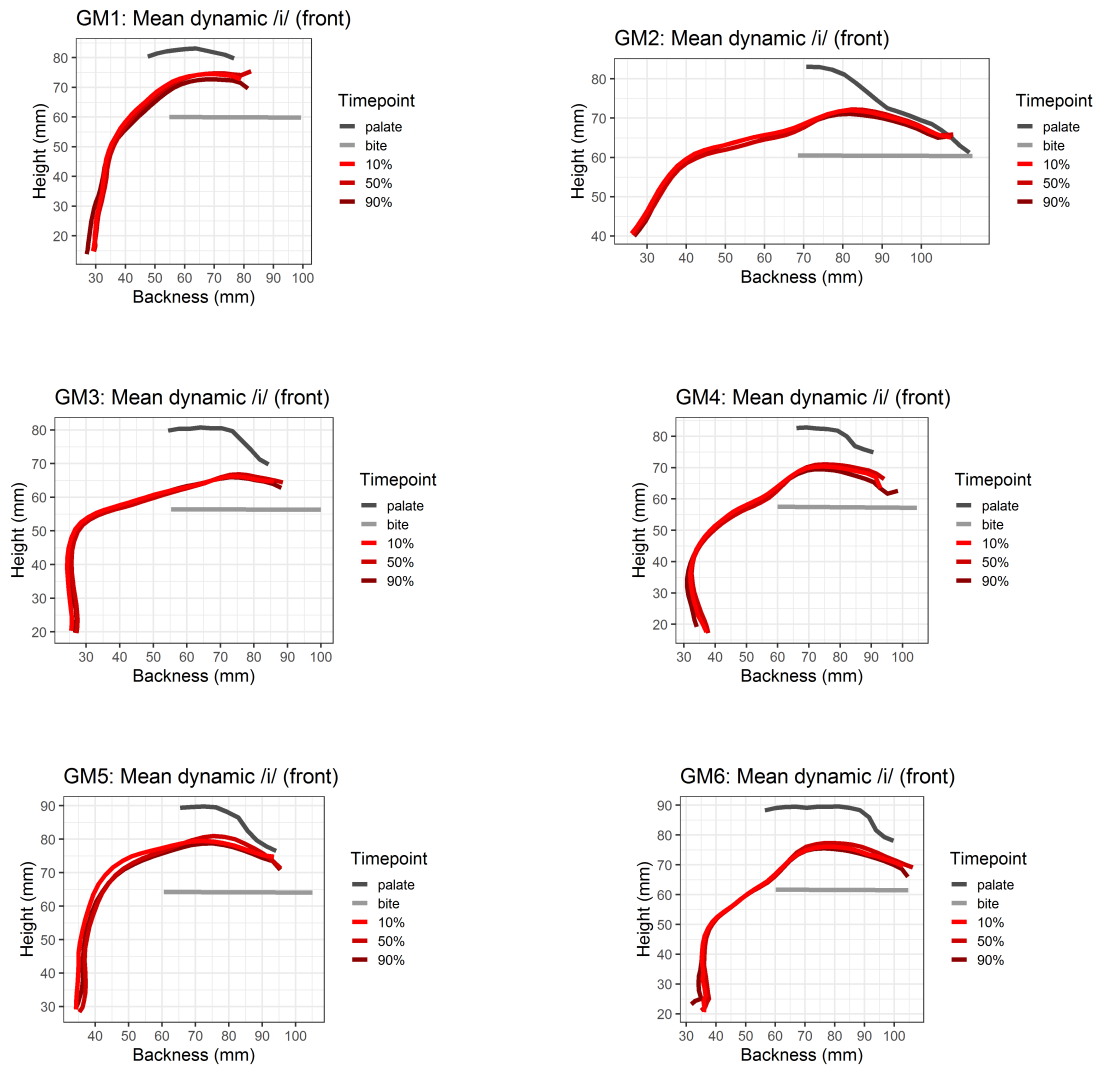
# Appendix I

## Dynamic mean splines for /i:/

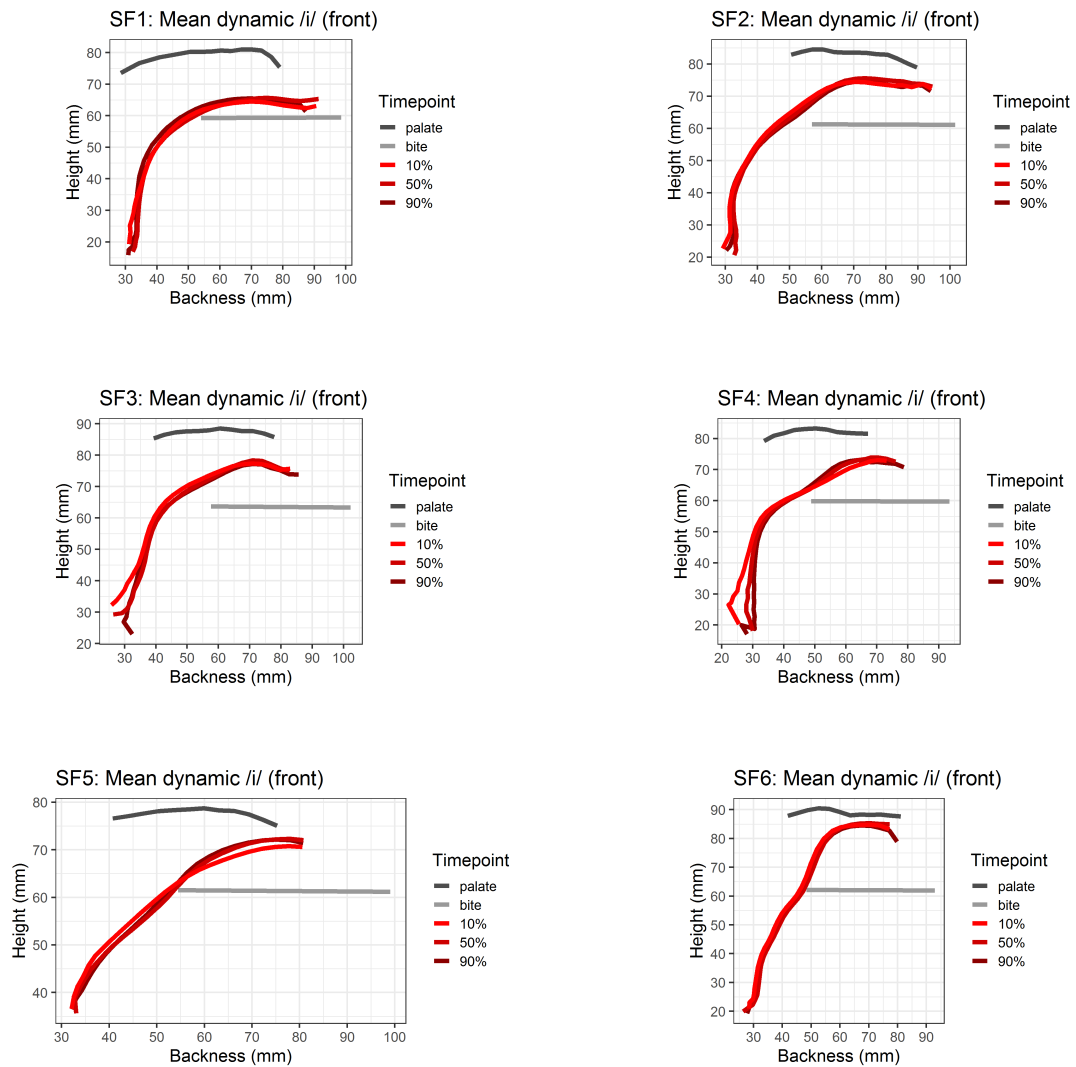
Mean dynamic ultrasound contours for individual speakers' production of /i:/ at the 10%, 50%, and 90% timepoint is provided on the following pages. Front contexts are presented in full first, followed by back contexts. The sample excludes SM3, whose ultrasound data was corrupted. The tongue is shown in midsagittal (profile) view, facing right.

**Gothenburg females: front**



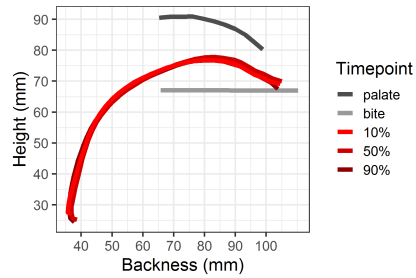
**Gothenburg males: front**

## Stockholm females: front

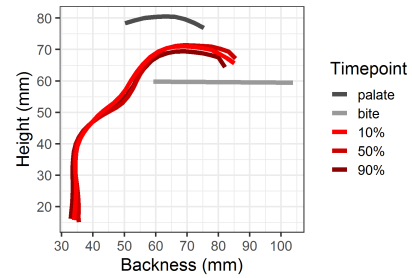


## Stockholm males: front

SM1: Mean dynamic /ɪ/ (front)

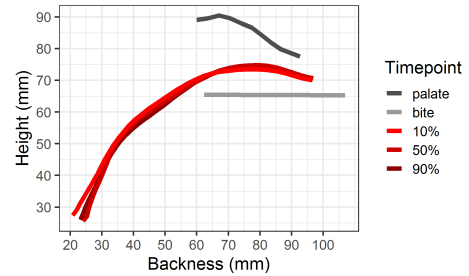


SM2: Mean dynamic /ɪ/ (front)

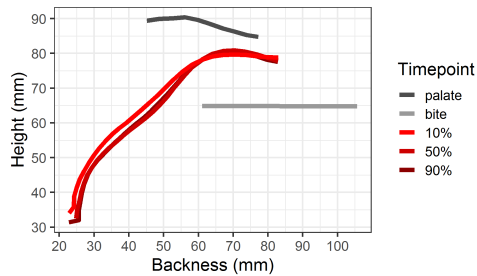


SM3: No ultrasound data

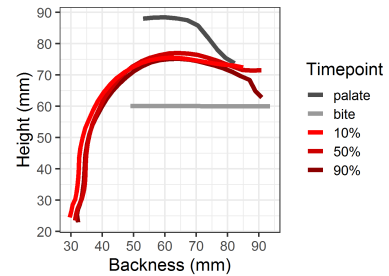
SM4: Mean dynamic /ɪ/ (front)



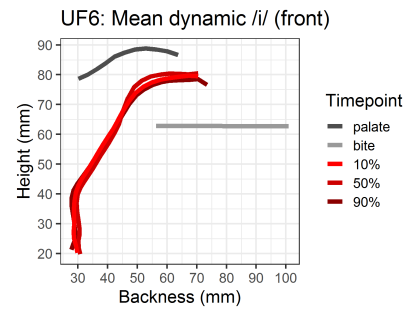
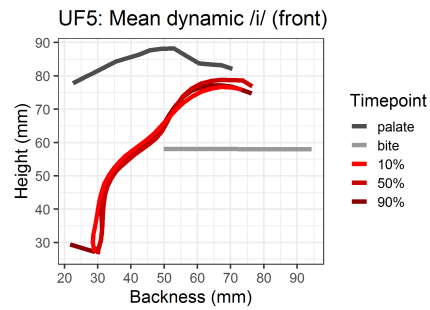
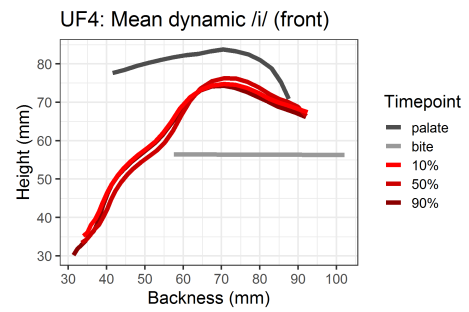
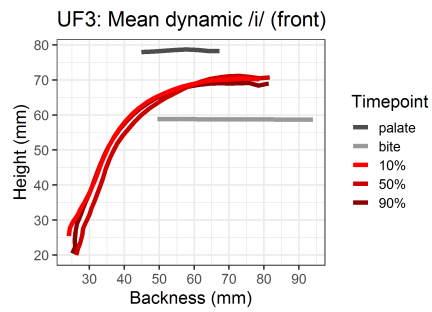
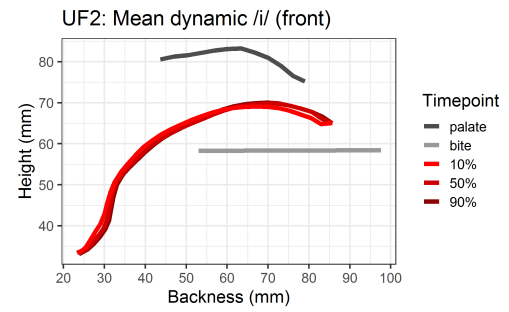
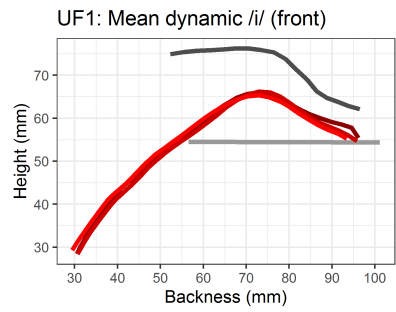
SM5: Mean dynamic /ɪ/ (front)



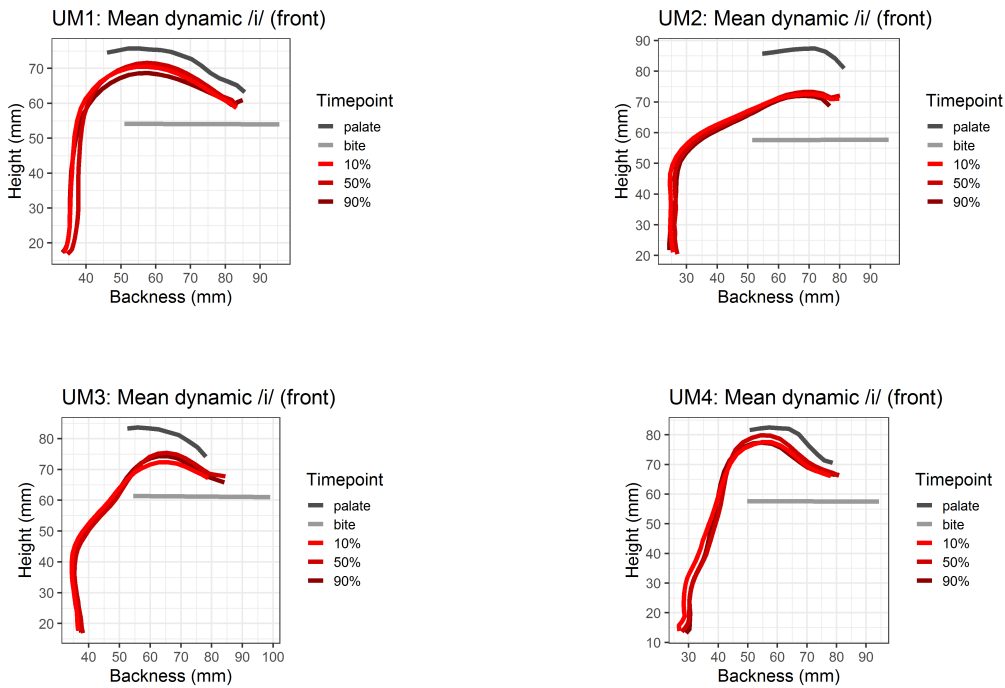
SM6: Mean dynamic /ɪ/ (front)



