

SWEDISH "VIBY-I": ACOUSTICS, ARTICULATION, AND VARIATION

Fabienne Westerberg

University of Glasgow
f.westerberg.1@research.gla.ac.uk

ABSTRACT

The paper investigates acoustic and articulatory variation in Swedish “Viby-i”, an /i:/ variant with an unusual vowel quality, characterised by a low F2. Viby-i appears to be part of a vowel shift, but its articulation is poorly understood, as speakers may be able to achieve it through multiple articulatory strategies. This study uses audio and ultrasound tongue imaging recordings of 34 speakers from Stockholm, Gothenburg, and Uppsala to explore variation in Viby-i acoustics, and its relation to socio-geographic and linguistic factors. It also examines the possibility of articulatory trade-off, comparing the relationship between tongue gesture and acoustic output. Viby-i is found to have a lower F2 in highly educated speakers, shorter vowel durations, and specific consonant environments. There is also a variable mapping between acoustics and articulation, such that F1 and F2 are partly disassociated from tongue height and backness. Possible compensatory behaviours are considered that could explain this phenomenon.

Keywords: Swedish, vowels, acoustic phonetics, articulatory phonetics, ultrasound tongue imaging

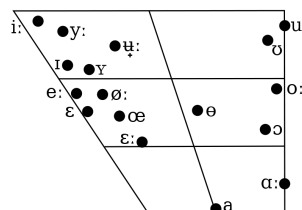
1. INTRODUCTION

Many vowel studies use acoustic data to make inferences about articulation, based on the finding that F1 and F2 can be used to predict tongue position [5, 19]. However, some vowels show the possibility of articulatory trade-off, whereby speakers use different strategies to achieve comparable acoustic outputs [13]. One vowel suspected of this is Swedish “Viby-i”, an /i:/ variant with an unusual, “thick”, “buzzing”, “damped” vowel quality [9]. This vowel is characterised by a low F2, and a high F1 and F3, but there are several articulatory strategies that could produce this effect [3, 15, 16, 17]. Social or linguistic factors could also condition variation in this vowel [4, 14]. This study explores such variation in both acoustics and articulation, to create a phonetic profile of Viby-i, and to contribute to a fuller understanding of vowel production.

1.1. Swedish vowels

Central Standard Swedish has 18 vowels (Fig. 1), divided into nine long/short pairs, which differ in duration and quality [8]. In the long vowels, contrast is maintained by three degrees of lip-rounding (e.g. [i:, y:, u:]), and various offglides: high vowels take a closing gesture, ending in frication, e.g. [i:j, u:β], while other vowels diphthongise towards schwa, e.g. [e:ə, o:ə] [8]. Despite this, the high to high-mid space is fairly crowded, and /ɛ:/ and /ø:/ are reported to be shifting [12]. Perceptual overlap has also been found between /i:/ and /y:/ [12].

Figure 1: Central Standard Swedish vowels [7].



1.2. Viby-i

Viby-i, also known as Lidingö-i [14], is an /i:/ variant found in many parts of Central Sweden [9]. In this paper, it is defined as an /i:/ which has a lower F2 than /e:/. The low F2 results in an unusual vowel quality, which may affect both /i:/ and /y:/ [1, 4, 6, 14]. While usually levelled in rural dialects (e.g. Viby), this vowel is associated with prestige and high socioeconomic class in the urban areas of Stockholm and Gothenburg. For this reason, it is said to be spreading rapidly, and Bruce [4] believes it will soon become the new standard variant. However, little is known about how common Viby-i is, or how it is realised in different parts of the country.

There are two experimental studies of Viby-i to date: Björsten & Engstrand [1] provide acoustic data for one older male speaker from Kräklinge (near Viby), and use an articulatory simulation model to find potential articulatory strategies. They suggest that Viby-i is a high central unrounded [i̠], which may be produced with a raised tongue tip to amplify

its “damped” quality. They also find that *Viby-i* is perceptually close to Turkish [i].

