The formant dynamics of long close vowels in three varieties of Swedish

Ewald, Otto; Asu, Eva Liina; Schötz, Susanne

Published in:
Interspeech 2017

DOI:
10.21437/Interspeech.2017-1134

2017

Document Version:
Publisher's PDF, also known as Version of record

The formant dynamics of long close vowels in three varieties of Swedish

Otto Ewald[1], *Eva Liina Asu*[2], *Susanne Schötz*[1]

[1]Lund University, Sweden
[2]University of Tartu, Estonia

otto.ewald@sol.lu.se, eva-liina.asu@ut.ee, susanne.schotz@med.lu.se

Abstract

This study compares the acoustic realisation of /iː yː ʉː uː/ in three varieties of Swedish: Central Swedish, Estonian Swedish, and Finland Swedish. Vowel tokens were extracted from isolated words produced by six elderly female speakers from each variety. Trajectories of the first three formants were modelled with discrete cosine transform (DCT) coefficients, enabling the comparison of the formant means as well as the direction and magnitude of the formant movement. Cross-dialectal differences were found in all measures and in all vowels. The most noteworthy feature of the Estonian Swedish long close vowel inventory is the lack of /yː/. For Finland Swedish it was shown that /iː/ and /yː/ are more close than in Central Swedish. The realisation of /ʉː/ varies from front in Central Swedish, to central in Estonian Swedish, and back in Finland Swedish. On average, the Central Swedish vowels exhibited a higher degree of formant movement than the vowels in the other two varieties. In the present study, regional variation in Swedish vowels was for the first time investigated using DCT coefficients. The results stress the importance of taking formant dynamics into account even in the analysis of nominal monophthongs.

Index Terms: vowels, formant dynamics, Central Swedish, Estonian Swedish, Finland Swedish, Vowel Inherent Spectral Change, Discrete Cosine Transform

1. Introduction

Swedish is unique among world's languages because of a number of phonologically distinct contrasts in the inventory of close vowels [1]. It has been shown that there exists a considerable variation in the realization of these contrasts depending on the variety of Swedish [2], [3]. The aim of the present study is to focus on long close vowels in three varieties of Swedish: Central Swedish (CS) as spoken in the outskirts of Stockholm, Finland Swedish (FS) as spoken in the Swedish speaking area outside Helsinki, and Estonian Swedish (ES), a nearly extinct lesser studied variety of Swedish once spoken in the coastal areas and islands off the west coast of Estonia but now surviving only in the speech of elderly immigrants residing mainly in the Stockholm area.

CS exhibits a phonological four-way contrast in the close vowels /iː yː ʉː uː/ where /iː/ and /yː/ are front vowels and /ʉː/ is a fronted vowel (usually transcribed as [ʉ̟ː]) with many similar articulatory and acoustic features, and /uː/ is a back vowel [4]. The vowels /iː/ and /yː/ display similar F1 and F2 values [5], and can be separated only by F3, which is lower for /yː/. The only relevant phonetic difference between /ʉː/ and /yː/ can be seen in the F2 and F3 values [6]. A characteristic of long close vowels in CS is that they tend to be diphthongised. According to Kuronen [3] the first three formants at the end of a diphthongised /ʉː/ and /uː/ are lowered, while the diphthongisation of /iː/ and /yː/ results in a lower F1 and higher F2 at the end of the vowel.

In FS, the close vowels differ somewhat from the CS ones, except for /uː/ which is rather similar in both varieties. The FS /iː/ and /yː/ are pronounced more open and further front than their CS counterparts; acoustically, these vowels are realised with higher F1 and F2 values than in CS [3]. Furthermore, /ʉː/ is pronounced further back in FS as compared to CS [3].

ES dialects are said to lack the rounded front vowel /yː/ [7], [8]. Words where in CS the /yː/ phoneme occurs are in ES instead realised with [iː], [eː], [ɛː], or [ʉː] or a diphthong [iœː] or [iʉː] (for examples see [8]) depending on the segmental context of the word and the dialect. A pilot study comparing ES long close vowels to those of CS and FS [9] showed that the Rickul variety of ES distinguishes between three long close vowels that were relatively evenly distributed in the vowel space. These results were, however, based on only one measurement point taken at the midpoint of each vowel, which did not make it possible to capture diphthongisation (that occurred in the test word used for eliciting /yː/). Therefore, the features of the ES inventory of long close vowels need further clarification.

The present study seeks to address the limitations of the pilot study by using a larger set of data as well as different methodology that would account for dialectal variation in the formant dynamics of these vowels. As vowel formants are not static but display vowel inherent spectral change (VISC) over time (e.g. [10]) a number of recent studies have sought to better understand this time-varying behaviour of formants. While some studies have investigated measures of the extent and speed of the formant change and found these to vary both between vowel categories as well as between dialects (e.g. [11], [12], [13]), others have used the coefficients of formant tracks fitted with the discrete cosine transform (DCT) to analyse formant means as well as the direction and magnitude of the formant change (e.g. [14], [15]).

