QUANTIFYING TONGUE-TIP TO UPPER INCISOR DISTANCE IN AN ULTRASOUND TONGUE IMAGING STUDY OF SWEDISH /i/

Fabienne Westerberg[†], Eleanor Lawson[‡], Jane Stuart-Smith[†]

[†] University of Glasgow

[‡] University of Strathclyde

ABSTRACT

The Swedish /iː/ variant 'Viby-i' has been described as 'thick', 'damped' and 'buzzing', implying the presence of frication. The articulatory basis for its unusual buzzing quality is unknown. We investigate the possibility that frication results from proximity of the tongue tip to the upper incisors during Viby-i production. Ultrasound tongue imaging recordings of word-list speech from 34 Central Swedish speakers was used to study articulatory-acoustic variation across six Swedish long vowels, including Viby-i. Measures of vertical distance between highest point of the tongue and tongue tip indicate that the /iː/ and /yː/ tongue shapes are more apical than the other vowels. /iː/ and /yː/ also have shorter distances between the tongue tip and upper incisors, and acoustic analysis showed greater levels of aperiodicity for /iː/ and /yː/ than for other vowels. However, a correlational analysis did not identify a significant negative relationship between tongue-tip-to-incisor distance and levels of aperiodic noise.

Keywords: Ultrasound Tongue Imaging, vowels, Swedish, Articulatory Phonetics, Sound Change.

1. INTRODUCTION

The Swedish /iː/ variant known as Viby-i (sometimes also called Lidingö-i) has an unusual quality, impressionistically characterised as 'thick', 'damped' and 'buzzing' [1], [2], [3]. Named after Viby parish in Central Sweden, its articulatory nature has long been disputed. [yː] is also found to have a similar buzzing quality [4], but is investigated only articulatorily in this paper. A key acoustic feature of Viby-i, determined from perceptual experiments, appears to be that its F2 is lower than that of /eː/ [5]. However, there is no consensus on how Viby-i is produced, with conflicting impressionistic accounts of the tongue movements involved [6], [7] ,[8], [9]. An ultrasound tongue imaging (UTI) and lip camera study of Viby-i in 2016 [10], identified that many speakers in the study produced Viby-i with an extremely fronted tongue, to the point where the tip could sometimes be seen protruding between the front teeth. A typical visual example is provided in Figure 1 (left), showing front-facing and profile lip-camera views of the midpoint of a Swedish female speaker's prodiction of [iː] and [e] (for comparison). For [iː], but not for [e], the tongue tip can see seen to be in close proximity to the upper incisors.

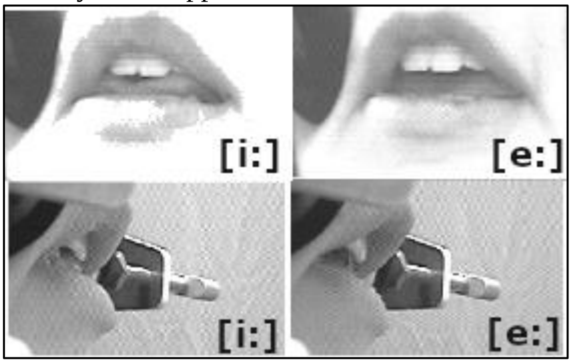

Figure 1. Comparison of tongue-tip postures for /iː/ (left) and /eː/ (right), produced by a female Swedish speaker [10].

We hypothesise that proximity of the tongue tip to the upper incisors is responsible for the 'buzzing', or fricated, quality of some variants of Viby-i, and use articulatory-acoustic methods to investigate aspects of Viby-i production compared to other Swedish long vowels.

2. METHOD

The existing subcorpus of Swedish word-list recordings used in this study was recorded in Central Sweden in 2016. Speakers were recorded in soundproofed studios, or in a quiet room, at the Universities of Gothenburg (12 speakers – 6M/6F), Stockholm (10 – 5M/5F), and Uppsala (8 – 3M/5F). Speakers were audio recorded and their midsagittal tongue movements were recorded with UTI, as they read aloud multiple repetitions of mono- and disyllabic words, containing a subset of the Swedish long vowels. Five long vowels were articulatorily and acoustically measured and analysed alongside /iː/, as a point of comparison, see Table 1.

