An auditory, acoustic, articulatory and sociophonetic study of Swedish *Viby-i*

Fabienne E. Westerberg

Submitted in fulfilment of the requirements for the Degree of MPhil

School of Critical Studies
College of Arts
University of Glasgow

June 2016

Abstract

The study investigates the acoustic, articulatory and sociophonetic properties of the Swedish /iː/ variant known as 'Viby-i' in 13 speakers of Central Swedish from Stockholm, Gothenburg, Varberg, Jönköping and Katrineholm. The vowel is described in terms of its auditory quality, its acoustic F1 and F2 values, and its tongue configuration. A brief, qualitative description of lip position is also included. Variation in /iː/ production is mapped against five sociolinguistic factors: city, dialectal region, metropolitan vs. urban location, sex and socioeconomic rating. Articulatory data is collected using ultrasound tongue imaging (UTI), for which the study proposes and evaluates a methodology.

The study shows that Viby-i varies in auditory strength between speakers, and that strong instances of the vowel are associated with a high F1 and low F2, a trend which becomes more pronounced as the strength of Viby-i increases. The articulation of Viby-i is characterised by a lowered and backed tongue body, sometimes accompanied by a double-bunched tongue shape. The relationship between tongue position and acoustic results appears to be non-linear, suggesting either a measurement error or the influence of additional articulatory factors. Preliminary images of the lips show that Viby-i is produced with a spread but lax lip posture. The lip data also reveals parts of the tongue, which in many speakers appears to be extremely fronted and braced against the lower teeth, or sometimes protruded, when producing Viby-i. No sociophonetic difference is found between speakers from different cities or dialect regions. Metropolitan speakers are found to have an auditorily and acoustically stronger Viby-i than urban speakers, but this pattern is not matched in tongue backing or lowering. Overall the data shows a weak trend towards higher-class females having stronger Viby-i, but these results are tentative due to the limited size and stratification of the sample. Further research is needed to fully explore the sociophonetic properties of Viby-i.

Table of contents

1. Introduction

This study explores the auditory, acoustic, articulatory and sociophonetic properties of *Viby-i*, a Swedish /iː/ variant which is often remarked upon for its unusual vowel quality. Viby-i has impressionistically been described as “thick”, “dark”, “buzzing” or “damped” (Engstrand et al 1998: 83-84), but although its existence is well-documented, the sound has not been extensively studied. The main controversy around this vowel concerns its articulation, an issue which has given rise to many contending theories, but has failed to be resolved due to a lack of articulatory data. Acoustic studies have found Viby-i to be associated with a low F2 (ibid: 87), suggesting tongue backing, but there appear to be several strategies for producing this vowel (Björsten & Engstrand 1999). It has been suggested that Viby-i is subject to regional or social variation, as the vowel occurs in several isolated pockets across Sweden, and seems to have different social functions in rural versus urban areas (Bruce 2010: 125). However it is also possible that there are different types of Viby-i, and that descriptions of the vowel vary for this reason. The purpose of this study is therefore to provide a description of Viby-i using auditory, acoustic and articulatory data, and to investigate whether geographic and social factors influence who uses this vowel, and how it is produced.

1.1. Aim and research strategy

The study aims to fill a gap in the literature, not only in terms of Viby-i as a component of Swedish phonology, but of vowel articulation in general. This is an under-explored field which is currently of growing interest within the phonetic community, as practical and affordable articulatory methods are becoming more accessible. In this study, articulatory data is collected using ultrasound tongue imaging (UTI), a relatively new technique which images the movements of the tongue surface during speech. Methods for collecting and analysing UTI data are still under development; thus the study seeks not only to provide data which can illuminate aspects of vowel articulation, but also to test and evaluate methods of handling such data. Finally, the study wishes to present articulatory, acoustic and auditory data together in order to provide more detailed insight into the production of this vowel, which may benefit our understanding of vowel production more broadly.

1.2. Research questions

The research questions which will be addressed are:

1. What are the auditory properties of Viby-i?
   → Is there a binary distinction from [iː], or is Viby-i gradient?
2. How is Viby-i defined acoustically?'
   → What does it look like in terms of F1/F2?
   → How does it relate to other vowels in the system?
   → Do its acoustic properties agree with its auditory properties?
3. How is Viby-i articulated with reference to the tongue and lips?
   → Is it characterised by a certain tongue height, advancement, or shape?
   → What is the role of the lips in Viby-i production?
   → Do the articulatory properties correspond to expectations created by F1/F2?
4. Do regional or social factors influence /iː/ production?
   → Do these factors determine whether or not speakers use Viby-i?
   → Do the auditory, acoustic or articulatory characteristics of Viby-i vary between different groups?

The working definition of Viby-i is an /iː/ vowel which differs in auditory quality from the standard [iː] used in other languages, such as English. All instances of /iː/ (Viby-i or standard) will be contextualised by the four long 'corner' vowels of the Swedish vowel system: /eː/, /ɑː/, /uː/ and /ʉː/.

1.3. Structure of thesis

The thesis will begin with a theoretical background chapter consisting of two main sections: the first covering relevant concepts and issues relating to phonetic vowel theory, and the second providing a linguistic and sociolinguistic background to Swedish, including a detailed subsection on Viby-i. These sections will be followed by a chapter describing the method used for data collection and analysis in the study, after which the results will be presented. The discussion chapter will then summarise and critically evaluate the results in relation to the research questions. The conclusion will summarise the main findings of the study and point towards future work.

2. Theoretical background

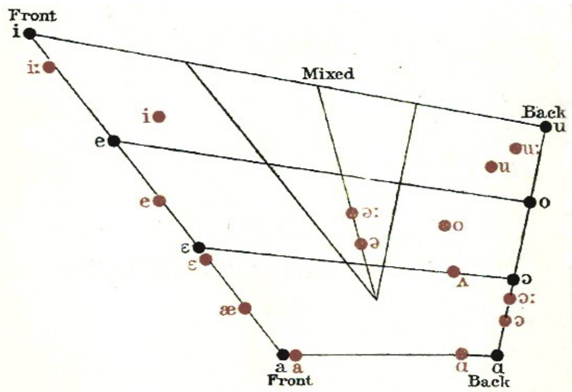
This chapter will provide context for the study, firstly by defining and discussing relevant aspects of traditional vowel theory, justifying the need for articulatory methods, and secondly by by providing a phonological, phonetic and sociolinguistic background to Swedish vowels. This section will also feature a summary of the literature on Viby-i to date, including results from the small number of studies that have looked at this vowel.

2.1. Theories of vowel production

Articulatory vowel study has historically been rare, partly due to a lack of safe, suitable methods, and partly due to the convenient correlation between tongue height and the first formant (F1), and tongue advancement and the second formant (F2) in the acoustic signal. This correlation has led to a simplified view of both acoustic and articulatory data, resulting in articulatory methods often being overlooked. Instead, F1 and F2 are often used as shorthand for articulatory parameters. There are a number of issues with these assumptions, and some of these issues will be raised here in order to emphasise the need for articulatory methods in vowel study.

2.1.1. The vowel space

Traditional (F1/F2) vowel description relies heavily on the idea of the vowel space. This is a conceptual space within which the tongue can move vertically and horizontally to produce vowels, bounded off by frication at the edges, i.e. when the corner vowels [i], [ɑ], [u] become the fricative consonants [j̞], [ʁ], [ɣ] (Catford 1994: 131). The vowel space is commonly represented in the form of the vowel quadrilateral, first developed by Daniel Jones (1956) for the cardinal vowel system (Fig. 1) and later adopted by the International Phonetic Association (IPA) for their vowel chart (Fig. 2).



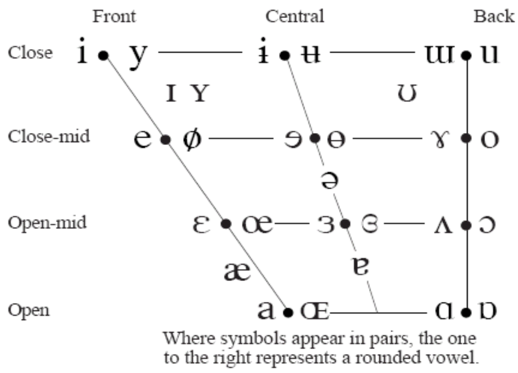


Fig. 1: Cardinal vowels (black), and English vowels (red) on the original quadrilateral (Jones 1917).

Fig. 2: Vowel chart for the IPA (International Phonetic Association 2005).

The cardinal vowels are plotted on the quadrilateral at equidistant steps, originally meant to represent equal differences between the vowels, both in auditory quality and in the position of the highest point of the tongue (Ladefoged 1967: 70). However, it has since been shown that the cardinal vowels are only "auditorily equidistant" (Ladefoged & Johnson 2015: 229), and that their articulatory properties are much more complex. Ladefoged, referring to a set of x-ray images of S. Jones producing the cardinal vowels in 1929, asserts that "the tongue does not move in a series of even approximately equidistant steps", and furthermore that "the tongue has such a different shape for the front and back vowels that it is meaningless to compare cardinal vowel number four with number five" (Ladefoged 1967: 71). For this reason, the vowel quadrilateral is not ideal for describing vowel articulation. Despite this, it continues to be the prevailing method for representing vowels, and is often used as a shorthand for articulation, with the articulatory terms 'high', 'low', 'front' and 'back' being used to describe the acoustic properties of F1 and F2, something which Lindblad refers to as "an unfortunate practice" (2010: 54).

Although it is true that a correlation exists between the position of the highest point of the tongue and the acoustic properties of F1 and F2 (Fig. 3), the vowel quadrilateral provides an idealised and unrealistic image of how vowels are produced in the vocal tract, and plotting acoustic data onto a seemingly articulatory space can therefore be misleading. To use the famous quote by G. Oscar Russell: "Phoneticians are thinking in terms of acoustic fact, and using physiological fantasy to express the idea" (in Ladefoged & Johnson 2015: 208).

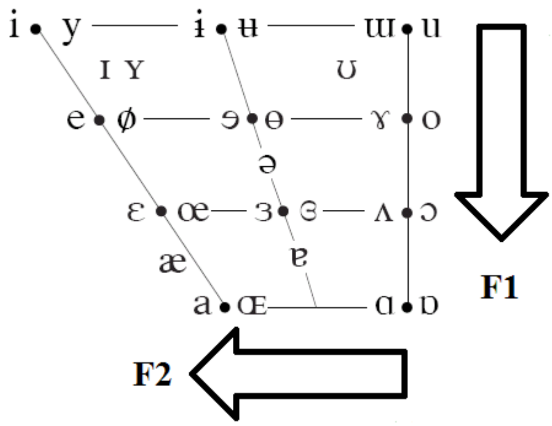


Fig. 3: How F1 and F2 values correspond to height and advancement on the vowel quadrilateral.

2.1.2. The role of the tongue in vowel production

The tongue is the most important articulator for vowel production, and its shape and position clearly influence vowel quality. However, quantifying tongue gestures for vowels is problematic because there is “no handy landmark on the tongue to serve as point of reference”, nor can we “locate a specific point of constriction or blockage in the oral cavity” (Clark et al 2007: 22) in the way that we can for most consonants. Instead, the articulatory characteristics of vowels are traditionally described in terms of the position of the highest point of the tongue, approximating the point of maximum constriction in the oral cavity. This system appears to work well due to the negative correlation between tongue height and F1, and the positive correlation between tongue advancement and F2. Thus, when vowels are plotted on a grid according to their F1 and F2 values (Fig. 4), their arrangement is highly reminiscent of the vowel quadrilateral. This effect is even more visible when the acoustic values are converted to a logarithmic perceptual scale (Fig. 5).

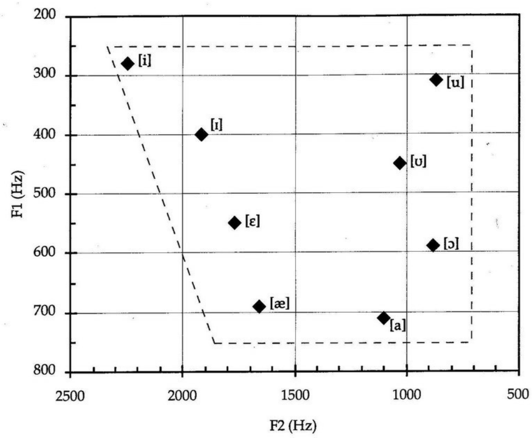


Fig. 4: F1 and F2 values (Hz) for an unspecified variety of English (Zhang 2006).

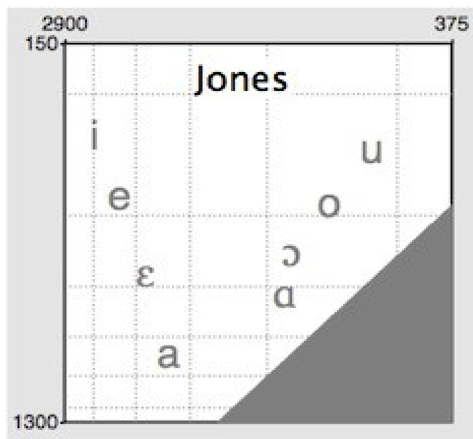


Fig. 5: Logarithmic F1 and F2 for the cardinal vowels (English Speech Services 2013).

As pointed out in the previous section, however, this system is mainly built on F1 and F2, and does not provide a full picture of vowel articulation. For instance, it vastly simplifies the role of the tongue in vowel production. Some of the assumptions underlying the vowel quadrilateral are illustrated in Fig. 6: the tongue is represented as moving in equidistant steps, with a more or less static shape, and is shown in midsagittal section with the highest point indicated. The problem with the assumption of equidistant steps has already been touched upon, but the static shape, midsagittal view and single-point reference tend to be less frequently discussed.

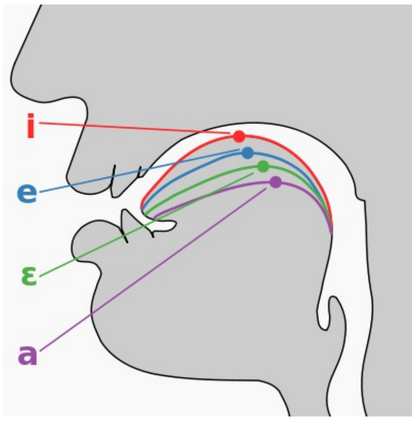


Fig. 6: Diagram illustrating the tongue positions for the front cardinal vowels (Wikimedia Commons 2008).

Ladefoged mentions a difference in shape between cardinal four and five (1967: 71), but does not state in which way they are different. We know that approximant consonants can use different tongue shapes to achieve virtually the same output; for instance, Scottish postvocalic /r/ can be produced either with a "bunched" or a "tip-up" gesture (Lawson et al 2013). Informal observations with UTI suggest that the tongue shape for a given vowel too can vary greatly, from a bunched or double-bunched articulation, to a sloped or even flat tongue shape, without the tongue shape variation necessarily being noticeable from the auditory quality.

Traditional vowel description also tends to rely on a midsagittal view of the tongue, but the tongue does not only move in two dimensions; three-dimensional properties such as bracing, grooving or lateralisation require a coronal perspective to be observed. Stone & Lundberg (1996) found in a three-dimensional study of English tongue shapes that different vowels display various types of both grooving and bracing, something which is often overlooked even in articulatory vowel data. Since the tongue is a "mobile and polymorphous mass" (Catford 1994: 132), and the acoustic properties of vowels "are based on the entire articulatory space" (Scobbie et al 2012: 105), it is important to consider overall tongue configuration in three dimensions, or acoustic analysis may be misinformed. For instance, in vowels with a flattened tongue shape, the highest point of the tongue may be averaged out over a large surface area. Similarly, it is possible for two vowels to have the same highest point, but for the tongue shapes to differ in other ways, resulting in different acoustic outputs. Thus, the position of the highest point of the tongue "is not a valid indicator of vowel quality (Ladefoged & Johnson 2015: 230), and it is therefore important to support conclusions drawn from acoustic analysis with articulatory data.

2.1.3. Variations in the vocal tract

One of the issues not taken into account by the vowel quadrilateral is the fact that vowel production relies not only on tongue configuration, but on various articulators and resonators, e.g. the lips, jaw, velum, tongue root, pharynx and larynx. When these articulators move, it affects the resonances of the vocal tract, and in turn the acoustic output. The impact of these effects is easily overlooked when articulatory data is not available. This section will discuss the impact of non-lingual articulators on vowel production.

The lips

The IPA vowel quadrilateral includes a dimension for lip-rounding by having rounded and unrounded vowels plotted in pairs, with the rounded version appearing to the right of the unrounded one (Fig. 7). However, lip-rounding can exist to different degrees. For instance, Swedish has three degrees of phonemic lip-rounding, found e.g. in the high vowels /iː/ (unrounded), /yː/ (outrounded) and /ʉː/ (inrounded) (Engstrand 1999: 141). It is thus possible that lip-rounding is more of a continuum than a binary parameter, as even small changes in rounding can affect the vowel quality.

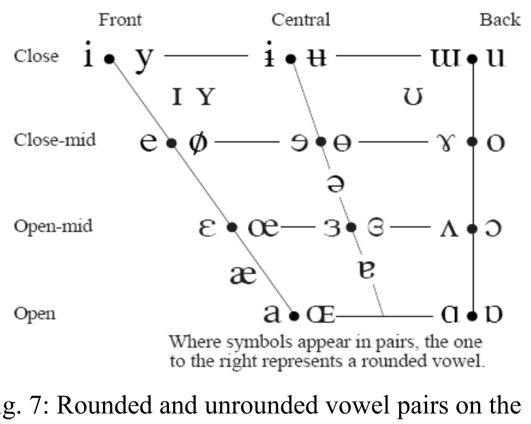


Fig. 7: Rounded and unrounded vowel pairs on the IPA chart (International Phonetic Association 2005).

The tongue and lips can also work together, seen e.g. in the tendency for back vowels to become more rounded as their height increases (Ladefoged & Johnson 2015: 230). Furthermore, the lips and tongue can compensate for each other's movements. For example, Ménard et al (2015) found in a study of blind and sighted speakers that when producing clear speech, sighted speakers exaggerated their lip movements to create greater contrasts between vowel sounds, whereas blind speakers relied more on tongue gesture.

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). The acoustic result is a lowering of the formants, particularly F2 (ibid: 42). But other articulatory changes, such as tongue backing, can also lower F2, meaning that the two articulations could not necessarily be told apart in an F1/F2 analysis. An inadequate understanding of the lips' movements can thus lead to acoustic changes being attributed to the wrong factors, which is why it is relevant to complement acoustic analysis with articulatory data.

The oral cavity

A great portion of the oral cavity – the teeth, the alveolar ridge, and the hard palate – are static and do not change the resonances of the vocal tract within individual speakers. However, the size of the oral cavity can still be affected by the degree of jaw opening. Even though the jaw "cooperates with the tongue in producing the desired area function" (Lindblom & Sundberg 1971: 1,166), something which has been documented in e.g. opera singers (Nair et al 2015), this nevertheless means that the space in which the tongue operates is fundamentally different for vowels with more jaw opening than for vowels with less. Low vowels which are prone to a greater degree of jaw opening, e.g. [ɑ], may thus require a different type of tongue configuration than high vowels, e.g. [i], to achieve the intended vowel quality.

Furthermore, UTI has shown that the shape and size of the palate and alveolar ridge differ greatly between speakers, creating unique resonances for each individual speaker's vocal tract. It is not yet known to what extent speakers use tongue gesture to compensate for these differences, but articulatory data can be used to investigate this question. This will help determine whether speakers actually employ the same articulatory techniques to achieve the same output, or if they use their articulators to compensate for physiological differences.

The pharyngeal cavity

The size and shape of the pharyngeal cavity is affected by several articulators, such as the velum, tongue root and larynx. Different vowels use these articulators in different ways. For instance, high back vowels are associated with velar constriction (Rosner & Pickering 1994: 42), which affects both the oral and pharyngeal cavities. According to Ladefoged & Maddieson (1996: 283-284), the tongue is closer to the roof of the mouth in low back vowels than it is in high back vowels, possibly because the jaw is hinged. This means that the height dimension for back vowels operates on a smaller scale than for front vowels, so that the difference in tongue height between [u] and [ɑ] is smaller than the difference between [i] and [a].

The size of the pharyngeal cavity can also be reduced by the retracted tongue root of back vowels (Lindblom & Sundberg 1971: 1,167) and vowels which employ jaw opening (Rosner & Pickering 1994: 24). The opposite effect is achieved by lowering the larynx. Larynx-lowering has a similar effect to lip-rounding, in that it lengthens the vocal tract (ibid: 39). All these changes affect the acoustic properties of the pharyngeal cavity, and in turn affect the vowels that resonate in this space. However, without articulatory data, it is difficult to know which articulators are involved, and how the acoustic signal is affected.

Summary

As these examples have shown, the vocal tract is in many ways dynamic, producing a great number of potential combinations for articulatory settings. This creates a complex relationship between articulation and acoustics, which can render acoustic data ambiguous. For example, Fant (1960), referring to conclusions about articulation reached through tube modelling, states that "*specific* shifts in a *specific* formant for a given vowel may occur when *either* the front *or* the back cavity changes" (in Rosner & Pickering 1994: 39). In other words, a tube model may be constricted either at the front or the back, and still produce a similar acoustic result. It is therefore difficult to know exactly where in the vocal tract a change occurs when the vowel quality changes. Fant adds that "alterations in cavity characteristics will not affect all vowels identically" (ibid), partly because vowels employ different overall vocal tract settings. In short, "the exact locations and configurations for [a] given speech sound are not absolute" (ibid: 22), but may be caused by several articulators at once, making acoustic data difficult to interpret. Articulatory data can make a valuable contribution to this discussion, particularly in relation to vowel production.

2.1.4. Articulatory trade-off and hidden articulations

As demonstrated in the previous section, "the relation between articulatory parameters and acoustic output is not linear" (Stevens 1972: 64). This is due to the phenomenon of articulatory trade-off: the possibility of producing virtually the same acoustic output using different articulatory settings. Due to the complexity of the acoustic signal, it is possible to compensate for changes made in one part of the vocal tract by modifying another, and speakers likely use some degree of articulatory trade-off to compensate for physiological differences. However, this also means that articulatory variation is often hidden, even when it is socially meaningful. Some articulatory studies suggest that articulatory tongue gestures used for specific sounds are not as uniform as previously believed, and this section will demonstrate how articulatory data can help to uncover such fine-grained phonetic variation.

Flipped vowels

In a recent paper, Noiray et al discuss "idiosyncratic patterns of articulation for contrasting the three [English] front vowel pairs /i-ɪ/ /e-ɛ/ and /ɛ-æ/" (2014: 272) in the form of "flipped" vowels: vowels where the F1 parameter is reversed, so that e.g. /ɪ/ has a lower F1 than /i/. The inversion of vowel height within these pairs was confirmed through both acoustic analysis and UTI, but despite this, speakers were able to convey the correct phonemic contrast to listeners. According to Ladefoged et al (1972), flipped vowels are not unusual, but "the highest point of the tongue for the vowels /ɪ/ and /e/ are flipped by many speakers" (in Noiray 2014: 274). Noiray et al's results support that idiosyncratic vowel productions are perhaps more common than we think, and that these patterns of articulation are very easily overlooked without articulatory data. The authors emphasise that the study does not address the "fine details of the articulatory-acoustic correspondence" (ibid: 281), but that the results indicate that "only the study of the articulation and acoustics together can locate the discrepancies found between phonological representations of vowels and their phonetic realizations" (ibid: 283).

Postvocalic /r/ in the Scottish Central Belt

Not all articulatory trade-off idiosyncratic, however; different articulatory strategies for achieving a specific sound can be socially or geographically stratified, despite the fact that the difference between them is barely audible. One example of socially indexical articulatory variation can be found in postvocalic /r/ in the Scottish Central Belt. Lawson et al (2011) found that Edinburgh, Livingston and Glasgow speakers were able to use either a "bunched" or "tip up" tongue gesture to produce postvocalic /r/, and that the articulations was socially stratified, with middle-class speakers being more likely to use a bunched tongue gesture, and working-class speakers more likely to use a tip-up gesture. The difference between the two articulations "has been claimed to be inaudible in other varieties of English" (Lawson et al 2011: 258), but may be acoustically distinguishable by the relationship between F4 and F5 (Zhou et al 2008). In the case of Scottish /r/, however, the question of articulation "was not resolved with acoustic analysis" (Lawson et al 2011: 257), but required articulatory data, in this case collected with UTI, to reveal a difference in tongue configuration. These studies demonstrate the usefulness of articulatory data as a means of discovering hidden articulatory variation which is masked by articulatory trade-off.

2.1.5. Section summary

The description and classification of vowels based solely on acoustic data is made problematic by a number of factors, such as articulatory trade-off, hidden articulations, and idiosyncratic production patterns, as well as the complex relationships between different articulators, and the under-studied properties of the tongue and lips. For this reason, we still know very little about vowel articulation and the effects of overall vocal tract settings on the acoustic signal. The lack of articulatory methods for vowel study has mainly been a historical problem, as technological advances are now making it much easier to collect and analyse articulatory vowel data. These technological advances will undoubtedly lead to new discoveries that change how we conceptualise vowel production.

2.2. Swedish phonology and sociolinguistics

In order to understand how Viby-i relates to the Swedish vowel system as a whole, a broader background on Swedish language and social structure is required. Swedish is known for its rich vowel system, but the realisations of these vowels are dependent on both geographic and social factors. This section will provide a brief background on the Swedish language, beginning with an overview of its main dialect groups, and touching upon aspects of sociolect. The issue of socioeconomic class in Sweden will then be discussed, as this is a sensitive topic which complicates sociolinguistic investigation of Swedish speakers. After this follows an introduction to the Swedish vowel system in general, followed by more detailed information about particular vowel properties in the studied dialect regions. Finally, a summary on the literature on Viby-i is provided, including what is known to date about the vowel's geographical and social spread, its auditory, acoustic and articulatory properties, and suggested similarities to sounds in other languages.

2.2.1. Swedish dialects and sociolects

Swedish can be broadly divided into three dialect groups: Southern Swedish, Central Swedish and Finland Swedish (Bruce 2010: 26). These three varieties are distinct in several ways, both in prosody and phonology, e.g. in terms of pitch accent type, the realisation of specific phonemes, and whether the vowel system is monophthongal or diphthongal. Some salient differences are summarised in Table 1.

	Southern Swedish	**Central Swedish**	**Finland Swedish**
Pitch accent type	1A	1B, 2A, 2B	0
Vowel system	Diphthongal	Mono > diphthongal	Monophthongal
Light or dark /l/	Light	Light > dark	Dark
Realisation of /r/	[ʀ], [ʁ]	[r], [ɾ], [ɹ]	[r], [ɾ]
Realisation of /ɧ/	[ɧ]	[ɧ], [ʂ]	[ç]

Table 1: Overview of some salient differences between the three main Swedish dialect groups (after Bruce 2010 & Elert 1995).

As Table 1 demonstrates, Central Swedish is subject to more variation than the others dialect groups, and for this reason it is usually further divided into four smaller subgroups: Northern Swedish, Eastern Central Swedish (including Dalaberg Swedish), Western Central Swedish and Gotland Swedish (Bruce 2010: 27) (Fig. 8).

Fig. 8: Distribution of Southern and Central Swedish dialects (after Bruce 2010: 27).

Apart from these regional varieties, Swedish has a non-regional variety known as *rikssvenska* or Central Standard Swedish. Although no official standard for this variety exists (Bruce 2010: 171) and it appears to take many different forms (Elert 1995: 35), Central Standard Swedish is generally said to be based on the dialect spoken in the Mälardalen region near Stockholm (Einarsson 2004: 140), and tends to sound like a dialect-levelled version of Eastern Central Swedish. This variety is generally linked to higher social status (Bruce 2010: 19).

There are also several multiethnolect varieties, primarily spoken in multilingual metropolitan and suburban areas, both by L1 and L2 Swedish speakers (Bodén 2007: 1-2). The most well-known subvarieties of multiethnolect are spoken in Sweden's three largest cities: Stockholm, Gothenburg and Malmö. These subvarieties differ from each other in many ways (e.g. Ekberg 2011), partly because they are located in three very different dialectal regions, however it is not uncommon for all multiethnolect varieties to be grouped together as *invandrarsvenska*, or 'Immigrant Swedish'. Because these varieties differ from the standard, and are prevalent in many low-income suburban areas (e.g. Rinkeby in Stockholm, Angered in Gothenburg and Rosengård in Malmö), they are often associated with low social status.

Linguistic variation in Swedish, as in most languages, is influenced by both geographic and social factors. As a general principle, varieties approaching standard language are usually associated with higher social status, younger speakers, and female speech, whereas non-standard varieties are associated with lower social status, older speakers, and male speech (Bruce 2010: 19-20). This is also the case in Sweden, however some sociolinguists (e.g. Alvtörn 2005, Breitholtz 2008) have argued that “distinctions between geolect and sociolect [in Sweden] are unclear” (Alvtörn 2005: 14). For instance, rural areas predictably display more regional features (Elert 1995: 34), and metropolitan areas display less, e.g. in the case of dark /l/, which is common in the areas around Stockholm and Gothenburg but not in the cities themselves (Bruce 2010: 158). At the same time, increased mobility and an increased awareness of the stigma which sometimes surrounds non-standard language has led to a great degree of dialect levelling, with “the overwhelming majority of Swedish speakers using regional standard varieties, which in turn have come to approach the national standard” (Elert 1995: 36). There is also a counter-movement brought on by a growing preference for casual spoken language and an awareness of the relation between language and identity (ibid). This means that in some settings non-standard varieties can be perceived as trendy, whereas in others they are perceived as unsophisticated. Sociolinguistic investigation may therefore benefit from investigating both geographical factors and attitudes towards the local speech variety.

2.2.2. Socioeconomic class in Sweden

Socioeconomic class is a complicated topic in Sweden, as “the very concept of class has become taboo over the years, and is not seen as something that is appropriate to talk about” (Håkansson & Norrby 2010: 93). Sweden is often described as having a flat social structure, associated with “conventional measures [of] … rapid social and economic mobility” (Clark 2012: 1), however differences in socioeconomic class are still present, and Clark suggests that “[t]rue rates of mobility in Sweden are similar to those of the supposedly more immobile economies of the UK and USA” (ibid).

Rather than using distinctions such as 'working-class', 'middle-class' or 'upper-class', socioeconomic categorisation in Sweden tends to be based on a combination of different indexical criteria relating to education, occupation and income (Håkansson & Norrby 2010: 93). The official statistics agency Statistiska Centralbyrån (1982) has presented a system known as SEI (*socioekonomisk indelning*, 'socioeconomic classification'), a complex framework consisting of 18 subgroups falling within three main groups: “manual workers”, “non-manual employees” and the “self-employed” (ibid: 6-7). However, fairly detailed information is required in order to use this index, and sociolinguists sometimes refer to simpler systems instead.

The present study uses a system known as SSYK (*standard för svensk yrkesklassificering*, 'standard for Swedish occupational classification'), also developed by Statistiska Centralbyrån (2012), to account for socioeconomic group. This slightly smaller index consists of 13 groups based on type of occupation and the associated level of education, expressed as “skill level” on a scale from 1 to 4 (Table 2).

Major group/ Sub-group	Name	Skill level
1 (excl. 17)	Managers	4
17	Operations managers in service industries	3
2	Occupations requiring advanced level of higher education	4
3	Occupations requiring higher education qualifications or equivalent	3
4	Administration and customer service clerks	2
5	Service, care and shop sales workers	2
6	Agricultural, horticultural, forestry and fishery workers	2
7	Building and manufacturing workers	2
8	Mechanical manufacturing and transport workers, etc.	2
9	Elementary occupations	1
01	Commissioned officers	4
02	Non-commissioned officers	3
03	Other ranks (privates. etc.).	2

Table 2: Major occupational groups, descriptions and associated skill levels in SSYK (Statistiska Centralbyrån 2012).