In the current study, formant tracks will be modelled as DCT coefficients, a method that has previously not been used for the analysis of Swedish vowels. Our main aim is to compare the formant dynamics of long close vowels in the three varieties of Swedish. Cross-dialectal differences between vowels could potentially be manifested both in the general location of the vowels in the vowel space as well as in the direction and extent of the formant movement. Based on earlier research (e.g. [9], [3]), we expect the three varieties of Swedish to potentially differ in the F1 and F2 of /iː/, /yː/, and /uː/, and in the F2 of /ʉː/. By and large, we hope to corroborate and specify earlier findings but also add important new knowledge about Swedish vowels in general.

2. Materials and method

2.1. Speech data

The materials consisted of test words produced by 18 elderly speakers of Swedish: six ES speakers (aged 83–89, median 86) originally from the Island of Vormsi (Swedish: Ormsö) which together with the Nuckö and Rickul sub-dialects forms the largest ES dialectal area; six CS speakers (aged 64–75, median 68) from two villages near Stockholm: Kårsta and Villberga; and six FS speakers (aged 56–88, median 71) from two villages in the Swedish speaking area of Finland near Helsinki: Borgå and Kyrkslätt. Only female speakers were chosen in order to minimise the need for formant normalisation.

The CS and FS speech material was taken from the Swedish dialect project SweDia 2000 (see e.g. [16]) research database, and the dialectal varieties were selected based on their geographical proximity to the cities of Stockholm and Helsinki respectively. The ES data were collected as part of the Estonian Swedish Language Structure (ESST) project. In both projects, a word list of about 100 words was elicited using questions in order to obtain semi-spontaneous speech.

All test words were monosyllabic with a CVC(C) structure. The SweDia material used for CS and FS consisted of three repetitions of four test words, one for each target vowel. Different test words were used for eliciting ES vowels (3–4 words per target vowel). As consonant context has been shown to have an effect on vowel formants [17], the ES test words were chosen to match the SweDia test words as closely as possible including a similar (mainly dental or alveolar) consonant context. As the words where /yː/ occurs in CS can have different phonetic realisations in ES, only test words with the [iː] realisation were chosen for elicitation in ES, since this is the most common equivalent of the CS /yː/. The following test words were used for CS and FS: *dis* [diːs] 'haze', *typ* [tyːp] 'type', *lus* [lʉːs] 'louse', and *sot* [suːt] 'soot', and for ES: *bita* [biːt] 'to bite', *dyr* [diːr] 'expensive', *liv* [liːv] 'waist', *lysa* [liːs] 'to shine', *ris* [riːs] 'rice', *syl* [siːl] 'awl', *tysk* [tiːsk] 'German', *vis* [viːs] 'manner', *duk* [dʉːk] 'tablecloth', *fågel* [fʉːɽ] 'bird', *sup* [sʉːp] 'drink', *bok* [buːk] 'book', *fot* [fuːt] 'foot', and *rot* [ruːt] 'root'. The number of tokens for each target vowel is shown in Table 1.

Table 1: *The number of tokens analysed for each target vowel in Central Swedish (CS), Estonian Swedish (ES) and Finland Swedish (FS).*

Target vowel	**CS**	**ES**	**FS**
/iː/	21	87	26
/yː/	19	-	28
/ʉː/	24	43	29
/uː/	29	35	27

The CS and FS speakers had been recorded in their homes with a Sony portable DAT recorder TCD-D8 and a Sony tie-pin type condenser microphone ECM-T140 at a 48 kHz/16 bit sampling frequency. The same microphone type was used with the ES speakers who were recorded in a quiet environment with a Roland R-09HR WAVE/MP3 recorder at a 44.1 kHz/16 bit) sampling frequency. The recordings were transferred to a computer and downsampled to 16 kHz.

2.2. Analysis

All acoustic analyses were carried out in Praat [18], and R [19] was used for the statistical analysis. The vowels in the CS and FS material (transcribed by SweDia 2000) were extracted using a script, and all the transcriptions were manually checked and corrected. The ES words were segmented and transcribed manually and the target vowel tokens were extracted automatically. After an auditory and visual analysis of the spectrograms, vowels with too weak formants or measurement errors due to non-modal voice qualities were eliminated from further analysis. The remaining vowel tokens were divided into four categories based on auditory analysis. The final data set consisted of 368 tokens of the four target vowels (see Table 1 for the distribution of the tokens).