Table 1: Swedish long vowels included in this study and number of tokens per vowel. N=2790.

vowel	iː	yː	eː	øː	oː	uː

N. tokens	1487	263	265	261	261	253

All /iː/ vowels in the study can be considered to be Viby-i variants, based on the acoustic criterion of having a lower F2 than /eː/.

2.1 Equipment

Recordings of the midsagittal tongue were made using a portable Telemed Echo-Blaster 128 ultrasound machine and a micoconvex probe with a 104º scan radius, and scanning at 67.19 frames per second. The probe was held in place under the chin using a stabilising headset to reduce/eliminate, probe movements in the coronal, axial and sagittal planes.

An Audio-Technica AT831b cardioid lapel microphone, sampling at 44,100 Hertz (16 bit), was clipped to the headset near the participant's mouth.

2.2 Articulatory measures

2.2.1 Occlusal/bite plane reference

The bite plane recording, see [12], is obtained using a medical-grade plastic bite plate (see Figure 2).

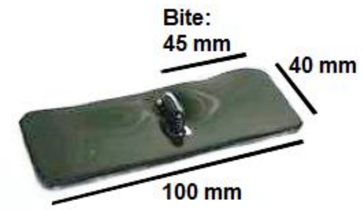

Figure 2. A plastic bite plate.

The bite plate is vacuum-moulded from a standard template with vertical protrusion part-way along its length. Participants are asked to place it inside their mouth, grip it with their molar teeth, with incisors positioned directly before the vertical protrusion. They are instructed to press their tongue flat against the underside of the bite plate, allowing imaging of a flat tongue surface that approximates the speaker's occlusal (bite) plane and helps determine the angle of the ultrasound probe relative to the occlusal plane, which can then be standardised to a 90º angle to improve inter-speaker comparison, see Figure 3.

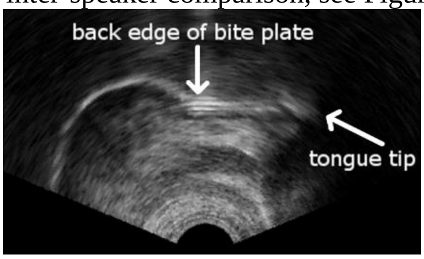

Figure 3. A midsagittal image of the "bite plane".

An additional potential benefit of using the bite plate is that it can help approximate the location of the biting edge of the upper incisors. The back of the bite plate is visible on the ultrasound image as the tongue bulges upwards, see Figure 3. It will be a fixed distance from the biting edge of the upper incisors allowing incisor position to be approximated.

We have estimated this fixed distance to be 45mm with a standard deviation of +/-1mm, based on measurement of 30 bite plates. Using AAA software [11], we fitted a scaled fiducial line, 45mm long, to the imaged bite plane surface, placing the left edge of the line at the point where the tongue's surface bulges upwards. The Cartesian coordinates of the termination of the line will approximate the location of the biting edge of the upper incisors.

2.2.2 Midsagittal tongue-surface measurement

A spline was fitted to each UTI-imaged midsagittal tongue surface, using AAA, at 10% of the vowel duration, as Central Swedish vowels reach their targets early, before moving into variable offglides [4]. Splines were then exported as sets of Cartesian coordinates for further analysis using R [13]. Automatic measures were taken of the highest point of the tongue and tongue tip location, and Lobanov normalised [14], to allow inter-speaker comparison.

2.2.4 Single-point tongue-surface measures

Single-point tongue measures can be used to (1) identify the relative tongue height of the vowels studied, and (2) characterise and quantify tongue shape by subtracting the tongue-tip height from the highest point of the tongue. If these values are identical, or very similar, resulting in a value close to 0, then it can be assumed that the vowel is produced with an apical (tongue-tip-raised) articulation; whereas greater positive vertical distance between the highest point and tongue tip indicates a more canonical convex vocalic tongue shape, with a lowered tongue tip and raised tongue middle, or dorsum.

2.2.3 Tip-to-incisor measure

The tongue-tip-to-biting-edge-of-incisor measure (in mm) was obtained automatically, using R, measuring the Euclidean distance between the (x,y) coordinates for (1) the tongue tip and (2) the first coordinate of the bite-plane fiducial.

2.3 Acoustic measures

Bandpass-filtered zero-crossing rate (bpZCR) quantifies the degree of aperiodicity in an acoustic signal, which may contain voicing [15]. This method filters out the vowels' periodic energy, and measures how often the waveform crosses the zero point in the remaining signal. Higher rates of zero-crossings indicates greater levels of aperiodicity/frication.