## Uppsala females: front

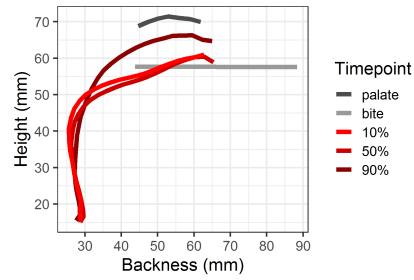


Uppsala males: front

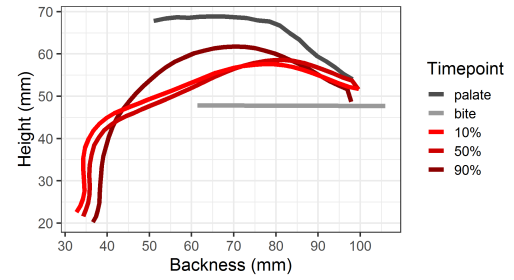


## Gothenburg females: back

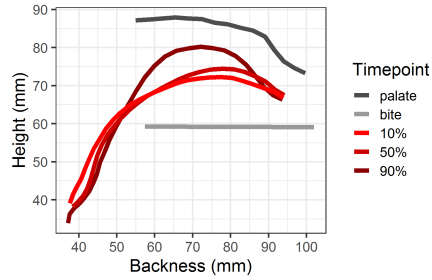
GF1: Mean dynamic /ɪ/ (back)



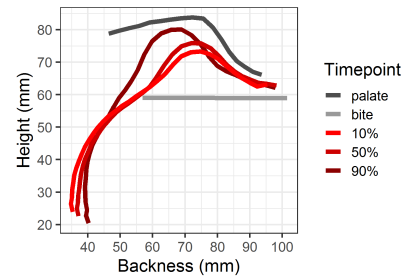
GF2: Mean dynamic /ɪ/ (back)



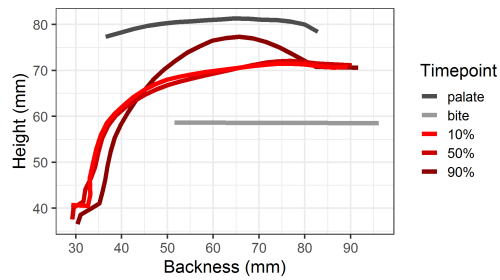
GF3: Mean dynamic /ɪ/ (back)



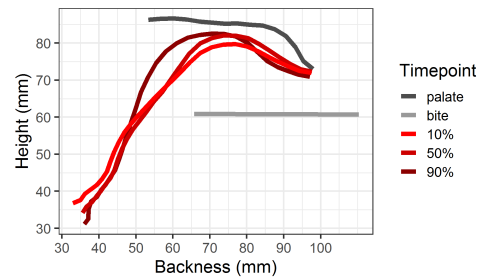
GF4: Mean dynamic /ɪ/ (back)



GF5: Mean dynamic /ɪ/ (back)

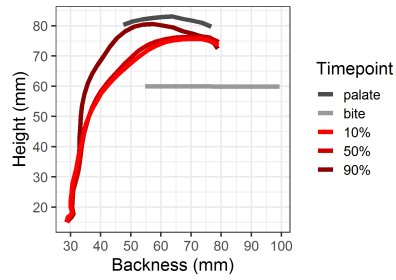


GF6: Mean dynamic /ɪ/ (back)

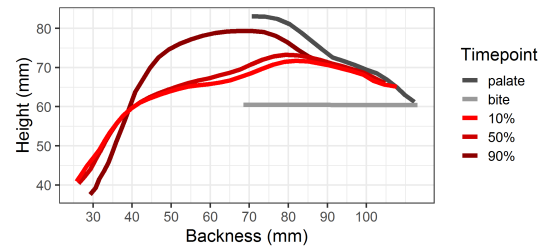


## Gothenburg males: back

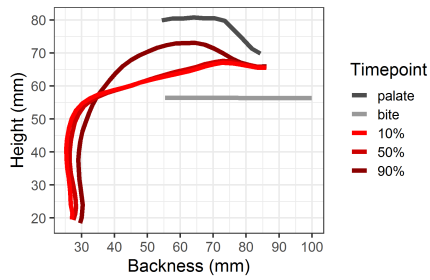
GM1: Mean dynamic /ɪ/ (back)



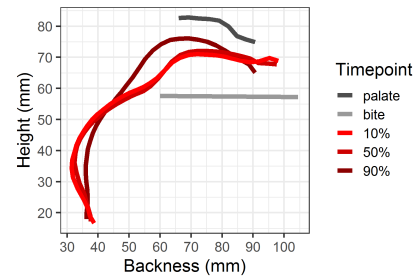
GM2: Mean dynamic /ɪ/ (back)



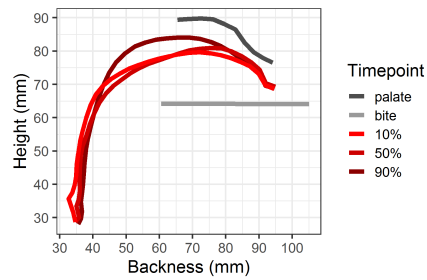
GM3: Mean dynamic /ɪ/ (back)



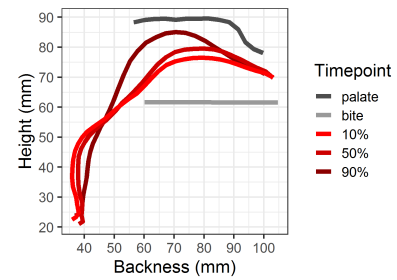
GM4: Mean dynamic /ɪ/ (back)



GM5: Mean dynamic /ɪ/ (back)

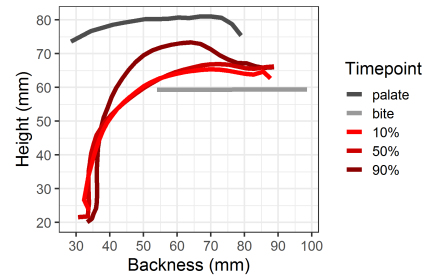


GM6: Mean dynamic /ɪ/ (back)

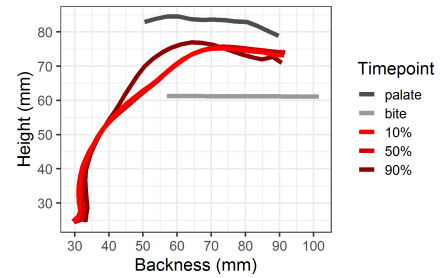


## Stockholm females: back

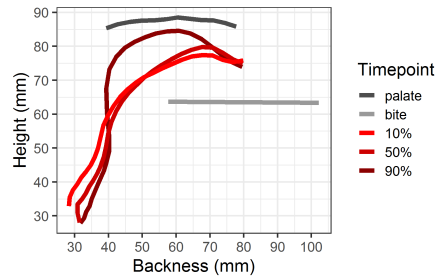
SF1: Mean dynamic /ɪ/ (back)



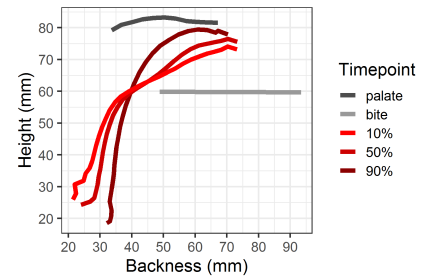
SF2: Mean dynamic /ɪ/ (back)



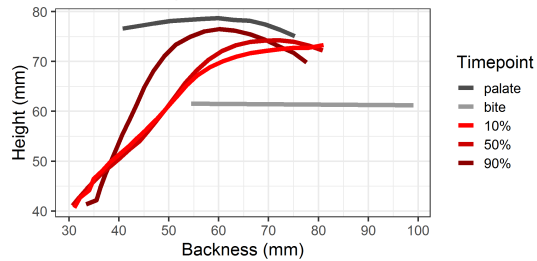
SF3: Mean dynamic /ɪ/ (back)



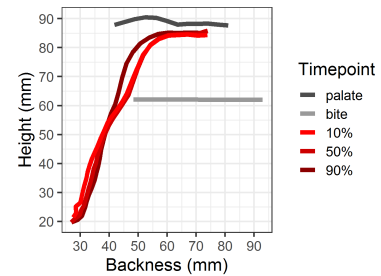
SF4: Mean dynamic /ɪ/ (back)



SF5: Mean dynamic /ɪ/ (back)



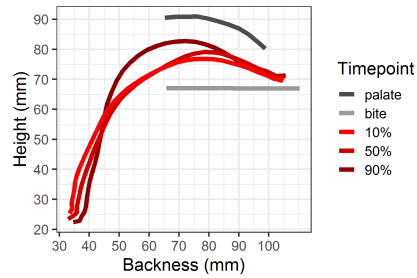
SF6: Mean dynamic /ɪ/ (back)



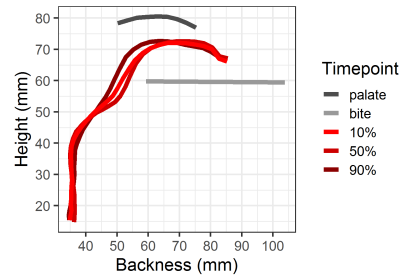


## Stockholm males: back

SM1: Mean dynamic /ɪ/ (back)

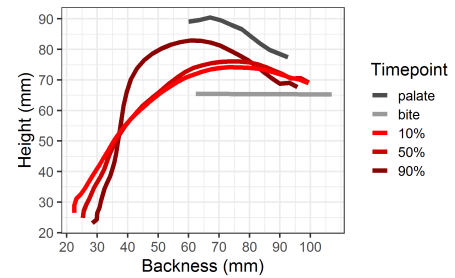


SM2: Mean dynamic /ɪ/ (back)

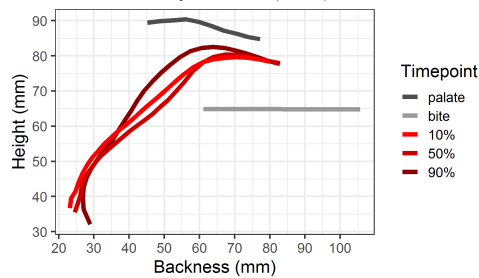


SM3: No ultrasound data

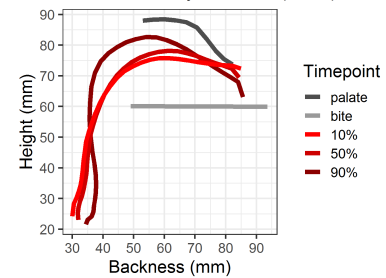
SM4: Mean dynamic /ɪ/ (back)



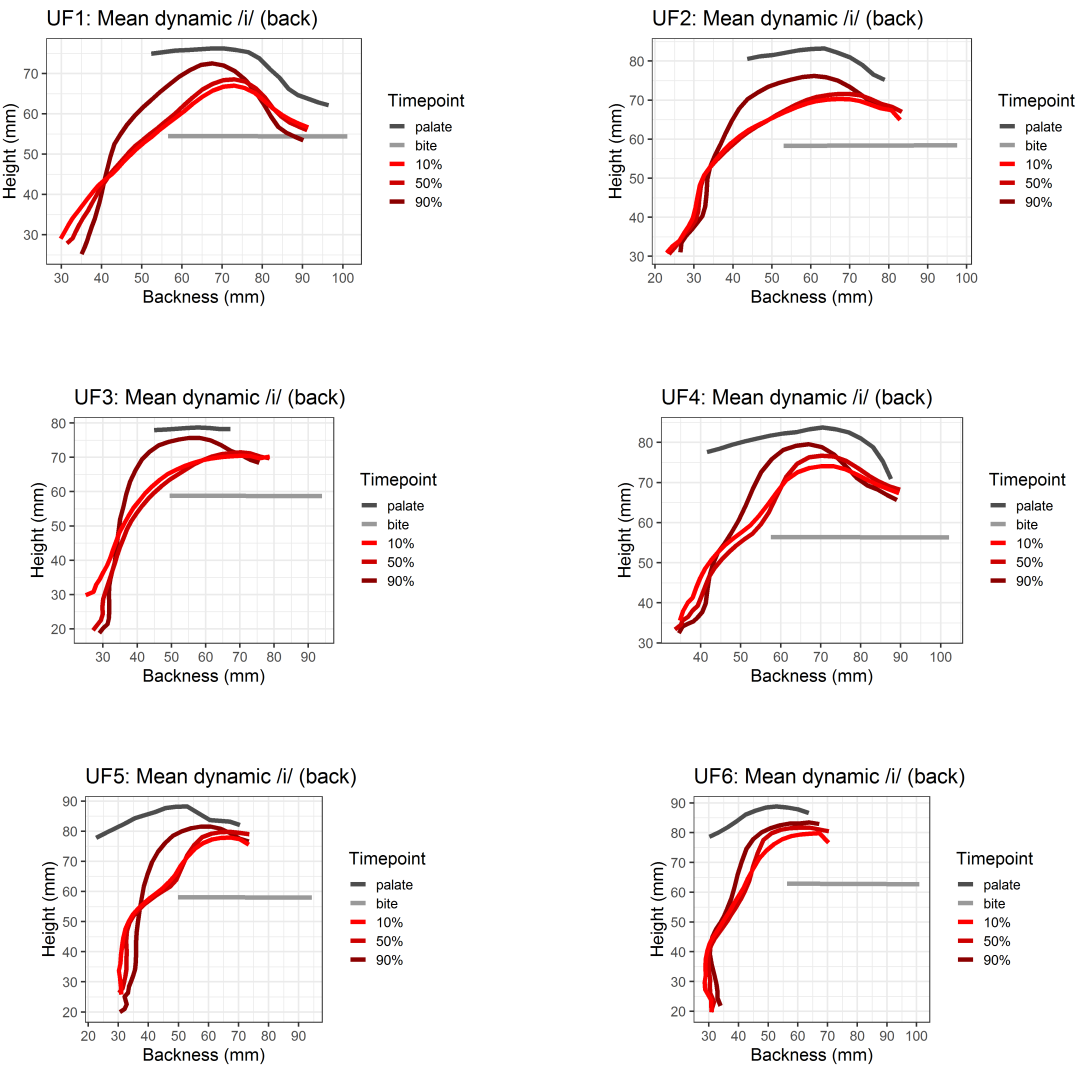
SM5: Mean dynamic /ɪ/ (back)



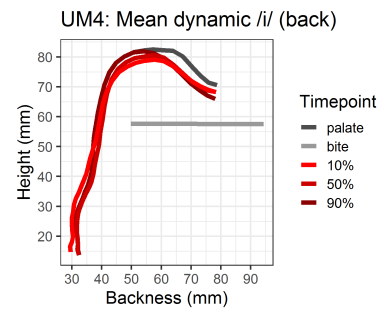
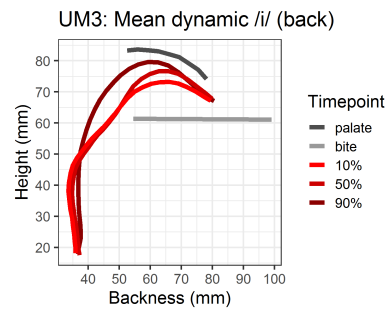
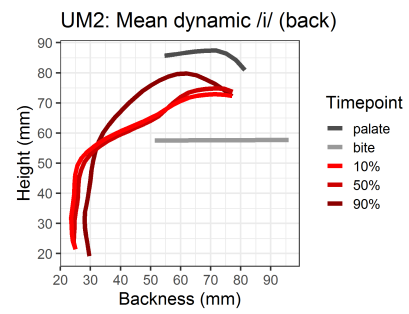
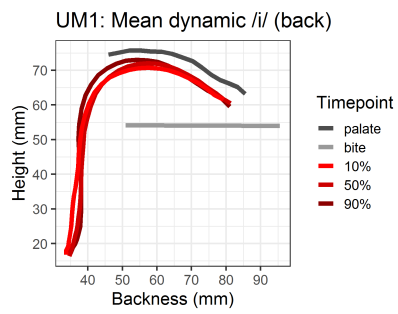
SM6: Mean dynamic /ɪ/ (back)