SSYK is used as a basis for SEI, but is simpler to use as its categories are broader and each category includes a list of occupations. For participants in this study, who were all relatively young, the parents' occupations were classified according to SSYK, and parents' education (skill level) was used as a shorthand for socioeconomic class, as education has been shown to be an important predictor of linguistic variation in Sweden (Håkansson & Norrby 93-94). It should nevertheless be acknowledged that young people are still likely to be socioeconomically mobile, and that may not fall into the same socioeconomic grouping as their parents.

2.2.3. Swedish vowels

Swedish has a densely populated vowel system, particularly in terms of high-front and mid-front vowels (Fig. 9). The vowels occur in phonemically long and short pairs, shown with examples in Table 3.

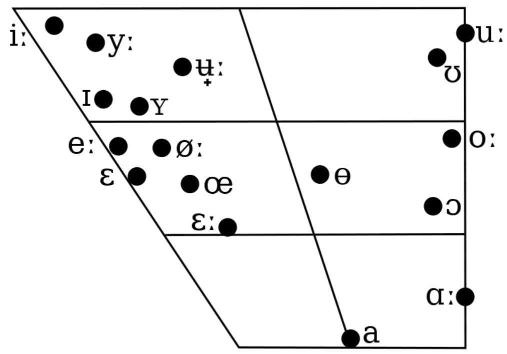


Fig. 9: Central Standard Swedish vowels (Engstrand 1999: 140).

Long vowels			**Short vowels**		
i:	*sil*	'strainer'	ɪ	*sill*	'herring'
y:	*syl*	'awl'	ʏ	*syll*	'sleeper'
ʉ:	*ful*	'ugly'	ɵ	*full*	'full'
e:	*hel*	'whole'	ɛ	*häll*	'flat rock'
ɛ:	*häl*	'heel'			
ø:	*nöt*	'nut'	œ	*nött*	'worn'
ɑ:	*mat*	'food'	a	*matt*	'feeble'
o:	*mål*	'goal'	ɔ	*moll*	'minor' (music)
u:	*bot*	'penance'	ʊ	*bott*	'lived' (perf)

Table 3: Long and short vowel pairs in Central Standard Swedish, after Engstrand (1999: 141).

There are three types of phonemic lip-rounding: unrounded, e.g. /i:/, /e:/, outrounded, e.g. /y:/, /ø:/ and inrounded, e.g. /ʉ:/, /u:/ (Engstrand 1999: 141) (Fig. 10). The degree of rounding is often said to be the only articulatory difference between e.g. /i:/, /y:/ and /ʉ:/ (e.g. Engstrand 2004: 297), however Schötz et al (2014) found that tongue body height can also differ between the three. As Fig. 10 illustrates, the low back vowel /ɑ:/ also tends to be rounded, but the degree of rounding depends largely on the dialect spoken.

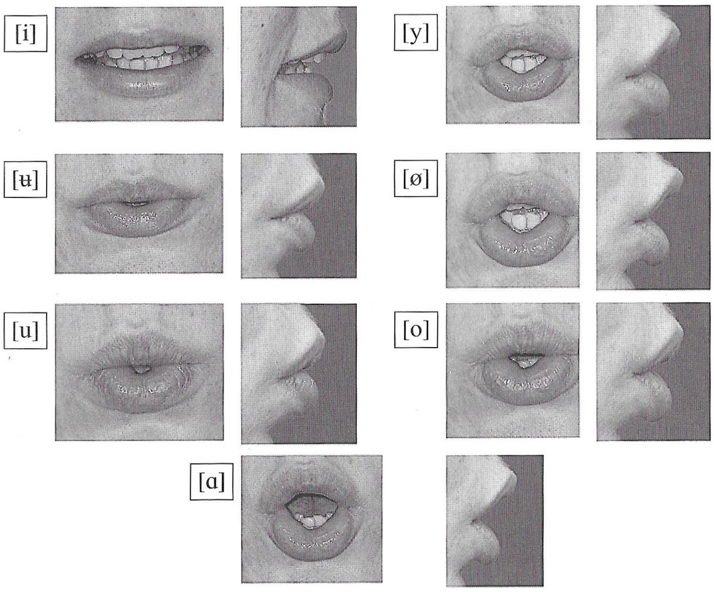


Fig. 10: Different types of lip-rounding in Swedish. Top left: unrounded. Remaining left: inrounded. Left column: outrounded. Bottom: rounded /ɑ:/ (Engstrand 2004: 98).

Most Swedish dialects are classified as monophthongal, but diphthongal varieties also exist, notably in Southern Sweden, in Mälardalen (Eastern Central Sweden) and on Gotland (Elert 1995: 38-43). However some degree of phonetic diphthongisation, e.g. [eːə] for /eː/, can be found in most varieties of Swedish, particularly in long vowels (Engstrand 1999: 141). Long high vowels are also commonly affected by end-frication, e.g. [i:j] for /iː/. This phenomenon will be discussed in more detail in the next section.

According to Bruce, Swedish vowels vary greatly between dialects a well as on a more local level, but a “collected, relatively complete description of Swedish vowel variation does not exist to date” (2010: 102). The following sections will therefore only provide a brief description of vowel variation relevant to this study, i.e. primarily of long vowels in the areas around Stockholm and Gothenburg, in the dialect regions of Eastern and Western Central Swedish.

Diphthongisation and end-frication

As mentioned above, diphthongisation and end-frication are both common in Swedish vowels. In most Central Swedish varieties, and Eastern Central Swedish in particular, these end-phases follow a particular pattern, with long high vowels experiencing end-frication, and the remaining long vowels having a centralised offglide (Bruce 2010: 126-127; Elert 1995: 40-41). This pattern is illustrated in Table 4. According to Elert, the high vowels can also be followed by a centralised offglide, either with or without end-frication, e.g. [iːjə] or [iːə] for /iː/ (1995: 40).

/iː/	→	[iːj]
/yː/	→	[yːj]
/ʉː/	→	[ʉːβ]
/uː/	→	[uːβ]
/eː/	→	[eːə]
/ɛː/	→	[ɛːə]
/øː/	→	[œːə]
/ɑː/	→	[ɒːə]
/oː/	→	[oːə]

Table 4: Central Swedish pattern of diphthongisation and end-frication (after Bruce 2010: 127).

For diphthongised vowels, the 'target' quality of the vowel is achieved from the outset and maintained for most of the production, with the centralised offglide happening much later (Bruce 2010: 127). Creak may also occur before the offglide (Elert 1995: 43). The most pronounced diphthongisation is found in the vowels /eː/ and /oː/, which serves to "sharpen the contrast between … /iː/ and /eː/, and between /uː/ and /oː/" (Bruce 2010: 127-128). Diphthongisation is most common when the vowel is in word-final position, or when it is followed by another vowel (Elert 1995: 43, Bruce 2010: 127).

Although both end-frication and diphthongisation "are so common in many parts of Sweden that they should be regarded as representative of natural speech" (Elert 1995: 43-44), both Bruce and Elert report that strong diphthongisation, especially of the /eː/ and /oː/ vowels, is usually associated with low social status. At the same time, Elert mentions that it is common among younger speakers in Stockholm (1995: 42), suggesting that it may be evaluated differently in this context.

Bruce also remarks that "diphthongisation, and thus the time dimension [of Swedish vowels, and of vowels in general] is undervalued and not sufficiently acknowledged" in phonetic study (2010: 121), and believes that diphthongised vowels are probably more common than traditional description suggests.

Realisation of /ɛː/ and /øː/

One of the most salient features of traditional Stockholm dialect is an /ɛː/-/eː/ merger, where both vowels are pronounced [eː] (Elert 1995: 47). However, this stereotypical feature appears to be becoming less common, with younger speakers instead displaying a lowering of /ɛː/ to [æː], as well as /øː/ to [œ̝ː] (Engstrand 2004: 116). Lowering of these two vowels traditionally only occurs before /r/ in Eastern Central Swedish (Bruce 2010: 118), but is now being produced in all environments. In Western Central Swedish, the same lowered realisations of /ɛː/ and /øː/ appear to have existed much longer (Elert 1995: 47, Bruce 2010: 140), but it is not clear whether Engstrand believes that they are undergoing further lowering in this region.

The lowering of /øː/ has also resulted in a merger between /œ/ (the short version of /øː/) and /ɵ/ (the short version of /ʉː/), with [ɵ] being used for both (Engstrand 2004: 116), or, more rarely, [œ] being used for both (Elert 1995: 48). The merger has been documented in both Stockholm and Gothenburg. However, it should also be mentioned that many speakers still retain the standard /ɛː/ and /øː/ qualities in both Eastern and Western Central Swedish.

Realisation of /ɑː/

As mentioned before, the Swedish /ɑː/ vowel is commonly produced as a rounded [ɒː], but the degree of lip-rounding varies depending on the region. According to Elert, "there is a tendency towards stronger lip-rounding in Stockholm and Gothenburg", with the vowel quality approaching [oː] (1995: 50). This is a particularly well-known feature of traditional Gothenburg dialect, especially in speakers of lower social status (Bruce 2010: 138), and Elert reports that a highly rounded /ɑː/ is often negatively evaluated in both Stockholm and Gothenburg (1995: 50).

In some varieties, e.g. Finland Swedish, /ɑː/ has a higher, more [aː]-like realisation. This feature was widespread in Sweden in the early 1900s, when it signified high social status (Bruce 2010: 138). Remnants of this realisation, and attitudes towards it, may still be present today.

Summary

Although Eastern and Western Central Swedish are easily distinguished by Swedish speakers, particularly when it comes to the Stockholm and Gothenburg dialects, the varieties also share a number of vowel properties, as this section has shown. Both dialect regions display diphthongisation and end-frication, lowering of /ɛː/ and /øː/, and strong /ɑː/-rounding, all of which appear to be stronger in metropolitan environments. End-frication and diphthongisation are more strongly associated with Stockholm, where /ɛː/-lowering coexists with /ɛː/-raising, whereas strong /ɛː/- and /øː/-lowering, as well as /ɑː/-rounding, are traditionally more characteristic of Gothenburg. The next section will discuss another vowel feature they have in common: Viby-i.

2.2.4. Viby-i

Viby-i is a Swedish /iː/ variant with an unusual auditory quality that has been described as “thick”, “dark”, “buzzing” or “damped” (Engstrand et al 1998). Audio samples of this vowel are available in the Appendix. The name 'Viby-i' stems from Viby parish near Örebro in Central Sweden where the vowel was first studied, but local names referring to the same vowel often spring up in areas where the sound is particularly salient, especially if /iː/ or /yː/ is present in the name, e.g. Lidingö in Stockholm or Lysekil in the Bohuslän region, north of Gothenburg. Although the phenomenon of Viby-i is well-known, its articulatory properties have long been debated, and very few phonetic studies have investigated this vowel. This section will outline the findings of the few studies that have been conducted on Viby-i, as well as summarise hypotheses and anecdotal evidence concerning its acoustic and articulatory properties, as well as its geographic and social distribution.

Phonemic and phonetic properties

Viby-i is an allophone of /iː/, and speakers tend to use either Viby-i or standard [iː] categorically. The vowel can nevertheless coexist with [iː] within a speech community, as is reported to be the case in Stockholm and Gothenburg (Schötz et al 2014; Bruce 2010: 133). Viby-colouring mostly affects long /iː/ and /yː/, but can also occur in short /ɪ/ and /ʏ/ (Bruce 2010: 136) in some dialects.

Viby-i is subject to many of the same speech processes as [iː], e.g. diphthongisation and end-frication, which often takes a [z] or [ð] rather than a [j] offglide (Bruce 2010: 135). It is not clear whether frication also occurs during the vowel portion, but it does not seem to be the case.

No set annotation exists for Viby-i, but is sometimes referred to with the IPA symbol for a high central unrounded vowel [ɨ] (e.g. Björsten & Engstrand 1999) based on its assumed articulatory properties. Other annotations include [i^z] (Elert 1995) based on the vowel's 'buzzing' quality, and [ʅ] (Lundell 1879), a symbol which was allegedly introduced into sinology by the linguist Bernhard Karlgren based on similarities in vowel quality between Viby-i and Chinese apical vowels. This similarity will be discussed in more detail in a later section on similar sounds in other languages. For the purposes of this paper, Viby-i will not be denoted by any specific phonetic symbol, as the use of particular symbols, particularly with diacritics, may imply articulatory characteristics which are not yet known.

Geographic distribution

According to Björsten & Engstrand, Viby-i can be found "in several scattered dialects, both in rural areas and in the city dialects of Stockholm and [Gothenburg]" (1999: 1,957). A dialect map produced by Elert in 1995 (Fig. 11) shows that the vowel is mainly present in Eastern and Western Central Sweden, but that its distribution is sporadic within these regions.

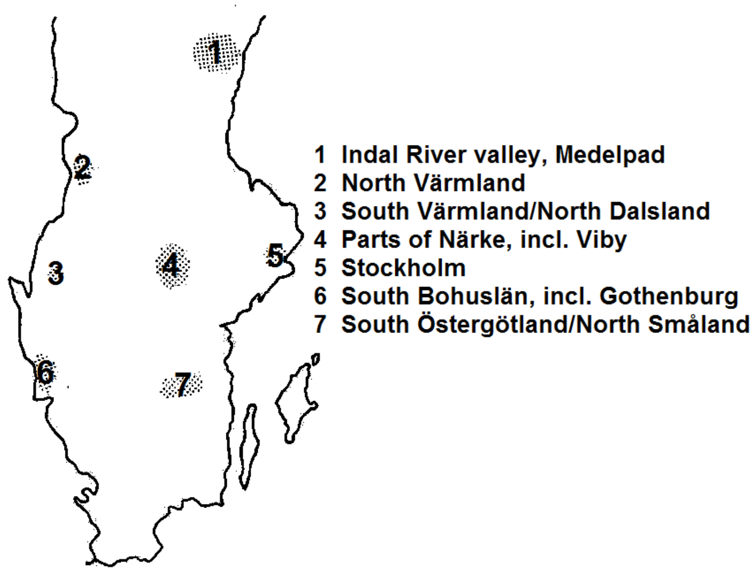


Fig. 11: Geographic distribution of Viby-i, after Elert (1995: 45).

Viby-i appears to be more salient in some areas than in others. Areas that are particularly well-known for this sound include Viby (number 4 on the map), the province of Bohuslän, including the seaside locations Orust, Tjörn, Smögen and Lysekil (number 6), as well as the Lidingö and Östermalm districts in Stockholm (number 5). The vowel quality in these places is often remarked upon, and speakers appear to be self-aware of using Viby-i, as demonstrated by this interview with a young man from Orust:

> **Younger male, Orust:**
> It's the same thing now when I'm working on Tjörn. They talk in quite a distinct way too, so then all the “i”s come out, and all these other expressions that we have out here. I was down in Partille [Gothenburg suburb] for work, and I went to get a haircut and I asked, “Do you know where I come from?” “No, I thought you were from here?” “No,” I said, “I'm from Orust.” “Where they do the 'i'-ing?” “Yup.” “You can't tell.” So I had adopted their [standard] accent, so to speak.
> (SweDia 2000, researcher's translation and brackets)

A popular theory presented by Lindström (2014) is that Viby-i spread from Bohuslän to Stockholm in the early 1900s, via wealthy Stockholm holidaymakers spending their summers on the West Coast. This could explain why Viby-i is perceived as a status symbol in Stockholm, something which will be discussed further in the section about social evaluation and distribution. The theory that Viby-i should have migrated from West to East has not been verified, however, and fails to account for the pockets of Viby-i which can be found in rural areas throughout Central Sweden.

Furthermore, anecdotal evidence suggests that weaker versions of Viby-i may be present to a much wider degree than is documented in the literature. Bruce believes that the vowel is spreading (2010: 136), and informal observations support this, for instance this quote by a Swedish learner in an online language forum:

> **mexerica feliz [senior member]:**
> Lidingö/Viby-i is so omnipresent on SVT [Sweden's public TV network] and TV3 newscasts it could be called a new “standard” Swedish /i/. For most people who have it it's the normal way of pronouncing /i/, and they fail to hear the difference with the “older” mainstream /i/, which makes it very funny when they start pronouncing English words with this [/i/]. Not only [“bimbos”] [use] it, but it's spreading like rapid fire, all across Western, Central and Eastern Sweden …
> (WordReference.com 2015, researcher's brackets)

If this is the case, the question is whether the strong and weak versions of Viby-i can be classified as the 'same' sound at different ends of a spectrum, or if the sounds are simply grouped together because they diverge from [iː] in a similar way.

Early auditory studies

One indication that there may be different versions of Viby-i is the fact that auditory accounts of the vowel's articulation tend to differ greatly. Early studies offer several, sometimes contradictory theories of how Viby-i is produced: Lundell believes that it is simply a high back [ɯ] (1879), while Noreen claims that it is produced with a constriction that is further front than for [iː], with either a tip-down (alveolar) or tip-up (apico-alveolar) tongue gesture (1903: 495). Borgström suggests a complex, lateralised, double-bunched articulation, where “the tongue tip rests against the lower teeth,” “the anterior sides of the tongue rest against the upper teeth”, and “the tongue [body] is somewhat lowered between its raised anterior and posterior parts, creating a confined resonance chamber” (1913: 33). Finally, Ladefoged & Lindau believe that Viby-i is achieved “by slightly *lowering* the body of the tongue while simultaneously raising the blade of the tongue” (1989, in Ladefoged & Maddieson 1996: 292).

What these theories have in common is the idea that the tongue blade or tip for Viby-i for is located near the alveolar ridge, similar to standard [iː], but the disagreement seems to be about whether the gesture is tip-up or tip-down, and whether the tongue body is lowered or not. Bruce, based on Engstrand et al (1998) and Björsten & Engstrand (1999), suggests a possible compromise, where “the basic articulation for Viby-i is a centralised (close unrounded) vowel”, but the buzzing quality can be exaggerated by “raising the blade or tongue tip towards the alveolar ridge” (2010: 133). He bases this on the fact that some reports describe Viby-i as more 'buzzing' or '/z/-like' than others (ibid). However it is still not clear whether the sounds themselves are actually different, or whether all these descriptions refer to different ways of producing the same sound, as a consequence of articulatory trade-off. In order to find this out, both acoustic and articulatory data is needed.

Acoustic properties

The main acoustic characteristic associated with Viby-i is a low F2 compared to standard [iː] (Engstrand et al 1998, Björsten & Engstrand 1999). In example data from Engstrand, a male speaker produces Viby-i with an F2 of about 1,500-1,600 Hz, whereas standard [iː] tends to have an F2 of over 2,000 Hz (2004: 119). In this example, Viby-i has similar F1 and F3 values to standard [iː], but Engstrand et al (1998) and Björsten & Engstrand (1999) found that F1 can also be higher in Viby-i than in standard [iː]. All this is based on data of one older male speaker from Kräklinge, near Viby. Low F2 values for Viby-i has also been reported by Schötz et al (2011).

A traditional interpretation of these acoustic results (low F2, high F1) would be that Viby-i is centralised, and Engstrand et al (1998) and Björsten & Engstrand (1999) indeed describe it as such, denoting it with the symbol [ɨ]. The issues of making this conclusion based solely on the first three formants has already been discussed, however Engstrand and Björsten & Engstrand partially mitigate this issue by using a speech synthesis model called APEX (Stark et al 1996) to attempt to reverse-engineer Viby-i, using computer technology to identify possible articulatory settings which produce a matching acoustic output . They find that Viby-i can be achieved with a centralised tongue body, with or without a tip-up gesture (apicalisation), and that

> … raising the tongue tip by 10mm leads to a further lowering of F2 (by approximately 100Hz). From an auditory point of view, this would result in further damping … It is possible that this dimension is exploited to different degrees in different dialects. Thus, the dialectologists' question whether the damped vowel type is produced with or without apicalisation could not be unambiguously settled – both possibilities still remain.
>
> (Björsten & Engstrand 1999: 1,959)

The term "damping" here is believed to refer to a darkening of the auditory quality, or divergence from standard [iː].

Björsten & Engstrand's study also includes a perceptual component, where Swedish and Turkish speakers were played instances of Turkish [ɨ], Viby-i, and standard [e] and [i], and were asked to rate whether these were acceptable instances of the 'damped' vowel in their language, i.e. Turkish listeners were asked if they heard [ɨ], and Swedish listeners asked if they heard Viby-i. The Swedish listeners did not have Viby-i in their own speech.

The results, shown in Fig. 12, show some overlap between [ɨ] and Viby-i, in that Swedish and Turkish speakers both gave positive responses for these vowels. However, the Swedish speakers consistently rated Viby-i higher than the Turkish speakers did, and the Turkish speakers rated [ɨ] higher than the Swedish speakers. Both groups were able to distinguish between damped and standard vowels. An interesting result is the gradient effect in the Swedish damped vowels, which were rated by both groups as the most acceptable in Orust, second in Kräklinge, and third in Gothenburg, suggesting an audible difference in the realisation of Viby-i between these places.

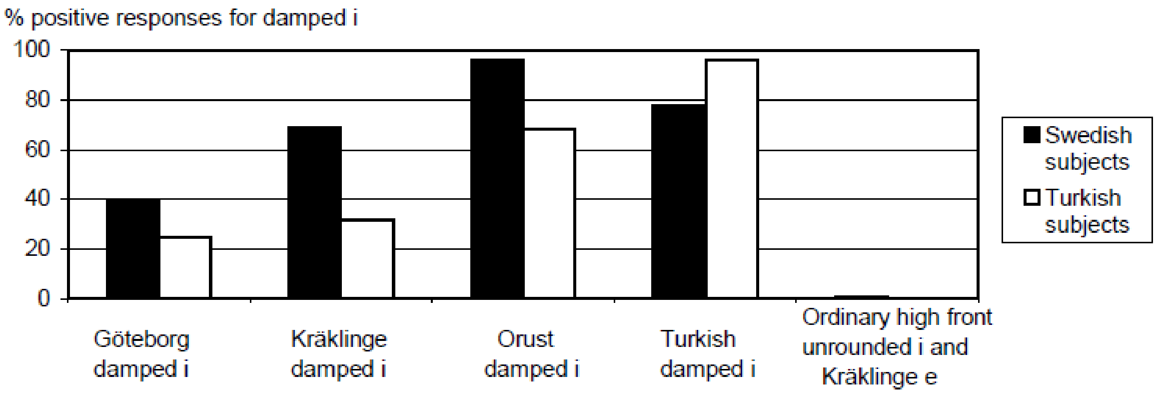


Fig. 12: Percentage of stimuli identified as "damped /i/" by Swedish and Turkish speakers (Björsten & Engstrand 1999: 1,958).

These results thus point towards the possibility of stratified regional variation in the production of Viby-i, perhaps caused by the optional tip-up gesture mentioned earlier. This would tie in with reports of Viby-i being 'buzzier' in certain locations, and would support the argument that the articulation of Viby-i is dependent on region.

Published articulatory studies

The only published articulatory study of Viby-i to date is conducted by Schötz and colleagues, and has been reported on in several papers, e.g. Schötz et al (2011, 2014). The study compares tongue positions and dynamics for Swedish /iː/, /yː/ and /ʉː/ in the Stockholm, Gothenburg and Malmö varieties of Swedish, using electromagnetic articulography (EMA). This method works by attaching magnetic sensors to the articulators and head to record their movement. In both studies, ten sensors were attached to the tongue, lips and jaw, and two were attached to immobile parts of the speaker's head (the bridge of the nose and behind the ear) to correct for head movement (Fig. 13). Only midsagittal tongue results (i.e. sensors 1-3) are reported on in the papers.

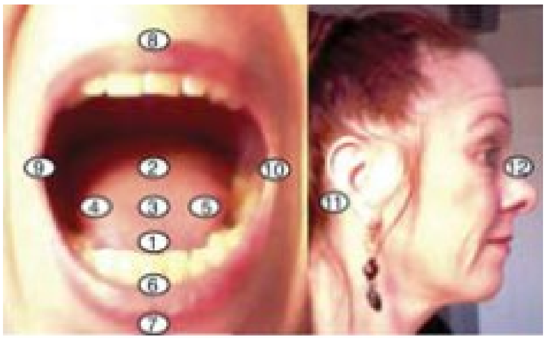


Fig. 13: Placement of EMA sensors in Schötz et al (2014).

The pilot study (Schötz et al 2011) investigates one Southern Swedish speaker without Viby-i and one Eastern Central Swedish speaker with Viby-i, measuring the steady-state portion of the vowel. This would entail taking the Southern Swedish speaker's measurement before diphthongisation occurred. The results, shown in Fig. 14, found that the Eastern Central Swedish speaker produced Viby-i with a low tongue dorsum, high tongue blade, and raised tongue tip, creating an “upward slope”, whereas the Southern Swedish speaker produced [iː] with a lowered tongue tip, resulting in a “downward slope” (ibid: 27).

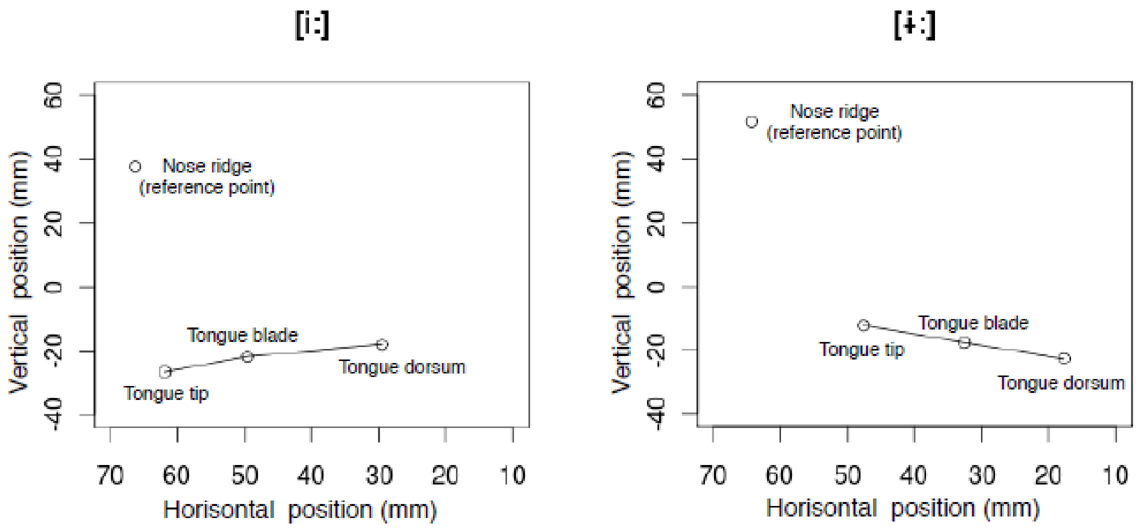


Fig. 14: Position of EMA sensors during steady-state production of [iː] (left) and Viby-i (right). Speakers seen in profile, facing left. (Schötz et al 2011: 27).

The subsequent study (Schötz et al 2014) follows up on these results by collecting similar data from nine speakers from Stockholm, Gothenburg and Malmö respectively. The methodology is similar, but the results are more focused on dynamics, and report the position of each EMA sensor separately. Overall, Schötz et al found that Stockholm and Gothenburg speakers could be divided into two groups, largely based on whether or not they had Viby-i in their speech (ibid: 20), and that this was related to geographical region. Type 2 speakers, who tended to have Viby-i, were usually from a more central part of the metropolitan area, and produced /iː/ with a significantly lower tongue body than Type 1 speakers, who tended to come from the outskirts of the metropolitan area (Fig. 15). The difference in tongue tip height between the two types were not significant (ibid: 19).

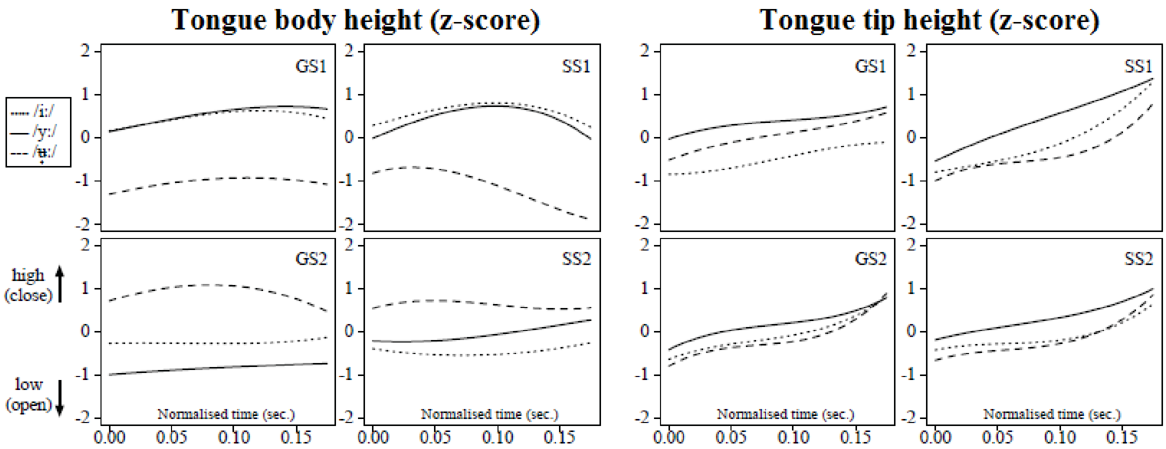


Fig. 15: Tongue body height (left) and tongue tip height (right) over time for /iː/ (dotted line), /yː/ (solid line), and /ʉː/ (dashed line). Type 1 speakers = top. Type 2 speakers = bottom. (Schötz et al 2014: 20)

The tongue body was also found to be more fronted in Type 1 speakers, whereas Type 2 speakers were more backed. The trajectories of the tongue tip and tongue body revealed different dynamic properties between Type 1 and Type 2 speakers, as well as between the Gothenburg and Stockholm dialects. The study thus concludes that regional variation exists in tongue gesture and dynamics, both between and within Stockholm and Gothenburg (ibid: 20).

The study acknowledges some limitations in the data in that it only analyses "two discrete points and two dimensions of the tongue" (ibid: 21), however lateral data was also collected. Nevertheless, a general problem with EMA is that it does not give a full picture of the tongue surface as a whole, but only reveals the positions of the sensors. This can limit a full understanding of the tongue shape. Schötz et al also omit lip-position from the analysis, but did record the lips for further study. An interesting finding is that /iː/, /yː/ and /ʉː/, which are commonly said to be distinguished only by lip-rounding, were found to have different tongue body heights (ibid: 21). The dynamic approach to vowel analysis is also very interesting and should be explored further.

Unpublished articulatory study

A small unpublished study of Viby-i using UTI (Westerberg 2013) also exists, investigating vowel articulation in three female Swedish speakers with Viby-i (two from Gothenburg, one from Varberg, 80 km south of Gothenburg) and three female Scottish English speakers with standard [iː]. The speakers were recorded reading a word list containing the vowels /iː/, /eː/, /aː/, /uː/ (Swedish) and /iː/, /ʉː/, /aː/, /oː/ (Scottish English) placed initially, medially or finally in non-lingual consonant frames, e.g. <_b>, <b_b>, <b_>. The Scottish speakers were also asked to listen back to the Swedish productions and try to mimic them.

UTI data was captured in midsagittal section, and tongue contours were manually extracted during the steady-state of the vowel, which was selected visually. Average tongue contours based on six repetitions of each vowel are presented in Fig. 16.