Frid et al. [10] investigate the tongue dynamics of /i:, y:, ɤ:/ in 27 speakers from Stockholm, Gothenburg, and Malmö, using electromagnetic articulography. They report that *Viby-i* is produced with a lower and backer tongue body, and different tongue tip behaviour, than [i:]. *Viby-i* was absent in Malmö speakers, and less common in speakers from the outskirts of Stockholm and Gothenburg. Different tongue gestures were also found for /i:/ and /y:/, despite previous literature stating that these vowels are only distinguished by lip rounding.

In addition to these studies, a perceptual pilot experiment for this study, relying on ratings from three non-native listeners, found that *Viby-i* was perceived as “stronger” in speakers with lower F2.

The sociolinguistic properties of *Viby-i* have not been examined in detail, but the presence or absence of this sound in young people was found to signal local identity in a suburb of Gothenburg [11], and gender and class in Stockholm [14]. It is also common in young Gothenburg speech [12].

2. METHOD

2.1. Speaker sample

34 native Swedish speakers from Gothenburg, Stockholm and Uppsala were recorded. Table 1 shows their distribution across city, age, and gender. Most participants had a university education, which was used as a proxy for socioeconomic class. The sample was largely homogeneous in terms of language background, ethnicity, and gender identity. The coding of the social variables is described in 3.2.

Table 1: Distribution of speaker sample.

Age	Gothenburg	Stockholm	Uppsala
18-29	2F, 2M	2F, 2M	2F, 2M
30-49	2F, 2M	2F, 2M	2F, 2M
50-80	2F, 2M	2F, 2M	2F, 0M

2.2. Materials

Speakers produced three randomised repetitions of a word list, comprising the long vowels /i:, y:, ɤ:, e:, ø:, ɛ:, ɑ:, o:, u:/ in varying contexts. All vowels varied in syllable context (*bi/bita*), and following consonant place (*bita/biga*). In addition, /i:/ varied in preceding consonant voicing (*bita/pita*), preceding consonant manner (*bita/vita*), and following consonant manner (*vita/vira/vila/vina/visa*). Each speaker

produced 51 tokens of /i:/, and 9 tokens of every other vowel. Demographic information, as well as information about social behaviours and attitudes, was collected through a questionnaire.

2.3. Equipment

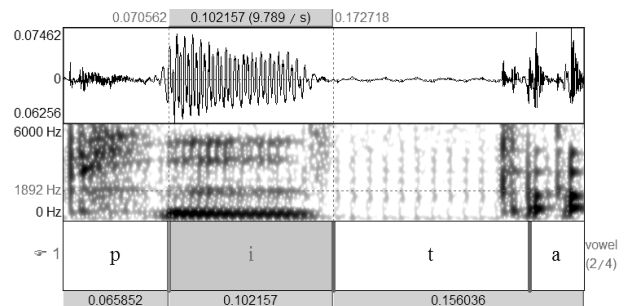
Recordings consisted of simultaneous audio, lip video, and ultrasound tongue imaging (UTI). UTI uses echolocation to generate an image of the tongue surface when an ultrasound probe is placed under the chin [20]. The output is a midsagittal image of the tongue contour. The hard palate can also be imaged when swallowing. This study used an Echo-Blaster 128 ultrasound with a 2-4 MHz convex probe, recording at 67.19 fps. The probe was held in place by a stabilising headset [22], and the probe angle was standardised using a bite plate [18].

Audio was collected using an Audio-Technica AT831b lapel microphone, sampling at 44,100 Hz. A profile view of the participant’s mouth was recorded with a custom-made headset camera [22], collecting NTSC video at 29.97 fps (interlaced).

2.4. Data preparation

The acoustic data was hand-segmented in Praat [2], and average F1 to F4 measures taken across each 10% of the vowel duration. Since the vowel segments included offglides as well as vocalic portions (Fig. 2), this analysis is based on the first 10%, due to the observation that offglides can start as early as 30% into the vowel. For brevity, only /i:, y:, e:, ɑ:, u:/ are presented in this paper.