Uppsala females: back



## Uppsala males: back

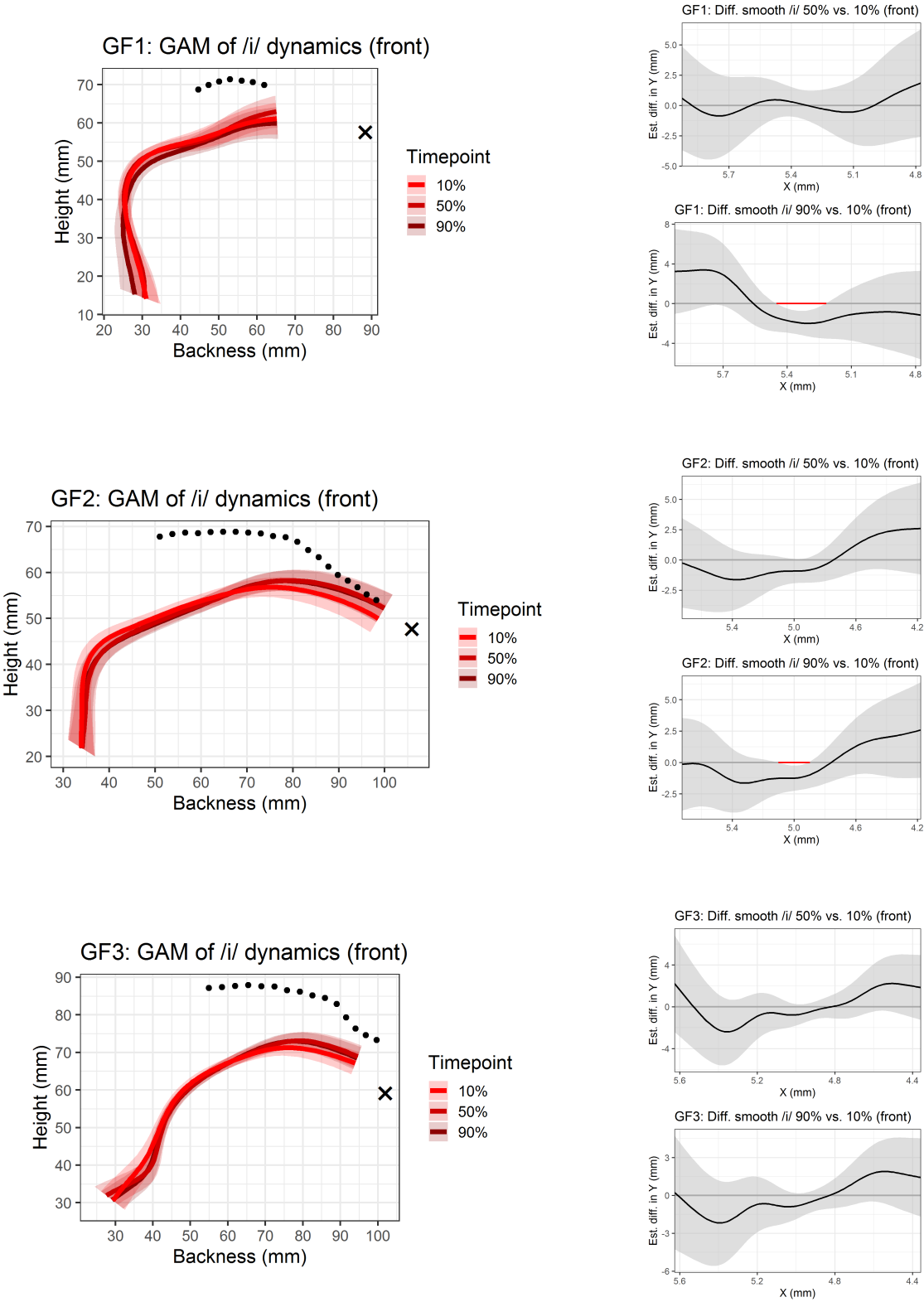


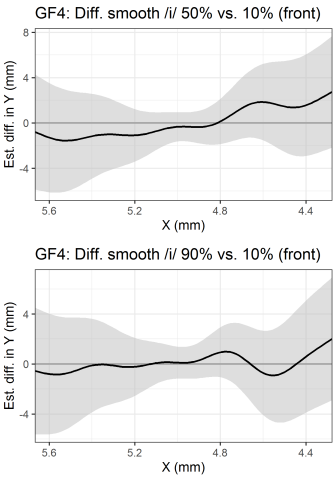
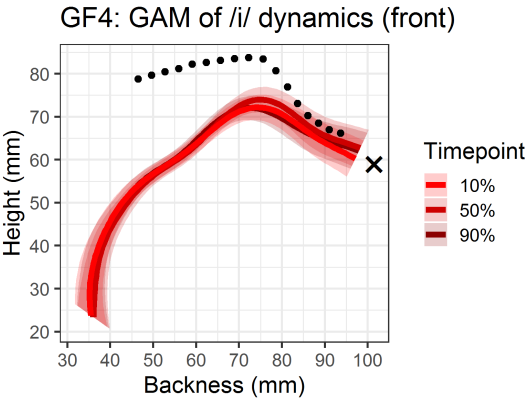
# Appendix J

## Individual dynamic GAMMs

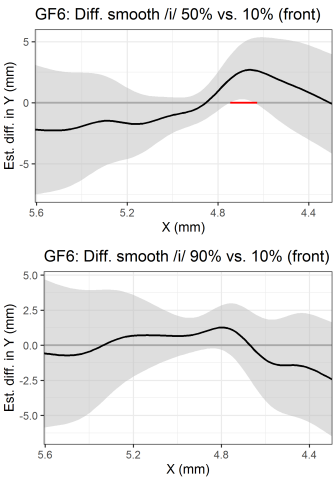
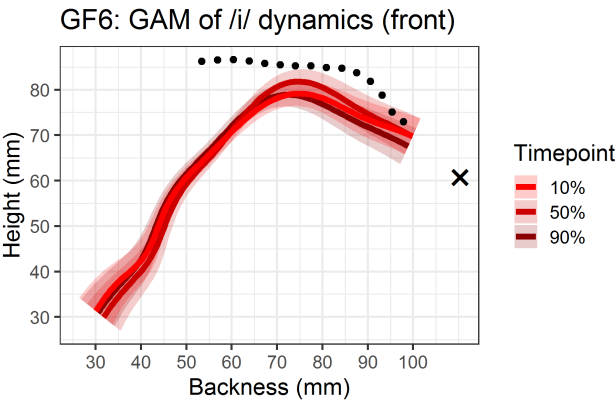
GAMM trajectories for individual speakers' tongue gestures for /i:/ at the 10%, 50%, and 90% timepoint are provided here. Difference smooths for comparisons between 10-50% and 10-90% are positioned beside each figure. The sample excludes GF5, SF4, SM1, SM3, UF3, UF4, UM1, and UM3 due to missing data.

Gothenburg females: front

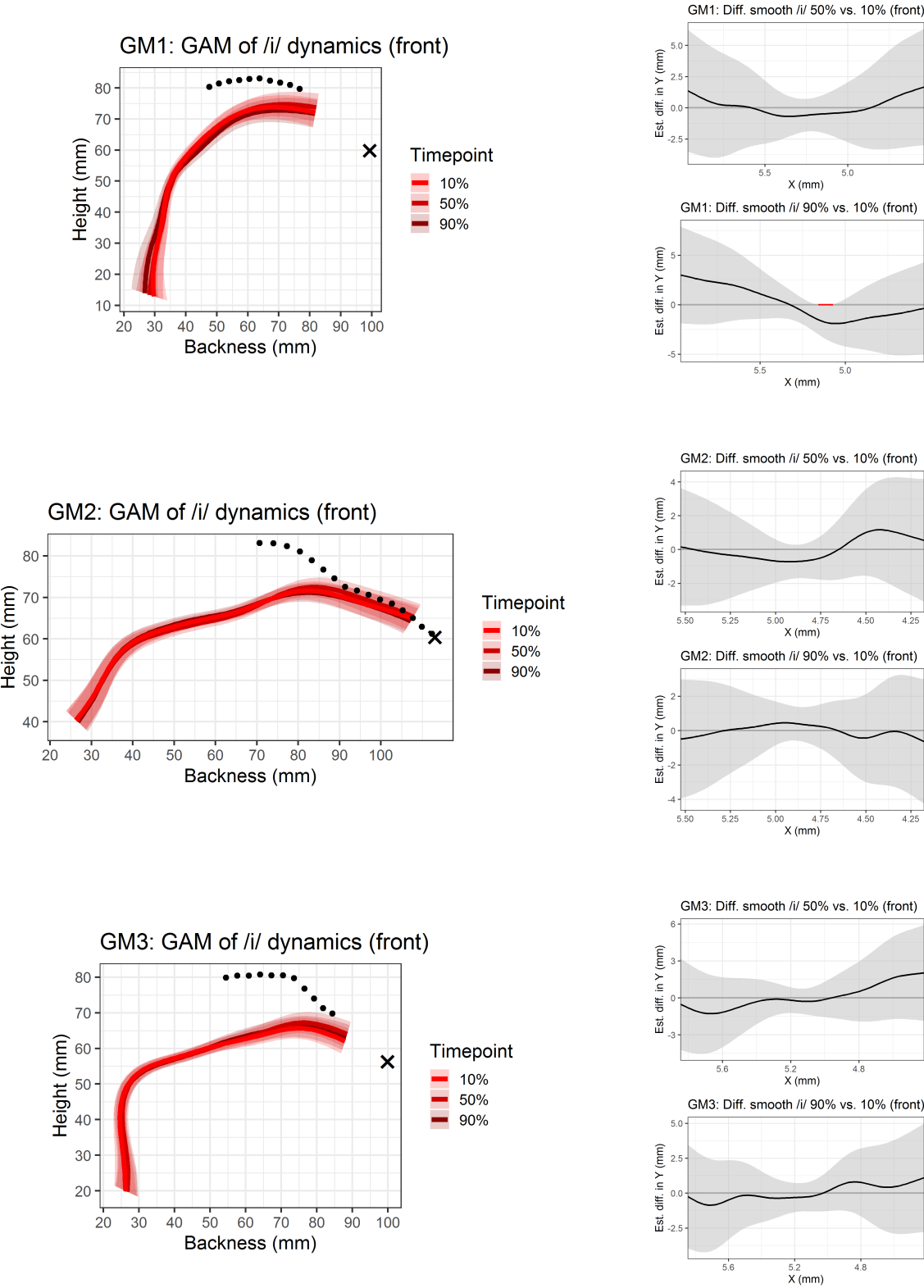


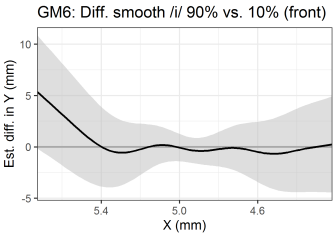
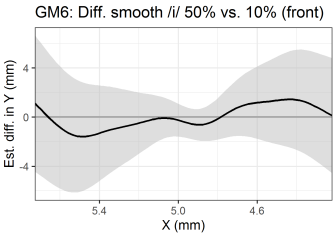
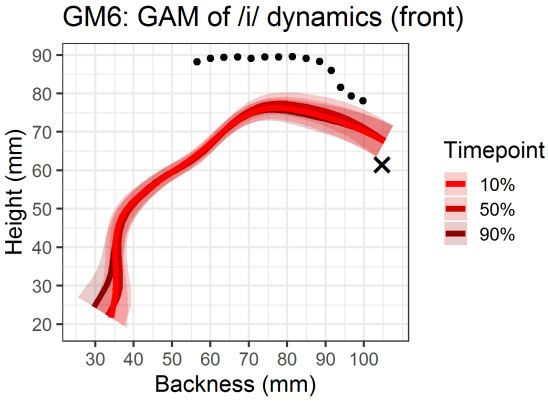
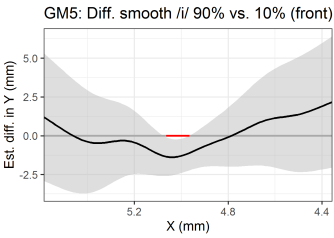
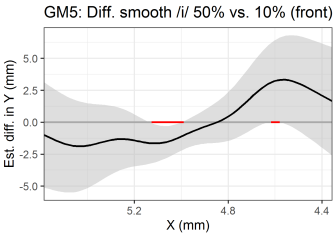
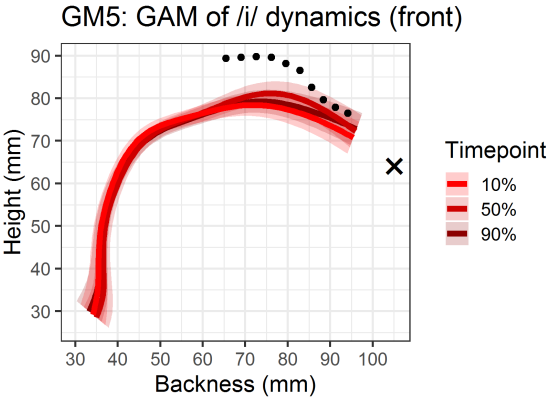
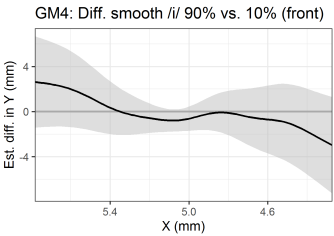
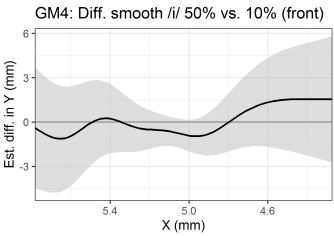
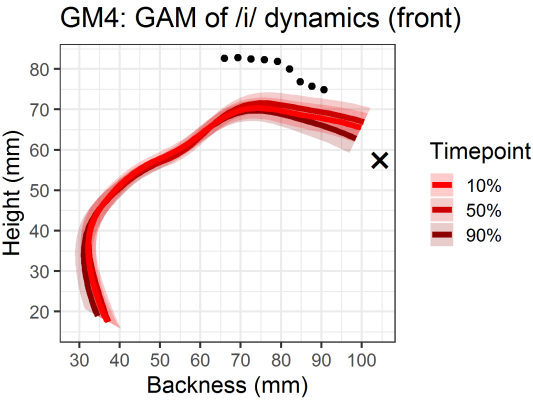