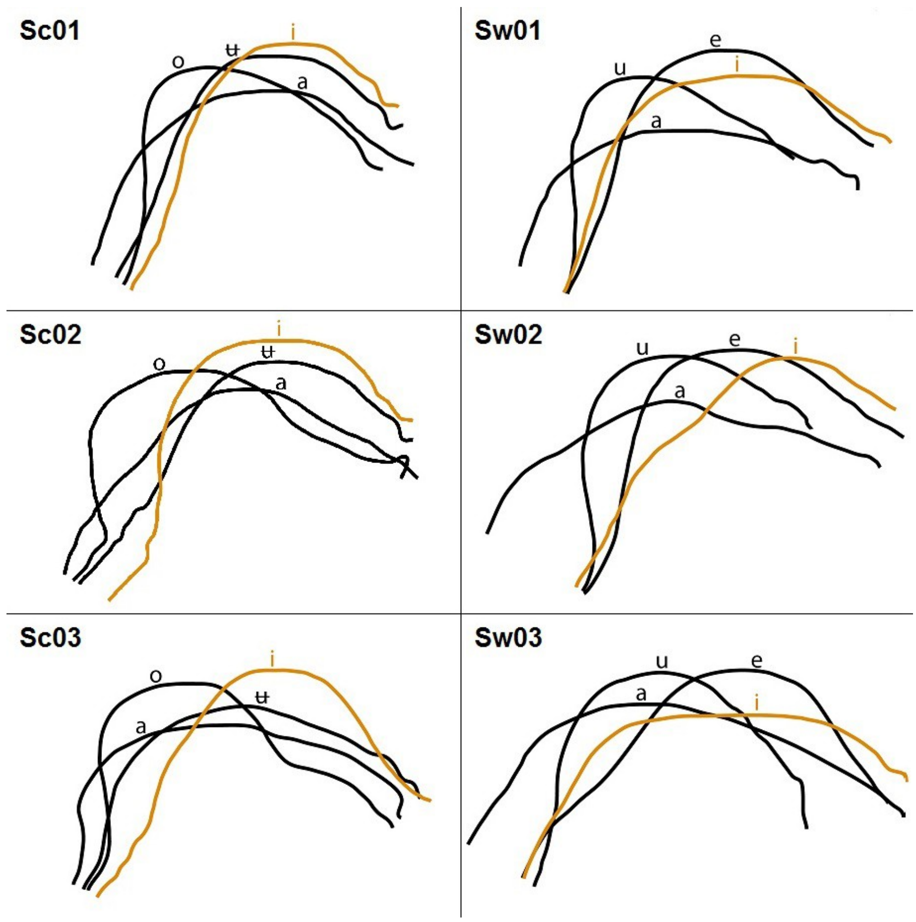


Fig. 16: Average tongue contours for Scottish and Swedish speakers in the production task. Seen in profile, facing right (Westerberg 2013).

The results show that Scottish /iː/ was consistently articulated as a high front vowel, but that Swedish speakers were highly variable, both in vowel quality and in tongue configuration. The Gothenburg speakers (Sw01, Sw03) were auditorily judged to have a stronger Viby-i than the Varberg speaker (Sw02), and also tended to have a lower tongue body. Sw03 had the lowest tongue body and was judged to have the strongest Viby-i. The difference in Viby-i strength between the more metropolitan area (Gothenburg) and the more peripheral urban area (Varberg) corresponds to the results of Schötz et al (2014), but unfortunately no further demographic information was collected. The study also lacks an acoustic components for measuring formant differences.

In the mimicry task, Scottish speakers were able to convincingly mimic Swedish /eː/ /aː/ /uː/, but only managed to produce a vowel which was auditorily similar to Viby-i in two instances. In the process, they attempted many unusual articulatory strategies, such as raised larynx, tongue protrusion, several high unrounded vowels, and lateral consonants. The Scottish speakers' inability to reproduce Viby-i could indicate that they found the articulation difficult to parse.

Social evaluation and distribution

The social evaluation of Viby-i appears to vary depending on the setting. In rural areas, Viby-i tends to be a "genuine dialectal feature" (Elert 1995: 45) and can sometimes be perceived as unsophisticated (Bruce 2010: 125). For this reason, Elert suggests that the vowel may be subject to dialect levelling, as it "tends to quickly be put aside when transitioning into more standard language" (1995: 45). This would be supported by the quote provided earlier from the young male speaker from Orust, who used Viby-i when he was close to home, but reverted to a more standard /iː/ when he was in a region that did not use this vowel. In Bohuslän, where Orust is located, Viby-i is strongly associated with the dialect of the old fishing community (Lindström 2014), and its usage may thus depend on the speaker's relation to this community (Elert 1995: 45).

In some urban and metropolitan areas, on the other hand, Viby-i is a "clear prestige marker", and a sign of high social status (Bruce 2010: 135-136). This is particularly true in Stockholm, but also seems to be the case in Gothenburg, despite the fact that Viby-i has existed in Gothenburg for much longer (Elert 1995: 45), possibly given its proximity to Bohuslän. In Stockholm, Elert stipulates, Viby-i may still be spreading (ibid), and Bruce believes that this is happening on a national level (2010: 136). According to Bruce, higher sociolects of Southern Swedish and Copenhagen Danish are now beginning to show Viby-colouring (2010: 136) and the 'damped' quality of the vowel may even be spreading to /j/ in words like *hej* 'hello' and *okej* 'okay' in places where Viby-i is used (2010: 216). Both Bruce and Elert report that the use of Viby-i in Stockholm and Gothenburg is socially stratified, with younger speakers using it more than older speakers, and women using it more than men (Bruce 2010: 136, Elert 1995: 45). This could be indicative of Viby-i becoming more common, as young females are often at the forefront of language change (e.g. Labov 2001). Bruce also reports that Viby-i does not seem to be present in multiethnolect varieties (2010: 136).

Despite the prestige attached to Viby-i in Stockholm and Gothenburg, Elert claims that “[t]hose who notice the sound generally don't appreciate it” (1995: 45), pointing to a discrepancy between evaluations of the sound itself and its social meaning. Bruce disagrees; he claims not only that Viby-i is perceived as “the more 'elegant' pronunciation” (2010: 136), but also that it is “highly contagious and difficult to resist, if one should wish to do such a thing. It would not be surprising if this pronunciation … became established as standard within a few decades” (ibid: 216). It would appear, then, that Viby-i is a case of change in progress.

Similar sounds in other languages

There are several languages which are said to feature sounds similar to Viby-i, e.g. Chinese Mandarin and Grasslands Bantu (Faytak & Merrill 2014), Turkish (Björsten & Engstrand 1999), progressive forms of Danish (Bruce 2010: 136) and, according to Engstrand, several “South American indigenous languages” (2004: 120), although no specific languages are mentioned. Viby-i is sometimes described an “exotic” sound (e.g. Schötz et al 2011, 2014; Engstrand 2004) in terms of its rarity in the world's languages, however the degree to which this is true depends on the actual articulation involved. As Schötz et al put it, “[a] fronted (alveolar) variant would seem to be more odd … while a retracted (central) variant would appear to be a vowel articulation which is less unusual and found in a fair number of the world's languages” (2011: 26). The link between Viby-i and similar sounds in other languages has not been specifically studied except in Björsten & Engstrand's perceptual study between Swedish and Turkish (1999), however it would nevertheless be difficult to assess articulatory similarities between these sounds, as articulatory vowel data for most languages is still rare.

Some articulatory data on apical vowels in Chinese Mandarin is provided by Faytak and Lin (2015) in an UTI study, and these show some interesting results. Mandarin [ɿ] and [ʅ] are sometimes auditorily judged to match Viby-i, in that they have a similar 'thick' or 'buzzing' quality, but they are strictly conditioned by consonant environment: [ɿ] only appears after [s] and [ʅ] only appears after [ʂ]. Both vowels have been said to have a “fricative-like tongue posture, [but] they are typically either weakly fricated or free of frication” (Faytak & Lin 2015: 1). This was confirmed by Faytak & Lin, who only saw occasional vowel frication in one of their five participants (ibid: 3), but at the same time found that the tongue gestures associated with [ɿ] and [ʅ] were extremely similar to the preceding fricative consonant (ibid: 2). Similar coarticulation was present in [i] following the fricative [ɕ], but not to the same extent.

Fig. 17 shows the mean midsagittal tongue curves for two speakers in the study: S1, who showed the greatest articulatory difference between fricative consonant and apical vowel, and S8, who showed the smallest difference. S8 has more or less complete overlap in tongue gesture between consonant and apical vowel, and this was also found to be the case in coronal section. Faytak & Lin conclude that the lack of frication in the vowel must be caused by modification in another part of the vocal tract (ibid: 4), for example widening of the pharynx or slowing the airflow during the vowel to reduce frication.

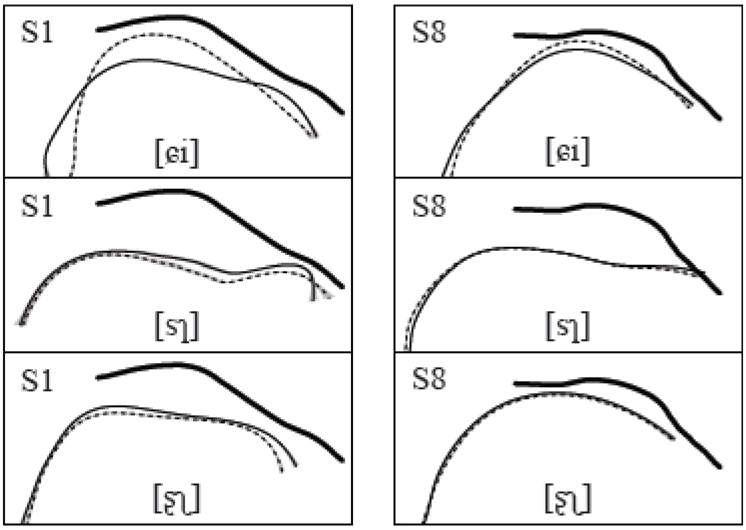


Fig. 17: Tongue diagrams (facing right) of two speakers producing fricative consonants (solid lines) and vowels (dotted lines). The thick solid line represents a palate trace (Faytak & Lin 2015).

Fig. 17 also reveals that the tongue gesture for the apical vowels [ɿ] and [ʅ] are lower than for the standard vowel [i], something which was typical for all speakers (ibid: 2). Faytak & Lin describe the tongue gesture for apical vowels as follows:

> During production of [ɿ] and [ʅ] the tongue dorsum is as low and retracted as [a], with the tongue posture differing mainly in the extension and raising of the tongue blade and tip.
>
> (Faytak & Lin 2015: 2)

In most speakers, the tongue was marginally lower for the vowel than for the consonant, however in the one speaker who produced fricated vowels, the tongue was marginally higher, and the tongue tip was closer to the palate. The description of tongue body lowering, and optional tip-raising to produce fricative noise, is highly reminiscent of some of the articulatory descriptions of Viby-i based on auditory and acoustic analysis (e.g. Bruce 2010).

There is in other words a possibility that Viby-i is both acoustically and articulatorily similar to Mandarin [ɿ] and [ʅ], however an important difference is that the Mandarin vowels are conditioned by consonant environment. According to Bruce, this is also the case in Copenhagen Danish, whereas Swedish Viby-i is “independent of consonant environment” (2010: 136).

3. Method

This chapter provides an account of the method used for data collection and analysis in this study, beginning with an outline of the recruitment strategy and participant sample, and continuing with a description of the equipment and recording procedure, including a technical description of how ultrasound tongue imaging (UTI) works, and how it was set up in this particular study. This is followed by a section on prompt and questionnaire design, and a discussion of ethical considerations. Finally, the chapter will outline how the different types of data were processed and analysed. Decisions made along the way will be justified and evaluated within each section.

3.1. Participants

This section outlines the recruitment strategy by which participants were obtained for the study, and provides a brief description of demographic information which was extracted through a questionnaire. The questionnaire design will be described in more detail in a later section.

3.1.1. Recruitment strategy

In order to investigate the linguistic and social properties of Viby-i, a recruitment strategy was required which could sample speakers with Viby-i in their speech without making them aware that this was the variable under study. Such an awareness could prime participants, causing them to change their linguistic behaviour. Another issue was the fact that the study was conducted in Scotland, which meant that access to Swedish speakers was very limited. With this in mind, the study began recruitment with the criteria that participants should be native Swedish speakers from either Stockholm or Gothenburg, where Viby-i is known to occur (e.g. Björsten & Engstrand 1999), as it was believed that these large cities would be fairly well-represented in the urban areas of Glasgow and Edinburgh. However, in order to maximise participant numbers, the criteria were later broadened to include the entire dialectal regions of Eastern and Western Central Swedish, again based on the documented presence of Viby-i (e.g. Elert 1995). The sampling region was set by the researcher to encompass the provinces Bohuslän, Halland, Värmland, Dalsland, Västergötland, Västmanland, Närke, Östergötland, Småland, Uppland and Södermanland, based on the approximate dialect borders with Southern and Northern

Swedish according to Bruce (2010). This region was indicated on a map of Sweden on the recruitment posters for the study (Fig. 18).

Fig. 18: Recruitment poster specifying the Eastern and Western Central Swedish dialectal regions.

Anyone who was over the age of 18 and a native Swedish speaker from the specified region was invited to participate. Speakers were not screened for Viby-i in their speech prior to their inclusion in the study, as participants without Viby-i could be used as a reference point for articulation and acoustics, and could also aid investigation of possible geographic or social factors influencing the presence or absence of this vowel.

Recruitment was carried out both in person and online, using various Scandinavian and Swedish networks. Posters and flyers were distributed in areas where Swedish speakers were likely to move, including university grounds, Swedish cafés and shops, IKEA, and a Nordic research conference. The study was also advertised online through local Swedish conversation and appreciation groups, student societies and expatriate networks, mainly through social media. Calls for participants were sent out to university students via internal email and advertisement systems. The sampling strategy thus consisted of a mixture between convenience and snowball sampling, which did not allow for much control over the distribution of participants, but did enable recruitment within a small, relatively hard-to-reach community for which probability sampling would not have been possible (Bryman 2012: 424).

Social media played an important role in the recruitment process, both in finding and communicating with potential participants, and for this reason it became as an early priority to establish an online presence for the project. This was done in the form of a Facebook page which provided general information about the study, tracked its progress, and acted as a convenient point of contact for interested parties. The page was updated regularly to ensure visibility, and the outreach of each post could be tracked through a feature on the website. After the recruitment process, the page was maintained in order to promote public engagement with the project.

Swedish speakers who expressed interest in participating were sent an information sheet outlining the purpose of the study, the data collection method, and the conditions for taking part (available in the Appendix). They were not informed about the exact focus of the research, but knew that it concerned tongue movements and Swedish language. Those who chose to participate were then able to schedule themselves in for recordings through an online calendar software.

3.1.2. Participant sample

18 participants were recorded in total, but five were excluded due to technical problems, or because they did not fulfil the recruitment criteria. The remaining 13 were native Swedish speakers from Eastern and Western Central Sweden, 10 females and 3 males, between the ages of 18 and 27. Seven of these were from Stockholm, two from Gothenburg, two from Varberg (80 km south of Gothenburg), one from Jönköping (145 km east of Gothenburg), and one from Katrineholm (140 km south-west of Stockholm). The geographic distribution of the participants is shown in Fig. 19.

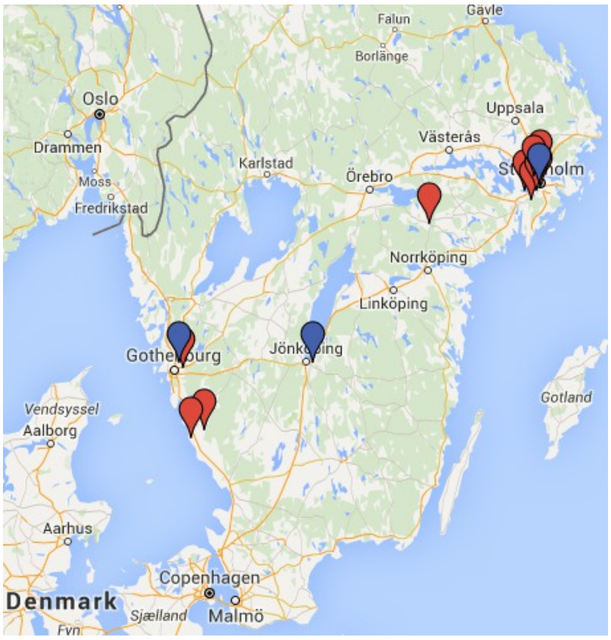


Fig. 19: Geographic distribution of participants.
Females in red, males in blue.

Almost all participants self-identified as having an accent that belonged in their native region. The exceptions were GBG_YM_01, who spoke Gothenburg multiethnolect, and VBG_YF_01 and KAT_YF_01, who were evaluated by the researcher as speaking regional standard varieties. Speakers from different cities within the same dialectal region, e.g. speakers from Gothenburg, Varberg and Jönköping, did not necessarily speak the same dialect, but their dialects could be classified as subgroups of the broader dialectal group, in this case Western Central Swedish. Six speakers reported having lived at other locations in Sweden, but for all but one speaker, the new location was in the same dialectal region and less than 60 km from their home. The speaker that had moved further had relocated from Gothenburg to Southern Sweden, but had still retained her native dialect.

Several participants reported other languages being spoken in the home during their childhood, e.g. English, French, Finnish and Lingala, but only one person stated that they themselves spoke a language that was not Swedish or English, and this language was not spoken in the home. All participants were fluent in English and were living in Scotland at the time of recording, with the time spent abroad ranging from 5 months to 8 years (average 3.5 years). All speakers had retained their fluency in spoken Swedish and did not have any phonetic interference from English, despite some problems recalling certain Swedish vocabulary items.

As a result of the recruitment strategy, most participants were international undergraduate students, forming quite a homogeneous group in terms of sex, age and socioeconomic group. For ethical reasons, participants had to be over 18, but were still classified as young, with the oldest speaker being 27 and the mean age being 22.2 years. Most speakers were from a high socioeconomic background, with an average score of 3.4 on a scale from from 1 (lowest) to 4 (highest). The score was arrived at by locating the parents' occupations in the SEI index (Statistiska Centralbyrån 2012), assigning them an associated “skill level” on a scale of 1-4, and averaging this number between the two parents in order to create a score that was also valid for single-parent households. The parents' occupations were used because most of the participants themselves had never worked in a position that they considered a part of their career.

A summary of the demographic distributions for city, dialectal region, metropolitan vs. urban location, age, sex and socioeconomic rating within the sample is provided in Table 5.

Speaker	City	Region	Metro/urban	Age	Sex	Socioeconomic score
GBG_YF_01	Gothenburg	West	Metropolitan	23	F	4.0
GBG_YM_01	Gothenburg	West	Metropolitan	20	M	1.5
JON_YM_01	Jönköping	West	Urban	22	M	3.8
KAT_YF_01	Katrineholm	East	Urban	27	F	3.5
STH_YF_01	Stockholm	East	Metropolitan	21	F	3.5
STH_YF_02	Stockholm	East	Metropolitan	19	F	4.0
STH_YF_03	Stockholm	East	Metropolitan	19	F	4.0
STH_YF_04	Stockholm	East	Metropolitan	24	F	3.3
STH_YF_05	Stockholm	East	Metropolitan	21	F	3.5
STH_YF_06	Stockholm	East	Metropolitan	21	F	4.0
STH_YM_01	Stockholm	East	Metropolitan	20	M	4.0
VBG_YF_01	Varberg	West	Urban	26	F	2.5
VBG_YF_02	Varberg	West	Urban	26	F	3.0

Table 5: Demographic distribution of participants. Class is ranked from 1 (lowest) to 4 (highest).

The small size and unbalanced nature of the sample means that conclusions drawn from the speech data must be tentative, however the data is still valid for an exploratory analysis and in order to pilot new articulatory methods involving ultrasound.

3.1.3. Materials

This section will describe the prompts and questionnaire questions used to elicit data in the study, and explain the rationale behind their design. Lists and images of the prompts, and a copy of the questionnaire can be found in the Appendix. The term 'vowel set', refers to the vowels /iː/, /eː/, /ɑː/, /uː/ and /ʉː/, with /iː/ being the focus of the study, and /eː/, /ɑː/, /uː/, /ʉː/ acting as corner vowels for context.

Picture description

Prompts were divided into three groups based on the speech task for which they were used. Spontaneous speech was collected through a picture description task, using custom-drawn pictures of different scenes, e.g. animals in a zoo, or things happening inside a house. In order to provide some control over the token count for each vowel, specific target words were 'planted' in the images, so that altogether the pictures contained 20 things with an /iː/ vowel, 20 things with an /eː/ vowel, and so on for the rest of the vowel set. The target words were not controlled for consonant environment, as a varying consonant environment was seen as more representative of natural speech. The words used were basic vocabulary items, represented by salient, stereotypical depictions in the images in order to elicit the intended word. There were five pictures in total, and participants were asked to describe them in as much detail as possible.

Midsagittal word list

The second set of prompts were used for the midsagittal word list. The list consisted of 35 words, seven for each vowel, with the vowels presented in isolated, word-initial, word-medial and word-final conditions, e.g. <i> (letter), <iver> 'eagerness', <bibel> 'bible' and <kemi> 'chemistry'. The words were mono- or bisyllabic with the vowel in stressed position, and were spoken in isolation rather than in a carrier sentence to save time. The consonant frames were non-lingual, using /p/, /b/, /f/, /v/, /m/ or /h/ to avoid coarticulation between the consonant and the vowel. Most of the words were relatively common, and some were related to specific themes, e.g. animals, emotions, or technology. The thematic links were a strategy for distracting participants from the vowels.

For context, each item on the word list was accompanied by an illustration, again custom-drawn for the study. The words were also selected to be as unambiguous in spelling as possible, so that the pronunciation would be obvious even if the participant did not know the word. In hindsight, the only word which caused difficulty was <emu> (bird), which by some speakers was pronounced /'eːmɵ/ rather than the target /ɛ'mʉː/. It was not clear whether this was caused by dialectal differences or unfamiliarity with the word, but non-target tokens were nevertheless excluded as the change in stress affected affected the vowel quality of the target vowel /ʉː/.

The word list order was determined by an online randomising software before it was loaded into the recording program, Articulate Assistant Advanced (AAA, Wrench 2014), as the program does not have a randomisation function. The words were thus presented in a set order, which will have increased the likelihood of list intonation and other order effects. On the other hand, the order of the words may have reinforced thematic links, distracting participants from the focus of the study.

The word list was repeated three times, giving a token count of 21 tokens of each vowel for each speaker, and 105 total vowel tokens per speaker.

Coronal word list

Coronal recordings used a shorter word list, consisting of nine words which were either real vocabulary items or nonsense words. Only the vowels /iː/, /eː/ and /ʉː/ were used, in order to keep the word list short while still maintaining some points of comparison for /iː/. The purpose of the coronal recordings was to observe whether tongue grooving or lateralisation was present. For reference, each of the vowels was therefore placed in a consonant frame with a combination of either /l/, which has lateralisation, or /s/, which has grooving, on either side, e.g. <lila> 'purple', <sisa> (nonsense word), <sila> 'to sift'. The words were bisyllabic with the vowel in stressed position, and were not accompanied by pictures as not all of them were real words. The coronal word list was also repeated three times, producing 9 tokens per vowel, and 27 tokens per speaker.

Questionnaire

The questionnaire collected demographic information about the participants in relation to their age, sex, regional background, language background, and socioeconomic background. Regional background was investigated in relation to the city where the participant grew up, which part of the city they grew up in, whether they had lived anywhere else, and if so, for how long. 'Part of the city' was intended to provide information about how central the area was to the city, and could also have been used for socioeconomic indexing. The distance and amount of time spent away from home was used as an indicator of the likelihood that the participant's dialect had changed. Participants were also asked if they self-identified as having an accent that belonged in their native region.

Language background was investigated in terms of whether the participant had

Swedish as their first language, whether they were fluent in any other languages, and whether there were any other languages spoken in the home when they grew up. This was partially used to confirm that participants were indeed native speakers, and partially to explain any possible interference from other languages or dialects.

The participant's socioeconomic background was mapped using questions about the participant's level of education, their occupation (if they considered the work to be part of their career), their parents' occupations, and their parents' highest level of education. The education options were based on the Swedish education system, but provided English translations as many participants were studying or had studied abroad. The parents' occupation and education were considered more important for younger speakers, who were likely to be more economically mobile.

Participants were able to answer questions in as much detail as they wished, and were asked to draw a line through questions that they did not wish to answer. This was to prevent confusion between questions being intentionally left blank, and questions where the participant did not know the answer. For the same reason, 'Don't know' options were included where relevant, and many questions were divided into two parts, with the first one asking for a 'yes/no' reply, and the second one asking for more details. The questionnaire was written in Swedish, but a translated version is included in the Appendix. The participant's name was never written on the questionnaire, only their anonymised code.

3.2. Equipment and recording procedure

3.2.1. Ultrasound tongue imaging

The primary data collection method in the study was ultrasound tongue imaging (UTI). This is an articulatory technique which generates an image of the tongue surface when an ultrasound probe is held under the chin. The method is growing in popularity both within clinical and phonetic research, as it is able to produce a contour of the entire tongue surface in either midsagittal (profile) or coronal (front-facing) section, making it particularly well-suited to study speech sounds that use the tongue, and vowels in particular. This section will provide a brief description of how UTI works, describe how UTI data was recorded in this study, and finally list some of the benefits and drawbacks of UTI compared to other methods.

How UTI works

The ultrasound probe contains a chemical compound called lead zirconate titanate, a piezoelectric ceramic material. These ceramic materials generate an electric charge when they are deformed, and inversely deform when an electric field is applied to them (APC 2014). When the ultrasound machine is on, an alternating electric current travels to the probe, "caus[ing] each element to deform, which pushes [the] air molecules nearby and sets up [a] high-frequency … pressure wave" (Lawson 2007: 3), also known as an ultrasound wave. Ultrasound is defined as sound with a frequency over 20,000 Hz (the upper limit of human hearing). For UTI, this soundwave will have a frequency of around 5-8MHz.

The ultrasound wave propagates well when travelling through soft tissue and fluid, but the signal is absorbed by hard tissue such as as bone and cartilage, and loses energy in air. Thus, when the probe is placed under the chin, the ultrasound signal travels through the soft tissue of the tongue, and is reflected by the tissue-air barrier at the tongue surface (Fig. 20). The reflected soundwave is received by the ultrasound probe, deforming the piezoelectric ceramic elements and generating an electrical current, whose "voltage will vary depending on the intensity of the reflected wave" (ibid). The ultrasound machine interprets this electrical signal and transforms it into a video image, on which the tongue surface appears as a white line (Fig. 21). The palate and alveolar ridge can be imaged in a similar way, by asking the participant to swallow some water, or to press their tongue against the roof of the mouth, thus eliminating the air pocket. Hard tissue, e.g. the jaw and hyoid bone, sometimes cast shadows on the image (Fig. 21).

Fig. 20: Schematic representation of ultrasound tongue imaging.

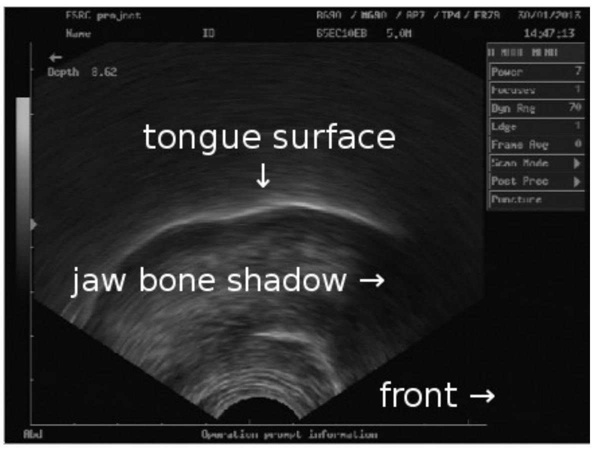


Fig. 21: Tongue surface and jaw bone shadow on the ultrasound image, speaker facing right.

Depending on the depth setting of the probe and the amount and type of tissue through which the ultrasound has to travel, the signal can vary in strength, affecting the quality of the image. Speakers with small heads and less fatty tissue under the chin usually produce better images, as the signal has to travel a shorter distance to reach the tongue surface. Young speakers also tend to image better than older speakers, partly because they have less tissue build-up. Speakers with beards can also be difficult to image, as the hair creates air pockets between the ultrasound probe and the skin. Ultrasound gel, which is used to maximise contact between the probe and the skin in all speakers, can mitigate against this as long as the beard is not too thick.

Probe stabilisation and standardisation

When recording UTI data for research, it is important to ensure that the probe stays in a set position in relation to the head, as probe movement will change the position of the tongue in relation to the ultrasound image, which means that multiple frames cannot be compared. It is also useful to standardise the angle of the probe between different speakers, so that their tongue surfaces are presented in the same orientation. In the present study, probe stabilisation was achieved using a stabilising headset, and the probe orientation was standardised using a bite plate. A palate trace, obtained by asking participants to swallow small sips of water, was also taken for each participant in order to gauge the tongue's proximity to the palate. The palate trace was obtained after any adjustment of the probe angle.

The stabilisation headset (Fig. 22) was manufactured by Articulate Instruments (Wrench 2008). It consists of an aluminium frame with padding against the cheekbones and sides and top of the head, and weighs approximately 0.8 kg (Wrench 2013). Most parts of the headset are adjustable to accommodate for different head sizes and shapes, and the strap which clips on at the back of the headset has a quick-release latch, making it easy to remove. The ultrasound probe is fastened in an adjustable holder near the base of the headset, and the headset used in this study was also mounted with brackets holding micro-cameras to record lip movement.

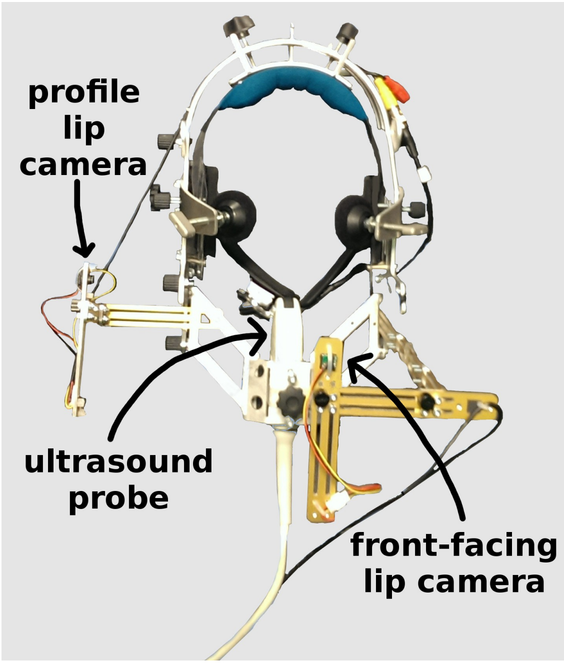


Fig. 22: UTI probe stabilisation headset with micro-cameras by Articulate Instruments (Wrench 2008).

The stabilisation headset "maintains the probe in the midsagittal plane and restricts rotational (yaw) and translational (roll) movement within the midsagittal plane" (Wrench 2013), while at the same time allowing for natural head and body movement (Scobbie et al 2008). While the headset itself is not uncomfortable, its weight combined with the weight of the probe can result in some discomfort after a prolonged period of time. Recording sessions with the headset were therefore kept to around 30 minutes.

Probe angle standardisation was achieved using custom-made bite plates of medical-grade plastic (Fig. 23). The bite plates measure approximately 100 x 40 x 2 mm, and have a small protrusion on the upper side near the middle of the plate. Participants were asked to insert half the bite plate into their mouth and bite down with their front teeth resting against the protrusion, and to press their tongue against the underside of the plate (Fig. 24).

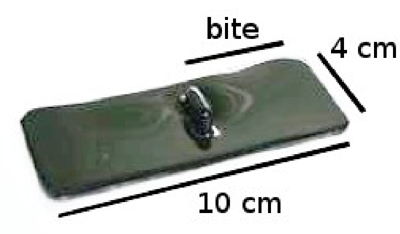


Fig. 23: Bite plate for probe angle standardisation.

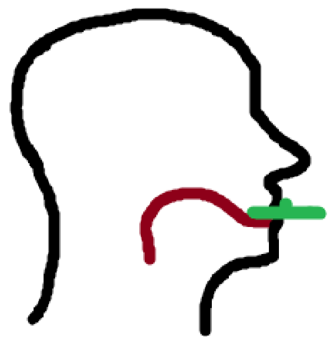


Fig. 24: Illustration of how the bite plate is used.