Figure 2: Waveform and spectrogram of *Viby-i* in a young female speaker saying *pita* [p^hiːtɑ]. The vowel segment includes the fricated offglide.



The UTI data was prepared in Articulate Assistant Advanced [21], importing time stamps from Praat, and manually drawing an outline of the tongue surface at 10%, 50%, 90% of the vowel duration. The 10% tongue splines for /i:/ are presented here, contextualised by the palate and the smaller

vowel set. Tongue contours are not normalised between speakers, as there is currently no generally accepted method for normalising UTI data. This paper presents preliminary ultrasound results from the female speakers only, and will not present lip data, or measures of vowel frication.

3. ACOUSTIC RESULTS

3.1. Viby-i in the vowel space

Viby-i is characterised by a markedly low F2, and a high F1 and F3. Fig. 3 shows the position of /i:/ in the (normalised) vowel space, and Table 2 provides reference values in Hz. Viby-i has a similar F1 to /e:/, but its F2 is considerably lower. Meanwhile, /e:/ does not appear to have shifted, leaving the high front part of the vowel space empty.

There is considerable overlap between /i:/ and /y:/, with /y:/ having slightly lower formant values overall, presumably as a result of lip-rounding. These values nevertheless suggest that /y:/ is Viby-coloured in these speakers, supporting findings of perceptual overlap between /i:/ and /y:/ [12].

Figure 3: Normalised F1/F2 for all speakers.

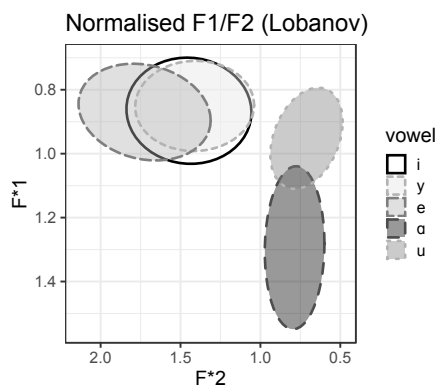


Table 2: Mean F1-F3 (Hz) by gender.

Females	/i:/	/y:/	/e:/
F1	398	394	398
F2	1946	1892	2368
F3	3209	3043	2945
Males	/i:/	/y:/	/e:/
F1	337	332	341
F2	1709	1660	1994
F3	2741	2576	2513

To determine if perceptual overlap was likely to occur with other nearby vowels, four linear mixed effects regression (LMER) models were run (one for

each formant), comparing the distributions of /i:/, /y:/, /e:/, /a:/, /u:/. The model showed no difference between any of these vowels in normalised F1, but significant differences between /i:/, /e:/, /a:/ in normalised F2 ($p < 0.01$), and significant differences between all vowels in normalised F3 and F4 ($p < 0.001$). Each vowel thus had its own distinct formant profile.

3.2. Social stratification

All speakers in the sample produced /i:/ with a lower F2 than /e:/, but the degree of F2 lowering was variable. To examine if this variation was conditioned by sociolinguistic factors, an LMER was run to investigate the effects on normalised F2 of:

- *City*, expecting Viby-i to be absent in Uppsala, and different between Gothenburg and Stockholm due to dialect differences.
- *Distance from city centre*, expecting speakers who grew up in central Stockholm or Gothenburg to have stronger Viby-i (lower F2).
- *Age*, expecting younger speakers have stronger Viby-i, if this is a change in progress.
- *Gender*, expecting women to have stronger Viby-i than men, due to prestige and possible change in progress.
- *Education*, expecting speakers with higher education (class) to have stronger Viby-i.
- *Local identity*, expecting locally “patriotic” speakers in Gothenburg and Stockholm to have stronger Viby-i, since it is a local marker.
- *Linguistic awareness*, expecting linguistically aware speakers in Gothenburg and Stockholm to have stronger Viby-i, due to prestige.
- *Social hobbies*, expecting highly social speakers in Gothenburg and Stockholm to have stronger Viby-i, due to prestige and possible change in progress.
- *Random factors*: Speaker and word, to account for effects of data clustering.