GF5: No GAMM data



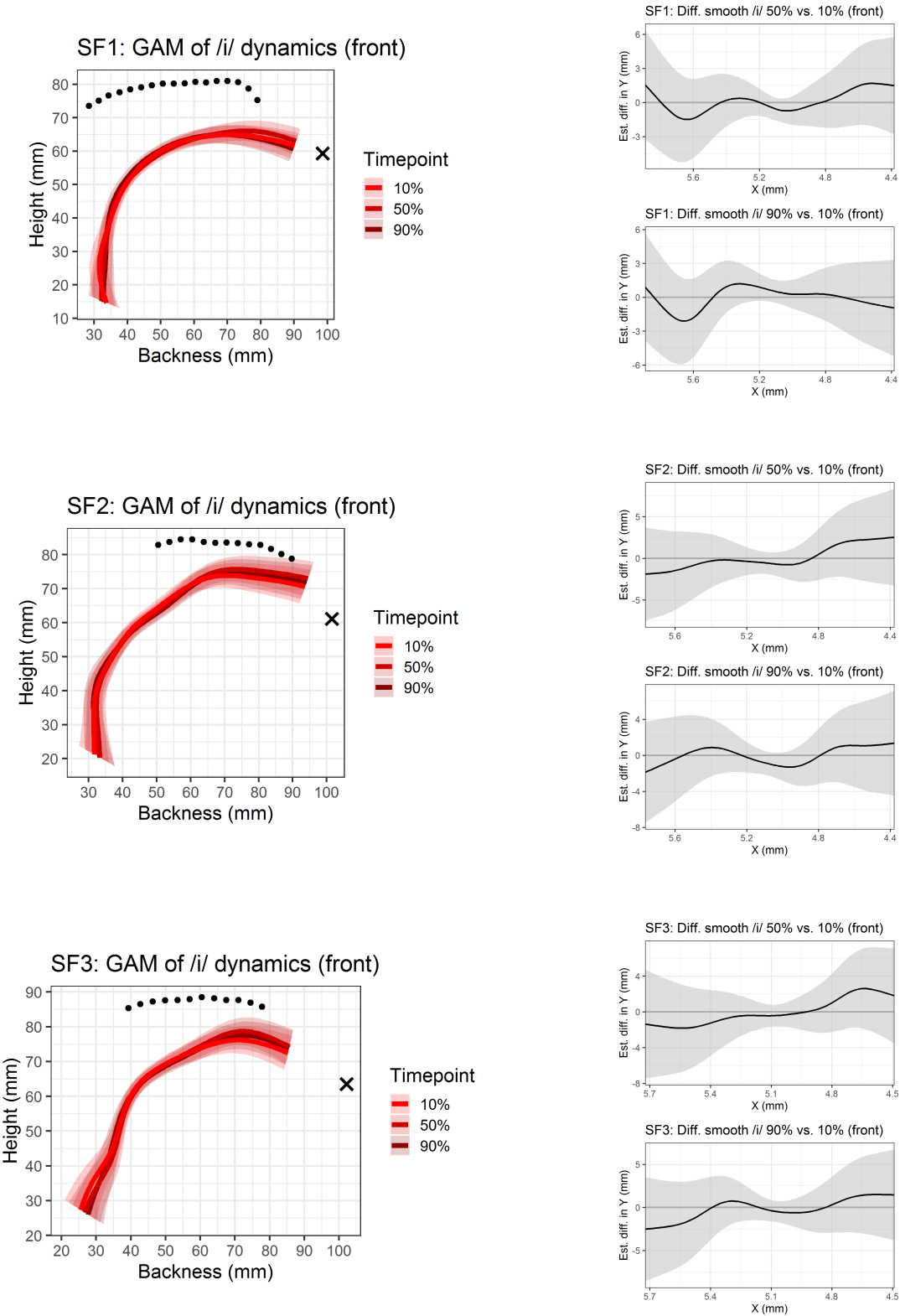
Gothenburg males: front



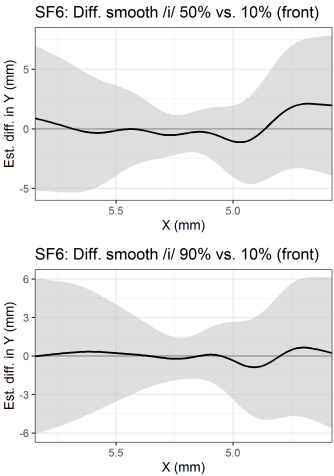
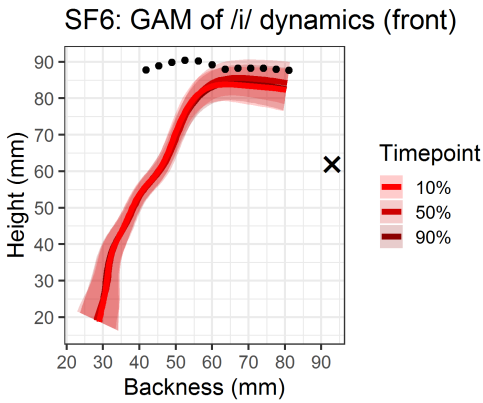
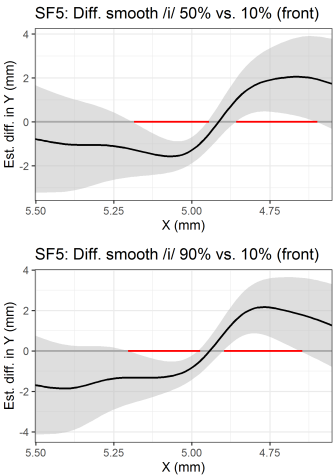
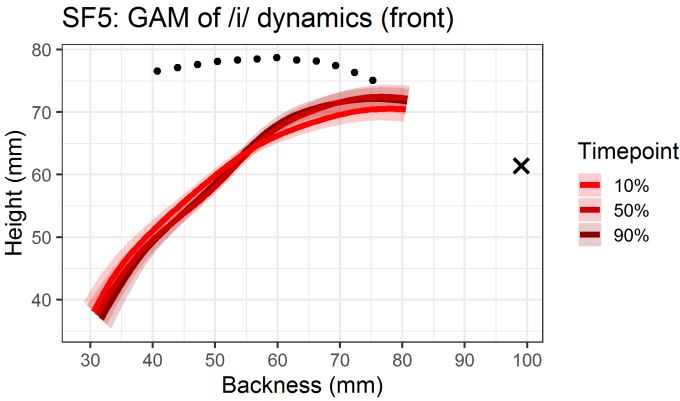




Stockholm females: front

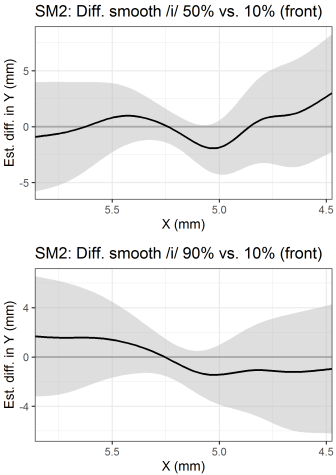
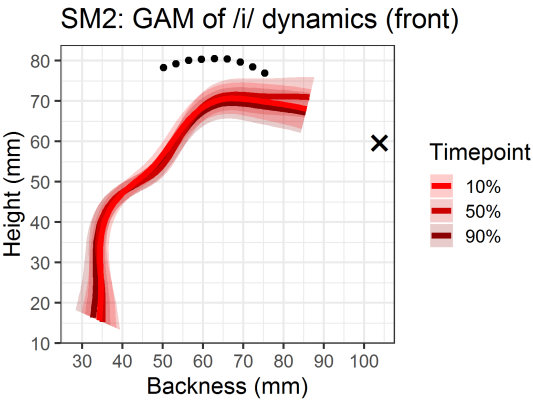


SF4: No GAMM data

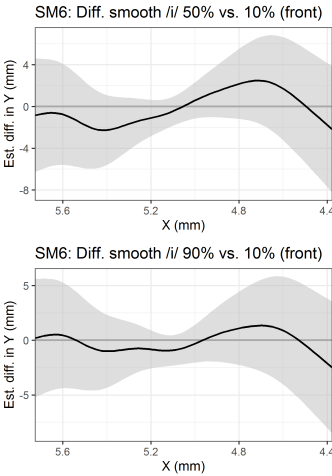
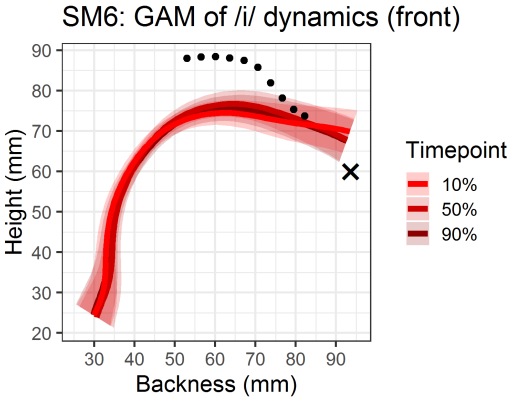
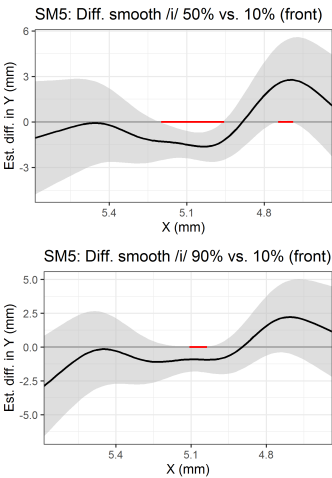
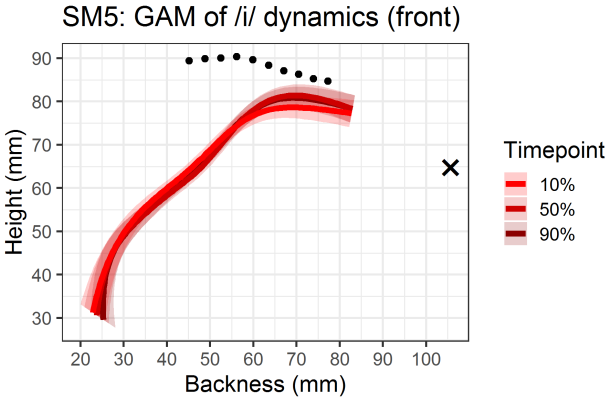
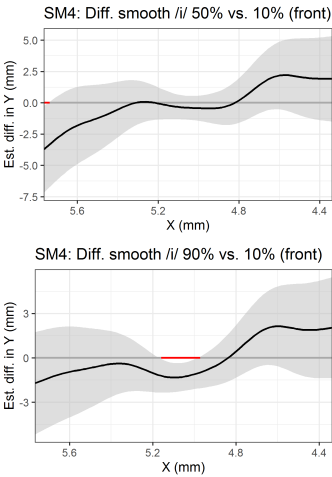
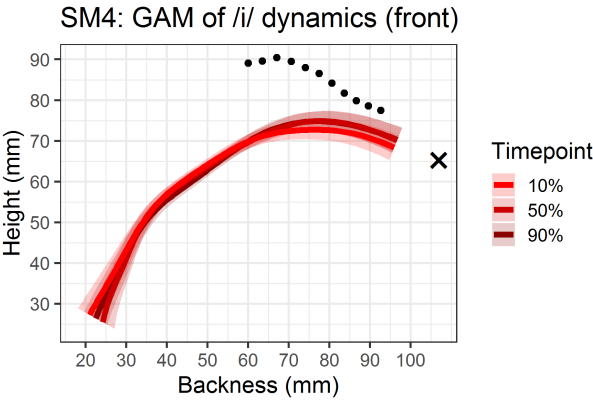


Stockholm males: front

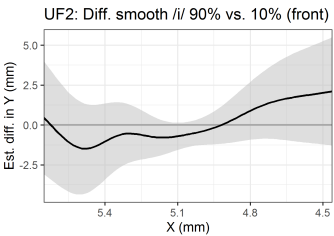
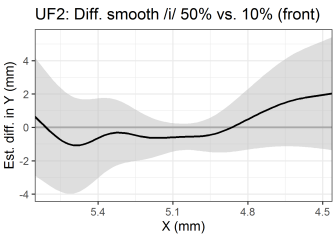
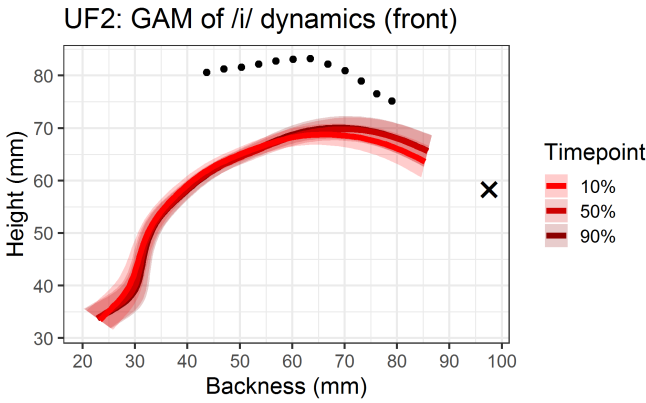
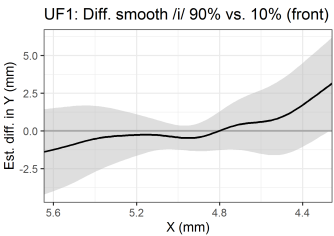
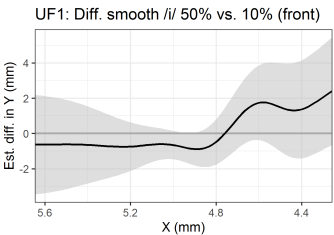
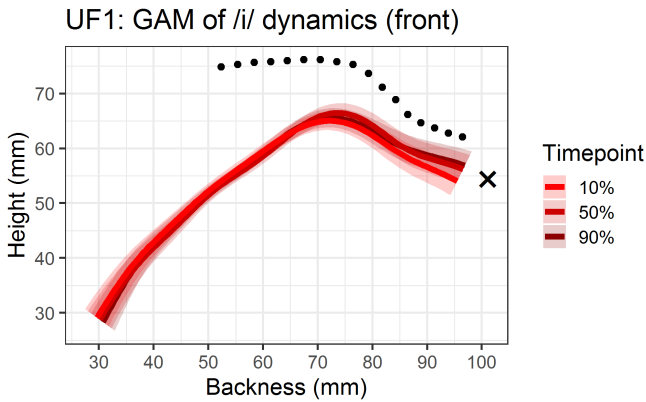
SM1: No GAMM data