The tongue pressing against the bite plate reveals a flat surface, showing the speaker's bite plane, i.e. a plane that can be considered horizontal for each speaker. The probe angle was then adjusted so that the bite plane was horizontal in relation to the video image. A brief recording was made of the tongue against the bite plate in case further rotation of the data would be required afterwards.

Recording software

UTI, lip video and audio recordings were made using the software Articulate Assistant Advanced (AAA) (Wrench 2014). UTI and audio were synchronised automatically during recordings, and lip video was synchronised using an external synchronisation unit which puts a sound pulse on a secondary audio channel and superimposes a white square in the corner of the lip video image. AAA is then able to batch synchronise audio and video.

AAA was also used to present text and image prompts to participants, and recordings were either started and stopped manually (word list data), or set to continuous recording (spontaneous speech). Due to the great processing load of saving longer UTI recordings, however, continuous recordings could not be made longer than 15 seconds, followed by a 10-second data-saving window, as this would run the risk of overloading AAA's memory allocation on the computer, resulting in data loss. Pilot testing of the software indicated that the specified sampling interval provided the best trade-off between recording length and processing load. The restrictions on recording length did not cause any problems for the study, as around 15-20 samples were still obtained for each participant.

Once the data was recorded, AAA was used to manually draw splines onto the contour of the tongue, palate, and bite plate. To fit a spline to the midsagittal tongue surface, an individual video frame is selected, and a fan template is added to the image. The template consists of a fan grid with 42 radial lines (Fig. 25). These are not seen by the user at the time. The spline is then drawn across this fan template by clicking and dragging the mouse cursor. As this happens, the software specifies a crossing point along each of the fan lines, and the points are joined together as a smooth curve (Fig. 26).

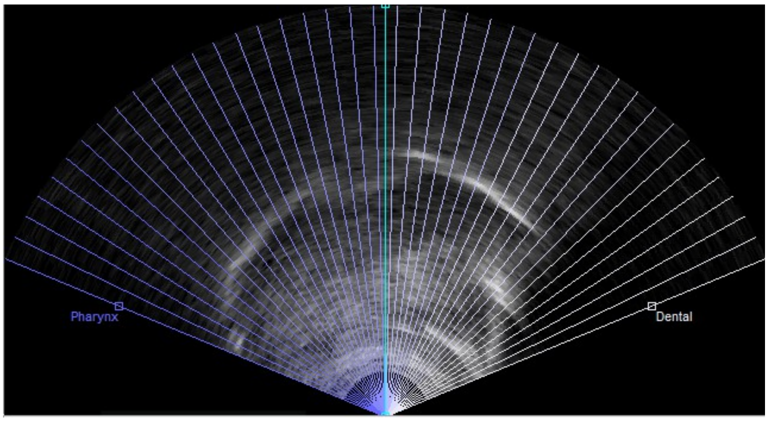


Fig. 25: Fan lines superimposed on the ultrasound image.

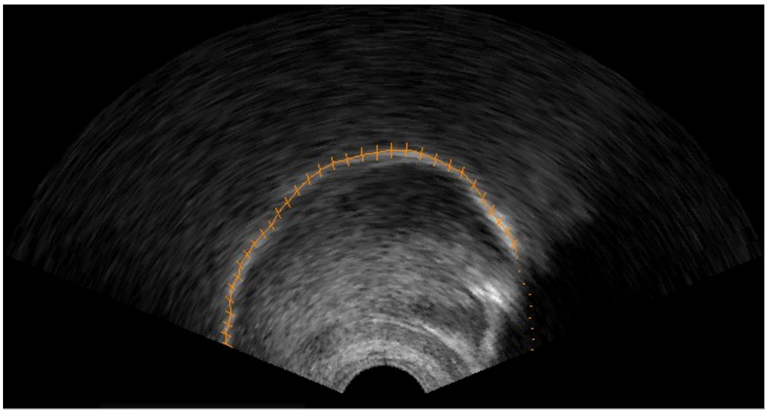


Fig. 26: Spline drawn along the tongue surface.

The splines were exported to a workspace, where they were rotated with reference to the bite plate to correct for probe angle (if necessary), and mean and standard deviation splines were generated for each vowel. Mean tongue surface splines were then exported as a series of x,y coordinates for further processing in the software R (R Core Team 2013). AAA also has a publisher function which can be used to plot and format tongue splines for export in image format.

For articulatory analysis, vowel tongue splines are usually drawn either on a selected video frame during the steady-state of the vowel, or at a specific timepoint corresponding to e.g. acoustic measurements. This study used the latter approach, firstly because it is possible for speech sounds to experience gestural delay (e.g. Lawson et al 2008), meaning that the steady-state or maximum tongue constriction does not coincide with the speech 'target', and secondly because the Swedish vowels were highly dynamic, and there was a possibility that the articulatory steady-state would have coincided with an acoustic end-phase, e.g. frication. To avoid this, annotations from Praat were imported into AAA, specifying the timepoint at which the acoustic measurements had been taken, and the tongue splines were then drawn at a corresponding timepoint.

Lip recordings

The lip data was not quantitatively analysed in this study due to time constraints; however it is possible to add fiducial (vertical) reference markers to the video image in AAA to quantify e.g. lip protrusion. As a preparatory measure, so that the lip data can be analysed in the future, a scale for the image was obtained by creating a short recording of each participant holding a ruler to their lips. A descriptive account of the lip data will be provided in the Results.

Coronal tongue recordings

Coronal tongue data was recorded for some speakers by rotating the ultrasound probe by 90º and following the same recording procedure as before; however, this data was not analysed either due to time constraints. An issue that arose during recordings was how to standardise the probe position during coronal data collection. As only a two-dimensional section of the tongue is visible at a time, it is difficult to know how far front or back in the mouth this section should be, and how to replicate this between different speakers. More work s thus needed to explore collection methods for coronal tongue data.

Benefits and drawbacks of UTI

UTI has several benefits compared to other articulatory methods. The main methods which will be used for reference here are electromagnetic articulography (EMA), which measures the movements of electrical sensors attached to the articulators, and magnetic resonance imaging (MRI), which creates cross-sectional images or 'slices', usually of the entire head, using a fluctuating magnetic field which measures the proportion of hydrogen atoms in different parts of the body.

One of the great benefits of UTI is that it can create an image of the entire tongue surface from root to tip, giving a holistic image of the tongue's shape. This is not the case with EMA, which only maps the few selected points along the tongue surface where sensors are attached. On the other hand, MRI images the entire vocal tract, including several articulators that cannot be seen with UTI, but it also tends to have a slow frame rate, causing image quality to deteriorate when recording fast movements. This can be a problem when recording spontaneous speech, which often features rapid or brief articulations. Depending on the type of machine used, UTI usually has a faster frame rate more suitable for studying spontaneous speech; the machine used in this study was set to 110 frames per second, which provided a good level of detail. There can sometimes be a trade-off between frame rate and image quality, with faster frame rates generating lower quality images, but this is generally not an issue with UTI.

UTI is not associated with any known health risks, as the machine is of the same type used in medical examinations and prenatal imaging. Anyone is eligible for imaging, and no preparation is required on the participant's part. For both MRI and EMA, a significant amount of time is required to prepare the equipment and the participant for recording, and not all speakers can be recorded with MRI, as any metal within the participant's body can pose a serious risk. The set-up time for UTI is much faster, as most preparations (fitting the headset, standardising probe orientation, obtaining a palate trace) can be carried out within a few minutes. Using UTI for live biofeedback (i.e. without the stabilising headset) requires virtually no preparation at all.

Although the standardisation process requires participants to use a bite plate, UTI in itself is a non-invasive technique. This makes it more comfortable for participants than EMA, which requires a small adjustment period for the participant to speak naturally with the sensors in their mouth. Compared to MRI, which requires participants to lie in a confined space where the noise level is high and where they may feel claustrophobic, UTI is less intimidating, and the participant can sit and move their body and head relatively normally. The supine position required by MRI also causes gravity to pull the tongue towards the back on the vocal tract, affecting both articulation and acoustics (Hoedl 2015).

UTI machines can be stationary or portable, allowing for data collection either in a laboratory setting or in the field. The visual output is intuitive to interpret and the method therefore has a strong public appeal. However, an important difference between UTI and EMA is that UTI only creates an image, whereas EMA tracks the absolute, physical movements of the articulators. This makes EMA data more quantifiable, with more objective and reliable measurements than the tongue splines used for UTI, which are often subject to the researcher's judgement. Although software algorithms are increasingly being used to draw tongue splines, they are not yet sophisticated enough to do so without manual correction.

Another issue with UTI is the difference in image quality based on physiological differences between speakers. In some speakers, particularly those with large vocal tracts, parts of the tongue can disappear from the image when sounds are produced that use a high tongue body or raised tongue tip. This can be because the ultrasound signal attenuates before it reaches the tongue surface, or due to a mandible shadow or an air pocket below the tongue. In some speakers it may similarly be difficult to image the palate due to its distance from the ultrasound probe.

Probe stabilisation can be an issue, but it is greatly mitigated by the use of the headset. However, if the probe is not fastened properly to the headset it may slip during recording, which usually means that the participant's data has to be discarded. In some cases, the data can be rescued by identifying palate traces whenever the speaker swallows between recordings. Some movement of the probe is inevitably present even when the stabilisation headset is worn, particularly when the participant is swallowing or pronouncing sounds with a great degree of jaw opening. If the participant swallows a large gulp of water for the palate trace, this can cause temporary displacement of the probe, increasing the distance between it and the palate. When comparing the palate trace to the tongue splines, it may then appear that the tongue is moving above and beyond the palate. This issue can be avoided by asking participants only to take small sips of water; in this study this was achieved by giving them a straw.

3.2.2. Recording equipment and set-up

Recordings took place at Queen Margaret University, Edinburgh, using two adjacent soundproofed rooms between which the equipment was divided. For clarity, a drawing of the set-up is provided in Fig. 27.

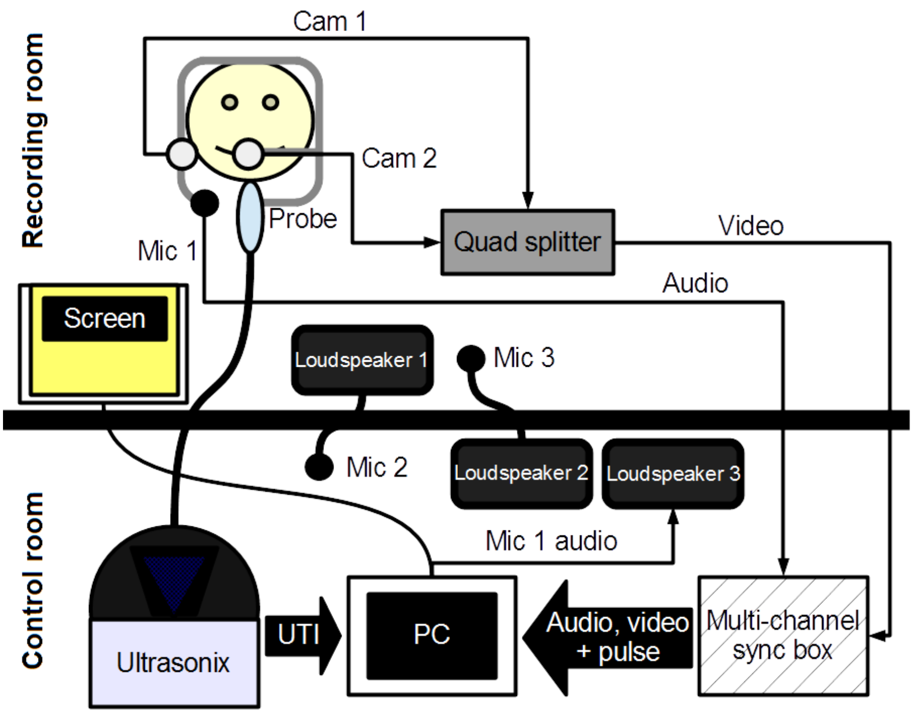


Fig. 27: Schematic drawing of the recording set-up.

The ultrasound machine was an Ultrasonix Sonix RP, with a probe frequency of 5MHz and a frame rate of 110 frames per second (non-interlaced). The machine was located in the control room, with the cable for the ultrasound probe running through a hole in the wall into the recording room, which was free of noise-making equipment. A 136º short-handle micro-convex probe was used, set to a depth of 80 mm. Also in the recording room were two micro-cameras recording in NTSC format with a frame rate of 29.97 frames per second, used for collecting lip video in profile and front-facing view. The two cameras were connected to a quad splitter, which enabled both camera inputs to be recorded by AAA as a single video source. For the audio recordings, an Audio-Technica clip-on condenser microphone was used. The microphone and cameras were attached to the ultrasound headset, and their input was processed by a multi-channel synch box before it reached the computer. As mentioned previously, the function of the synch box was to put a pulse on an alternative audio channel and a white square in the corner of the video image, which AAA could then use to synchronise audio and video post hoc.

Apart from the recording equipment, there was also a communication system between the two rooms, consisting of two microphones, one in each room, directly connected to a loudspeaker in the opposite room. This system enabled the researcher and participant to communicate with each other throughout the recording session. Prompts were presented to the participant on a computer screen which showed the same image as the computer in the control room. The recording room screen was partially covered so that the participant could only see the prompt, and not other aspects of the AAA recording window, e.g. video of their tongue and lips, which could potentially distract them and make them feel self-conscious.

3.2.3. Recording procedure

Before each recording session began, the participant was given an information sheet (see Appendix) and asked to sign a consent form (see Appendix). They were then assigned a code name based on their regional background, age, gender, and order of inclusion in the study, e.g. STH_YF_01 for a young female speaker from Stockholm. They were then asked to fill in the questionnaire.

Participants were told not to bring any potential noise-making objects, e.g. mobile phones, into the recording room. They were fitted with the ultrasound headset by the researcher, and the probe angle was adjusted using the bite plate. A palate trace was also obtained by recording the participant taking a few sips of water. The researcher self-piloted the use of sparkling water for this purpose, as the air bubbles would cause more reflections of the ultrasound signal, resulting in a brighter image. The ultrasound image was indeed strongly illuminated by the sparkling water, however it also caused some glare which was not useful for palate tracing. This glare could depend on the degree of carbonation of the water. The final step was to obtain a scale measure on the lip video, which was done by recording the participant holding a ruler against their lips in both profile and front-facing view.

After this, the researcher introduced the spontaneous speech task, which consisted of picture description. The spontaneous speech was collected before the word list speech in order to encourage more naturalistic production, based on Di Paolo & Yaeger-Dror (2011):

> Given the importance of the vernacular to sociolinguistic analysis, the tasks which focus on pronunciation should always be placed as late in the session as possible so that the conversation itself will be as untrammeled with self-conscious speech as possible.
>
> (Di Paolo & Yaeger-Dror 2011: 16)

To further promote informal, conversational speech, the researcher presented the pictures for the spontaneous speech task in person to the participant. This also served the purpose of being able to prompt them if they found the task challenging. The guideline for the task was "Describe everything you see in the picture in as much detail as possible", which was sometimes complemented with questions about depicted objects, colours, etc., or requests for the participant to "describe the picture like a story". This was usually in response to participants struggling with the task or developing list intonation, e.g. "I see X, I see Y, I see Z". During the spontaneous speech task, AAA was set to automatic continuous recording, sampling 15 seconds of speech at 10-second intervals. The participant was unaware of the sampling procedure and could carry on speaking normally.

The spontaneous speech was followed by a word list task, which was carried out with the researcher in the control room. This arrangement was due to the fact that each individual recording had to be started and stopped manually, which required a full view of the AAA recording screen. The participant was given initial instructions in person, and then received feedback through the microphone-loudspeaker system. Each word in the word list appeared on the screen in front of the participant accompanied by an illustration, and after a short delay the background turned green, prompting the participant to speak. The word list was randomised but presented in the same order for each participant, as AAA does not have a randomising function. The word list was read three times, and after each reading the participant was asked if they were happy to continue, as the weight of the ultrasound headset can cause tension after prolonged wearing.

If the participant was also happy to record a coronal word list, this was carried out by rotating the ultrasound probe by 90º, once again making bite plate and palate trace recordings, and then using the same procedure as for the previous recordings, recording three readings of the coronal word list. Once this was finished, the session was ended and the headset removed. Some participants were asked what they thought the focus of the study was. Generally, answers were linguistics-related but not oriented towards vowels or particular speech sounds. The experimental setting may have caused participants to produce careful speech with more stereotypical vowel qualities, however "vowel variation generally occurs below the level of conscious awareness" (Di Paolo et al 2011: 87), and it is not believed that participants significantly altered their vowel qualities.

3.2.4. Ethics

Ethical approval was obtained from the University of Glasgow before recruitment began. This ensured that the study met ethical requirements in relation to participant safety, data collection and handling, anonymity and confidentiality, acquisition of informed consent, and compensation. Each of these points will now be covered briefly.

Participant safety

No compromises to the health and safety of participants were expected or experienced during the study. UTI is a safe, non-invasive method which is not associated with any known risks, and recordings took place at university premises with a supervisor present. Participants were also required to be signed in as visitors at the university. The ultrasound stabilisation headset can sometimes cause discomfort after prolonged use due to its weight, but participants' comfort was ensured by limiting recording sessions to around 30 minutes, and asking regularly if they were comfortable. The headset had a quick-release catch, which meant that it could be removed easily at any time. The ultrasound image was never used to comment on the participant's health, as the researcher does not have relevant training for this. All communication with the participants was in Swedish, to ensure that they were at ease and able to express themselves clearly.

Data collection and handling

Some personal demographic data was collected in the questionnaire, with the purpose of identifying potential correlations between demographic factors and pronunciation. The questionnaire data was therefore kept anonymous and confidential, and individuals were only associated with labels of age, sex, geographic background, language background and socioeconomic background. Participants were not obliged to answer questions on the questionnaire, and were free to provide as much or as little detail as they liked. Once digitally recorded, the filled-in questionnaires were destroyed.

The recordings did not collect any personal information in terms of speech content, neither is UTI sufficient to identify individuals. Lip videos only recorded the participant's mouth, not the face as a whole. It would hypothetically be possible for someone who knows a participant well to identify them from their lip video and/or audio recordings, but this is unlikely, and it is also unlikely that this would affect the participant in a negative way.

Anonymity and confidentiality

Participants were anonymised from the outset of data collection. Data was stored securely on password-protected university drives, and the anonymous recordings and questionnaire data was kept for future academic and educational use. During the recruitment process, the anonymity of participants was preserved by blind-copying multiple participants into group emails and moving conversations on social media to a private medium as soon as possible. Participants' contact details were deleted once the participant had been recorded.

Informed consent

Informed written consent was obtained from all participants, who were first given information in Swedish about the purpose and procedure of data collection, including information about the method, data handling, anonymity and conditions for taking part. Participants were also given the researcher's contact details, and the details of the project's supervisors. Participants were reminded that their participation in the study was voluntary, and that they were free to withdraw at any time without having to state a reason.

Compensation

Recording sessions were approximately 45 minutes long, however some participants travelled for up to 1½ hours to be recorded. For this reason, they were offered compensation for their return travel costs, including petrol costs if applicable. Participants were also given a gift of £5 as an inducement to take part in the study and to thank them for their time.

3.2.5. Data preparation and analysis

Due to time constraints, only a selection of the collected data could be analysed in this study. The selection consists of UTI, audio and lip video recordings of the midsagittal word list, alongside demographic information from the questionnaire. The spontaneous speech and coronal word list data will remain available for analysis at a later date. This section will describe the methods used to prepare and analyse the data selected for the study.

Acoustic data

Audio recordings were manually segmented in Praat (Boersma & Weenink 2015) using two interval tiers: a broader tier encompassing the entire vowel unit, including e.g. diphthongisation and end-frication, and a narrower tier specifying a section within the vowel where the formants were relatively clear and stable. An example of the two tiers is provided in Fig. 28. For compatibility reasons, the vowels were not annotated with IPA symbols, but followed the system shown in Table 6. This system will be used consistently throughout the thesis.

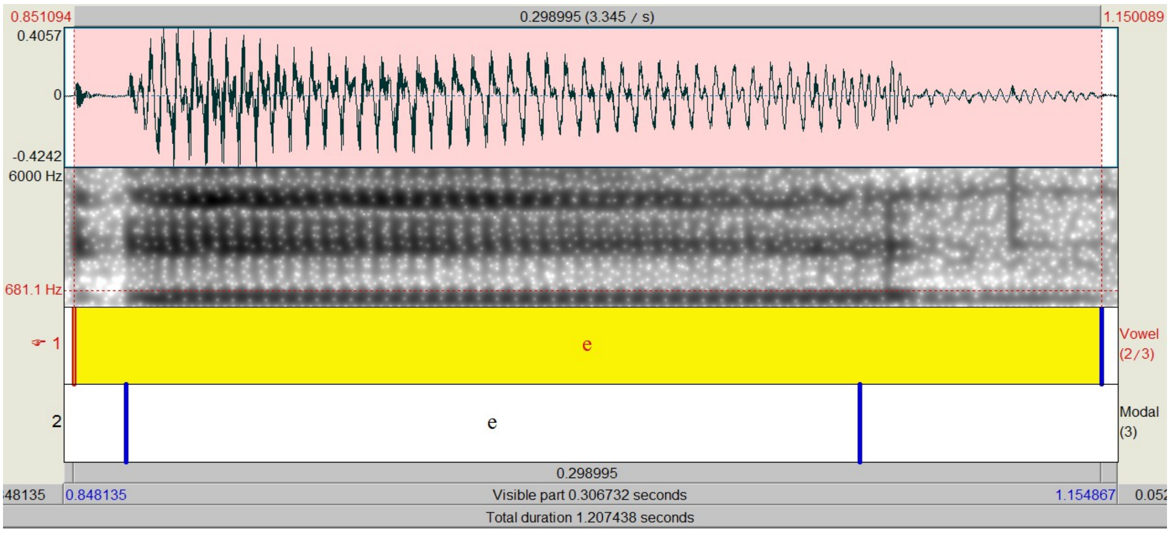


Fig. 28: Example of broad (top tier) and narrow (bottom tier) vowel annotation in Praat.

Vowel	Annotation
/ɑː/	a
/eː/	e
/iː/	i
/uː/	u
/ʉː/	uu

Table 6: Annotation system used for vowels in this study.

Segmentation in the broad vowel tier followed strict and consistent rules. Boundaries were always placed at the zero-intercept of the whole waveform cycle. In initial position, the vowel was said to begin with the first pulse on the waveform, excluding clicks and other productions unrelated to the word. After the fricatives /f/, /v/, /h/, the vowel began when the waveform developed regular cycles and the spectrogram showed formants rather than fricative noise. After the plosives /p/, /b/, the vowel began after the burst and after any post-burst aspiration, again when the waveform became regular and the spectrogram developed formants. After /m/, the vowel began when the waveform cycles developed a more vowel-like shape with several peaks, and when the formants (primarily F2) had stabilised after the upwards formant transition from /m/.

When vowels were followed by /f/, /v/, /h/, segmentation depended on whether or not the vowel was prone to end-frication. If it was, the waveform and spectrogram were closely observed for changing patterns, and listening to the sample was a key part of determining the boundary between end-frication and the following fricative consonant. The intensity curve in Praat was also used as an indicator of this boundary. If the vowel was not prone to end-frication, the vowel ended when the waveform ceased to be cyclical and when formants were no longer visible on the spectrogram. When /p/, /b/ followed the vowel, the vowel ended when the closure (silent phase) of the plosive began. When /m/ occurred after the vowel, the vowel ended when the cycles on the waveform developed a more nasal-like shape with fewer, smoother peaks, and after the downward formant transition (primarily of F2) into /m/. The transition was included as the /m/ tended not to be audible during this portion. In isolation, the vowel ended when the intensity curve approached zero, but waveform shape and formant visibility were also consulted.

The narrow vowel tier was segmented largely based on the stability and strength of the formants. The purpose of the tier was to provide a stable portion of the vowel which was representative of the target vowel quality, as the vowels overall were found to be highly dynamic. However, the dynamic properties also made it difficult to determine boundaries for the narrow vowel segments, due to the variable presence and duration of certain features in some tokens or speakers but not in others. For example, some tokens had very stable and visible formants, while others were weaker and more irregular (Fig. 29). Some speakers had dynamic properties which others did not, e.g. /eː/ always had a diphthongal offglide, but in some cases it also seemed to have an onglide phase (Fig. 30). For some tokens, end-frication could also be extremely long (Fig. 31), whereas other tokens had virtually no end-frication at all.

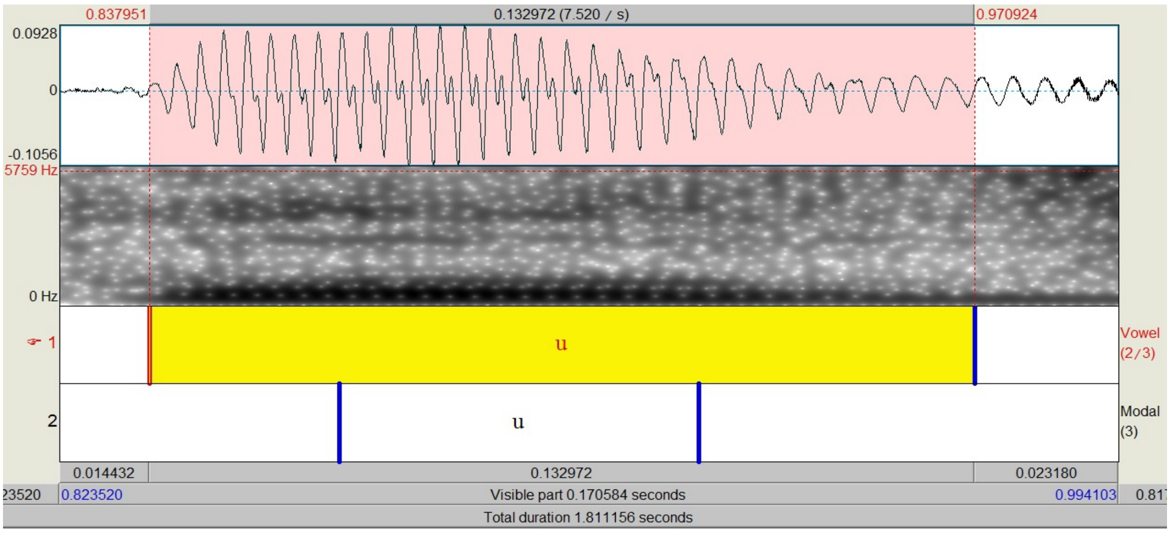


Fig. 29: /u:/ token with weak and irregular formants.

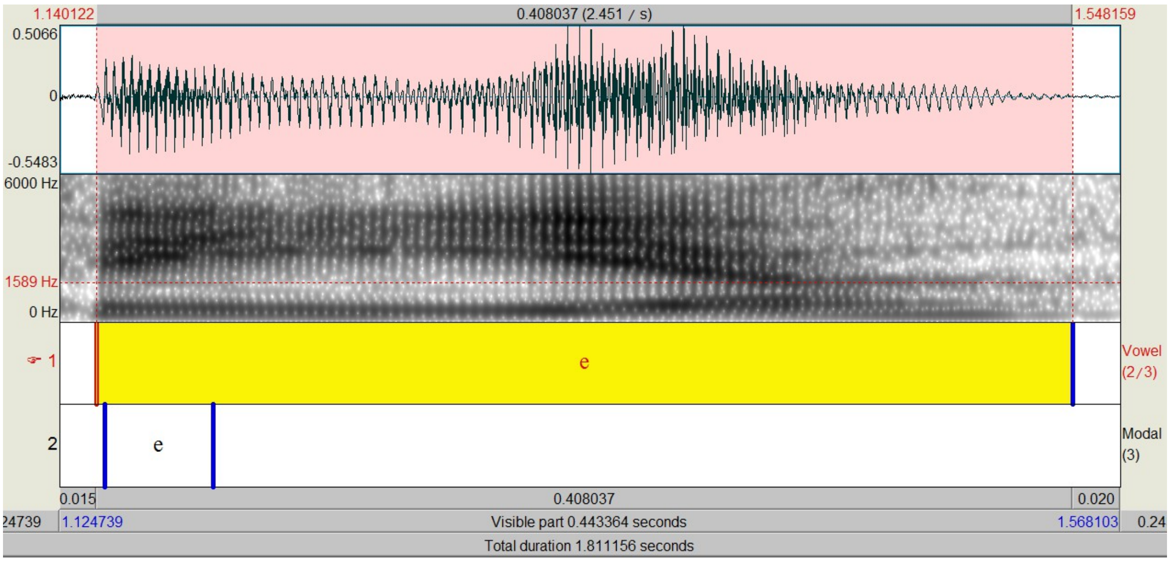


Fig. 30: /e:/ token with apparent onglide, offglide, and a stable portion in between.

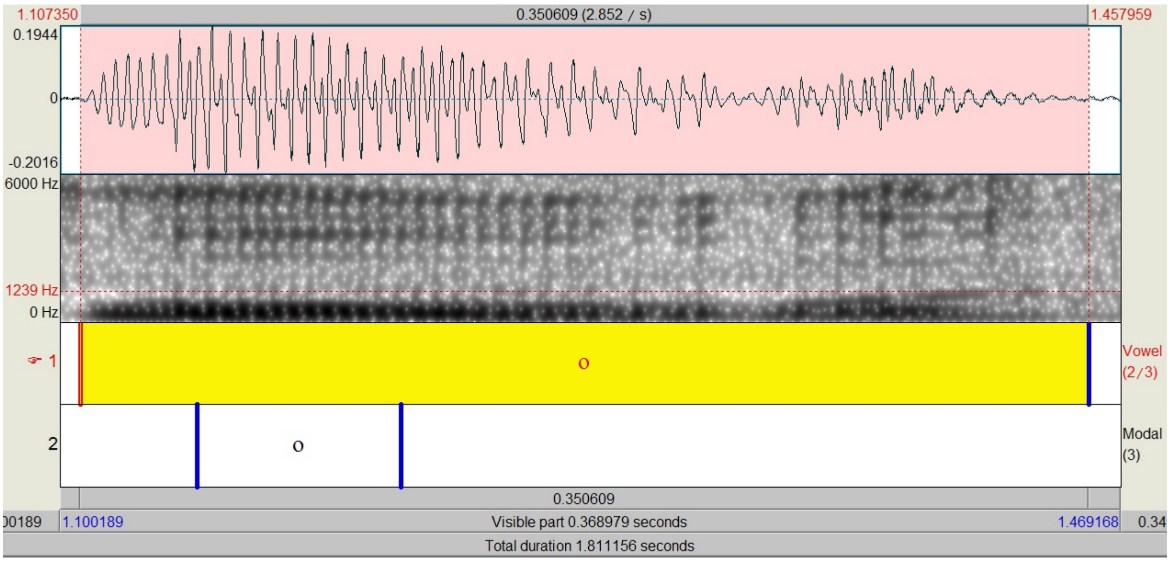


Fig. 31: /o/ token with long fricative end-phase. The visible creak is typical for the speaker.

The variable dynamic properties of the vowels meant that the narrow vowel segments often differed greatly in length and position in relation to the broader vowel segment. In order for vowel measurement to be more consistent, the narrow vowel segments were therefore discarded, and the broader ones were used instead. In both tiers, recording errors sometimes resulted in vowels being cut off. These vowels were annotated with a dash, e.g. “o-” and were also discarded, as they too varied both in duration and in how much of the vowel they included, which would have made acoustic measurements less consistent.

Once segmentation was finished, a Praat script (see Appendix) was used to take acoustic measurements for F1, F2 and F3 at a single point in the vowel. The single-point measure was used in order to save time, and to prevent inclusion of highly dynamic elements in the vowel measure. Measurements were made one-fourth into the broad vowel segment for /iː/, /eː/, /uː/ and /ʉː/, which gave the vowel enough time to stabilise, but not to develop end-frication or diphthongisation. For /ɑː/, the measurement was taken at the midpoint of the broad vowel segment, as its formants were generally more stable. Measurement points were manually inspected by the researcher in a small sub-sample of the data to ensure that it was reliable. The Praat script was set to a frequency range of 0-4,500 Hz for males and 0-5,500 Hz for females. These ranges were found to produce fewer outliers in the formant measurements compared to other ranges that were tested. The script also produced TextGrids containing a point tier which specified the time at which measurements had been taken, so that these could be matched with the UTI data, or used for remeasurement if necessary.