The only significant effect was education, in that normalised F2 decreased when education increased ($\beta = -0.35$, $p < 0.05$). In other words, highly educated speakers had a stronger Viby-i. However, this is a fairly weak effect, and the sample was not normally distributed. Nevertheless, given the scaling of Fig. 3, it is likely that this difference would be audible to listeners.

3.3. Linguistic environment

The same method was used to investigate the effects of linguistic environment on normalised F2, using the following factors:

- *Word frequency*, to account for some words in

the list being more frequent than others.

- *Vowel duration*, expecting longer vowels to have stronger Viby-i, since this phenomenon is mainly documented in long vowels [4].
- *Preceding fricative*, expecting Viby-i to be stronger after frication, as coarticulation might enhance its alleged “buzziness”.
- *Preceding voicelessness*, esp. in plosives, as aspiration might have similar effects to above.
- *Following backness*, expecting Viby-i to be stronger before back consonants, if it too is produced with a backed tongue position.
- *Following complexity*, expecting Viby-i to be stronger before complex consonants, if it too is produced with a complex articulation.
- *Random factor*: Speaker.

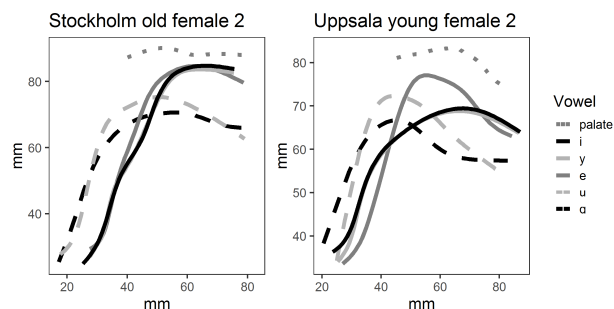
Normalised F2 decreased significantly ($p < 0.001$), i.e. Viby-i was stronger, when vowel duration was short ($\beta = -0.0002$), when followed by a complex consonant ($\beta = -0.02$), and when preceded by a fricative rather than a plosive ($\beta = -0.08$). The difference between fricatives and plosives was smaller when the consonant was voiced ($\beta = -0.03$). Based on the small coefficients, however, it is questionable to what extent these differences are audible.

4. ARTICULATORY RESULTS

The acoustic results suggest that Viby-i is produced with a backed, somewhat lowered tongue gesture, but the female UTI data shows that this is not the case in most speakers. Instead, Viby-i is mainly characterised by tongue lowering, as exemplified in Fig. 4 (right speaker). Only five out of 18 speakers had similar tongue heights for /i:/ and /e:/ (c.f. left speaker), but in these cases, the tongue body was usually fronted. There is thus a variable mapping between tongue position and F1/F2. For example, the speakers in Fig. 4 have very different tongue heights, but their F1 values are similar (460 vs. 520 Hz), while their F2 values differ greatly (2180 vs. 1890 Hz). This phenomenon persists after normalisation. In the right speaker, tongue lowering causes some tongue root retraction, but it is unclear whether this could have such a disproportionate pull on F2.

There is also inter-speaker variation in the tongue shape used for Viby-i. The tongue may be convex, steeply sloping, plateaued, or bunched (constricted at the front and the back, with a dip in the middle), while still producing similar acoustic output. Some of these shapes are more reminiscent of a consonant (e.g. /ɹ/) than a vowel. Note that /y:/ takes an almost identical tongue gesture to /i:/, which explains their acoustic overlap.

Figure 4: Mean tongue contours of two female speakers with similar F1, facing right.



5. DISCUSSION & CONCLUSIONS

The study supports previous findings that Viby-i is characterised by a low F2, and high F1 and F3. However, speakers are variable in acoustics, and the reasons behind this variation are not yet clear. The correlation between low F2 and high education suggests that Viby-i may be a prestige marker, but a more balanced sample is required to confirm this.