First, all frequencies below 1,000 Hz were removed. Filtering was carried out using the 'Hann Band filter' function in Praat [16], set to 100 Hz smoothing. Then the 'Zero Point Process' function was used to identify all (falling and rising) zero-crossings. The average zero-crossing rate was calculated for a 10% window the vowel onset, as acoustic analysis of Central Swedish vowels showed them to be very dynamic and much more variable after the first 10% of the vowel [17].

2.4 Articulatory and articulatory-acoustic analysis

Using spline and fiducial Cartesian coordinates, R was used to automatically quantify (1) the highest point of the tongue, (2) vertical distance between highest point of the tongue and tongue-tip, (3) Euclidean distance between the tongue tip and biting edge of the upper incisors. Praat was used to measure (4) mean bpZCR 10% of the vowel. Finally, using R, we carried out (5) a correlational analysis of measures (3) and (4), hypothesising that a significant negative correlation would be found.

3. RESULTS

3.1 Highest point of the tongue

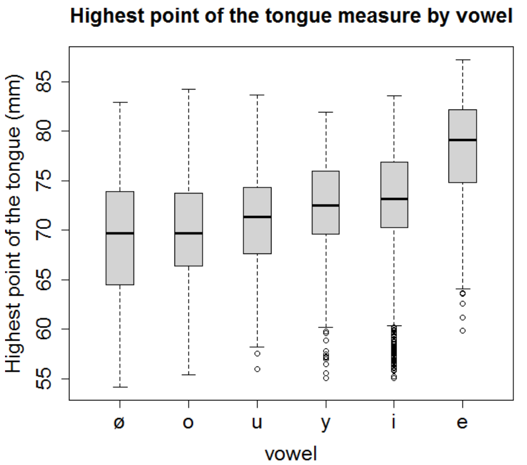

Figure 4. Highest point of the tongue measure (mm – distance from probe surface), by vowel, ordered by mean. N=2790.

Figure 4 shows the highest point of the tongue for all vowels. An ANOVA identified significant variation across these vowels, F=80.99, p<0.001. Figure 4 shows that /eː/, not /iː/, is the highest vowel in articulatory space. Post-hoc Tukey tests showed significant differences between /eː/ - /iː/ and /eː/ - /yː/, p<0.001.

3.1 Highest point of the tongue to tongue-tip height

Figure 5 shows the vertical distance between the tongue tip and the highest point of the tongue. An ANOVA identified significant variation across the vowels studied F=528.9, p<0.001. Post-hoc Tukey tests showed no significant differences between /iː/ and /yː/, which can therefore be said to have the most tip-raised articulation of the vowels studied. The tongue tip was identified as the highest point of the tongue (indicating an apical articulation) for 15% and 13% of /yː/ and /iː/ tokens respectively; whereas the tongue tip was the highest point of the tongue for /øː/ and /eː/ in only 7% and 4% of tokens respectively, and never for /oː/ and /uː/. Post-hoc Tukey tests found significant differences between /yː/, /iː/ and all other vowels studied at p<0.001.

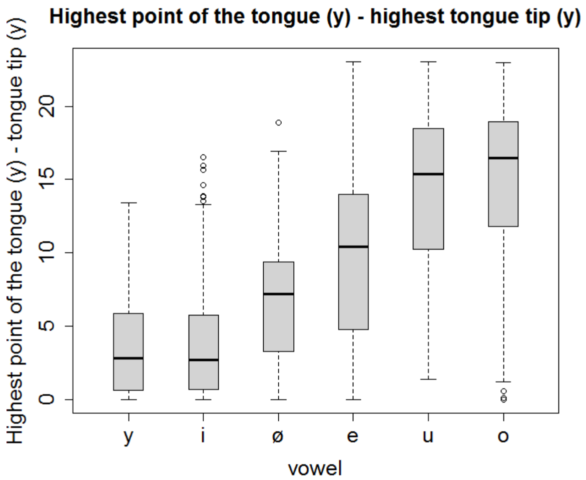

Figure 5. Vertical distance measure between the highest point of the tongue and tongue tip (mm), by vowel, ordered by mean. N=2790.