The F1, F2 and F3 values were collated in a spreadsheet and checked for statistical outliers by calculating the lower and upper bound for the outliers (+/- 1.5 x the interquartile range) within each formant, vowel, and speaker, e.g. within F1 for /eː/ in GBG_YF_01. Lower and upper bound were calculated using the following formulas:

$$\text{Lower bound} = \text{Quartile } 1 - ((\text{Quartile } 3 - \text{Quartile } 1) \times 1.5)$$
$$\text{Upper bound} = \text{Quartile } 3 + ((\text{Quartile } 3 - \text{Quartile } 1) \times 1.5)$$

Values that fell below the lower bound value, or above the upper bound value, were classified as outliers. These were manually remeasured in Praat using the formant listing function at the same timepoint as the previous measurement. If the F1 or F2 values were still outliers, the entire token was discarded. If F3 was still an outlier, the F3 value was deleted from the data set, but F1 and F2 for the token were kept, in order to maximise the token count for F1/F2 analysis.

The spreadsheet with the F1, F2 and F3 values was saved in CSV format, in the layout required for analysis with the NORM vowel normalisation suite (Thomas & Kendall: 2007) and the associated 'vowels' package in the software R. A script developed for the 'vowels' package was then used to plot the acoustic data and to generate descriptive statistics, which will be presented in the results chapter.

The acoustic data was normalised using the Watt & Fabricius method (2002), which was found to perform very similarly to Lobanov, but was more successful in reducing scatter in the high front section of the vowel space. Watt & Fabricius normalises by comparing vowel tokens to a centroid measure, which is located in the centre of a triangle connecting the three corners of the vowel space, usually /i/, /ɑ/ and /u/ (ibid). The 'vowels' package automatically identifies the corner vowels, which in this study were found to be /e/, /ɑ/ and /u/.

Auditory data

As a complement to the acoustic analysis, the study also wanted to provide auditory ratings for the different speakers' /iː/ vowels, in order to give an auditory view of whether speakers used Viby-i, and if so, whether their Viby-i was weak or strong. These auditory ratings could then be compared to the acoustic and articulatory data. In order to obtain auditory ratings, a Praat multiple forced choice experiment was created in which listeners were played Swedish /iː/ tokens and asked to rate the /iː/ on a scale from 0 (no Viby-i) to 4 (strong Viby-i). The /iː/ tokens which were presented as stimuli were three separate productions of the word /'biːbɛl/ 'bible' for each speaker, making up a total of 39 tokens. A more comprehensive experiment could not be carried out due to the limited time scope of the study. Auditory ratings were provided by four listeners who had phonetic training but did not speak Swedish, as they were believed to be more sensitive to the difference between [iː], which existed in their first languages, and Viby-i, which was unfamiliar to them. There was also the constraint of finding available Swedish listeners in time.

Before the experiment, listeners were played examples of weak and strong Viby-i, and the degree of Viby-colouring was discussed impressionistically in terms of the vowel's 'thickness' or 'darkness'. The stimuli were presented in randomised order, but listeners were able to replay each word as many times as they liked. Once the experiment was complete, the auditory ratings were collated, and the researcher inspected the results to ensure that inter-listener rating was relatively consistent, and that there was no intra-listener bias. Mean auditory ratings were then calculated for each Swedish speaker's /iː/ vowel.

Articulatory UTI data

AAA was used to draw tongue splines, palate traces and bite plate traces on the UTI data, as described in the 'Recording software' section above. The TextGrids indicating the timepoint of the acoustic measurements were imported into AAA, and tongue splines were drawn on the UTI data at a corresponding time. Due to a bug in AAA, tokens which had unusual characters in the prompt, such as *é, å, ä*, could not be imported. One environment for /ʉː/, one environment for /uː/ and two environments for /eː/ therefore had to be excluded from the articulatory data. A small number of tokens were also excluded due to recording errors. A full token count is provided in the summary at the end of this chapter.

When spline-fitting was completed, each speaker's tongue, palate and bite splines were exported to the AAA workspace and rotated where necessary, ensuring that participants tongue surfaces were presented in the same orientation. Mean and standard deviation splines were created for each vowel using an automatic function in AAA. The Cartesian coordinates of these tongue splines were then exported as a text file, and an R script was run which plotted the mean tongue curves on a scale representing the absolute distance from the ultrasound probe to the tongue surface in mm. The script was also used to calculate the highest point of the /iː/, /eː/ and /ɑː/ splines, and the backest points of the /iː/, /eː/ and /uː/ splines. The normalised position of /iː/ could then be expressed as a proportion of the distance between /eː/ (the highest vowel) and /ɑː/ (the lowest vowel) in the vertical dimension, and as a proportion of the distance between /eː/ (the frontest vowel) and /uː/ (the backest vowel) in the horizontal dimension. These normalised measures were preferred over raw distances in the analysis, as the various speakers' vocal tracts were very different in size. The R script for plotting, measuring and normalising tongue splines is available in the Appendix.

Articulatory lip data

Due to time limitations, the lip data could not be analysed in full for this study, but a small selection of still frames for each vowel will be presented as a qualitative, exploratory supplement to the UTI data. The images were taken from the point of maximum lip constriction for the words <bibel> /iː/, <feber> /eː/ and <puma> /ʉː/. A larger set, including <fabel> /ɑː/ and <hovar> /uː/, is available in the Appendix.

Demographic data

The demographic data was collated from the questionnaires and used to create groupings within the data so that auditory, acoustic or articulatory phenomena could be investigated in relation to demographic factors. The geographical factors were dialect region (Eastern Central Swedish or Western Central Swedish) and metropolitan vs. urban environment. Stockholm and Katrineholm were classified as East, whereas Gothenburg, Varberg and Jönköping were classified as West. Gothenburg and Stockholm were classified as metropolitan, and Varberg, Jönköping and Katrineholm were classified as urban. The cities in the urban category all have populations under 150,000 people and are located outside the immediate vicinity of a larger city, falling into the Swedish category *tätort*, roughly equivalent to 'small city'. The distinction between metropolitan and urban was consciously made so as not to include rural areas where Viby-i would be a genuine dialectal feature.

The social factors under investigation were sex and socioeconomic group. The latter was either expressed as a continuous variable, or in terms of groupings based on socioeconomic score, where 1-1.99 = lower class, 2-2.99 = middle class, and 3-4 = higher class. These groupings are not necessarily indicative of the full range of socioeconomic class in Sweden, nor are they translatable to systems in other countries, and so they should only be taken to indicate relative positions on the 1-4 scale. Age was not used for social grouping, as all participants in the sample were young (aged 18-27).

3.2.6. Summary of data sample

In summary, the data analysed in this study consists of UTI, audio and lip recordings of the midsagittal word list for 13 speakers. The acoustic data set was largest, with 1,199 tokens, the distribution of which are shown in Table 7. The articulatory UTI data set was slightly smaller due to data collection and software issues, consisting of 1,121 tokens, shown in Table 8. The distributions of both acoustic and UTI tokens in terms of number of tokens per vowel and speaker are relatively balanced.

The auditory ratings consisted of a small subset of the data, with 3 /iː/ tokens per speaker, all produced in the word /'biːbɛl/ 'bible'. There were thus were 39 stimuli tokens, each rated by four different listeners.

The articulatory lip data also used a sub-sample, consisting of one token per speaker for each of the vowels /iː/, /eː/ and /ʉː/, spoken in the words /'biːbɛl/ 'bible', /'fɑːbɛl/ 'fable' and /'pʉːma/ 'puma'.

Speaker/Vowel	a	e	i	u	uu	Total
GBG_YF_01	15	16	17	18	17	83
GBG_YM_01	15	19	18	17	19	88
JON_YM_01	20	22	17	18	16	93
KAT_YF_01	17	18	17	18	16	86
STH_YF_01	19	21	17	18	18	93
STH_YF_02	20	21	21	19	21	102
STH_YF_03	18	22	21	19	19	99
STH_YF_04	20	20	22	17	16	95
STH_YF_05	15	21	14	19	17	86
STH_YF_06	20	20	17	17	13	87
STH_YM_01	20	21	21	21	22	105
VBG_YF_01	18	15	14	17	18	82
VBG_YF_02	19	22	20	21	18	100
Total	236	258	236	239	230	**1199**

Table 7: Token count for speakers and vowels in the acoustic data set.

Speaker/Vowel	a	e	i	u	uu	Total
GBG_YF_01	18	13	16	14	14	75
GBG_YM_01	18	15	21	14	17	85
JON_YM_01	20	16	20	18	15	89
KAT_YF_01	18	14	17	16	15	80
STH_YF_01	18	14	19	14	18	83
STH_YF_02	22	16	21	18	18	95
STH_YF_03	18	16	21	18	17	90
STH_YF_04	21	15	21	17	14	88
STH_YF_05	20	15	21	13	17	86
STH_YF_06	20	14	19	17	12	82
STH_YM_01	21	15	21	18	19	94
VBG_YF_01	18	15	17	17	15	82
VBG_YF_02	19	17	20	18	18	92
Total	251	195	254	212	209	**1121**

Table 8: Token count for speakers and vowels in the UTI data set.

3.2.7. Statistical analysis

Statistical analysis was carried out using three different types of tests. Comparisons between group means were made using a Welch two-sample independent t-test, which measures whether the mean values of two independent groups are significantly different from each other. Correlations were investigated using Pearson's product-moment correlation coefficient (Pearson's r), which investigates the relationship between two variables by calculating a coefficient, showing the direction of the relationship, and by determining whether the relationship is statistically significant. Finally, a stepwise mixed-effects linear regression model was used in one instance to find out whether the the distributions of the different vowels in the acoustic data were significantly different from one another, in other words, that /iː/, /eː/, /ɑː/, /uː/ and /ʉː/ all clustered into distinct groups, regardless of speaker variability and linguistic environment.

Statistical significance was determined using the following values: $p>0.1$ = not significant, $p\leq0.1$ = marginally significant, $p\leq0.05$ = significant, $p\leq0.01$ = highly significant.

4. Results

This chapter presents the results of the study, beginning with the auditory results, and continuing with the acoustic and articulatory results. Within each section, an overview will be presented first, and the results will then be explored in more detail with reference to phonetic, regional and social factors. A summary will be presented for each section before moving on to the next, and a chapter summary will be provided at the end.

4.1. Auditory results

4.1.1. Overview

The mean auditory 'strength' ratings for each Swedish speaker, as produced by four non-Swedish speaking listeners, is shown in Fig. 32. Ratings ranged from 0 (no Viby-i) to 4 (strong Viby-i). The results have here been classified as follows: Values <1 = no Viby-i, values between 1 and 1.99 = weak Viby-i, values between 2 and 2.99 = moderate Viby-i, values between 3 and 4 = strong Viby-i.

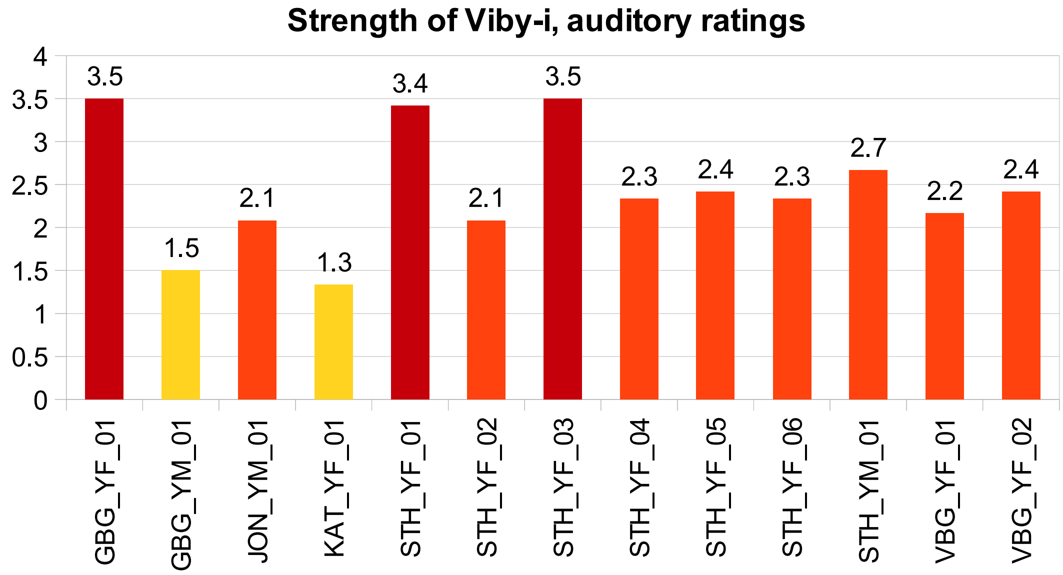


Fig. 32: Mean Viby-i 'strength' ratings for each speaker. Yellow = weak, orange = moderate, red = strong.

The ratings show that all speakers used some form of Viby-i, but that it varied in its perceived strength. The speakers can roughly be divided into three different groups: weak (yellow), moderate (orange), and strong (red). The variation seen in the sample supports the hypothesis that there are several types of Viby-i, and that some productions may be perceptually closer to standard [iː] than others.

4.1.2. Auditory strength by geographical groupings

Investigating the effects of dialect region, a comparison between speakers of Eastern vs. Western Central Swedish is provided in Fig. 33. The Eastern group is shown to have a higher mean auditory rating than the Western group, however a t-test for the two groups returns $t = 0.38827$, p-value $= 0.7074$. This means that the difference between the two groups is not statistically significant. In other words, there is no effect of dialect region on Viby-i strength.

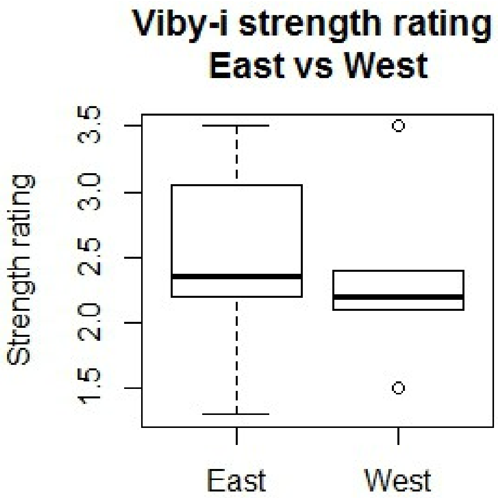


Fig. 33: Mean auditory strength ratings for the Eastern and Western Central Swedish groups.

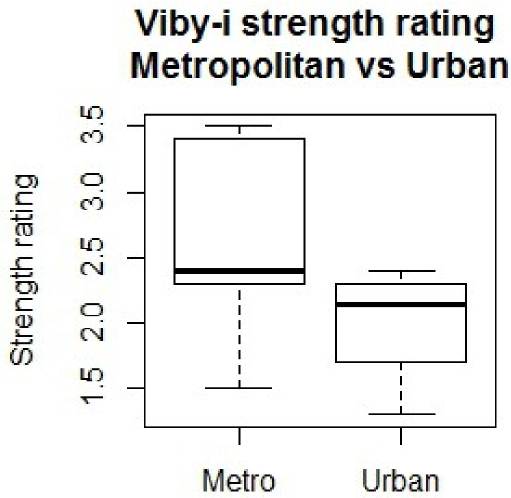


Fig. 34: Mean auditory strength ratings for the metropolitan and urban groups.

If the speakers are instead compared as metropolitan vs. urban groups, a different picture emerges. The mean auditory ratings for these groups are provided in Fig. 34. The metropolitan group has a higher mean rating than the urban group, and the difference is marginally significant: $t = 1.8836$, p-value $= 0.09432$. This suggests a trend for metropolitan speakers to have a stronger Viby-i than urban speakers. A similar trend can be observed in the summary auditory results (Fig. 32), where all three strong Viby-i users are from a metropolitan area. However, it should also be noted that GBG_YM_01 had a weak Viby-i despite belonging in the metropolitan group. Having said this, this participant was a multiethnolect speaker, and was noted during recordings to be variable in his /iː/ productions, using both weaker and stronger Viby-i. There may thus be an additional effect of ethnic speech variety which is not visible in this sample.

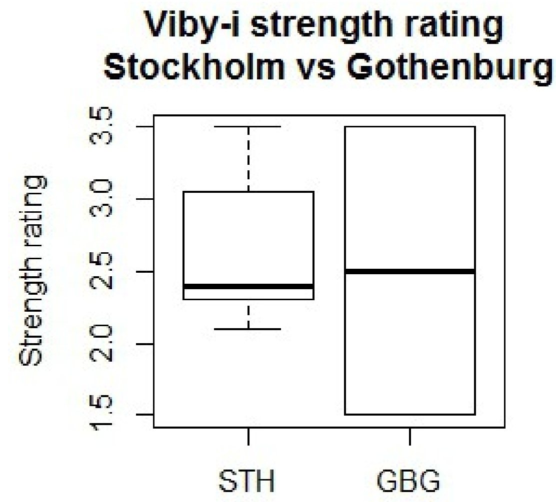


Fig. 35: Mean auditory strength ratings for the Stockholm and Gothenburg groups.

A comparison between Stockholm and Gothenburg speaker means (Fig. 35) shows that the Gothenburg speakers have a higher mean strength rating than the Stockholm speakers. However, a statistical test for this difference is not suitable as the groups are highly imbalanced. There are only 2 speakers in the Gothenburg group, versus 7 in the Stockholm group, and the Gothenburg mean value therefore likely to be disproportionately skewed by GBG_YF_01, who is a strong Viby-i user. For this reason, we cannot conclude that there is any difference between Stockholm and Gothenburg in Viby-i strength.

4.1.3. Auditory strength by social groupings

The mean auditory ratings for men and women are displayed in Fig. 36. Although the women are shown to have a higher mean rating than the men, it would again be unsuitable to carry out a statistical test, as the groups consist of 10 women and only 3 men. Looking back to the summary results for the auditory ratings (Fig. 32), all three of the strong Viby-i users were women, but this effect could be a result of other factors, such as the fact that a great number of the women in the sample were from metropolitan areas.

Finally, Fig. 37 illustrates the relationship between Viby-i strength rating and socioeconomic score. The results of the correlation test were $r = 1.890$, $p = 0.0854$, meaning that there is a positive relationship which is marginally significant. This suggests a trend for speakers from higher socioeconomic backgrounds to have a stronger Viby-i, but this conclusion must be tentative as the sample mainly consisted of speakers from higher socioeconomic backgrounds.

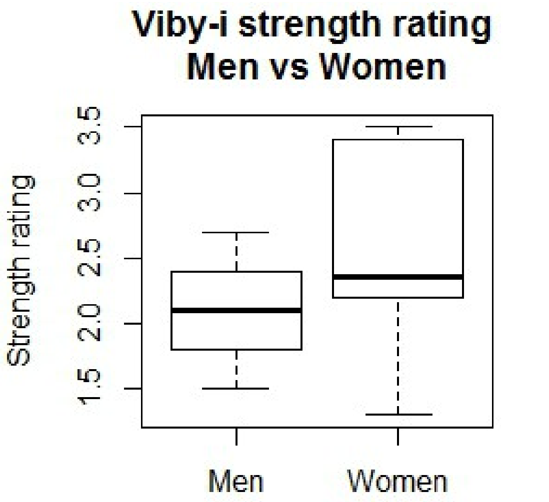


Fig. 36: Mean auditory strength ratings for men and women.

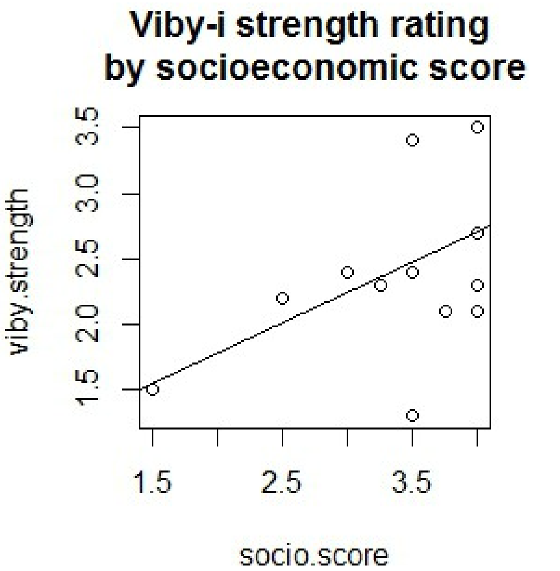


Fig. 37: Correlation between auditory strength rating and socioeconomic score.

4.1.4. Summary: Auditory results

The auditory results show that Viby-i was used by all speakers in the sample, but that it varied in auditory strength. The perceived strength of Viby-i was not affected by city (Stockholm vs. Gothenburg) or dialectal region (East vs. West), but a marginally significant effect was found for metropolitan vs. urban location, with metropolitan speakers having a higher mean rating for Viby-i strength. Similarly, a marginally significant result was found found for speakers of higher socioeconomic background to have a stronger Viby-i. No effects of sex were found; qualitatively, the highest auditory ratings were received by women, but it should be remembered that women also make up a vast majority of the sample as a whole.

4.2. Acoustic results

4.2.1. Overview

The normalised acoustic results (F1/F2) for all speakers and vowels are shown Fig. 38. Looking at /iː/, the most noticeable acoustic property is that it has a low F2 compared to /eː/, with a small degree of overlap with some /eː/ tokens, and a greater degree of overlap with /ʉː/. The F1 value of /iː/ is similar to /eː/, and slightly lower than /ʉː/.

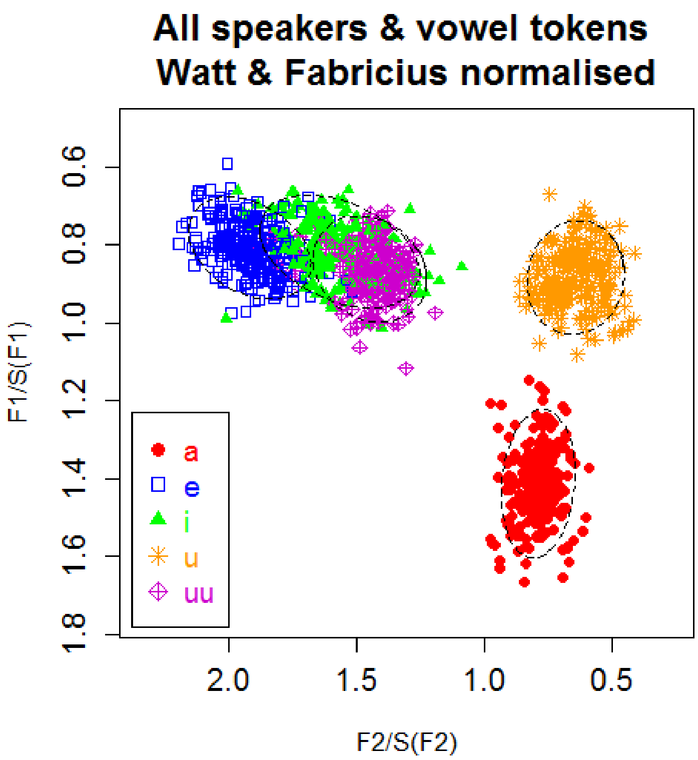


Fig. 38: Normalised F1/F2 values for all speakers and vowels. Ellipses indicate standard deviation.

As all speakers in the sample use some form of Viby-i, it is not clear how these formant values compare to those of a standard [iː]. However, a qualitative comparison can be made using a small sample of male data from Fant et al (1969). Fig. 39 shows mean F1 and F2 values in Hz for /iː/ and /eː/, with Fant et al's speakers plotted against the male speakers in the present study. The comparison serves both to give context for /iː/, and to investigate whether the movement of /iː/ has resulted in a chain shift, whereby /eː/ might have moved into the low F1, high F2 space originally occupied by [iː].

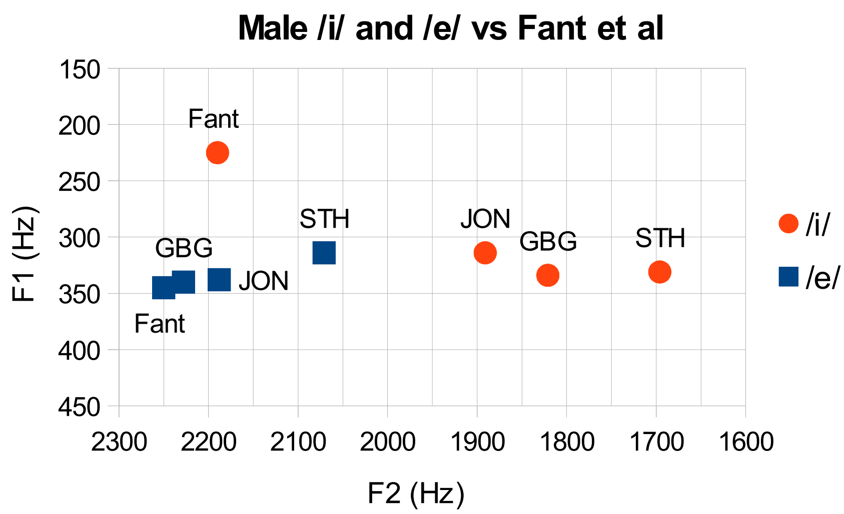


Fig. 39: Mean F1/F2 values (Hz) for male /i/ and /e/ in the current sample compared to Fant et al (1969).

The comparison shows that the male F1/F2 values for /iː/ in the present study are drastically different from those of Fant et al, with both a lower F2 and a higher F1. In both samples, /iː/ is distinct from /eː/. Although some lowering of F2 can also be seen in /eː/, it has not traded places with Fant et al's /iː/. It should be noted that these F1/F2 values have not been normalised, however it is unlikely that the difference between /iː/ in the present study and in Fant et al's data is caused by physiological differences. Compared to standard [iː], then, Viby-i can be said to be characterised by a high F1 and low F2 compared to what we would normally expect.

The other issue raised by the vowel plot in Fig. 38 is the overlap between /iː/, /eː/ and /ʉː/, prompting the question whether the distribution for /iː/ is significantly different from the other two, or whether there is acoustic overlap, which in turn could result in perceptual overlap. To investigate this, two mixed-effects linear regression models were run, with either F1 or F2 as the dependent variable, vowel (/iː/, /eː/ or /ʉː/) as the independent variable, and speaker and word environment (initial, medial, final) as random effects. The model thus answers the question: 'Disregarding the variation produced by individual speakers and word environment, are the distributions of these vowels on the F1/F2 plane significantly different from each other?' The results for F1 and F2 are shown in Tables 9 and 10 respectively.

F1: Random effects:	Chi.sq	Chi.DF	elim.num	p.value	
Speaker	42.67	1	kept	< 1e-07	
Context	50.42	1	kept	< 1e-07	
Differences of LSMEANS:	**Estimate**	**Standard Error**	**DF**	**t-value**	**p-value**
Vowel a - e	0.6	0.0136	28.6	44.22	<2e-16 ***
Vowel a - i	0.6	0.0137	29.2	43.29	<2e-16 ***
Vowel a - u	0.5	0.0137	29	38.55	<2e-16 ***
Vowel a - uu	0.5	0.0138	29.5	39.79	<2e-16 ***
Vowel e - i	0	0.0137	28.6	-0.66	0.512
Vowel e - u	-0.1	0.0136	28.4	-5.52	<2e-16 ***
Vowel e - uu	-0.1	0.0137	28.8	-4.07	3e-04 ***
Vowel i - u	-0.1	0.0137	29.1	-4.83	<2e-16 ***
Vowel i - uu	0	0.0138	29.5	-3.39	0.002 **
Vowel u - uu	0	0.0137	29.3	1.42	0.166

Table 9: Summary of mixed-effects results for F1 and vowel.

F2: Random effects:	Chi.sq	Chi.DF	elim.num	p.value	
Speaker	39.84	1	kept	< 1e-07	
Context	210.08	1	kept	< 1e-07	
Differences of LSMEANS:	**Estimate**	**Standard Error**	**DF**	**t-value**	**p-value**
Vowel a - e	-1.1	0.0309	29.9	-36.42	<2e-16 ***
Vowel a - i	-0.8	0.031	30.2	-24.96	<2e-16 ***
Vowel a - u	0.1	0.031	30.1	4.69	1e-04 ***
Vowel a - uu	-0.7	0.031	30.3	-21.19	<2e-16 ***
Vowel e - i	0.4	0.0309	29.9	11.39	<2e-16 ***
Vowel e - u	1.3	0.0309	29.8	41.14	<2e-16 ***
Vowel e - uu	0.5	0.0309	30	15.14	<2e-16 ***
Vowel i - u	0.9	0.031	30.1	29.67	<2e-16 ***
Vowel i - uu	0.1	0.031	30.3	3.75	8e-04 ***
Vowel u - uu	-0.8	0.031	30.2	-25.89	<2e-16 ***

Table 10: Summary of mixed-effects results for F2 and vowel.

The results show that all vowels were significantly different from each other on the F1 plane except /eː/-/iː/ and /ʉː/-/uː/ (Table 9), and that all vowels were significantly different from each other on the F2 plane (Table 10). There was in other words some overlap between /iː/ and /eː/ in F1, but no significant overlap between /iː/ and /ʉː/. It may however be the case that the /iː/ tokens which overlap with /ʉː/ on the vowel plot are auditorily stronger, something which will be investigated in the next section.

4.2.2. Effect of F1/F2 on auditory strength

If Viby-i is characterised by a high F1 and low F2, the question is whether a greater degree of F1-raising and F2-lowering is associated with an auditorily stronger Viby-i. An F1/F2 plot of the mean values associated with weak, moderate and strong Viby-i (Fig. 40) suggests that this is the case. In the F1 dimension, strong Viby-i has a higher value than weak and moderate Viby-i, and in the F2 dimension, there is a gradual movement from weak to strong, with F2 lowering throughout. The graph also shows overlap between strong Viby-i and /ʉː/ which is not seen in the weak and moderate versions of the vowel.

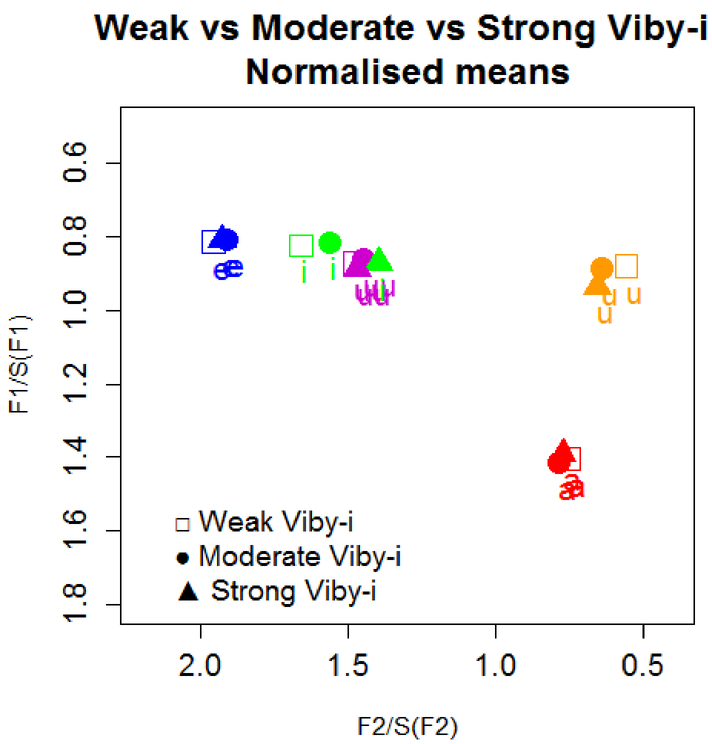


Fig. 40: Normalised mean F1/F2 values for weak, moderate and strong Viby-i groups.