SM3: No GAMM data

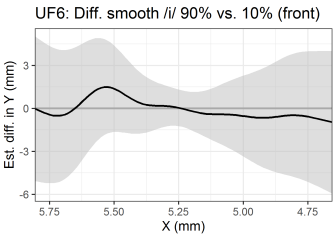
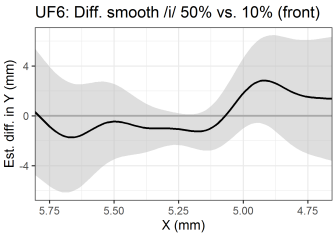
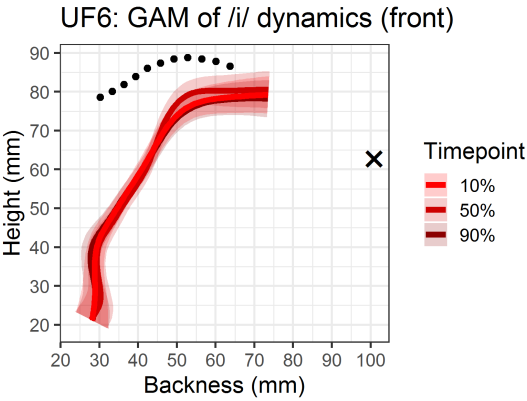
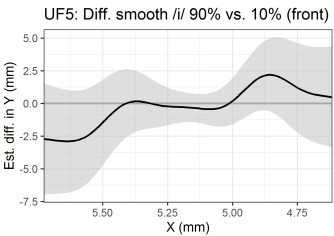
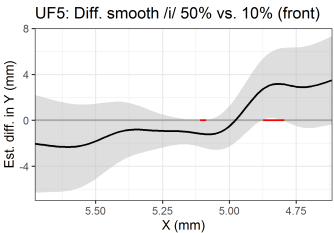
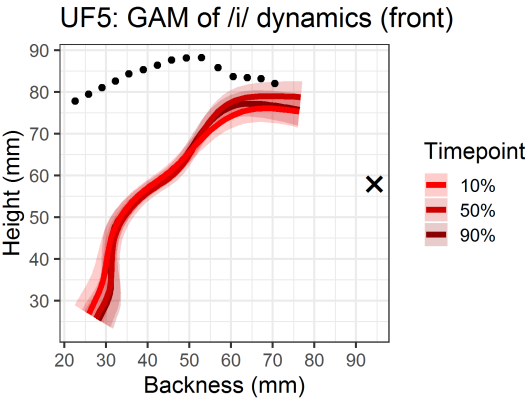


Uppsala females: front



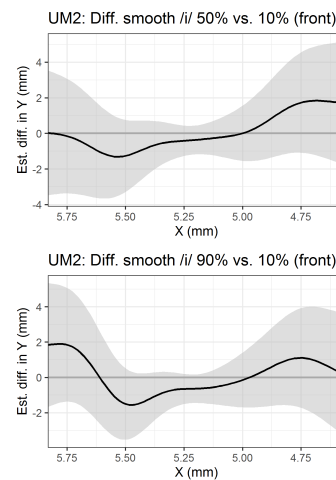
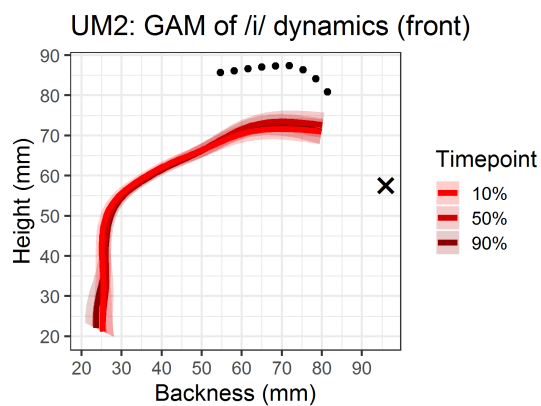
UF3: No GAMM data

UF4: No GAMM data

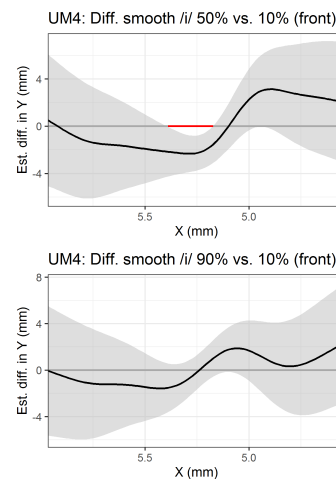
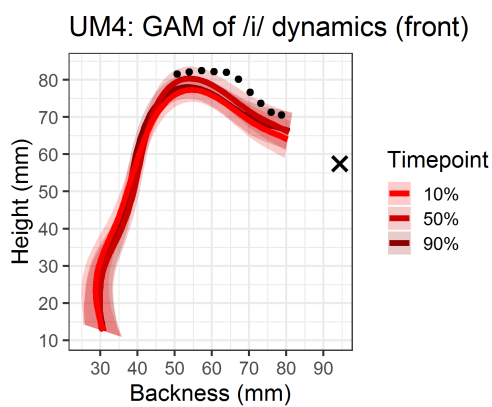


## Uppsala males: front

UM1: No GAMM data

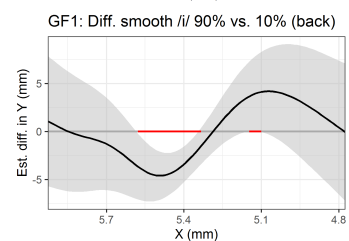
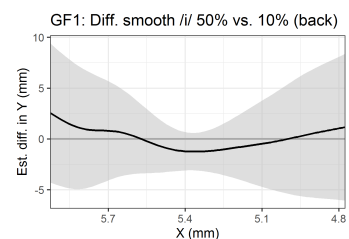
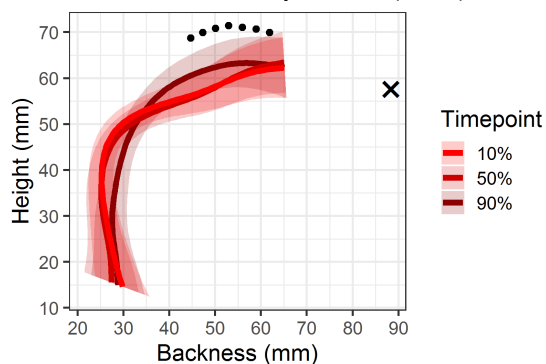


UM3: No GAMM data

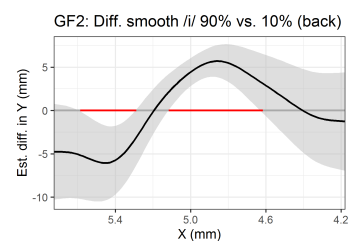
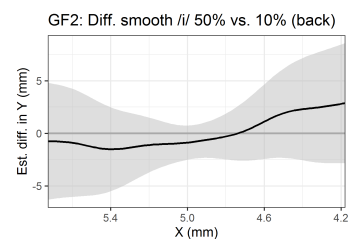
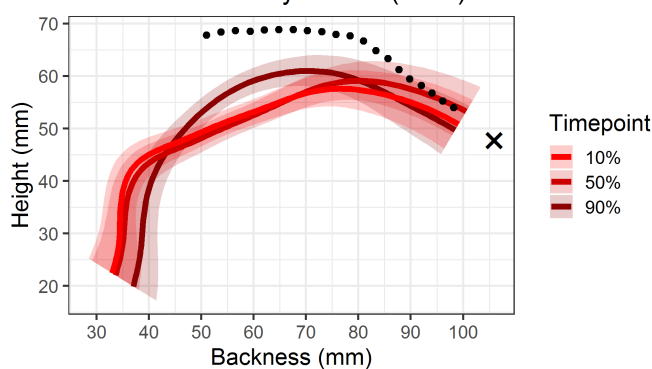


## Gothenburg females: back

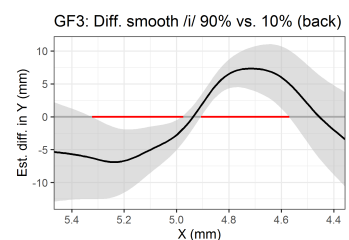
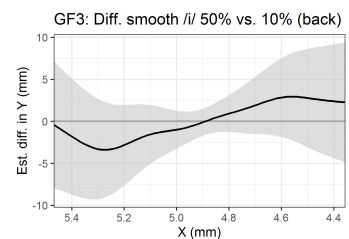
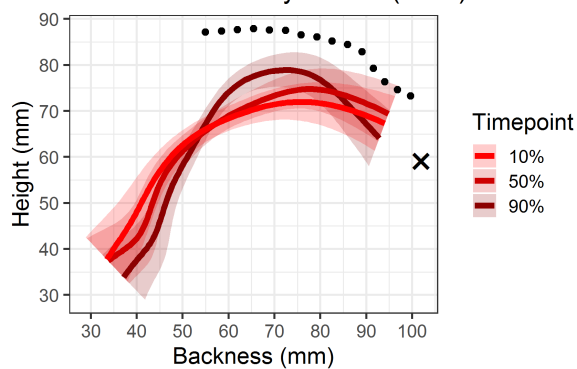
GF1: GAM of /i/ dynamics (back)



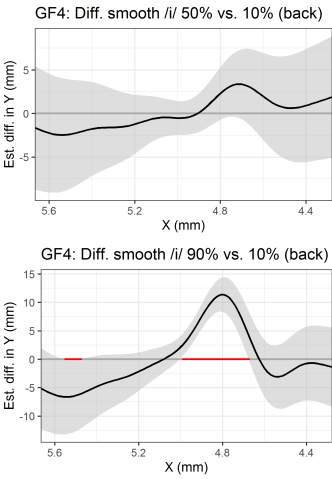
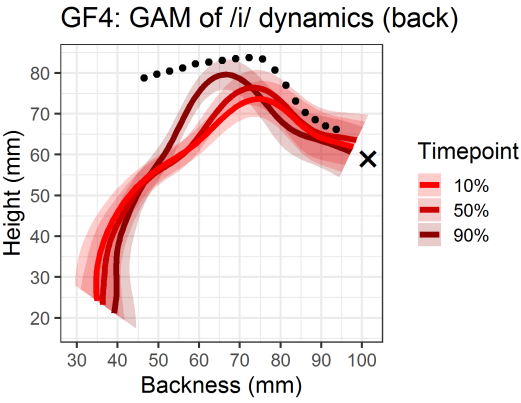
GF2: GAM of /i/ dynamics (back)



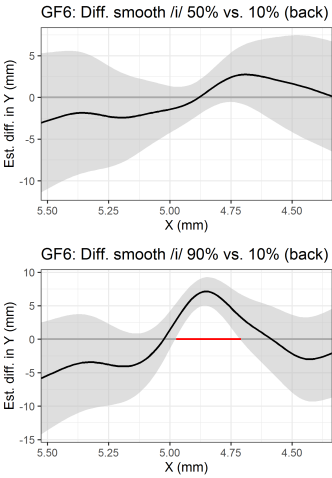
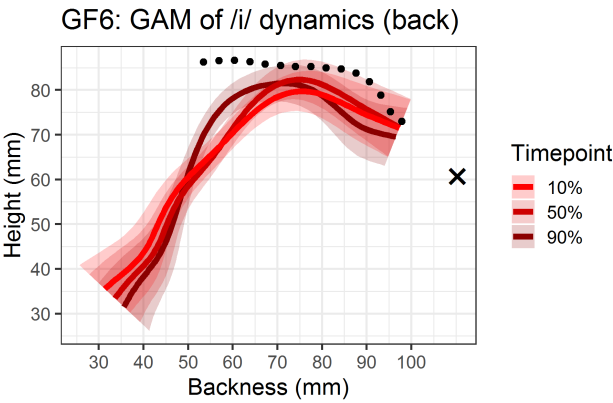
GF3: GAM of /i/ dynamics (back)