3.1 Articulatory analysis – tongue-to-teeth measure

Figure 6 shows the tongue-tip-to-biting-edge-of-incisor measure. An ANOVA found significant variation across vowels F=71, p<0.001. Tukey posthoc tests showed no significant differences between /iː/, /yː/ and /øː/, but significant differences between these vowels and all other vowels at p<0.001. The significant difference between /e/ and /ø/ was at p<0.05.

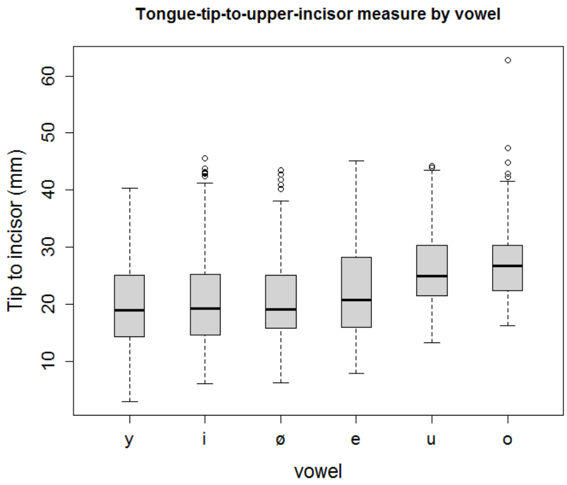

Figure 6. Tongue-tip-to-upper-incisor Euclidean distance (mm), by vowel, ordered by mean. N=2790.

3.2 Acoustic analysis – bandpass-filtered zero-crossing-rate measure

Figure 7 shows the mean bandpass-filtered zero-crossing rate (bpZCR) across the first 10% of the vowel.

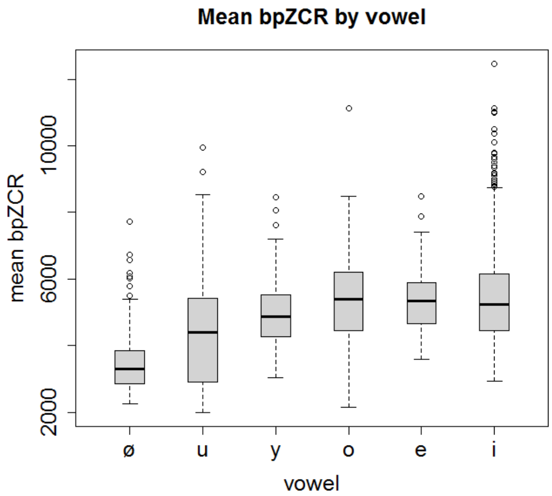

Figure 7. Average bpZCR by vowel, ordered by mean. N=2790.

A higher zero-crossing rate indicates greater levels of aperiodicity in the acoustic signal. An ANOVA found significant differences in teeth-to-incisor measures across vowels F=126.9, p<0.001. Post-hoc Tukey tests showed no significant differences in mean bpZCR between /iː/, /eː/ and /oː/, but significant differences between /iː/ and /yː/, p<0.001, /eː/ and /yː/ p<0.01, and /iː/ and /oː/, p<0.001.

3.3 Acoustic and articulatory correlational analysis

A Pearson's correlation test was carried out on the dependent variables (1) Euclidean distance between tongue tip and biting edge of the upper incisors, and (2) mean bpZCR value across the first 10% of the vowel. There was a significant positive correlation r_P=0.15, p<0.001 (see Figure 8), which was contrary to our hypothesis that a narrower tongue-tip-to-upper-incisor gap would result in greater frication (indicated by bpZCR).

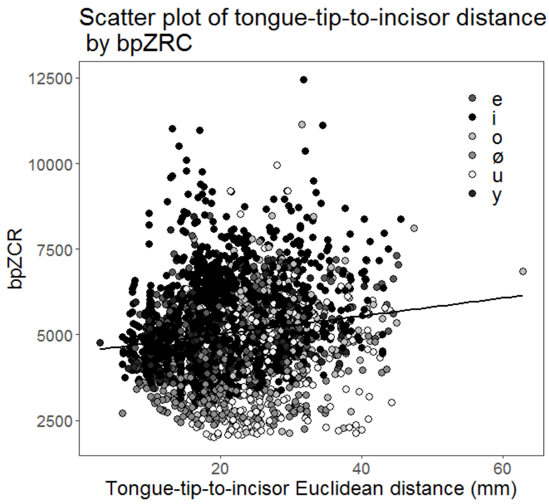

Figure 8. Tongue-tip-to-incisor Euclidean distance (mm) by mean bpZCR (aperiodic noise measure). N=2790.