These results were further explored with two Pearson's correlation tests, investigating the relationship between auditory strength and F1 and F2 respectively. The correlation coefficients are shown in Fig. 41 and Fig. 42. F1 was found to have a marginally significant positive correlation with Viby-i strength, $r = 2.066$, $p = 0.0632$, and F2 was found to have a highly significant negative correlation with Viby-i strength, $r= -5.313$, $p = 0.000247$. In other words, the higher the F1 and lower the F2, the stronger Viby-i becomes, although the influence of F2 appears to be much stronger than the influence of F1. The weaker correlation with F1 can be confirmed visually by the fact that the coefficient does not capture the distribution of the data points for F1 as well as it does for F2.

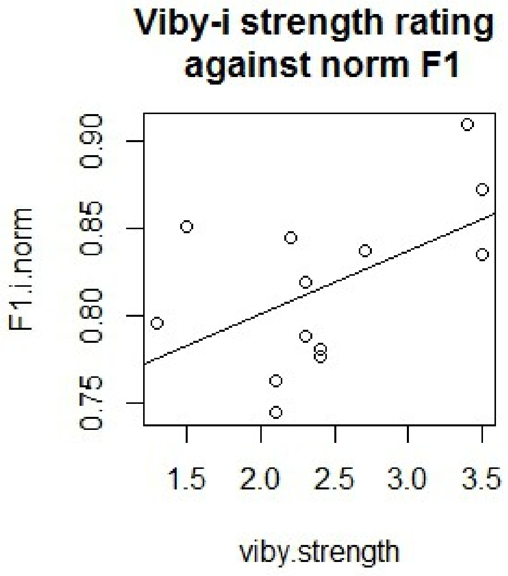


Fig. 41: Correlation between auditory strength rating and normalised F1 for /i/.

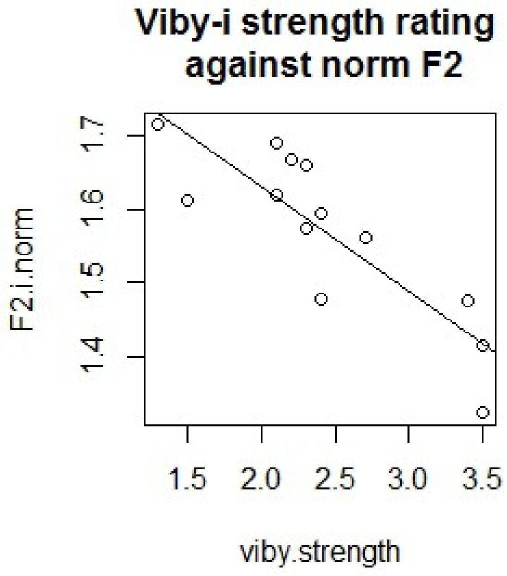


Fig. 42: Correlation between auditory strength rating and normalised F2 or /i/.

To summarise the results so far, Viby-i is acoustically characterised by a low F2 and high F1 compared to standard [iː]. There is some overlap in F1 between /iː/ and /eː/, but in F2, all vowels in the set are significantly different from each other. Nevertheless, stronger instances of Viby-i appear to have some overlap with /ʉː/, as they display a significantly greater degree of F2-lowering. F1-raising also had a marginal effect on the strength of Viby-i. This partial overlap with /ʉː/ may account for some of the 'dark' quality associated with Viby-i. However, /iː/ and /ʉː/ are still auditorily distinct from each other, which could be attributed to a number of features not accounted for here, e.g. F3.

4.2.3. F1/F2 by geographical groupings

The relationship between region and acoustic properties is investigated in this section by comparing group means for dialect region (East vs. West), metropolitan vs. urban location, and city (Stockholm vs. Gothenburg).

Beginning with dialect region, Fig. 43 shows the normalised acoustic mean values for Eastern and Western Central Swedish vowels. The position of /i:/ does not seem to differ between the two groups, and indeed a t-test returns t = 0.81202, p-value = 0.418 for F1 and t = -0.81949, p-value = 0.4135 for F2. Neither of these results are significant, meaning that there is no statistical difference in F1 or F2 between the Eastern and Western groups. This result corresponds to the lack of significant difference between these groups in the auditory results.

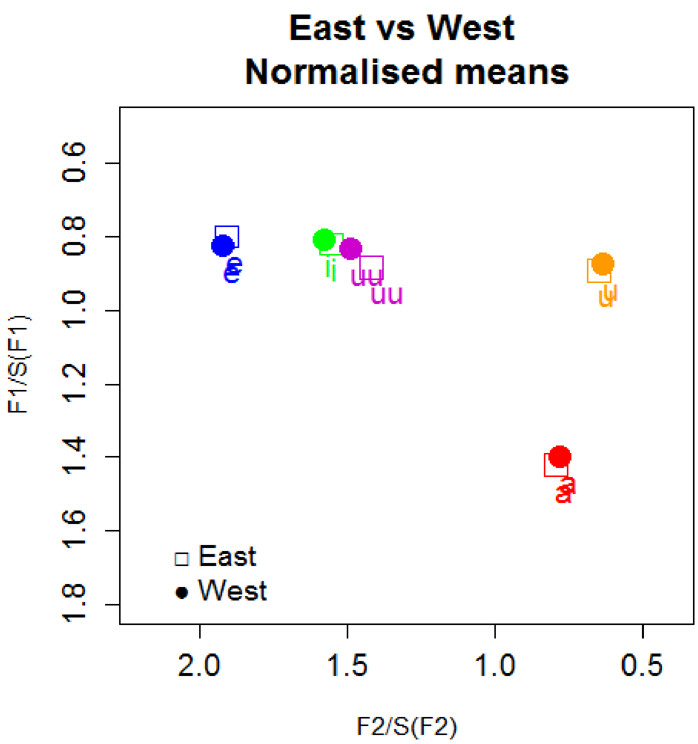


Fig. 43: Normalised mean F1/F2 values for Eastern and Western Central Swedish.

Moving on to a comparison between metropolitan and urban areas, an F1/F2 plot comparing mean values between the two is shown in Fig. 44. The plot shows that whereas most vowels in the system overlap almost perfectly, /i:/ has a much higher F1 and lower F2 in the metropolitan group than in the urban group. This finding also supports the trend in the auditory results for metropolitan speakers' Viby-i to be stronger. A t-test comparing the mean values for /i:/ in the metropolitan and urban groups reveals a highly significant difference both in F1 (t = 3.997, p-value = 0.0001075) and F2 (t = -6.8981, p-value = 5.193e-11). Thus, the metropolitan speakers in the sample are shown to experience a greater degree both of F1-raising and F2-lowering than urban speakers, which is in turn associated with an auditorily stronger Viby-i.

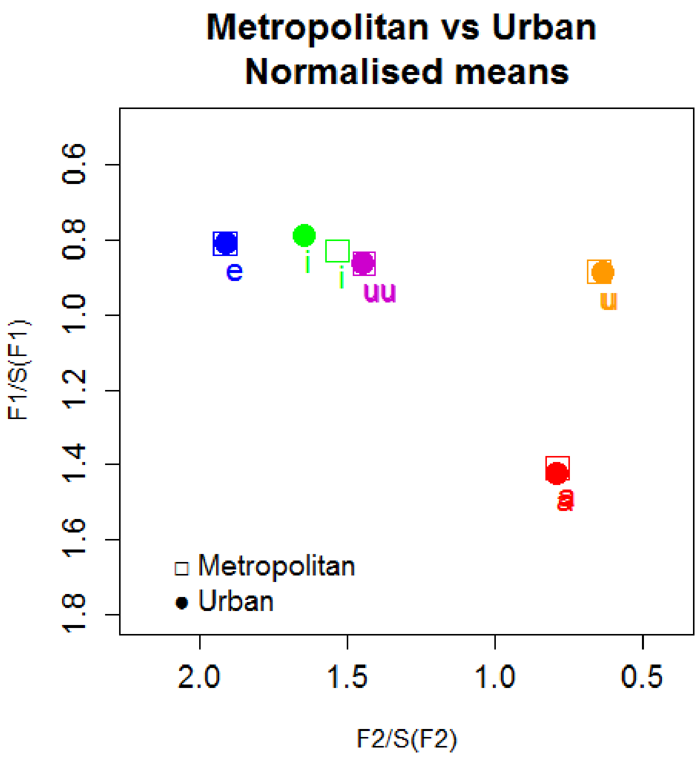


Fig. 44: Normalised mean F1/F2 values for the metropolitan and urban groups.

The final regional comparison for the acoustic data is made between Stockholm and Gothenburg, with group means displayed in Fig. 45. Once again, due to the unbalanced sample, no statistical comparison will be made between these groups, but a descriptive comparison of the data will be provided.

Looking at the graph, the Stockholm and Gothenburg groups show almost the opposite pattern of the metropolitan vs. urban data: /iː/ is very similar between the two groups, but the other vowels in the set are different. Gothenburg /iː/ has a marginally higher F1 and lower F2 than Stockholm /iː/, the reverse of what was found in the East vs. West data. This may seem odd as Stockholm, being geographically East, would be expected to pattern with this region. However, the reversal is most likely caused by the disproportionate influence of GBG_YF_01 on the small Gothenburg group. This speaker had a very strong Viby-i rating, and as such is likely to also display strong F1-raising and F2-lowering, which probably affects the group mean. Overall, the F1/F2 properties of /iː/

are very similar between the cities, suggesting that Viby-i is produced similarly in Stockholm and Gothenburg, despite other differences between the two vowel systems.

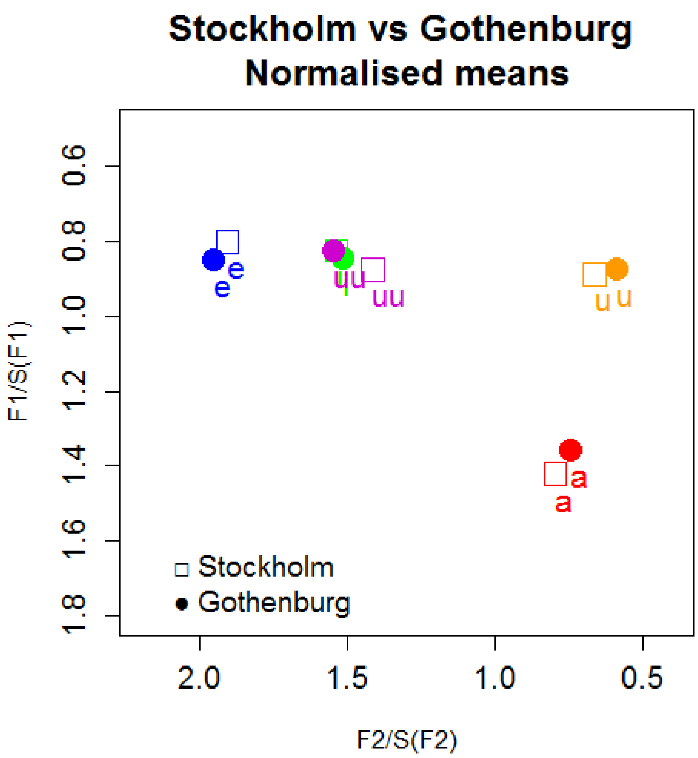


Fig. 45: Normalised mean F1/F2 values for the Stockholm and Gothenburg groups.

In summary, the regional groupings show no significant difference in F1 or F2 for /iː/ between East and West, and only a very small qualitative difference between Stockholm and Gothenburg. These findings are nevertheless interesting as acoustic differences between East and West, and between the two cities, did exist for the remaining vowels /eː/, /ɑː/, /uː/ and /ʉː/. Highly significant differences in both F1 and F2 were found between the metropolitan and urban groups, however, with the metropolitan group having a much higher F1 and lower F2, which suggests a stronger Viby-i.

4.2.4. F1/F2 by social groupings

The final comparison of the acoustic data is based on the social factors of sex and socioeconomic group, and will again be qualitative due to the unbalanced sample.

The mean F1/F2 values for men and women is presented in Fig. 46. The graph shows great similarity between the two groups, with a high degree of overlap which also extends to /iː/. The women have a slightly lower F2 for /iː/, corresponding to an average higher strength rating in the auditory results. As mentioned previously, however, this is not necessarily an effect of sex, as two of the three male speakers also belonged to groups less associated with Viby-i (multiethnolect speakers and peripheral urban speakers). Thus there does not appear to be any systematic difference in the acoustics of /iː/ between men and women.

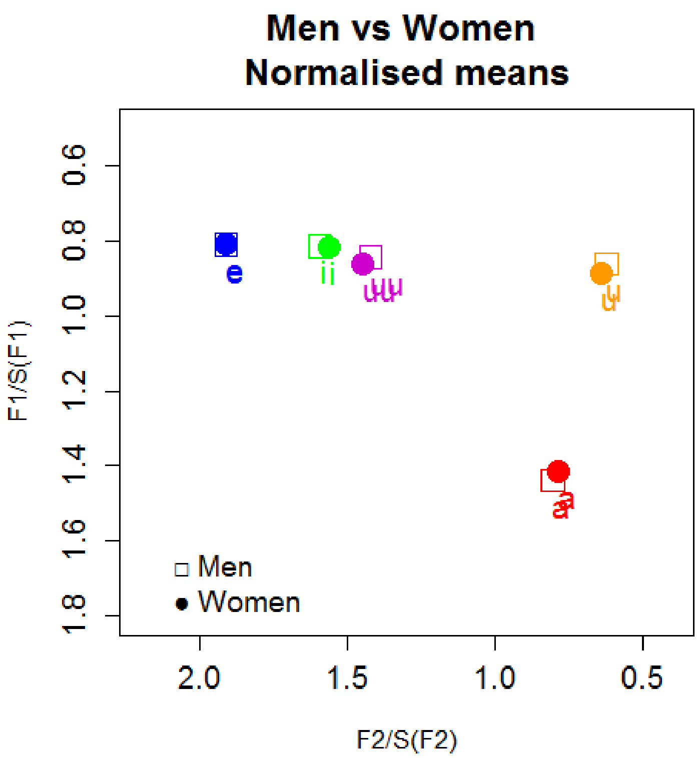


Fig. 46: Normalised mean F1/F2 values for men and women.

Finally, the acoustic results are compared for different socioeconomic groups. For the purpose of generating group means, broad class groupings have been made based on socioeconomic score, dividing speakers into lower class (1-1.99), middle class (2-2.99) and higher class (3-4). It should be pointed out again that these labels are used for convenience to denote ranges, and not as a reflection of the Swedish class system as a whole. As most speakers in the sample are from a higher socioeconomic background, the comparison between these groups will once again be descriptive.

The mean F1/F2 values for the socioeconomic groups is shown in Fig. 47. Visually, the degree of F1-raising appears to be stronger in lower and middle class speakers, who cluster together, whereas F2-lowering is strongest in higher class speakers. Interestingly, F2-lowering does not have a linear relationship with class, but lower class speakers are intermediate to the other two groups. This non-linearity could explain why socioeconomic score was only a marginal predictor of Viby-i strength, particularly given that F2 was shown to be more influential than F1 for the auditory ratings. Perhaps it is possible that Viby-i is a feature associated with lower and higher class speakers, in a similar way as it is associated with metropolitan as well as remote rural areas. However, the sample is too small to verify this. Furthermore, the low numbers may be the reason why F1 and F2 in this instance do not move together in the expected direction (higher F1 with lower F2). As the results may thus not be reliable, we cannot establish any clear difference in F1/F2 between the socioeconomic groups.

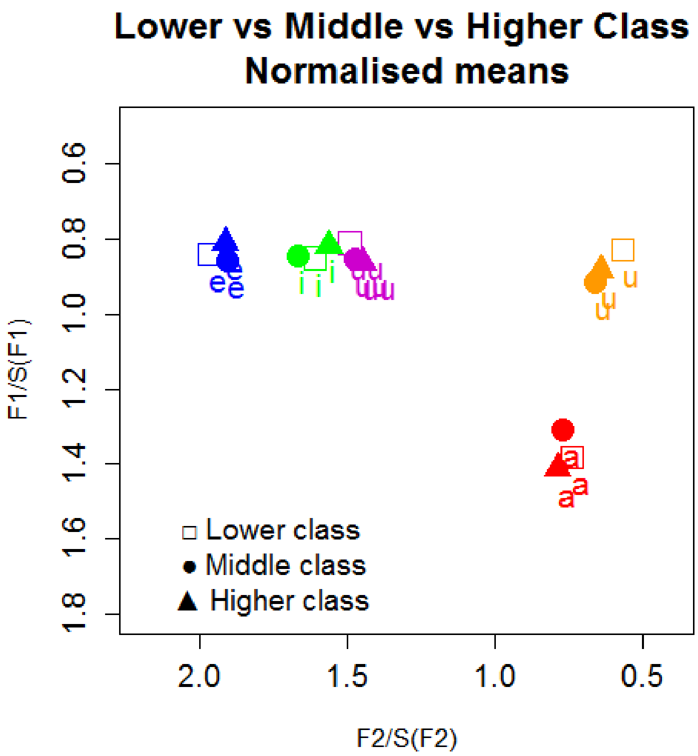


Fig. 47: Normalised mean F1/F2 values for lower, middle and higher class speakers.

4.2.5. Summary: Acoustic results

The acoustic results show that Viby-i is characterised by F1-raising and F2-lowering, rendering it similar in F1 to /eː/, but with a significantly lower F2. The higher the F1 and the lower the F2, the stronger the Viby-i is auditorily, and strong instances of Viby-i may overlap with /ʉː/ in the F1/F2 space. This overlap may be related to the vowel's 'dark' quality. The metropolitan group are found to produce /iː/ with a significantly higher F1 and significantly lower F2 than the urban group. There are no major differences in F1/F2 for /iː/ between dialectal regions, cities, men and women, or socioeconomic groups.

4.3. Articulatory UTI results

Tongue splines for the UTI results can be found in Figs. 63-75. For convenience, these have been placed at the end of the section. Higher resolution images are also available in the Appendix. In this section, tongue gesture will first be described impressionistically, and then more quantitatively in terms of its position for /iː/ in relationship to /eː/. This relationship is expressed as a proportional measure of lowering and backing from /eː/, based on the points marked X on the tongue diagrams. Tongue shape will be described qualitatively.

4.3.1. Overview

The most visible articulatory characteristic for /iː/ is that it has a lower tongue body position than /eː/ in all speakers, often having a similar height to /ʉː/. Contrary to what the low F2 suggests, the tongue is also fronted, at least in terms of its highest point. Going by this measure, /iː/ is actually the frontest vowel in all speakers. The tongue tip appears to be particularly high and fronted compared to the other vowels, suggesting that it is very close to the palate. But despite the fronted tongue tip and body, the back of the tongue tends to be more retracted for /iː/ than for /eː/, meaning that there is usually a crossover between the splines for these two vowels at some point along the tongue body. This retraction of the tongue back may be what results in the low F2. It is also this measure that will be referred to when discussing tongue backing, as opposed to the highest point of the tongue.

A summary of all speakers' proportional lowering and backing of /iː/ in relation to /eː/ is produced in Fig. 48. There is a great deal of variation in the sample, and one speaker in particular, STH_YF_01, appears to be doing something different. A quick visual inspection confirms that her extreme values for lowering and backing are indeed matched by a distinctly lowered and backed tongue gesture for /iː/ (Fig. 67). The remaining speakers' values fall within a much smaller range, and most display a pattern where tongue lowering is greater than tongue backing. One speaker, STH_YF_03, shows the reverse pattern, with more backing than lowering, and three speakers, GBG_YM_01, KAT_YF_01 and STH_YF_04, have negative values for backing, meaning that the back of the tongue is fronted in comparison to /eː/.

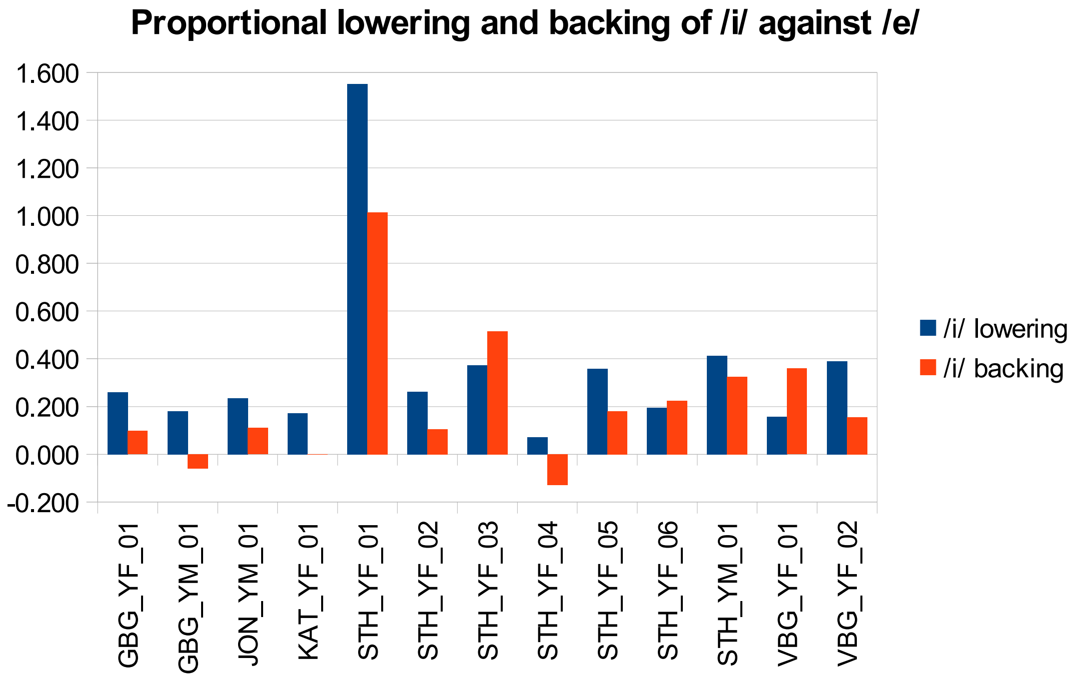


Fig. 48: Articulatory lowering and backing of /i/ expressed as a proportion against /e/.

Two distinct tongue shapes for /iː/ also appear in the data. Some speakers, e.g. JON_YM_01 (Fig. 65), produce /iː/ with a familiar, bunched, 'upward slope' shape, possibly similar to that observed by Schötz et al (2011), whereas other speakers, e.g. STH_YF_03 (Fig. 69), produce /iː/ with a double-bunched shape, reminiscent of what Borgström (1913) suggests for Viby-i. Some speakers fall within an intermediate category, with weak bunching or flattened slopes (e.g. STH_YF_04, Fig. 70). It is possible that these differences in tongue shape contribute to a difference in vowel quality, and may be stratified within the sample. This will be investigated in relation to the geographical and social factors discussed in the following sections.

4.3.2. Effect of articulatory properties on auditory strength

Possible correlations between tongue lowering and backing with Viby-i strength were investigated by plotting the proportional tongue values against Viby-i strength rating, generating coefficients for these plots, and running a Pearson's correlation test to investigate if the correlations were statistically significant. The effect of tongue lowering on Viby-i strength is seen in Fig. 49, and the effect of tongue backing on Viby-i strength in Fig. 50.

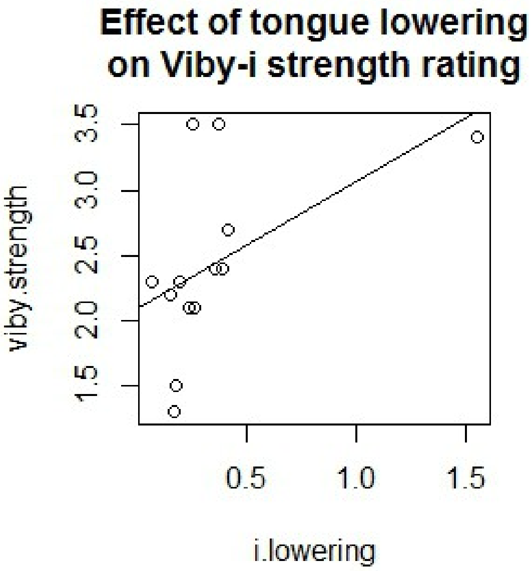


Fig. 49: Correlation between proportional /i/ lowering and auditory strength rating.

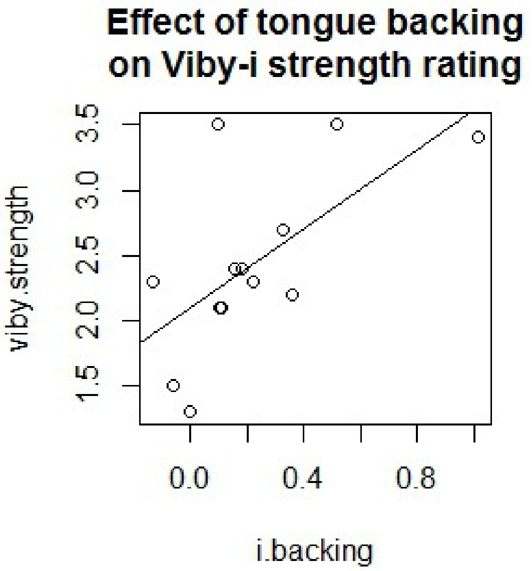


Fig. 50: Correlation between proportional /i/ backing and auditory strength rating.

There was a highly significant positive correlation between /i:/ backing and Viby-i strength, $r = 2.816$, $p = 0.0168$, and a marginally significant positive correlation between /i:/ lowering and Viby-i strength, $r = 2.051$, $p = 0.0648$. Tongue backing thus seems to have a greater influence on the auditory quality than lowering, but both can be said to be associated with a stronger Viby-i.

Notably, the two weak Viby-i users in the sample both have negative values for tongue backing (i.e. their tongues are fronted). However, STH_YF_04, who has an even greater negative value for tongue backing, is still rated as having a moderate Viby-i. The strongest Viby-i users form a mixed group, with STH_YF_01 showing extreme backing and lowering, STH_YF_03 also being very backed and lowered, but being the only speaker in the sample with more backing than lowering, and GBG_YF_01 having relatively little backing and lowering in comparison to the sample as a whole. It thus appears that these speakers are using different articulatory strategies to achieve a strong Viby-i.

With regards to tongue shape, sloped shapes seem to be relatively common in moderate Viby-i users, but double-bunched articulations are found across the sample, including both weak and strong speakers. Thus, double-bunching does not seem to have any particular effect on the auditory strength of the vowel. It is nevertheless possible that double-bunching is characteristic of some types of Viby-i as opposed to standard [iː].

In summary, there is a strong relationship between auditory Viby-i strength and tongue backing, and a weak relationship between Viby-i strength and tongue lowering. Weak Viby-i users tend to produce an /iː/ that is fronter than /eː/, whereas strong Viby-i users generally have a greater degree of both tongue backing and tongue lowering. Double-bunched tongue shapes are used by both weak and strong speakers, and thus does not seem to affect Viby-i strength. However, it could be the case that trade-off is taking place between tongue shape and tongue position, which could explain the great degree of variation in the sample.

4.3.3. Effect of articulatory properties on F1/F2

The effect of tongue backing on F2 and tongue lowering on F1 are investigated in a similar way to the auditory ratings, in order to see whether the results map onto what is already known about the relationship between auditory and acoustic properties (F2 contributes more to auditory strength than F1), and between auditory and articulatory tongue properties (tongue backing contributes more to auditory strength than tongue lowering).

The correlations are plotted in Fig. 51 and Fig. 52, with the coefficients moving in the expected directions: Tongue lowering is associated with an increase in F1, and tongue backing with a decrease in F2. However, only tongue lowering was statistically significant, $r = 2.266$, $p = 0.0446$, whereas tongue backing was not, $r = -1.614$, $p = 0.135$. Thus, tongue lowering here has a much stronger influence on F1 than tongue backing does on F2. This result creates a discrepancy between the articulatory and acoustic results, in that Viby-i is mainly characterised by tongue lowering in the physical dimension, but a low F2 in the acoustic dimension, which would suggest backing. The discrepancy could stem from a measurement error in the articulatory data, or alternatively, it may be a case of articulatory trade-off, with the low F2 being produced by another part of the vocal tract, e.g. the lips.

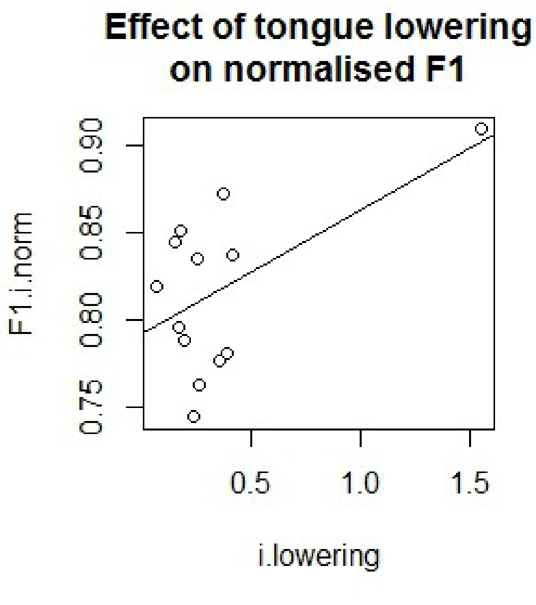


Fig. 51: Correlation between proportional /i/ lowering and normalised F1.

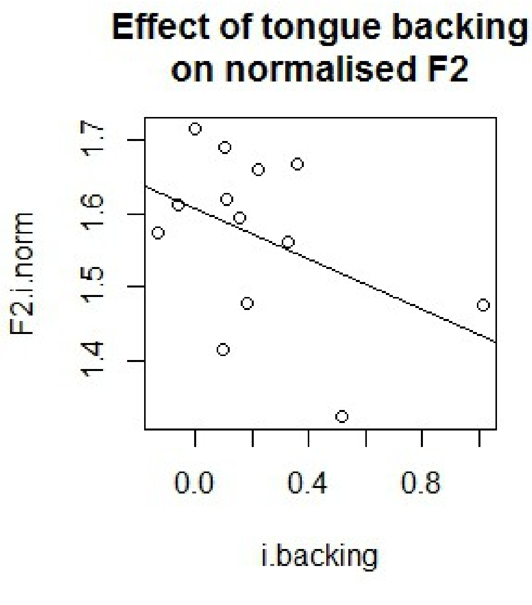


Fig. 52: Correlation between proportional /i/ backing and normalised F2.

4.3.4. Articulatory properties by regional groupings

Tongue properties will now be investigated according to regional factors, beginning with dialectal region (East vs. West), continuing with metropolitan vs. urban environment, and finishing with a comparison between cities (Stockholm vs. Gothenburg).

The mean values for proportional tongue lowering and backing in the Eastern and Western Central Swedish groups are compared in Fig. 53 and Fig. 54. Although there are visible differences between the groups, with tongue lowering and backing being more prevalent in Eastern Central Swedish, these differences are not significant. A t-test for lowering returns t = 1.0545, p-value = 0.3232, and for backing t = 1.0269, p-value = 0.3284. The significance did not change even if the large outlier values of STH_YF_01 were excluded. There is in other words no significant difference in tongue lowering or backing for /iː/ depending on dialectal region.

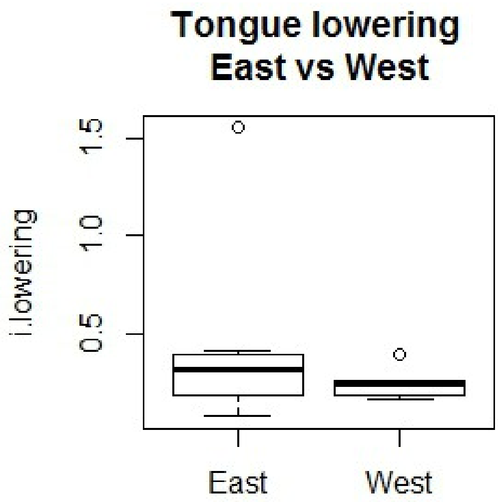


Fig. 53: Mean proportional lowering of /i/ for the Eastern and Western Central Swedish groups.