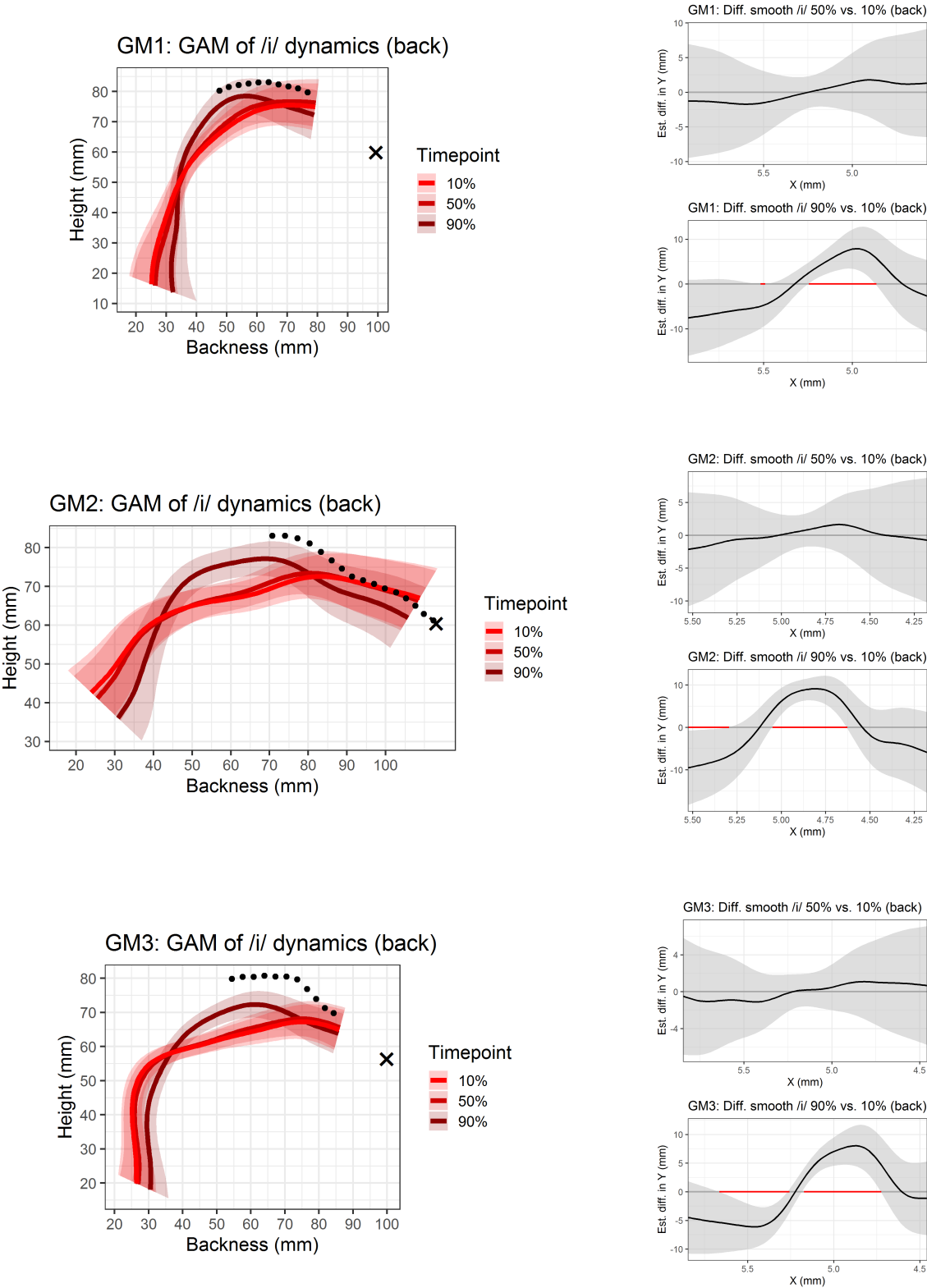


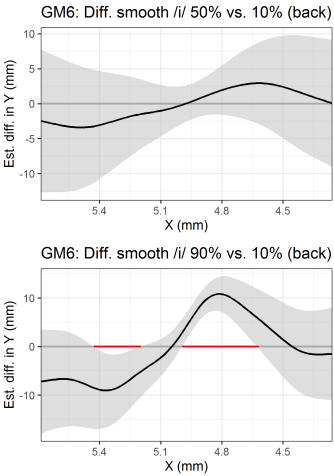
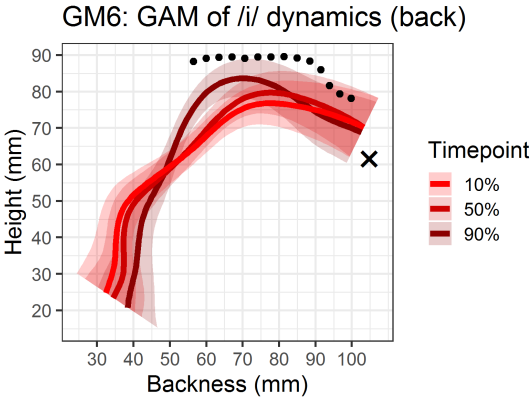
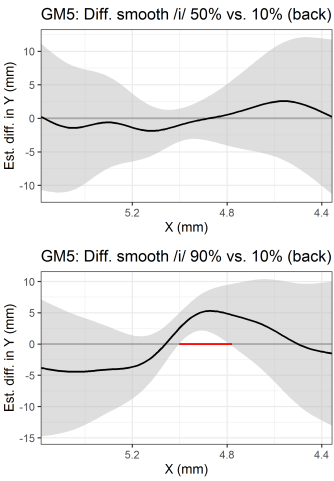
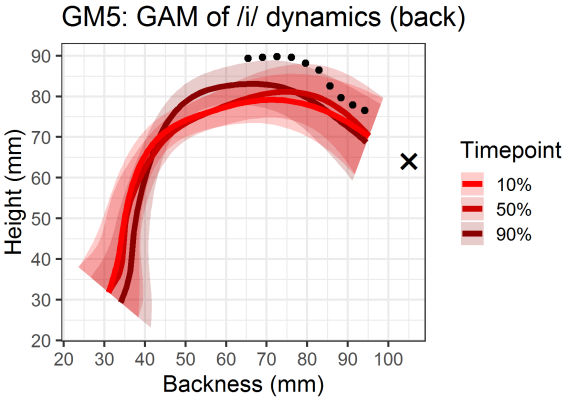
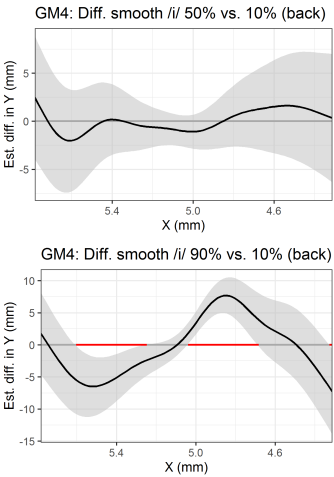
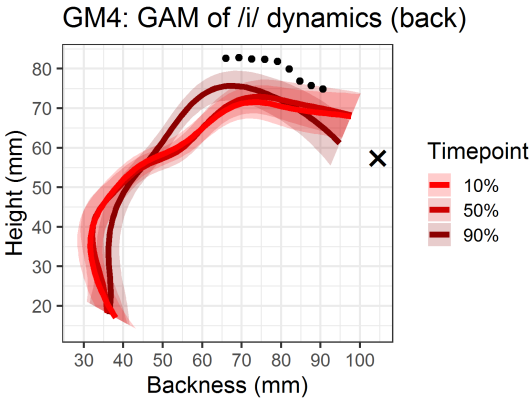


GF5: No GAMM data

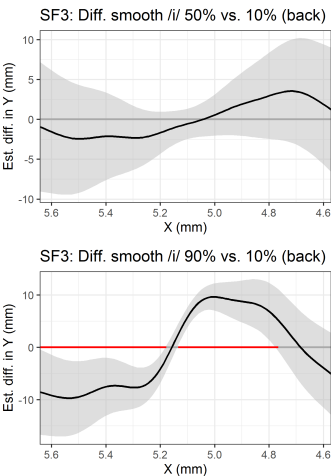
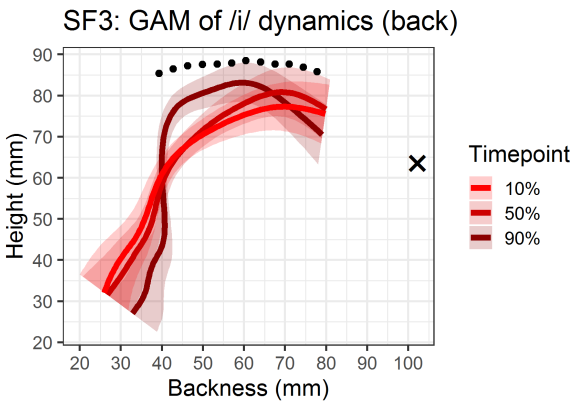
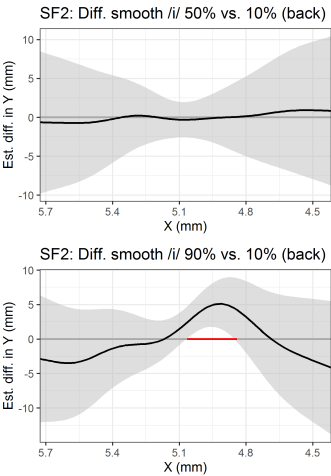
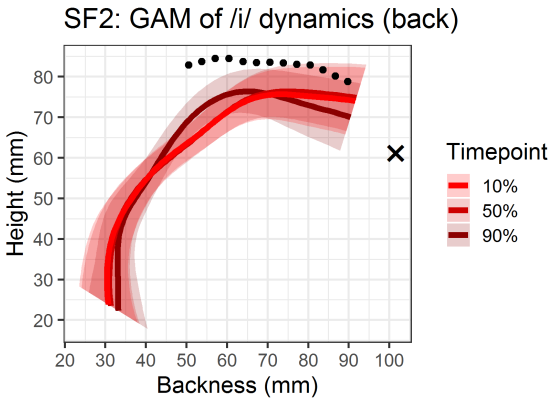
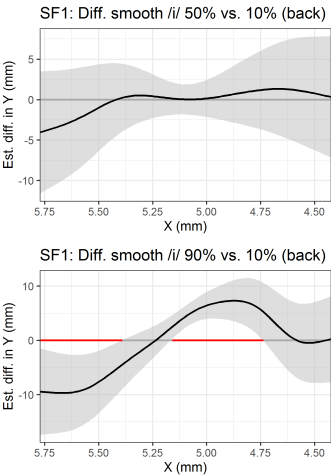
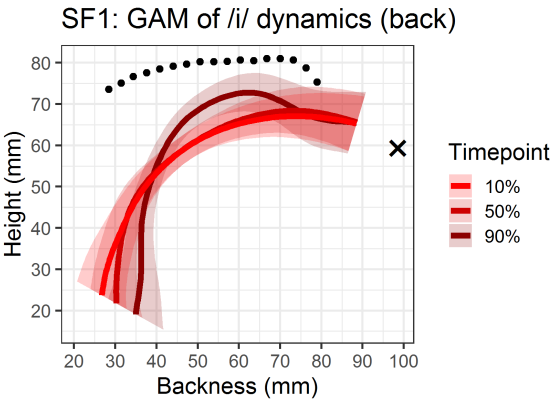


Gothenburg males: back

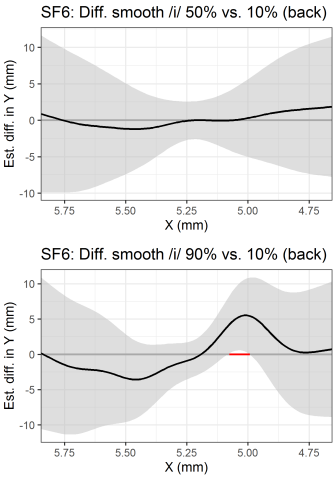
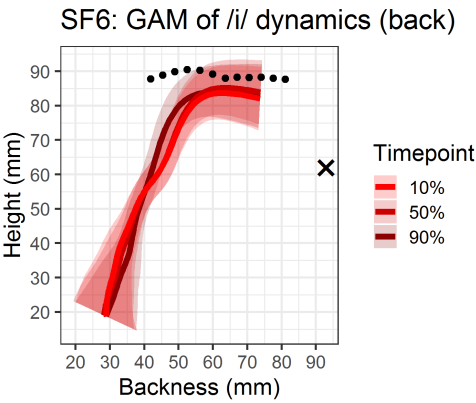
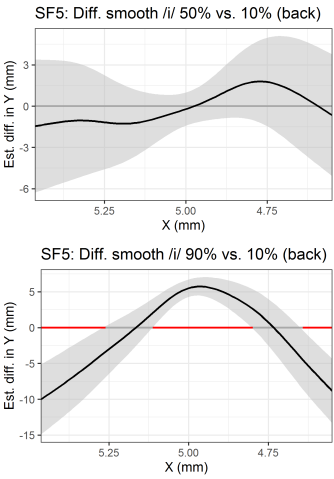
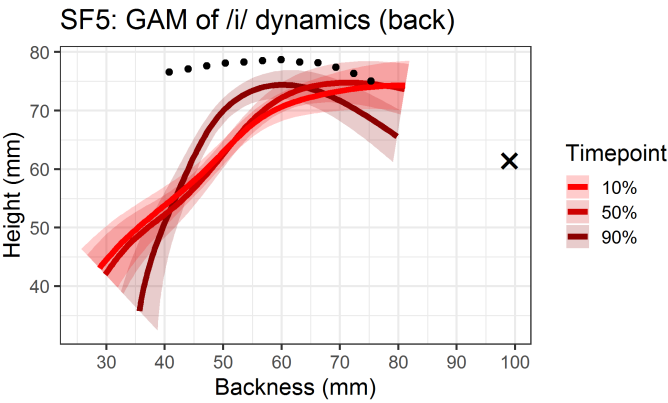




Stockholm females: back

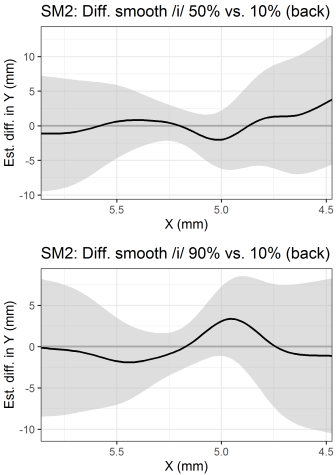
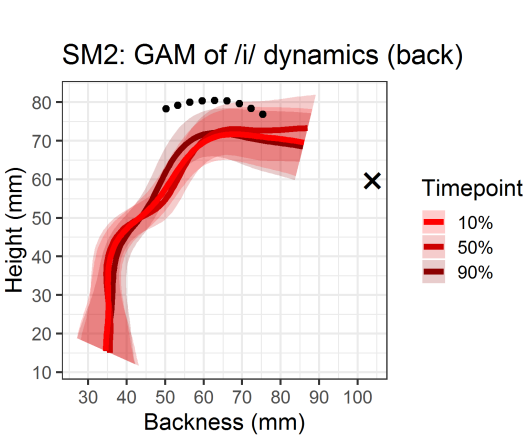


SF4: No GAMM data

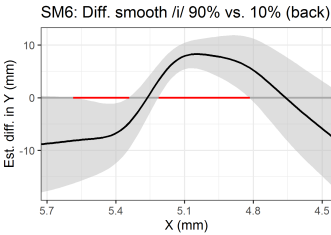
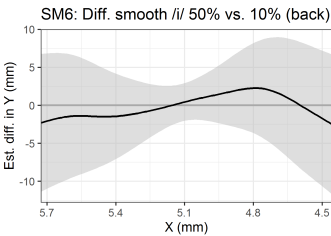
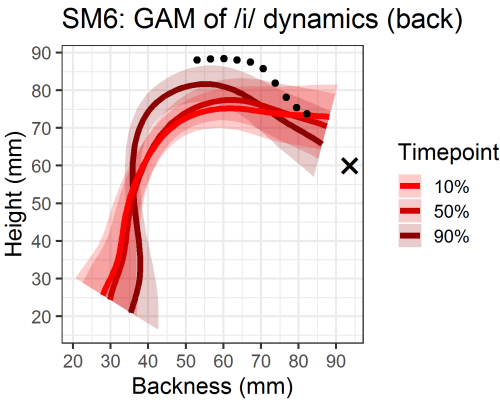
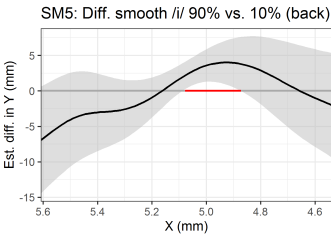
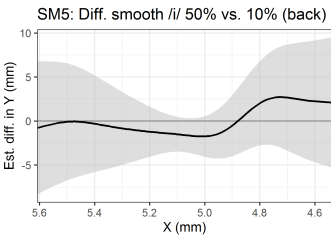
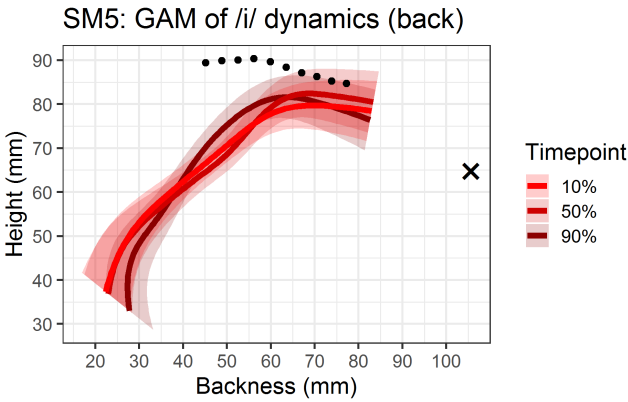
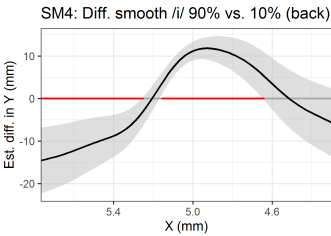
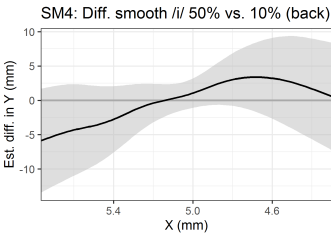
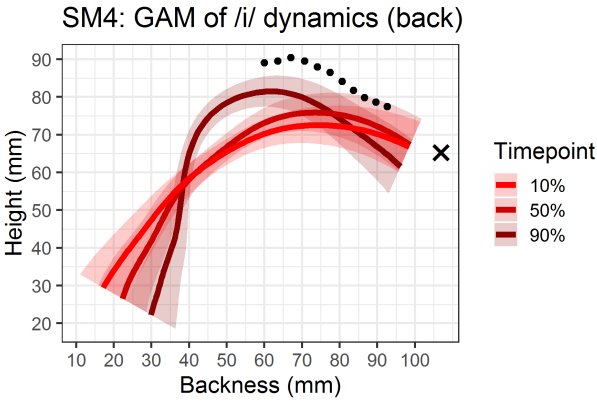