4. CONCLUSION

This study investigated whether the unusual "buzzy", or fricated, quality of Viby-i could be attributed to the proximity of the tongue tip to the upper incisors.

Our results show that while /eː/ has, on average, a higher tongue-body position in the oral cavity than /iː/, /iː/ has, on average, a higher tongue-tip position, and often an apical tongue shape. We found that /yː/ is also often produced with this articulatory strategy, and, indeed, has also been found to have a similar buzzing quality [4]. /iː/ had the greatest degree of aperiodic energy of the vowels studied, but not significantly greater than /eː/, or /oː/. Our hypothesis of a significant negative relationship between tongue-tip-to-incisor distance and bpZCR value (as a measure of aperiodic noise) was disproved. Despite this negative result, we have established that Viby-i is often produced with a raised tongue tip – an unusual vocalic tongue posture. Additionally, the method of approximating the position of the biting edge of the upper incisors using a bite plate could be of use to future UTI-based speech studies, providing another fixed landmark from the upper surface of the vocal tract with which to compare tongue position.

7. REFERENCES

[1] Engstrand, O., Björnsten, S., Lindblom, B., Bruce, G., & Eriksson, A. (1998). Hur udda är Viby-i? Experimentella och typologiska observationer. Folkmålsstudier, 39, 83–95.

[2] Kotsinas, U.-B. (2007). Ungdomsspråk (3rd ed.). Uppsala: Hallgren & Fallgren.

[3] Schötz, S., Frid, J., & Löfqvist, A. (2011). Exotic vowels in Swedish: An articulographic and acoustic pilot study of /i:/. In Proceedings of the 17th International Congress of Phonetic Sciences (pp. 1766–1769). Hong Kong.

[4] Bruce, G. (2010). Vår Fonetiska Geografi: Om Svenskans Accenter, Melodi och Uttal. Lund: Studentlitteratur.

[5] Björsten, S., & Engstrand, O. (1999). Swedish "damped" /i/ and /y/: Experimental and typological observations. In Proceedings of the 14th International Congress of Phonetic Sciences (pp. 1957–1960). San Francisco.

[6]Lundell, J. A. (1878). Det svenska landsmålsalfabetet. Svenska landsmål och svenskt folkliv, 1(2), 11–158.

[7] Borgström, M. (1913). Askermålets ljudlära. Svenska landsmål och svenskt folkliv, B(11).

[8] Ladefoged, P. & Lindau, M. (1989). Modeling articulatory-acoustics relations: A comment on Stevens' "On the quantal nature of speech". Journal of Phonetics, 17, 99–106.

[9] Noreen, A. (1903). Vårt språk (Vol. 1). Lund: Gleerups.

[10] Westerberg, F. (2016). An Auditory, Acoustic, Articulatory and Sociophonetic Study of Swedish Viby-i (Masters dissertation). University of Glasgow.

[11] Articulate Instruments Ltd. (2019). Articulate Assistant Advanced, version 2.17. Retrieved 2019-03-26,from http://www.articulateinstruments.com/downloads

[12] Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of articulatory and acoustic parameters for the GOOSE vowel across British Isles Englishes. The J. Acoust. Soc. Am., 146(6). 4363–4381.

[13]R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria. Retrieved 2019-11 28, from http://www.r-project.org/

[14]Lobanov, B. M. (1971). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. 49(2), 606–608.

[15]Gordeeva, O. B., & Scobbie, J. M. (2010). Preaspiration as a correlate of word-final voice in Scottish English fricatives. In S. Fuchs, M. Toda, & M. Zygis (Eds.), Turbulent sounds: An Interdisciplinary Guide (pp. 167–208). New York: Walter de Gruyter.

[16]Boersma, P., & Weenink, D. (2019). Praat: Doing Phonetics by Computer, version 2.0.29. Retrieved 2019-03-26, from http://www.praat.org/

[17] Westerberg, F. (2020) Heavens, what a sound! The acoustics and articulation of Swedish Viby-i. Unpublished PhD thesis, University of Glasgow.