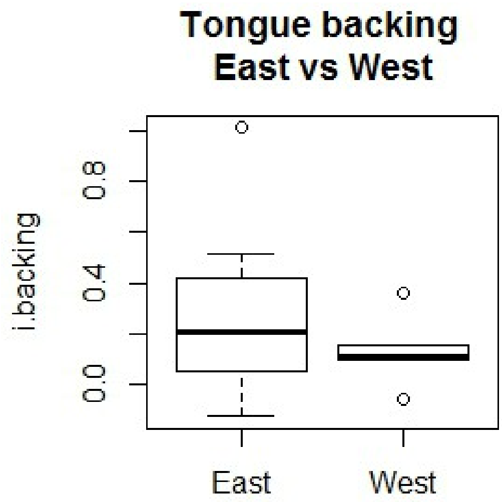


Fig. 54: Mean proportional backing of /i/ for the Eastern and Western Central Swedish groups.

With regards to tongue shape, there is a proportionally similar number of double-bunched and sloped tongue shapes in the Eastern and Western groups, with most speakers in the sample overall using a double-bunched articulation. No clear stratification of tongue shape based on dialectal region could thus be established.

Moving on to the metropolitan vs. urban groups, a comparison of the mean values for lowering and backing are shown in Fig. 55 and Fig. 56. Once again, there is a discernible difference between the group means, with lowering and backing being more prevalent in the metropolitan group. But neither of the results were significant, with lowering giving a t-test result of t = 1.0745, p-value = 0.3084 and backing t = 0.70486, p-value = 0.4956. This is an unexpected result given that these groupings were significant in both the auditory and acoustic data. However it may simply be the case that small articulatory differences (or differences not measured here) have strong effects on the acoustic signal.

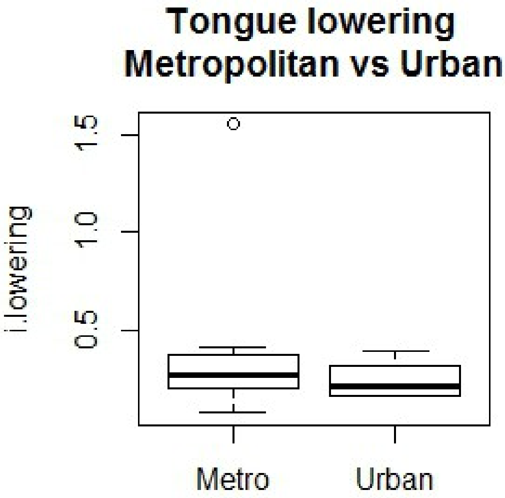


Fig. 55: Mean proportional lowering of /i/ for the metropolitan and urban groups.

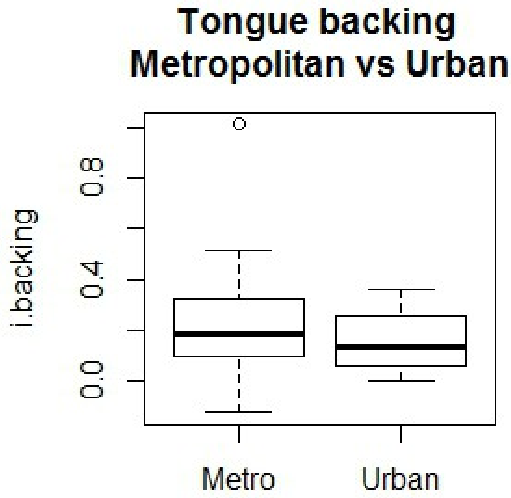


Fig. 56: Mean proportional backing of /i/ for the metropolitan and urban groups.

The distribution of tongue shape seems to be somewhat stratified however, with a greater proportion of speakers using double-bunched or weak double-bunched articulations for /iː/ in the metropolitan group, and a greater proportion of speakers using sloped articulations in the urban group. This is not necessarily linked to auditory Viby-i strength, as previous results have shown, but it could be a signal that region could have an influence on tongue gesture.

The final regional comparison is between Stockholm and Gothenburg, with the mean values for tongue lowering and backing for these groups shown in Fig. 57 and Fig. 58. The uneven sample sizes render the means unsuitable for statistical testing, so a qualitative description will be provided instead. Judging by the mean values, the Stockholm group are more prone to both tongue lowering and backing than the Gothenburg group. However, as the Gothenburg group only consists of two speakers, the mean value can be misleading. For example, both Gothenburg speakers show tongue lowering, but only GBG_YF_01 has tongue backing, whereas GBG_YM_01 produces a fronted /iː/ in comparison to /eː/.

The two speakers also use different tongue shapes, with GBG_YM_01 using a stronger double-bunched articulation than GBG_YF_01. The Stockholm group also varies greatly in both tongue position and shape, with about half the group using double-bunching. The complexity of these articulatory properties in combination with the small and unbalanced sample stand in the way of any conclusions about tongue gesture stratification between Stockholm and Gothenburg.

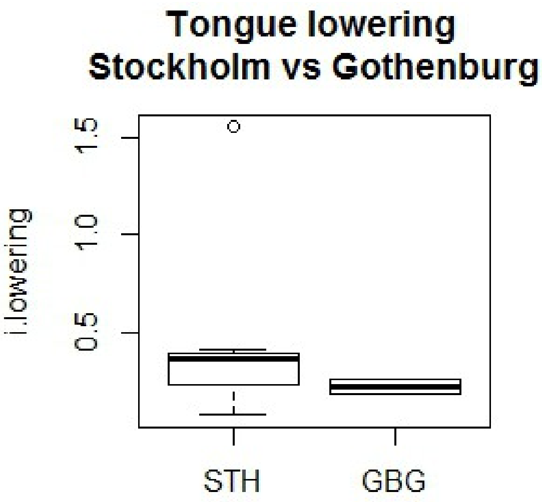


Fig. 57: Mean proportional lowering of /i/ for the Stockholm and Gothenburg groups.

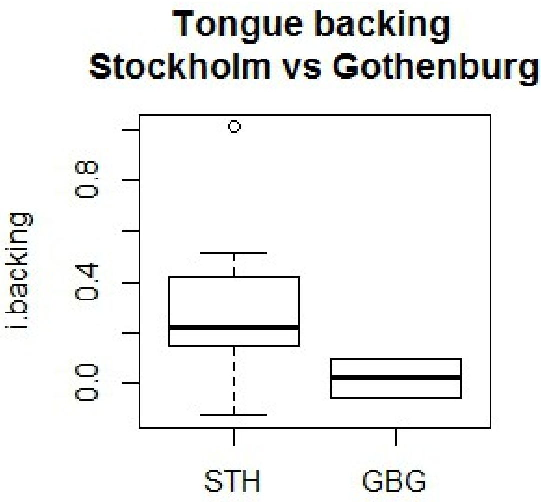


Fig. 58: Mean proportional backing for /i/ for the Stockholm and Gothenburg groups.

In summary, region was not found to correlate with any specific characteristics of tongue lowering or backing for /iː/, nor was there any clear stratification of tongue shape, except a possible tendency for metropolitan speakers to use double-bunching more than urban speakers. A larger and more balanced sample might be able to provide greater insight into regional aspects of Viby-i articulation.

4.3.5. Articulatory properties by social groupings

Finally, the tongue properties for /iː/ will be analysed in relation to the social factors of sex and socioeconomic group. The mean values for tongue lowering and backing in men and women are shown in Fig. 59 and Fig. 60. Once again, these results will be compared qualitatively due to the small number of men in the sample.

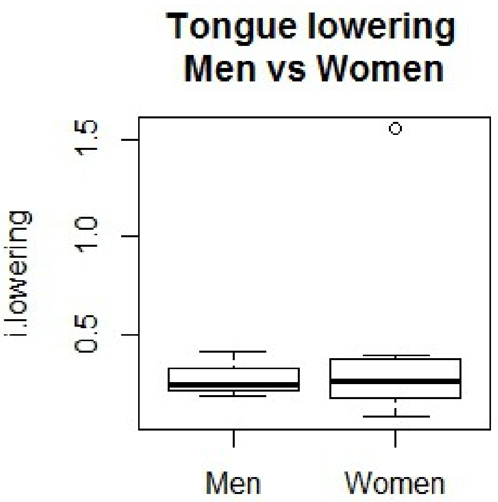


Fig. 59: Mean proportional lowering of /i/ for men and women.

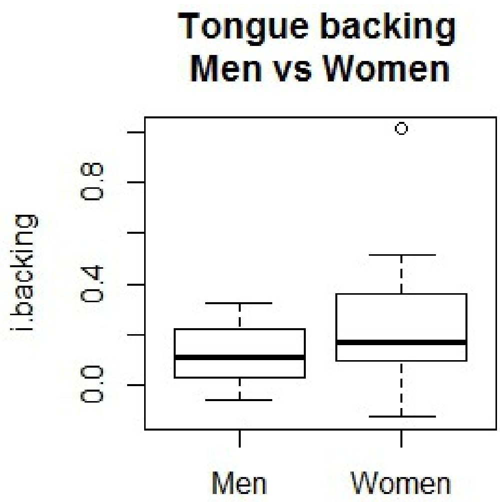


Fig. 60: Mean proportional backing of /i/ for men and women.

Comparing the mean values for men and women, the women appear to produce /iː/ with more tongue backing and slightly more tongue lowering than the men. This result is in line with the finding that women in the sample produce a stronger Viby-i, and one that has a somewhat lower F2 than the men's. Nevertheless, as the unbalanced sample means that none of these results can be tested statistically, any difference between the male and female groups should be treated as highly tentative.

The final investigation concerns the relationship between socioeconomic score and articulatory properties. Fig. 61 shows the correlation between socioeconomic score and tongue lowering, and Fig. 62 the correlation between socioeconomic score and tongue backing.

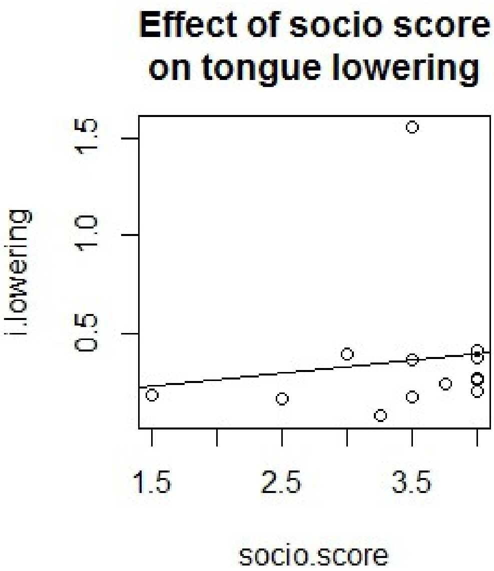


Fig. 61: Correlation between proportional /i/ lowering and socioeconomic score.

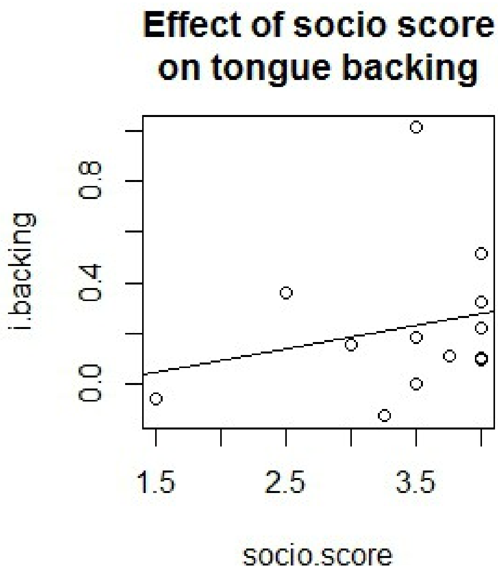


Fig. 62: Correlation between proportional /i/ backing and socioeconomic score.

As the graphs show, there is a positive relationship between a higher socioeconomic score and a greater degree of tongue lowering and backing for /iː/, but the results are not significant. A Pearson's correlation test gives r = 0.439, p = 0.669 for tongue lowering, and r = 0.783, p = 0.450 for tongue backing. Part of the reason why the results are not significant could be the fact that most speakers in the sample are from a higher socioeconomic background. However, it is also clear from the data points that there is quite a lot of variation within the samples. Great variability is also seen in the tongue shapes, with no clear trend emerging in terms of socioeconomic score. The low number of participants in the sample make this variation more difficult to interpret.

4.3.6. Summary: UTI results

The tongue splines show that /iː/ in the sample is primarily associated with tongue lowering, and, to a lesser extent tongue backing, in comparison to /eː/. The highest point of the tongue was nevertheless found to be very fronted, with the tip appearing to be high and front in all speakers. A key finding is that /iː/ can be produced using different tongue shapes, either sloped, double-bunched, or intermediate with light double-bunching.

Tongue lowering in comparison to /eː/ was found to be marginally correlated with a stronger Viby-i, and correlated with an increase in F1. Tongue backing, on the other hand, was highly correlated with a stronger Viby-i, but not correlated with a decrease in F2. The main physiological parameter associated with Viby-i was tongue lowering, despite the fact that F2-lowering was the main acoustic parameter. This discrepancy between acoustics and articulation could be caused by a number of factors, such as small degrees of tongue backing having a disproportionate effect on F2, trade-off in other parts of the vocal tract (e.g. the lips), or a measurement error.

There were no significant or distinct patterns of regional or social stratification in the articulation of Viby-i. More data is needed from a wider variety of backgrounds in order to make conclusive judgements about articulatory stratification in this vowel.

4.3.7. Tongue graphs

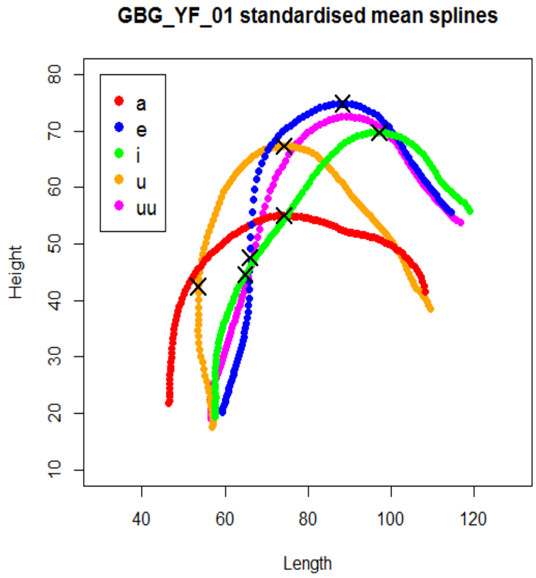


Fig. 63: Tongue splines for GBG_YF_01.

Tongue lowering: 0.2594336

Tongue backing: 0.09814862

Weak double-bunching

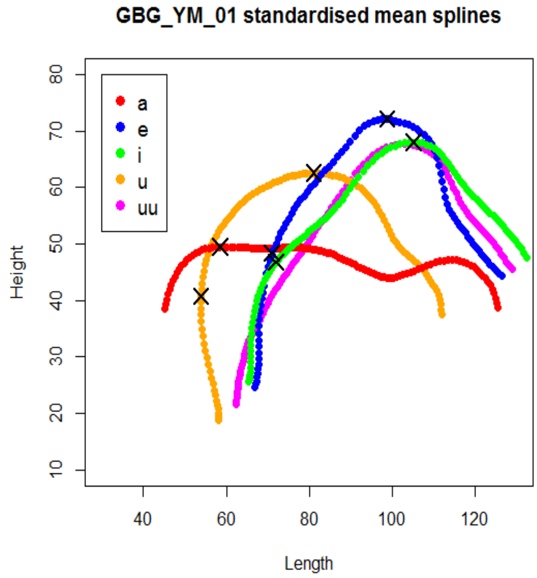


Fig. 64: Tongue splines for GBG_YM_01.

Tongue lowering: 0.1805158

Tongue backing: -0.06002231

Double-bunching

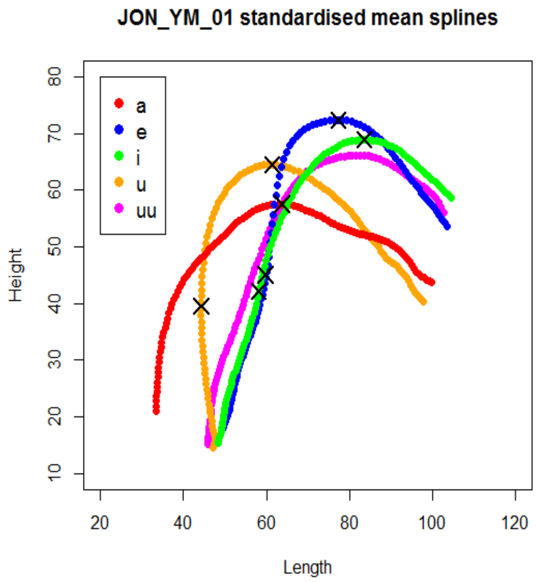


Fig. 65: Tongue splines for JON_YM_01.

Tongue lowering: 0.2353012

Tongue backing: 0.110692

Slope

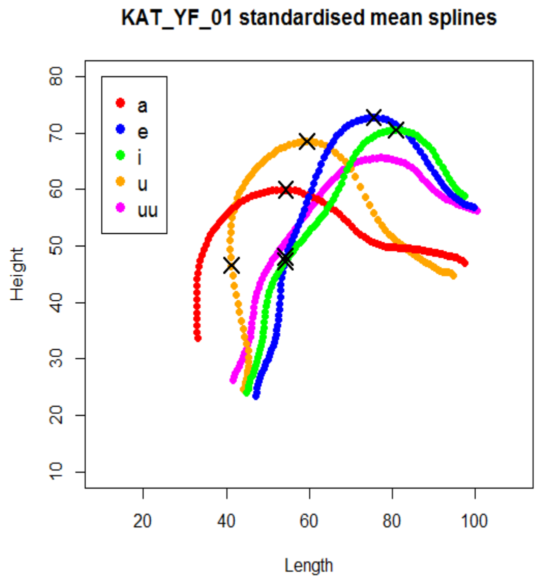


Fig. 66: Tongue splines for KAT_YF_01.

Tongue lowering: 0.173112

Tongue backing: -0.001861378

Double-bunching

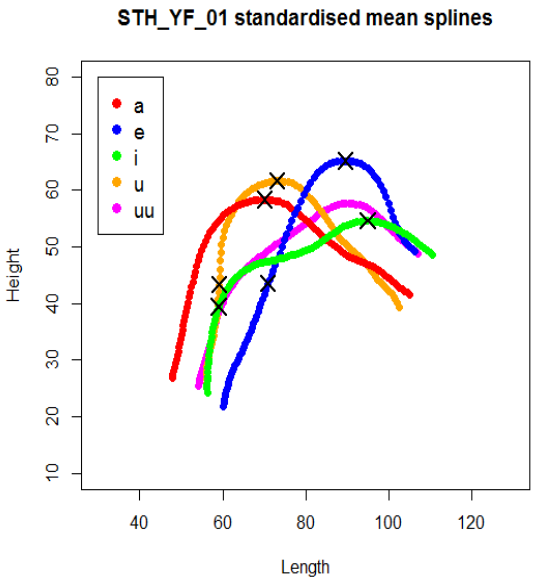


Fig. 67: Tongue splines for STH_YF_01.

Tongue lowering: 1.551823

Tongue backing: 1.013482

Double-bunching

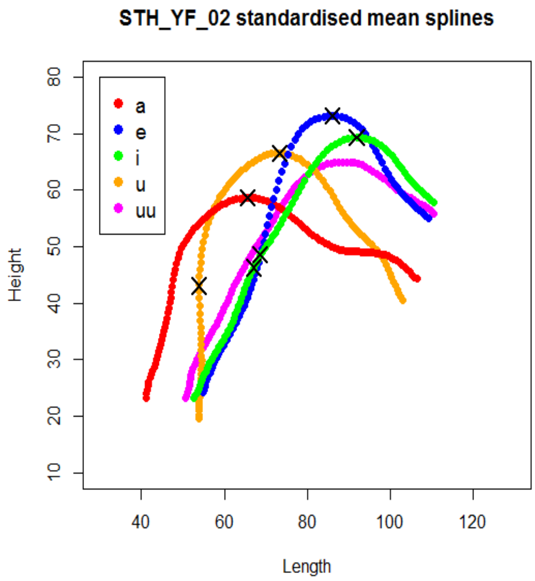


Fig. 68: Tongue splines for STH_YF_02.

Tongue lowering: 0.2616939

Tongue backing: 0.1055852

Slope (with dip?)

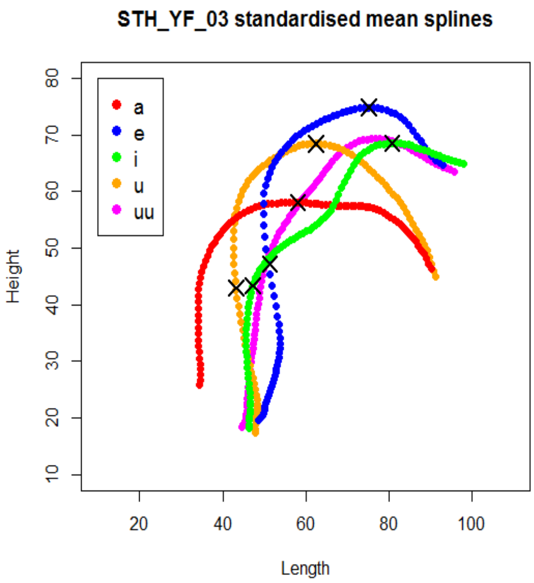


Fig. 69: Tongue splines for STH_YF_03.

Tongue lowering: 0.3733524

Tongue backing: 0.5159733

Double-bunching

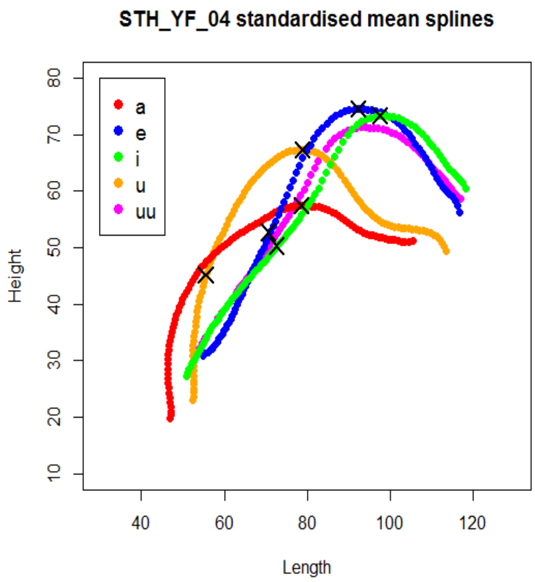


Fig. 70: Tongue splines for STH_YF_04.

Tongue lowering: 0.0712544

Tongue backing: -0.1292654

Weak double-bunching

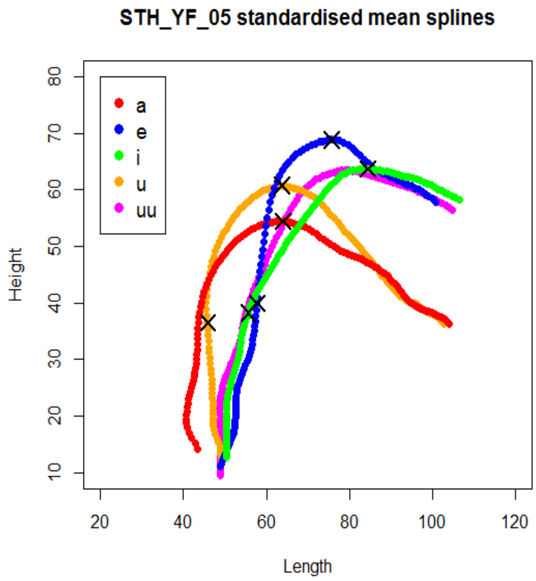


Fig. 71: Tongue splines for STH_YF_05.

Tongue lowering: 0.3578671

Tongue backing: 0.1804903

Flat slope (concealed dip?)

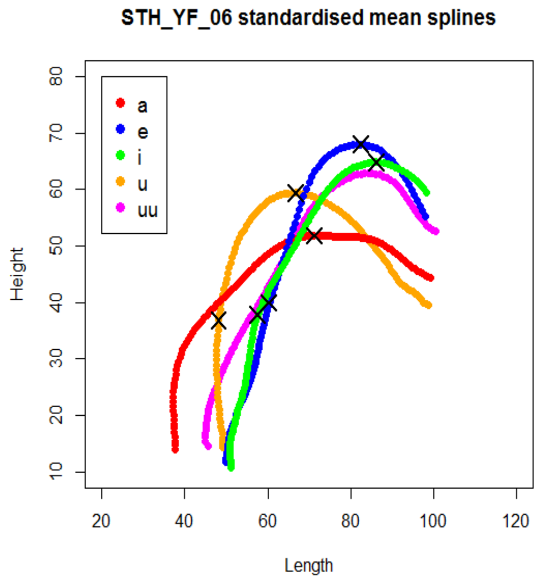


Fig. 72: Tongue splines for STH_YF_06.

Tongue lowering: 0.1956571

Tongue backing: 0.2237251

Slope

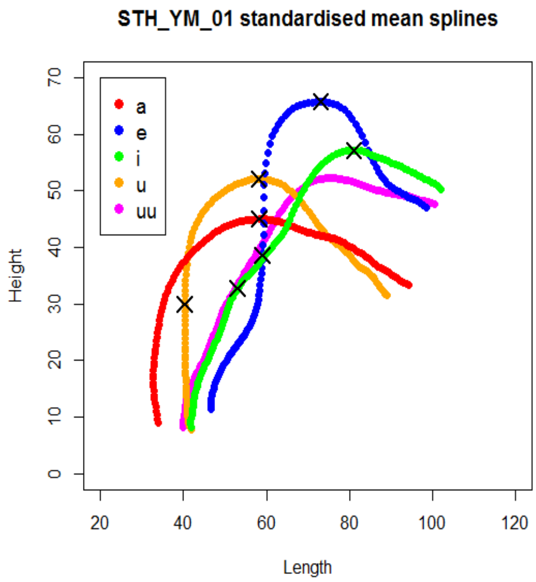


Fig. 73: Tongue splines for STH_YM_01.

Tongue lowering: 0.4131561

Tongue backing: 0.3241951

Double-bunching

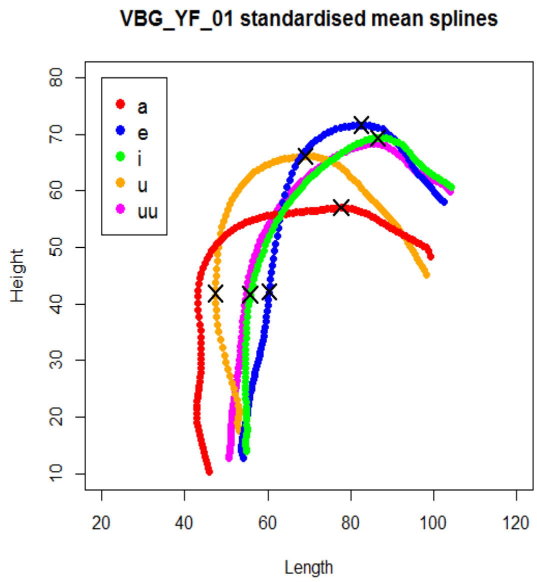


Fig. 74: Tongue splines for VBG_YF_01.

Tongue lowering: 0.1576448

Tongue backing: 0.3590782

Flat slope

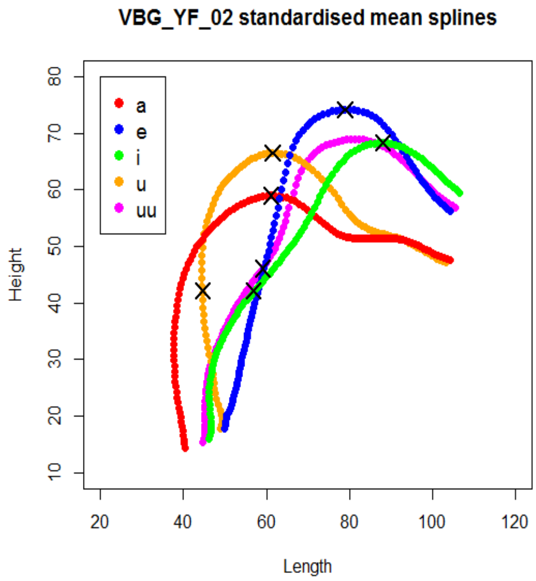


Fig. 75: Tongue splines for VBG_YF_02.

Tongue lowering: 0.3893383

Tongue backing: 0.1548299

Double-bunching

4.4. Articulatory lip results

A small sample of the lip data for each speaker is presented in Figs. 76-114, which for convenience have been placed at the end of this section. The only vowels which are presented here are /iː/, /eː/ and /ʉː/, however lip data for the full vowel set are available in the Appendix. This section will point out some of the general characteristics of the lips during /iː/ production in comparison to /eː/ and /ʉː/. The data will not be investigated in great detail, but is instead presented as a point of interest for further study.

An initial observation about the three vowels is that there is a clear difference in lip-rounding and protrusion between /ʉː/ and the other two vowels. In comparison to the rounded vowels in the set, /iː/ and /eː/ can definitely be classified as unrounded, however they do not necessarily conform to the lip positions normally associated with this category. The corners of the mouth are not very retracted, and the mouth opening is still relatively large, particularly for /eː/. This points towards a more lax lip position for /iː/ and /eː/ in these speakers. The lack of tense lip-spreading may have a slight 'rounding' effect on /iː/ and /eː/, which may result in some of the characteristics of Viby-i, such as the low F2 and the different overall vowel quality.

The lips tend to be somewhat more spread for /iː/ than for /eː/, resulting in a smaller mouth opening, however in some cases the size and shape of the mouth opening are very similar for the two vowels (e.g. JON_YM_01, STH_YF_04).

However, in these speakers and in many others, we can also gauge an interesting detail about the tongue from these images: the front of the tongue is often visible, showing that it is fronted in relation to the other articulators, something which is not visible in the UTI data. In most cases, the tongue tip appears to be lowered and braced against the lower front teeth, and the centre line of the tongue is grooved. In some speakers, the tongue tip is more or less protruded between the teeth (e.g. STH_YF_05, VBG_YF_01, VBG_YF_02). This extremely fronted tongue position may be discernible in the vowel quality, as it may be remembered that some of the Scottish speakers asked to mimic Viby-i in the precursor to this study tried to produce it with tongue protrusion.

The lip results, and what they tell us about tongue articulation, should act as a prompt for further study, as measurements of the lips are needed to estimate the effects of the lax lip posture found for Viby, particularly in comparison to /eː/ and standard [iː]. Coronal tongue data also seems to be essential, given the visible grooving during /iː/ production in many of the participants.