Stockholm males: back

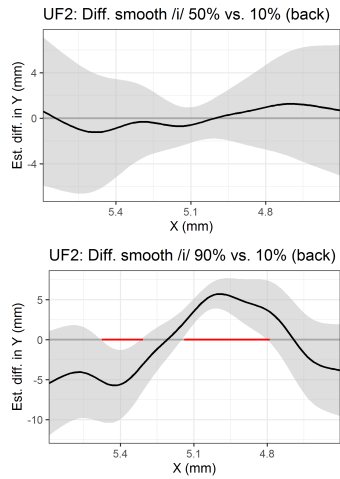
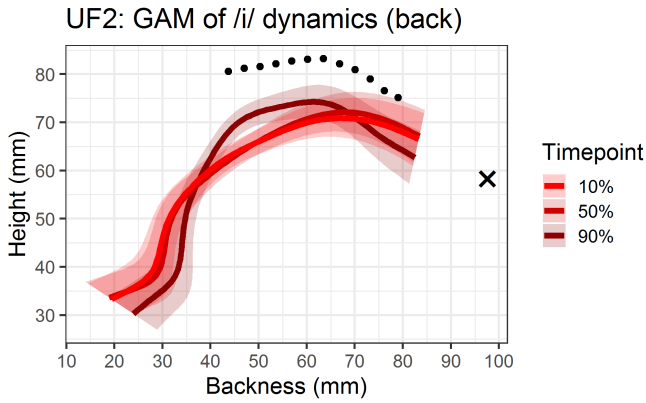
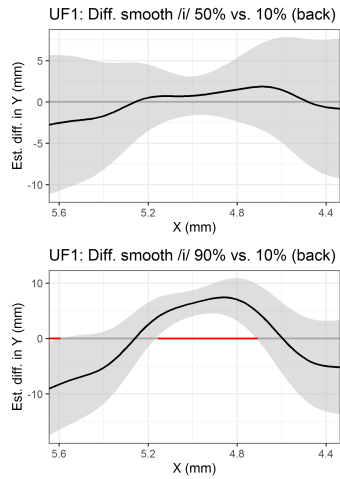
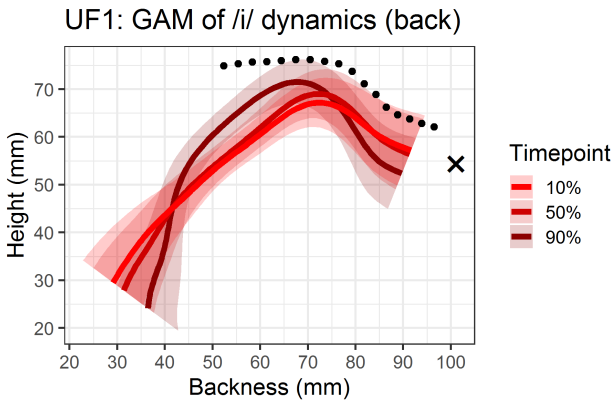
SM1: No GAMM data



SM3: No GAMM data



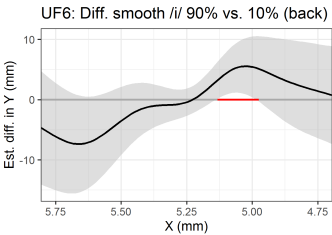
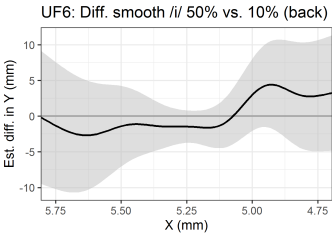
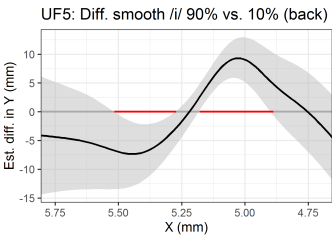
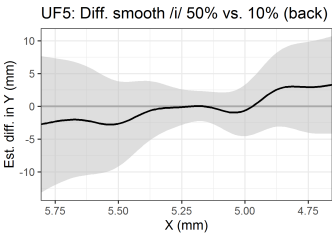
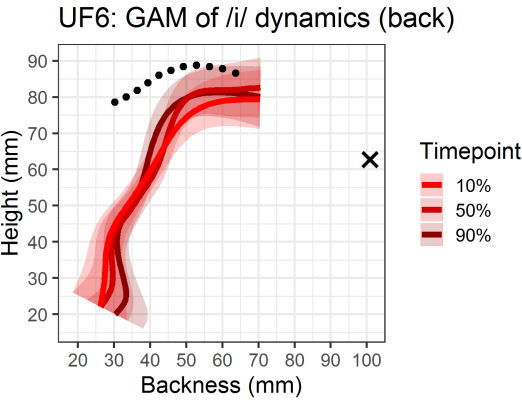
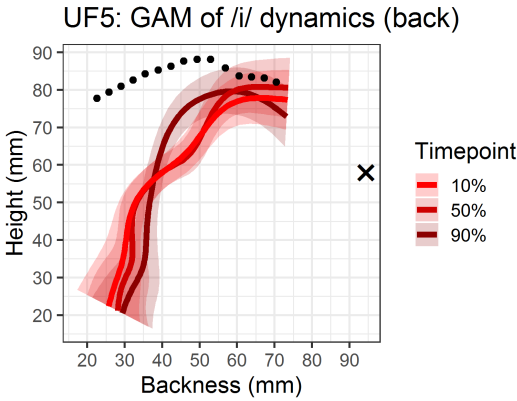
Uppsala females: back



UF3: No GAMM data

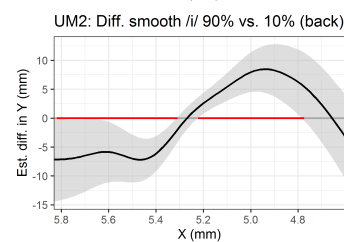
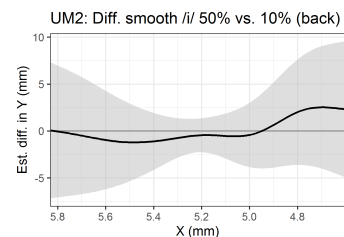
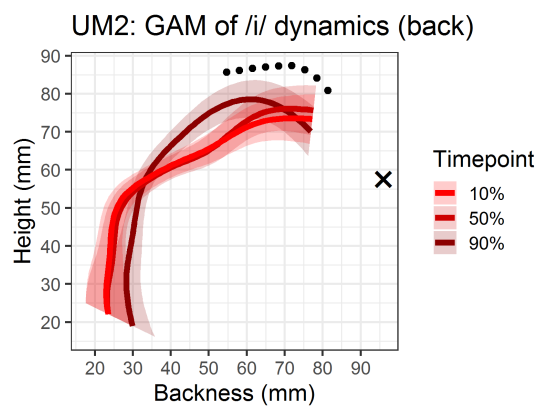


UF4: No GAMM data

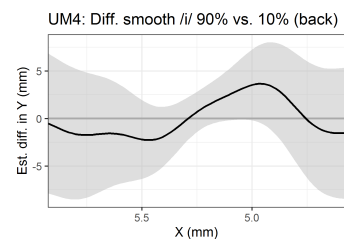
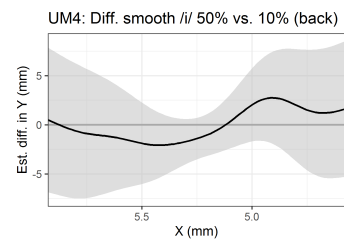
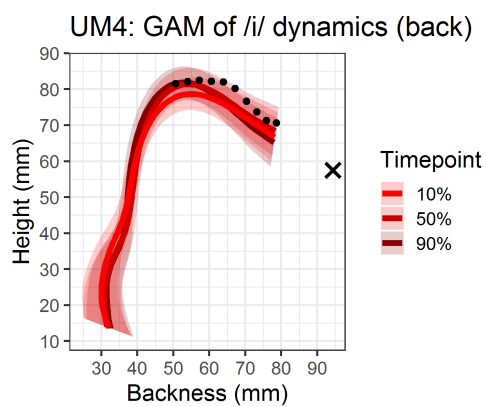


## Uppsala males: back

UM1: No GAMM data



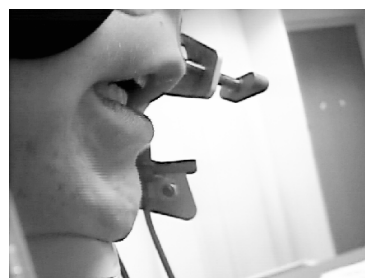
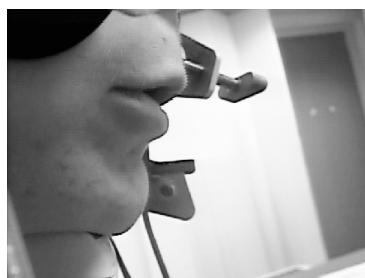
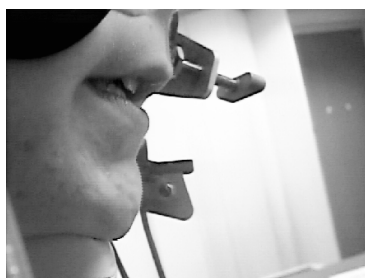
UM3: No GAMM data



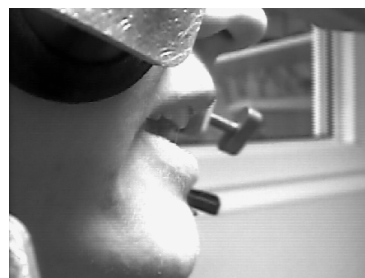
# Appendix K

## Lip images

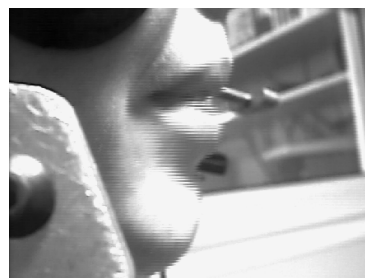
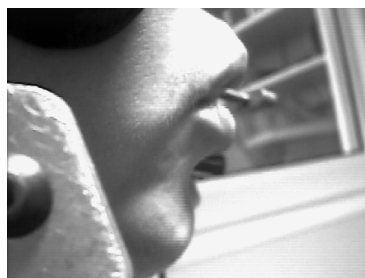
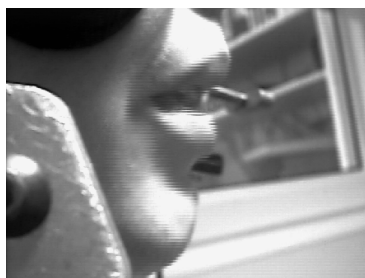
Lip images for individual speakers' productions of /i:, y:, e:/ are provided below. Speaker SF3 is excluded due to missing data.



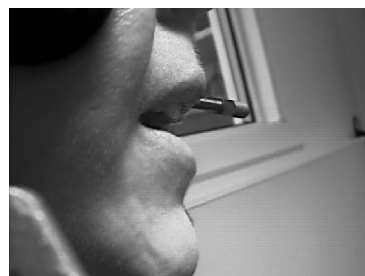
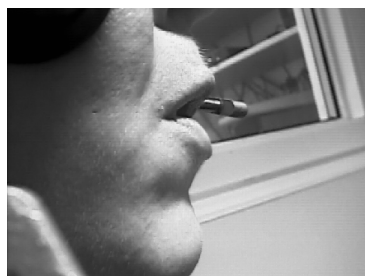
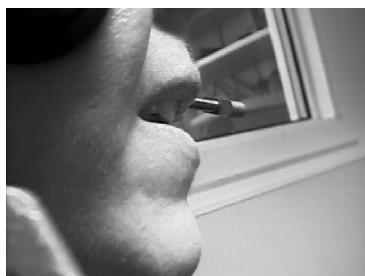
Speaker GF1: /i:, y:, e:/



Speaker GF2: /i:, y:, e:/



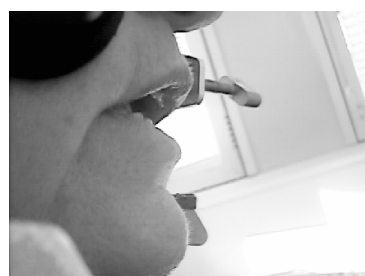
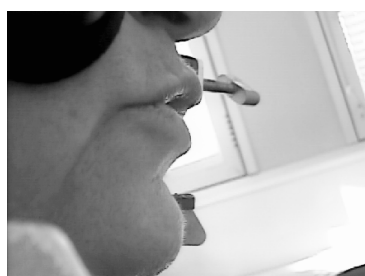
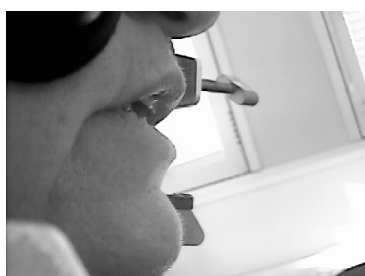
Speaker GF3: /i:, y:, e:/



Speaker GF4: /i:, y:, e:/



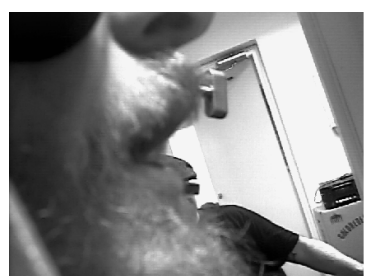
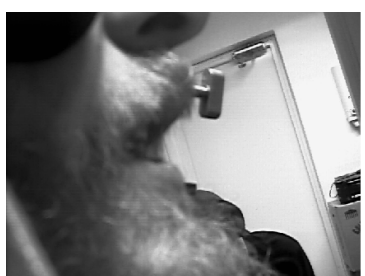
Speaker GF5: /i:, y:, e:/



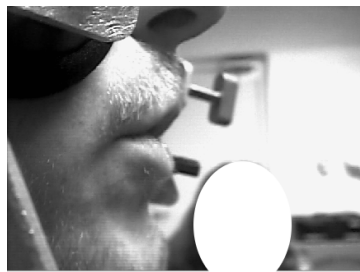
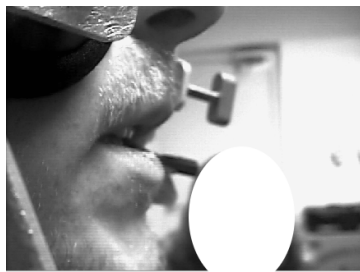
Speaker GF6: /i:, y:, e:/



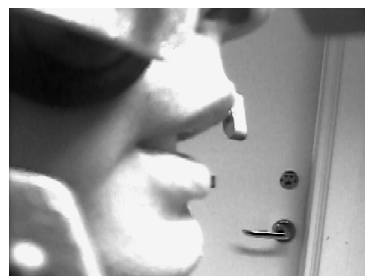
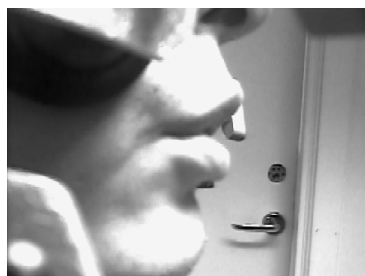
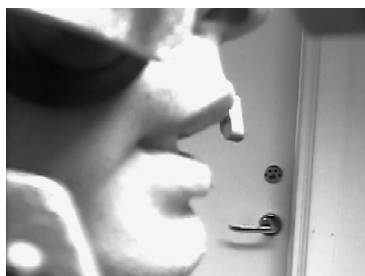
Speaker GM1: /i:, y:, e:/