The fact that all speakers in the sample used Viby-i is surprising, as this vowel has not previously been documented in Uppsala, and is usually absent in phonetic descriptions of Swedish. Based on these results, Viby-i is likely to be more common in Central Sweden than previously believed. The prevalence of Viby-i in older speakers could also mean that, at least in larger cities, this vowel shift may already have taken place.

The linguistic effects are difficult to explain before learning more about the articulation of this vowel, but the link between shorter duration and lower F2 ties in with previous reports of Viby-i as a centralised vowel. Similarly, the link between low F2 and complex consonant environment suggests that complex tongue shapes may help explain the quality of this vowel. The role of the tongue in producing frication or “buzziness” will be explored further in future work.

The variable mapping between acoustics and articulation raises many questions, not only about Viby-i, but other vowels as well. How common is this kind of variation? Do speakers use it to compensate for vocal tract differences, or do they simply have preferred strategies? How do we investigate the relationships between complex tongue shapes and acoustics? For Viby-i, it seems that the traditional measures of tongue height and backness are not sufficient; instead, overall tongue gesture, and its relation to the acoustic chamber, seem to create the impression of this unusual vowel.

6. REFERENCES

- [1] Björsten, S., Engstrand, O. 1999. Swedish "damped" /i/ and /y/: Experimental and typological observations. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco.
- [2] Boersma, P., Weenink, D. 2017. Praat: Doing phonetics by compute, version 6.0.29 [software]. <http://www.praat.org/>.
- [3] Borgström, M. 1913. Askermålets ljudlära. *Svenska landsmål och svenskt folkliv* 2.
- [4] Bruce, G. 2010. *Vår fonetiska geografi: Om svenskans accenter, melodi och uttal*. Lund: Studentlitteratur.
- [5] Delattre, P. 1951. The physiological interpretation of sound spectrograms. *Publications of the Modern Language Association of America* 66, 864–875.
- [6] Elert, C. 1995. *Allmän och svensk fonetik*. Stockholm: Norstedts.
- [7] Engstrand, O. 1999. Swedish. In: *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 140–142.
- [8] Engstrand, O. 2004. *Fonetikens grunder*. Lund: Studentlitteratur.
- [9] Engstrand, O., Björsten, S., Lindblom, B., Bruce, G., Eriksson, A. 1998. Hur udda är Vi-by-i? Experimentella och typologiska observationer. *Folkmålsstudier* 39, 83–95.
- [10] Frid, J., Schötz, S., Gustafsson, L., Löfqvist, A. 2015. Tongue articulation of front close vowels in Stockholm, Gothenburg and Malmöhus Swedish. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- [11] Grönberg, A. 2004. *Ungdomar och dialekt i Alingsås*. Doctoral thesis, University of Gothenburg.
- [12] Gross, J. 2018. *Mapping vowels: Variation and change in the speech of Gothenburg adolescents*. Doctoral thesis, University of Gothenburg.
- [13] Guenther, F., Espy-Wilson, C., Boyce, S., Matthies, M., Zandipour, M., Perke, J. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105, 2854–2865.
- [14] Kotsinas, U. 2004. *Ungdomsspråk*. Uppsala: Hallgren & Fallgren.
- [15] Ladefoged, P., Lindau, M. 1989. Modeling articulatory-acoustic relations. *Journal of Phonetics* 17, 99–106.
- [16] Lundell, J. 1879. Det svenska landsmålsalfabetet. *Svenska landsmål och svenskt folkliv* 1.
- [17] Noreen, A. 1903. *Vårt språk*. Lund: Gleerups.
- [18] Scobbie, J., Lawson, E., Cowen, S., Cleland, J., Wrench, A. 2011. A common co-ordinate system for mid-sagittal articulatory measurement. *QMU CASL Working Papers* 20.
- [19] Stevens, K., House, A. 1955. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America* 27, 484–493.
- [20] Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics* 19, 455–501.
- [21] Wrench, A. 2004–2018. Articulate Assistant Advanced (AAA), version 2.17.05 [software]. <http://www.articulateinstruments.com/aaa/>.
- [22] Wrench, A. 2013–2018. Articulate Instruments Ltd.