4.4.1. Lip screenshots


Fig. 76: GBG_YF_01 /iː/


Fig. 77:GBG_YF_01 /eː/


Fig. 78:GBG_YF_01 /ʉː/


Fig. 79:GBG_YM_01 /iː/


Fig. 80:GBG_YM_01 /eː/


Fig. 81:GBG_YM_01 /ʉː/


Fig. 82:JON_YM_01 /iː/


Fig. 83:JON_YM_01 /eː/


Fig. 84:JON_YM_01 /ʉː/

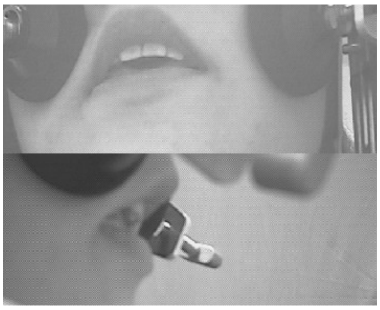


Fig. 85:KAT_YF_01 /iː/

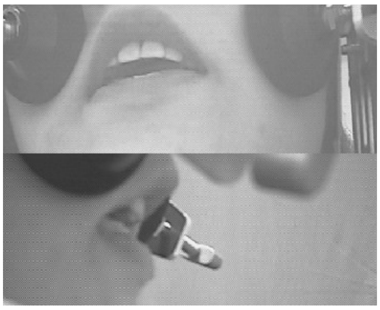


Fig. 86:KAT_YF_01 /eː/


Fig. 87:KAT_YF_01 /ʉː/


Fig. 88:STH_YF_01 /iː/


Fig. 89:STH_YF_01 /eː/


Fig. 90:STH_YF_01 /ʉː/

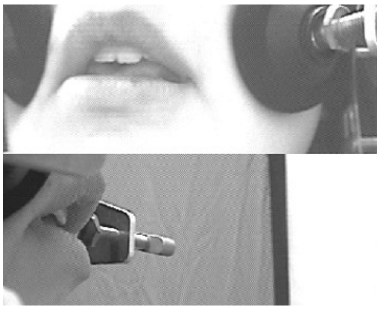


Fig. 91:STH_YF_02 /iː/


Fig. 92:STH_YF_02 /eː/


Fig. 93:STH_YF_02 /ʉː/


Fig. 94:STH_YF_03 /iː/


Fig. 95:STH_YF_03 /eː/

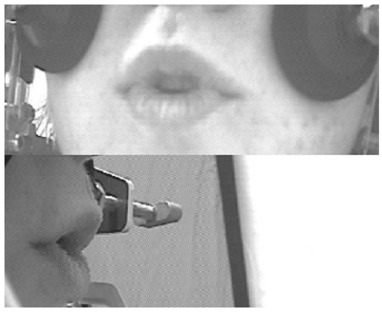


Fig. 96:STH_YF_03 /ʉː/


Fig. 97:STH_YF_04 /iː/


Fig. 98:STH_YF_04 /eː/


Fig. 99:STH_YF_04 /ʉː/

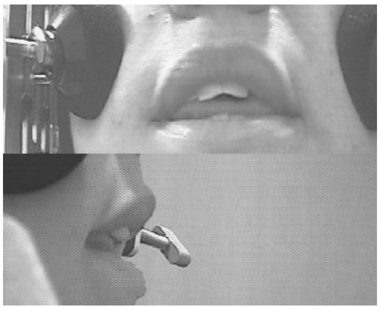


Fig. 100:STH_YF_05 /iː/


Fig. 101:STH_YF_05 /eː/

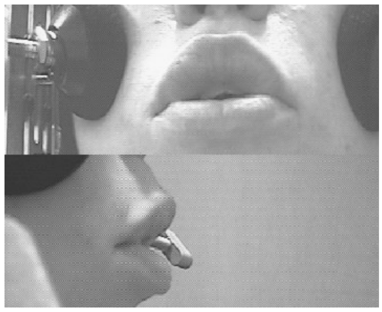


Fig. 102:STH_YF_05 /ʉː/


Fig. 103:STH_YF_06 /iː/


Fig. 104:STH_YF_06 /eː/


Fig. 105:STH_YF_06 /ʉː/

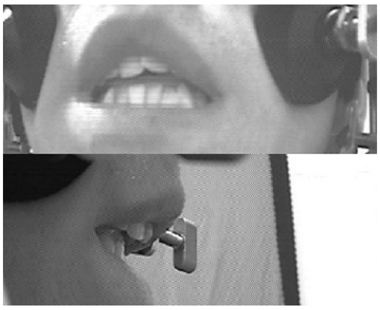


Fig. 106:STH_YM_01 /iː/

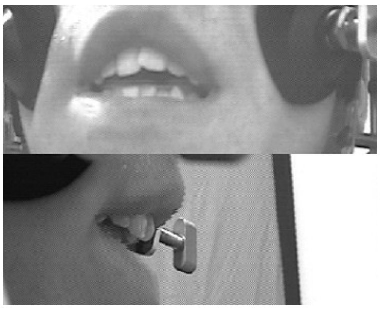


Fig. 107:STH_YM_01 /eː/


Fig. 108:STH_YM_01 /ʉː/


Fig. 109:VBG_YF_01 /iː/

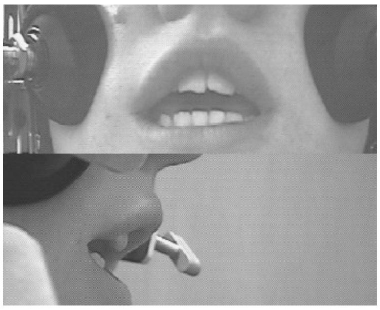


Fig. 110:VBG_YF_01 /eː/


Fig. 111:VBG_YF_01 /ʉː/


Fig. 112:VBG_YF_02 /iː/


Fig. 113:VBG_YF_02 /eː/


Fig. 114:VBG_YF_02 /ʉː/

4.5. Summary of results

To summarise and triangulate, the auditory, acoustic and articulatory results have shown that Viby-i can generally be characterised by a low F2 and high F1, resulting in a more 'centralised' vowel than standard [iː], which reflects the choice of Björsten & Engstrand to refer to it as [ɨ] (1999). Viby-i is also auditorily distinct from [iː], but the distinction is not binary; there are different strengths of Viby-i, and stronger instances of this vowel tends to be characterised by a lower F2, and to some extent a higher F1.

The acoustic and auditory results tended to covary, and to some degree they also covaried with articulation. However, a discrepancy was found in that the main articulatory characteristic of Viby-i was tongue lowering, and to some extent tongue backing, whereas the acoustic results would suggest the opposite pattern. This difference could be caused by a measurement error or a case of articulatory trade-off.

A key finding was that Viby-i is frequently produced with a double-bunched tongue shape, however more familiar sloped tongue shapes were also used, and there did not appear to be a difference in Viby-i strength between the two.

A brief overview of the lip data showed that Viby-i is produced with a lax lip posture, the acoustic and auditory effects of which should be investigated in further studies.

Regional and social stratification for Viby-i were found particularly in comparisons between metropolitan and urban areas, with metropolitan speakers tending to use a stronger Viby-i, with a higher F1, lower F2, and a greater degree of tongue backing and lowering. Comparisons between the Eastern and Western Central Swedish dialect regions, Stockholm and Gothenburg, men and women, and different socioeconomic backgrounds did not reveal any clear patterns of stratification. The lack of findings here may be an effect of the small and unbalanced sample, an issue which will be treated further in the Discussion.

5. Discussion

This chapter will begin by revisiting the research questions presented in the introduction and summarising the results of the study in relation to these questions. It will then link the findings to existing literature, reflect upon how successfully the research questions were answered, and discuss directions for further research. Potential issues raised by the results, or by the design of the study, will also be discussed.

The following research questions will now be addressed in turn:

1. What are the auditory properties of Viby-i?
   → Is there a binary distinction from [iː], or is Viby-i gradient?
2. How is Viby-i defined acoustically?'
   → What does it look like in terms of F1/F2?
   → How does it relate to other vowels in the system?
   → Do its acoustic properties agree with its auditory properties?
3. How is Viby-i articulated with reference to the tongue and lips?
   → Is it characterised by a certain tongue height, advancement, or shape?
   → What is the role of the lips in Viby-i production?
   → Do the articulatory properties correspond to expectations created by F1/F2?
4. Do regional or social factors influence /iː/ production?
   → Do these factors determine whether or not speakers use Viby-i?
   → Do the auditory, acoustic or articulatory characteristics of Viby-i vary between different groups?

5.1. Auditory description of Viby-i

According to previous literature, Viby-i is characterised by a “thick”, “buzzing” or “damped” auditory quality (Engstrand et al 1998: 83-84) which distinguishes it from standard [iː]. However, auditory descriptions of Viby-i and theories about its articulation sometimes vary greatly, suggesting that there may be more than one type of Viby-i. The auditory section of this study explored whether these types could be expressed on a gradient scale, and found that this was indeed the case. The results raise a problem in the Viby-i literature, as a strength dimension has not really been specified before. It is therefore unclear whether previous studies have applied the Viby-i label only to stronger versions of this vowel, or if they, like this study, have defined Viby-i by its difference in vowel quality from [iː].

A related problem, touched upon in the literature review, is that weaker types of Viby-i may not be perceived by Swedish speakers as different from standard [iː], as they are phonemically equivalent. This could have consequences for the way Viby-i is described; for example, it could be the case that weak Viby-i is more widespread than the literature suggests. The sample used in this study would support this theory, as all speakers were found to have Viby-i, even those from areas not directly associated with it, i.e. Varberg, Jönköping and Katrineholm. The Gothenburg multiethnolect speaker also used Viby-i, despite Bruce (2010: 136) specifically describing this as a variety that does not use this vowel. However, it is possible that Viby-i has simply become much more widespread since most of the literature was compiled.

Something which is not explored in this study is intra-speaker variability of Viby-i, as conditioned by e.g. duration or phonetic environment. Informal observations of the participants' recordings suggest that some speakers varied internally in Viby-i strength, sometimes between repetitions of the same word, and sometimes between words. Word-medial and word-final /iː/ tokens seemed to be particularly affected by stronger Viby-i, which could relate to differences in vowel duration or specific consonant environments. The degree of variability could also have to do with speech variety, as the multiethnolect speaker in the sample was observed to vary audibly in Viby-i strength from token to token. Social performance could also be a factor, as the experimental setting of the study may have caused participants to use more careful speech, resulting in the speakers selectively, but not necessarily consciously, adjusting their vowel qualities towards more canonical productions.

The auditory description of Viby-i in this study is quite brief, due to the fact that the auditory ratings were based on a small and homogeneous audio sample, which did not allow for more detailed comparison. The most important finding is that Viby-colouring does exist do different degrees, and that Viby-i was found throughout the sample, even where it was not expected. For future study, it would be interesting to look at the role of phonetic environment for intra-speaker variation, as well as the awareness and social evaluation of this vowel in Swedish speakers.

5.2. Acoustic description of Viby-i

The acoustic properties of Viby-i in this study correspond to its description in previous literature (e.g. Engstrand et al 1998, Björsten & Engstrand 1999), in that it is characterised by a low F2 and high F1 in comparison to standard [iː]. In the sample, /iː/ was located between /eː/ and /ʉː/ in the F1/F2 plane, and although it overlapped with /ʉː/ in particular, the overall distributions were different for each vowel. The most prominent parameter for Viby-i in the data was F2-lowering, but in the wider context of standard [iː] it was also found to undergo F1-raising. Only a rudimentary comparison could be made between Viby-i and standard [iː] however, as none of the speakers in the sample used [iː]. Instead, sample data from Fant et al (1969) was used for reference.

The study also investigated whether the unusual acoustic properties of Viby-i has resulted in a chain shift. According Engstrand et al, "/e/ tends to fill the gap in the acoustic vowel space created by the unusual vowel quality of [Viby-i]" (1998: 87). This would entail that as /iː/ undergoes F1-raising and F2-lowering, the opposite pattern would occur in /eː/ until the two vowels change places. A shift of this kind could enhance the perceptual contrast between /iː/ and /eː/. The comparison with Fant et al (1969) did not support this theory, however, as /eː/ seemed to undergo F2-lowering alongside /iː/. The acoustic 'high-front' space normally occupied by standard [iː] is thus left empty in Viby-i users. Nevertheless, it appears that sufficient perceptual contrast is still maintained between /iː/ and /eː/.

Another issue of contrast was posed by the partial overlap between /iː/ and /ʉː/ in the F1/F2 plane. This overlap was related to stronger instances of Viby-i, with a greater degree of F1-raising and F2-lowering. The unusual, 'dark' auditory quality of Viby-i may thus be related to its similarity in F1 and F2 to /ʉː/. The /iː/ and /ʉː/ vowels are nevertheless auditorily distinct, most likely because /ʉː/ is subject to strong lip-rounding and protrusion, which may be more visible in F3. Although data for F3 was collected, it was not investigated in this study, and it is possible that Viby-i displays some interesting characteristics in the higher formant ranges which could not be discerned here.

Overall, the acoustic parameters of F1-raising and F2-lowering coincided with stronger auditory strength ratings for Viby-i, indicating that these factors are important for its unique vowel quality. For example, significant correlations between Viby-i strength and demographic factors also tended to be significant in acoustics. An investigation of how Viby-i patterned acoustically in individual speakers could not be undertaken due to time constraints, but again it would be interesting to examine intra-speaker variation and the effects of phonetic context on Viby-i acoustics.

Another dimension which should be explored further is the dynamic properties of Viby-i, as well as the other vowels in the Swedish system. The highly dynamic nature of many of these vowels made it difficult to choose consistent measurement points on the acoustic signal, as the different 'phases' of each vowel varied in proportion and duration compared to the segment as a whole. Future work will attempt to incorporate the dynamic nature of the vowels into the analysis, as this appears to be a key characteristic of many Swedish vowels, including /iː/.

5.3. Articulatory description of Viby-i

Articulatorily, Viby-i was generally produced with a lowered tongue body and retracted tongue back compared to /eː/. The peak and tip of the tongue, however, tended to be extremely fronted; more so than most other vowels in the set. This illustrates one of the problems of only using the highest point of tongue (or indeed any single point) as a reference for the tongue as a whole. In height, /iː/ was often similar to /ʉː/, which may be connected to the overlap sometimes found between the two in the F1/F2 plane.

In line with the findings of Schötz et al (2011), Viby-i tended to be produced with an “upward slope” tongue shape. However, in many cases the shape also had a dip, resulting in a double-bunched articulation similar to that suggested by Borgström (1913). This shape was prevalent in both strong and weak Viby-i users. The tongue shape for Viby-i could not be compared to standard [iː] as there were no [iː] users in the sample, but future study should investigate whether the tongue shapes found in this sample are exclusive to Viby-i, and if so, what influence they have on the acoustic and auditory properties of the vowel. According to Fant (1960) “F1 of the vowels [e], [i], and [ɨ] is almost completely determined by the [oral] back cavity volume and the narrowest section of the mouth cavity” (1960: 121), both of which are likely to be affected by tongue shape. A further investigation of tongue shape in Viby-i is therefore needed.

Overall, the articulatory results corresponded to the acoustic values, however tongue backing seemed to have a stronger pull on acoustic values than tongue lowering, which was the main articulatory parameter. This again illustrates the complex relationship between articulation and acoustics, and could be caused by several articulatory factors, e.g. tongue shape and lip-rounding. Articulatory trade-off could potentially explain some of the discrepancy between acoustics and articulation, and could also be part of the reason why speakers who produced /iː/ with very different tongue gestures sometimes sounded similar.

An example of this could be seen in the strong Viby-i users, who used different articulatory strategies to achieve virtually the same auditory quality. However, it is also possible that there was fine-grained articulatory variation within the sample which was not visible due to the small and homogeneous sample group.

5.4. Regional and social distribution of Viby-i

Demographic information was collected in order to investigate if regional or social factors influenced whether or not speakers used Viby-i, as well as which type of Viby-i they used. As all speakers in the sample did use Viby-i, the regional distribution of this vowel as opposed to standard [iː] could not be addressed, although it could be stated that Viby-i occurred to a wider extent than expected. Rather than focusing on the distribution of the feature, then, the different types of Viby-i were instead investigated for differences based on regional and social groupings.

No clear influences of region were found when comparing speakers of Eastern and Western Standard Swedish, nor when comparing Stockholm and Gothenburg. However, speakers from metropolitan areas did differ from speakers from smaller, more peripheral urban areas. The results showed that the metropolitan speakers tended to have an auditorily stronger Viby-i, with a greater degree of F1-raising, F2-lowering, and possibly with a more backed and lowered tongue body than the urban speakers. The difference between the two groups corresponds to Schötz et al's (2014) finding that Viby-i was more common in central Stockholm and Gothenburg than in more peripheral locations. As the present study involved no rural speakers, no comparison could be made between the metropolitan version of the vowel and the more dialectal, rural Viby-i frequently described in the literature (e.g. Elert 1995; Björsten & Engstrand 1999).

There were no significant effects of sex or socioeconomic score on Viby-i production, although there seemed to be a weak trend towards women and speakers from higher class backgrounds to use a stronger version of the vowel. This could suggest that Viby-i is undergoing change, or that it is used as a status marker, as Bruce (2010) stipulates. However, the unbalanced sample made it very difficult to carry out a reliable analysis of these results, and a larger, stratified sample would be required to meaningfully investigate both regional and social relationships in the data.

5.5. Methodological issues

Finally, there are several theoretical problems that should be addressed when handling UTI data. For instance, in this study, tongue splines were drawn at the corresponding timepoint to the acoustic measurement, but this does not necessarily correspond to the articulatory steady-state of the vowel. Drawing the tongue splines at this timepoint was nevertheless a deliberate decision due to the dynamic nature of the vowels, as the steady-state might not occur until after diphthongisation or end-frication has set in. Again, the dynamic properties of these vowels are very interesting, and will be of greater focus in future work. Such a focus would also allow for investigations of coarticulatory and prosodic factors.

Another question which arose during the study was how articulatory measurements should be quantified. For example, where on the tongue spline should measurements be taken, and what should they be compared to? As seen in the acoustic results, the vowel system as a whole can differ between two dialects, and this is most likely reflected in articulation. This makes it very difficult to select a universal reference point, especially if this point is another vowel. Because every speaker's vocal tract is different, absolute measurements against e.g. the palate can also be unreliable, and can be more difficult to obtain with UTI.

However, the tongue is not the only articulator influencing vowel quality, and apart from UTI data the study also includes a qualitative overview of the lip positions associated with Viby-i. Although quantitative measurements were not made due to time constraints, the data nevertheless provided important insights into both lip and tongue behaviour that could not have been observed from UTI data alone. What we can learn from this is that different types of data are sometimes needed to form a full picture, especially of an issue as complex as vowel articulation. Future work will therefore continue to incorporate lip video, and also attempt to collect coronal tongue data, which could be used to explore aspects like tongue grooving.

6. Conclusion

The study has provided an exploratory description of the auditory, acoustic, articulatory and sociophonetic properties of Viby-i in 13 Central Swedish speakers, contributing to the small body of literature which currently exists on this vowel. Data provided by the study has shown that Viby-i is gradient in its auditory quality, existing in weaker and stronger forms, and that stronger forms of the vowel are associated with an increasing degree of F1-raising and F2-lowering, as well as tongue lowering and tongue backing. Many speakers also produced Viby-i with an unusual double-bunched tongue shape. The effects of this tongue shape and the possible effect of lip-rounding will act as points of interest for future work.

The relationship between articulation and acoustics in the study was found to be non-linear, emphasising the importance of combining acoustic and articulatory data, both here and in vowel research in general. The study has also shown that more work is needed on articulatory methodology, so that articulatory properties can be quantified more reliably.

The regional and social distribution of Viby-i could not be investigated in detail due to limitations in the sample, however some degree of stratification was nevertheless found in that metropolitan (Stockholm and Gothenburg) speakers tended to use a more pronounced version of Viby-i than urban speakers. More fine-grained stratification may become apparent in a larger and more balanced sample; the regional and social factors influencing Viby-i will therefore continue to be investigated. Future work also aims to look into the social evaluation of this vowel.

Further linguistic investigation of Viby-i will include a dynamic view of the vowel in the context of the broader Swedish vowel system, in order to investigate whether its 'buzzing' quality is somehow related to the fricated end-phase of /iː/. The researcher also wishes to explore the influence of factors such as prosodic environment, consonant environment, and vowel duration on Viby-i.

Apart from providing data on Viby-i, the study has outlined and evaluated methods for collecting and analysing articulatory data using ultrasound tongue imaging. With UTI, the study has been able to illuminate articulatory aspects of Viby-i that were previously unexplored, such as the presence of a double-bunched tongue shape, and the position of the tongue in the physical vowel space. The development of UTI methodology is nevertheless very much a work in progress, which will continue as the search for Viby-i goes on.

7. References

Alvtörn, L. (2005) 'En människas språkhistoria: I ljuset av ett vidgat lektbegrepp.' Dissertation submitted to the Institution of Nordic Languages, Lund University.

APC (2014) 'What is "PZT"?' [online]. APC International Ltd. https://www.americanpiezo.com/piezo-theory/pzt.html 2015-09-22

Björsten, S. & Engstrand, O. (1999) 'Swedish "damped" /i/ and /y/: Experimental and typological observations'. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 99)*, San Francisco, August 1999, 1,957-1,960.

Bodén, P. (2007) 'Språket hos ungdomar i en flerspråkig mijö i Malmö: "Rosengårdssvensk" fonetik och fonologi.' *Småskrifter från Nordiska språk vid Lunds universitet* 27. Centre for Languages and Literature, Nordiska språk, Lund University.

Boersma, P. & Weenink, D. (2015). Praat: Doing phonetics by computer, v. 5.3.66 [software]. http://www.praat.org/ 2014-03-09

Borgström, M. (1913) 'Askersmålets Ljudlära'. *Svenska Landsmål och Svenskt Folkliv*, B(11). Stockholm: P.A. Norstedt.

Breitholtz, E. (2008) 'Språklig variation: Dialekt, sociolekt, sexolekt, kronolekt, register och stil' [online]. Course material for *Språket, individen och samhället* at the University of Gothenburg. http://www.ling.gu.se/~/ellen/teaching/sisvt08/sissprakligvariation.pdf 2015-08-17

Bruce, G. (2010) *Vår fonetiska geografi: Om svenskans accenter, melodi och uttal.* Lund: Studentlitteratur.

Bryman, A. (2012) *Social Research Methods, 4th ed.* New York: Oxford University Press.

Catford, J. (1994) *A Practical Introduction to Phonetics.* New York: Oxford University Press.

Clark, G. (2012) 'What is the true rate of social mobility in Sweden? A surname analysis, 1700-2012' [online]. University of California, Davis. 16th August 2012. http://www.econ.ucdavis.edu/faculty/gclark/papers/Sweden%202012%20AUG.pdf 2015-08-18

Clark, J., Yallop, C. & Fletcher, J. (2007) *An Introduction to Phonetics and Phonology, 3rd ed.* Oxford: Blackwell.

Di Paolo, M., Yaeger-Dror, M., & Beckford Wassink, A. (2010) 'Analyzing vowels'. *Sociophonetics: A Student's Guide*, eds. Di Paolo, M. & Yaeger-Dror, M. New York: Routledge.

Einarsson, J. (2004) *Språksociologi.* Lund: Studentlitteratur.

Ekberg, L. (2011) 'Joint attention and cooperation in the Swedish of adolescents in multilingual settings: The use of *sån* “such” and *såhär* “like”'. *Ethnic Styles of Speaking in European Metropolitan Areas*, eds. Kern, F. & Selting, M. Amsterdam: John Benjamins.

Elert, C-C. (1995) *Allmän och svensk fonetik.* Stockholm: Norstedts.

English Speech Services (2013) 'The vowel space' [online]. Published 27 March 2013. http://englishspeechservices.com/blog/the-vowel-space/ 2015-03-15

Engstrand, O. (1999) 'Swedish'. *Handbook of the International Phonetic Association*, 140-142. Cambridge: Cambridge University Press.

Engstrand, O. (2004) *Fonetikens grunder.* Lund: Studentlitteratur.

Engstrand, O., Björsten, S., Lindblom, B., Bruce, G. & Eriksson, A. (1998) 'Hur udda är Viby-i? Experimentella och typologiska observationer'. *Folkmålsstudier* 39, 83-95.

Fant, G. (1960) *Acoustic Theory of Speech Production.* The Hague: Mouton.

Fant, G., Henningson, G. & Stålhammar, U. (1969) 'Formant frequencies of Swedish vowels'. *Quarterly Progress and Status Report: Department for Speech, Music and Hearing, TL-QPSR* 10(4), KTH, Stockholm, 26-31.

Faytak, M. & Lin, S. (2015) 'Articulatory variability and fricative noise in apical vowels'. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, August 2015.

Faytak, M. & Merrill J. (2014) 'Bantu spirantization is a reflex of vowel spirantization'. *Proceedings of Sound Change in Interactive Human Systems*, University of California, May 2014.

Hoedl, P. (2015) 'Defying gravity: Formant frequences of English vowels produced in upright and supine body position'. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, August 2015.

Håkansson, G. & Norrby, C. (2010) *Introduktion till sociolingvistik, 2nd ed.* Lund: Studentlitteratur.

International Phonetic Association (2005) IPA: Vowels [online]. https://www.internationalphoneticassociation.org/content/ipa-vowels 2015-03-14

Jones, D. (1917) *An English Pronouncing Dictionary.* London: J.M. Dent & Sons.

Jones, D. (1956) *The Pronunciation of English*. Cambridge: Cambridge University Press.

Jones, S. (1929) 'Radiography and pronunciation'. *British Journal of Radiology* 2, 149-150.

Kendall, T. & Thomas, E. (2010) Vowels: Vowel manipulation, normalization, and potting in R, v. 1.1 [software]. http://ncslaap.lib.ncsu.edu/tools/norm/ 2015-08-25

Labov, W. (2001) *Principles of Linguistic Change, vol. 2: Social Factors*. Oxford: Blackwell.

Ladefoged, P. (1967) *Three Areas of Experimental Phonetics*. London: Oxford University Press.

Ladefoged, P. (2006) UCLA Phonetics Lab Data [online]. *A Course in Phonetics, 5th ed.* Thomson/Wadsworth. http://www.phonetics.ucla.edu/course/chapter9/cardinal/cardinal.html 2015-03-14

Ladefoged, P. & Johnson, K. (2015) *A Course in Phonetics, 7th ed.* Stamford, CT: Cengage Learning.

Ladefoged, P. & Lindau, M. (1989) 'Modeling articulatory-acoustic relations'. *Journal of Phonetics* 17, 99-106.

Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World's Languages.* Oxford: Blackwell.

Ladefoged, P., DeClerk, J., Lindau, M. & Papçun, G. (1972) 'An auditory-motor theory of speech production'. *Working Papers in Phonetics, UCLA* 22, 48-75.

Lawson, E. (2007) 'Using ultrasound tongue imaging in sociophonetic research' [resource booklet]. Edinburgh: Queen Margaret University.

Lawson, E., Scobbie, J. & 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. & Stuart-Smith, J. (2013) 'Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation'. *Journal of Phonetics* 41(3-4), 198-210.

Lawson, E., Stuart-Smith, J. & Scobbie, J. (2008) 'Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English'. *Selected Papers from NWAV 35, University of Pennsylvania Working Papers in Linguistics* 14, 102-109.

Lindblad, P. (2010) *Fonetikens grunder* [compendium]. Linguistics Department, Lund University.

Lindblom, B. & Sundberg, J. (1971) 'Acoustical consequences of lip, tongue, jaw, and larynx movement'. *Journal of the Acoustical Society of America* 50, 1,166-1,179.

Lindström, F. (2014) 'Därför surrar stekaren som fiskaren från Smögen'. [online]. *Språktidningen* August 2014. http://spraktidningen.se/artiklar/2014/08/darfor-surrar-stekaren-som-fiskaren-fran-smogen 2015-08-24

Lundell, J. (1879) 'Det svenska landsmålsalfabetet'. *Svenska landsmål och svenskt folkliv 1*(2). Stockholm.

Ménard, L., Trudeau-Fisette, P., Côte, D. & Turgeon, C. (2015) 'Acoustic and articulatory correlates of speaking condition in blind and sighted speakers'. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS15)* [digital], Glasgow, August 2015.

Nair, A., Schellenberg, M., Gick, B. (2015) 'A case study on the efficacy of ultrasound biofeedback in voice pedagogy'. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS15)* [digital], Glasgow, August 2015.

Noiray, A., Iskarous, K. & Whalen, D. (2014) 'Variability in English vowels is comparable in articulation and acoustics'. *Laboratory Phonology 2015* 5(2), 271-288.

Noreen, A. (1903) *Vårt språk, vol. 1.* Lund: Gleerups.

R Core Team (2013) R: A language and environment for statistical computing [software]. R Foundation for Statistical Computing. http://www.R-project.org

Rosner, B. & Pickering, J. (1994) *Vowel Perception and Production*. Oxford: Oxford University Press.

Rosner, B. & Pickering, J. (1994) *Vowel Perception and Production.* Oxford: Oxford University Press.

Schötz, S. Frid, J., Gustafsson, L. & Löfqvist, A. (2014) 'Tongue articulation dynamics of /iː, yː, ʉ̟/ in Stockholm, Gothenburg and Malmöhus Swedish'. *Proceedings of FONETIK 2014*, Department of Linguistics, Stockholm University.

Schötz, S., Frid, J. & Löfqvist, A. (2011) 'Exotic vowels in Swedish: A project description and an articulographic and acoustic pilot study of /i/:'. *Proceedings of FONETIK 2011: Speech, Music and Hearing, TMH-QPSR,* KTH, Stockholm, 25-28.

Scobbie, J., Stuart-Smith, J. & Lawson, E. (2012) 'Back to front: A socially stratified ultrasound tongue imaging study of Scottish English /u/'. *Italian Journal of Linguistics/Rivista di Linguistica* 24(1), 103-148.

Scobbie, J., Wrench, A. & van der Linden, M. (2008) 'Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement'. *Proceedings of the 8th International Seminar on Speech Production: Models and Data*, 373-376.

Stark, J., Lindblom, B. & Sundberg, J. (1996) APEX: An articulatory synthesis model for experimental and computational studies of speech production. *Fonetik 96: Papers presented at the Swedish Phonetics Conference, TMH-QPSR* 2, 45-48.

Statistiska Centralbyrån (1982) 'Meddelanden i samordningsfrågor: Socioekonomisk indelning (SEI)', 4th reprint 1984 [online]. http://www.scb.se/statistik/_publikationer/OV9999_1982A01_BR_X11%C3%96P8204.pdf 2015-08-18

Statistiska Centralbyrån (2012) 'Meddelanden i samordningsfrågor för Sveriges officiella statistik: SSYK 2012 – Standard för svensk yrkesklassificering 2012.' [online] http://www.scb.se/statistik/_publikationer/OV9999_2012A01_BR_X70BR1201.pdf 2015-08-18

Stevens, K. (1972) 'The quantal nature of speech: Evidence from articulatory-acoustic data'. *Human Communication: A Unified View*, eds. Denes, P. & Davis, E. New York: McGraw-Hill, 51-66.

Stone, M. & Lundberg, A. (1996) 'Three-dimensional tongue surface shapes of English consonants and vowels'. *Journal of the Acoustical Society of America* 99(6).

SweDia (2000) 'Yngre man, Orust' [online]. SweDia project website, eds. Engstrand, O., Bruce, G. Elert, C-C., Eriksson, A. and Strangert, E. http://swedia.ling.gu.se/Gotaland/Bohuslan/Orust/ym.html 2015-08-24

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(9), 159-173.

Westerberg, F. (2013) 'An articulatory study of Swedish /i/ using ultrasound tongue imaging' [unpublished]. Undergraduate dissertation submitted to the Department of English Language, University of Glasgow.

Wikimedia Commons (2008) File: Cardinal vowel tongue position-front [online]. https://commons.wikimedia.org/wiki/File:Cardinal_vowel_tongue_position-front.svg 2015-09-07

WordReference.com (2015) Post #12, by user mexerica feliz on 15 June 2015 in 'Swedish: Pronouncing the long i sound' [online]. WordReference Language Forums. http://forum.wordreference.com/threads/swedish-pronouncing-the-long-i-sound.3029643/ 2015-08-24

Wrench, A. (2008) Ultrasound stabilisation headset [equipment]. Edinburgh: Articulate Instruments Ltd.

Wrench, A. (2013) 'Articulate Instruments: Ultrasound products – Headset' [online]. http://www.articulateinstruments.com/ultrasound-products/ 2015-09-22

Wrench, A. (2014) Articulate Assistant Advanced, v. 2.16 [software] Edinburgh: Articulate Instruments Ltd.

Zhang, M. (2006) Huma 300K: Topics in Chinese grammar, spring 2006 [online]. Hong Kong University of Science and Technology. http://teaching.ust.hk/~huma300k/suppl2.htm 2015-03-15

Zhou, X., Espy-Wilson, C., Boyce, S., Tiede, M., Holland, C. & Choe, A. (2008) 'A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/'. *Journal of the Acoustical Society of America* 123, 548-557.

8. Appendix

The appendix is supplied in digital format. To access it, please visit: http://tinyurl.com/WesterbergAppendix, or see enclosed CD.