Speaker GM2: /i:, y:, e:/



Speaker GM3: /i:, y:, e:/



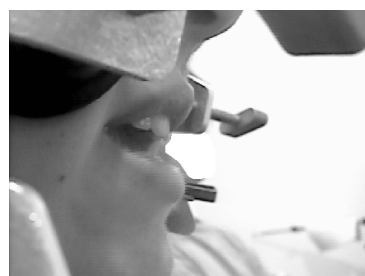
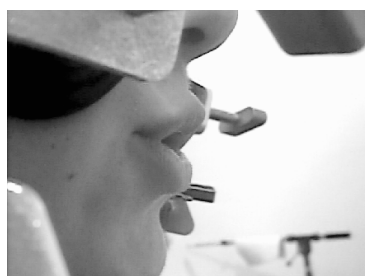
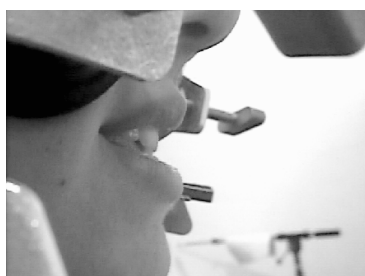
Speaker GM4: /i:, y:, e:/



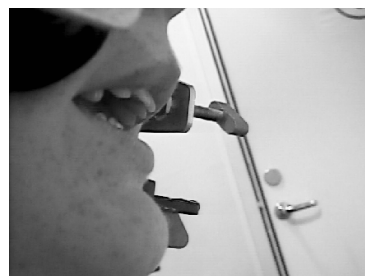
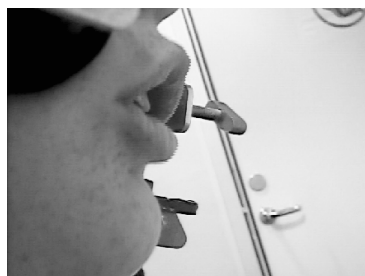
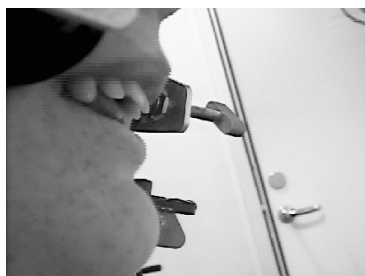
Speaker GM5: /i:, y:, e:/



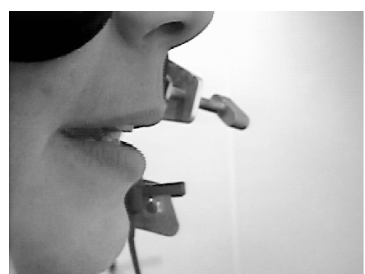
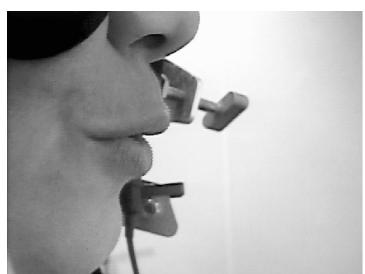
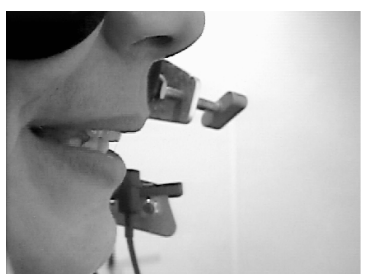
Speaker GM6: /i:, y:, e:/



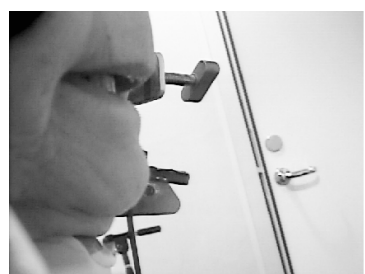
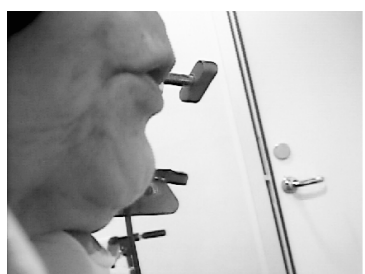
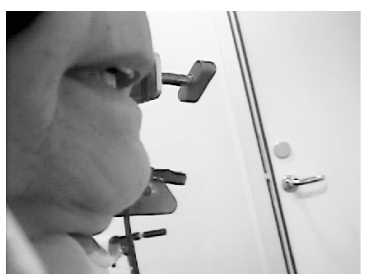
Speaker SF1: /i:, y:, e:/



Speaker SF2: /i:, y:, e:/



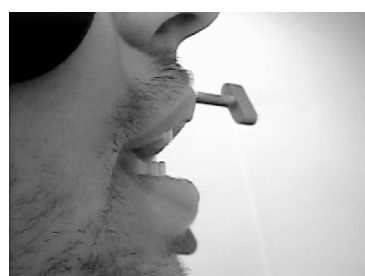
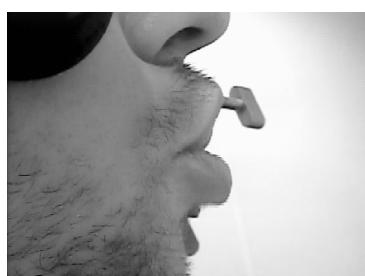
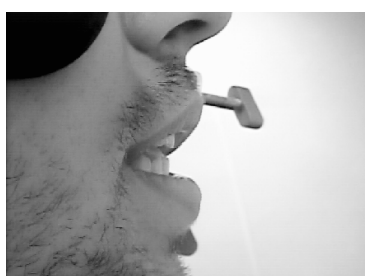
Speaker SF4: /i:, y:, e:/



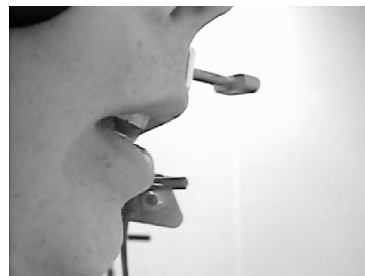
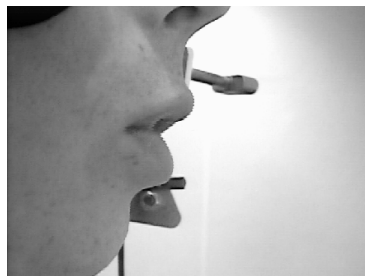
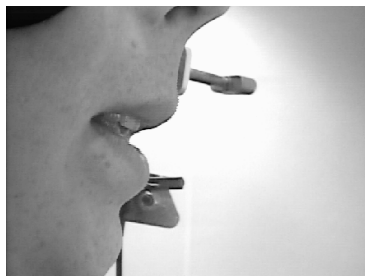
Speaker SF5: /i:, y:, e:/



Speaker SF6: /i:, y:, e:/



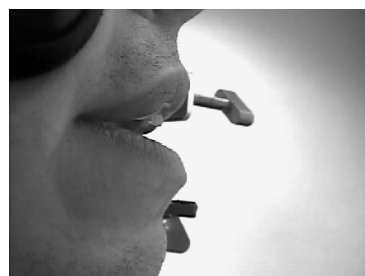
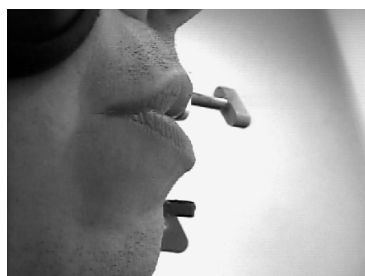
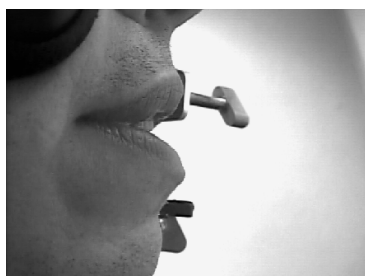
Speaker SM1: /i:, y:, e:/



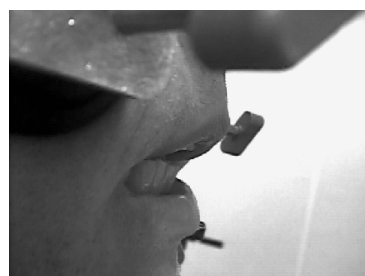
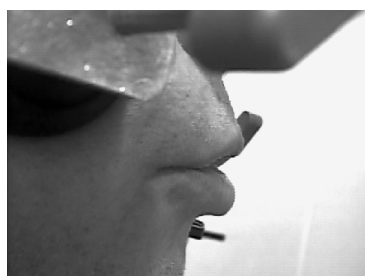
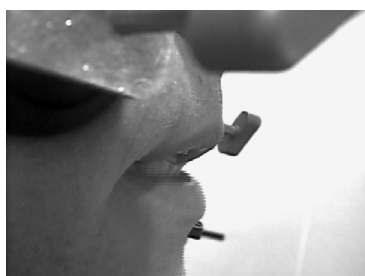
Speaker SM2: /i:, y:, e:/



Speaker SM3: /i:, y:, e:/



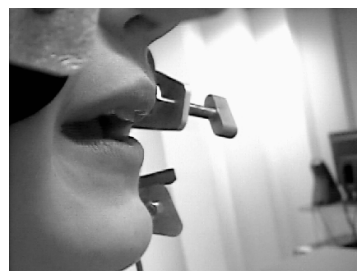
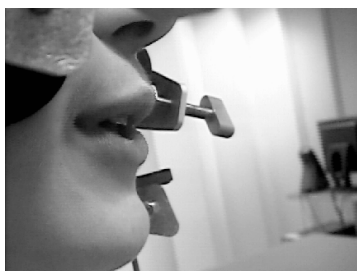
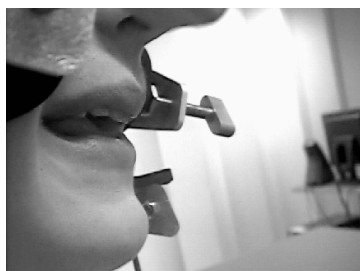
Speaker SM4: /i:, y:, e:/



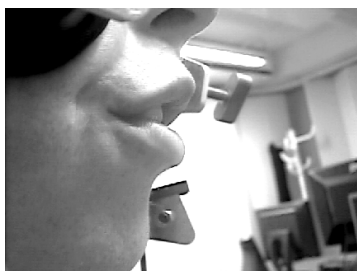
Speaker SM5: /i:, y:, e:/



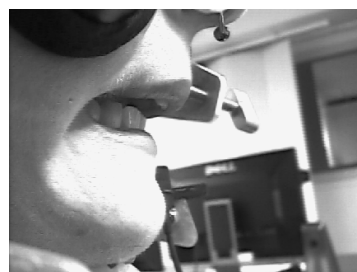
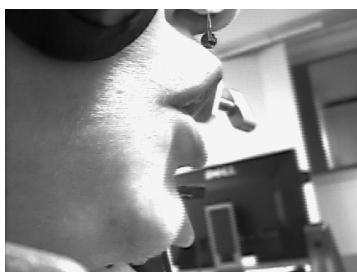
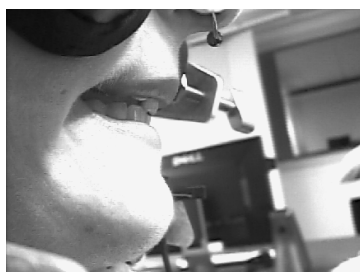
Speaker SM6: /i:, y:, e:/



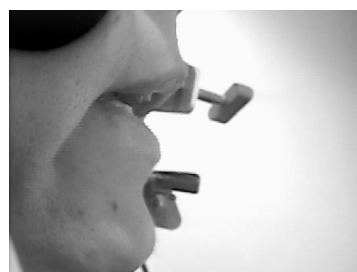
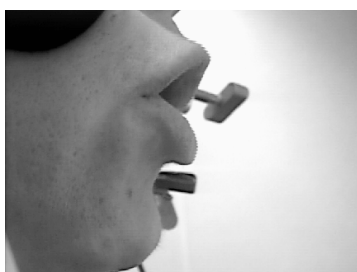
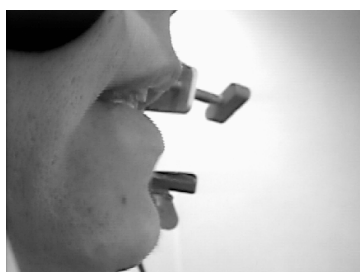
Speaker UF1: /i:, y:, e:/



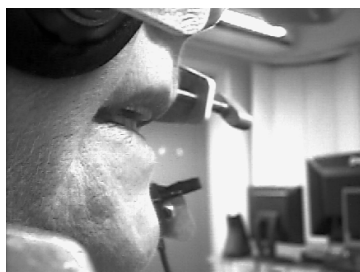
Speaker UF2: /i:, y:, e:/



Speaker UF3: /i:, y:, e:/

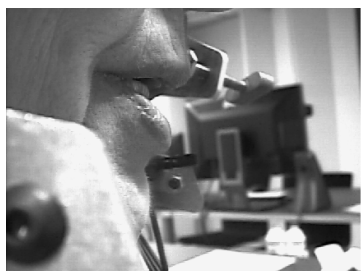


Speaker UF4: /i:, y:, e:/

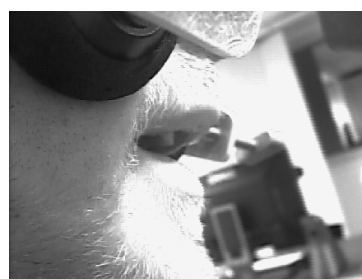
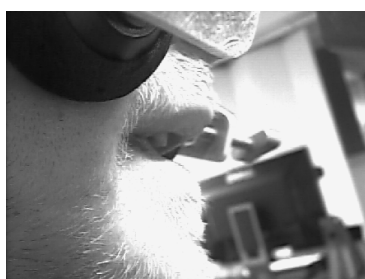


Speaker UF5: /i:, y:, e:/





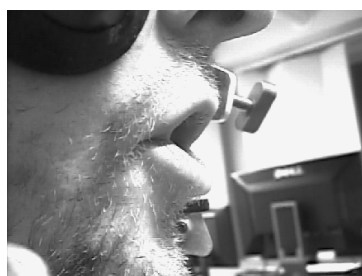
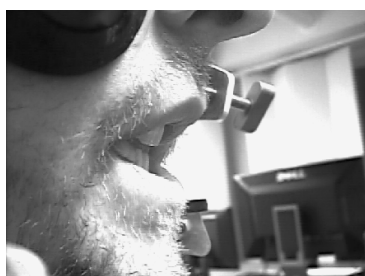
Speaker UF6: /i:, y:, e:/



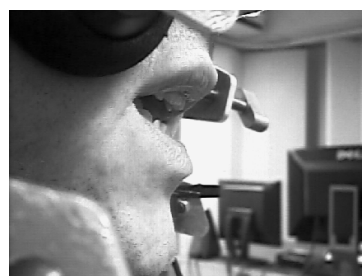
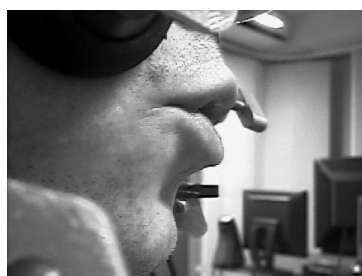
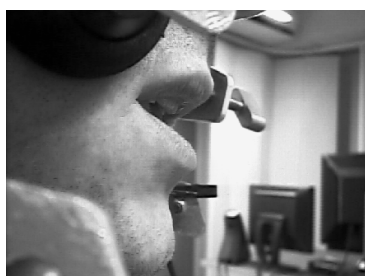
Speaker UM1: /i:, y:, e:/



Speaker UM2: /i:, y:, e:/



Speaker UM3: /i:, y:, e:/



Speaker UM4: /i:, y:, e:/