Using several Praat scripts the first three formant frequencies were measured at 30 equidistant points over the central 60% of each vowel. Based on [20] and [13], optimal formant ceilings for each vowel in each dialect were calculated by measuring the first five formant frequencies at each sampling point 151 times, increasing the formant ceiling in 10 Hz increments from 5000 to 6500 Hz. For each point, the optimal ceiling was subsequently selected as the one that yielded the smallest total variance in F1-F3 (calculated along a logarithmic scale) in that vowel and dialect. The formant tracks were then fitted with parametric curves using DCT in order to produce smoother trajectories as well as to obtain coefficient values for analysis, using the formula in [14].

The analysis focused on the 0^{th} and 1^{st} DCT coefficients (C0 and C1) of F1-F3, where C0 is proportional to the formant mean, while C1 describes both the direction and the magnitude of the formant movement. The 2^{nd} DCT coefficient (C2) was excluded since it contains little additional information as compared to C0 and C1 (e.g. [14], [15]). In order to lessen the effect of measurement errors from the automatic formant analysis, outliers were removed from all measures with an R script [21] using the Tukey's method [22].

As a preliminary step in order to verify our classification of /iː/ tokens in ES that correspond to the CS and FS phoneme /yː/, the realisation of such tokens was compared to the remaining ES /iː/ tokens using a Welch's t-test. As the results were non-significant for all measures, all these tokens were subsequently treated as /iː/ in the cross-dialectal analysis.

The cross-dialectal variation was analysed by fitting a linear mixed-effects model with a fixed effect for dialect and a random effect for speaker to each of the six measures in each of the four vowels using the lme4 package [23], and testing the fixed-effect term in each model using the anova() function from the lmerTest package [24]. Pairwise comparisons of least-squares means of the levels of the dialect factor were conducted using the lsmeans package [25]. The p-values were adjusted for multiple comparisons using the multivariate *t* (mvt) method.

3. Results

Figure 1 shows F1-F2 and F2-F3 plots with corresponding formant movements of /iː yː ʉː uː/ in the three dialects. The formant trajectories have been smoothed using DCT.

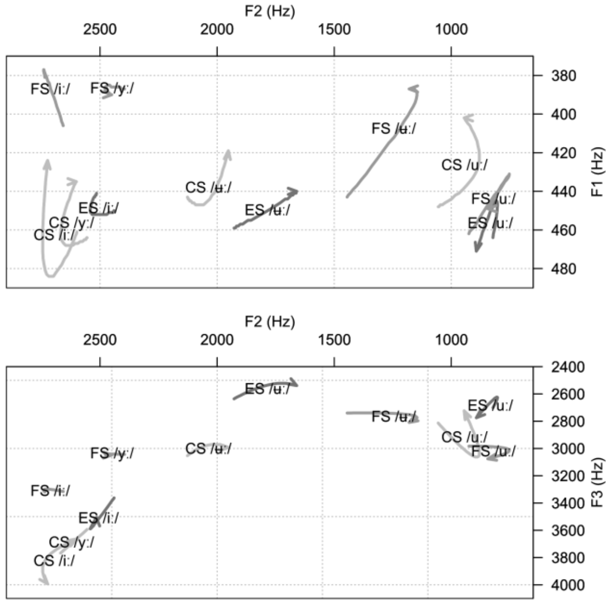

Figure 1: *F1-F2 (top) and F2-F3 (bottom) plots of the long close vowels /iː yː ʉː uː/ in Central Swedish (CS, light grey arrows), Estonian Swedish (ES, dark grey arrows), and Finland Swedish (FS, medium grey arrows). Labels are located at the formant means, while arrows represent the formant movement.*

The statistical analysis revealed a significant main effect of dialect in /iː/ on F1C0 ($F(2,14) = 4.88$, $p = .025$), F3C0 ($F(2,13) = 4.49$, $p = .032$), and F3C1 ($F(2,127) = 7.27$, $p = .001$). The effect of dialect was not significant on the remaining three measures. The results of the pairwise tests, displayed in Table 2, showed that significant differences between the dialects are manifested as a lower F3C1 in CS than in ES and FS (i.e. more formant movement towards lower F3 frequencies in CS), as well as lower F1C0 and F3C0 values in FS compared to CS.

There was a significant main effect of dialect in /yː/ on F1C0 ($F(1,9) = 5.55$, $p = .042$), F3C0 ($F(1,9) = 20.36$, $p = .001$), and F1C1 ($F(1,45) = 8.12$, $p = .006$), with all three measures having lower values in FS than in CS, where the difference in F1C1 indicates more formant movement towards higher F1 values in CS. The effect of dialect was not significant on the other measures.

For /ʉː/ the statistical analysis showed a significant main effect of dialect on F2C0 ($F(2,13) = 45.38$, $p < .001$), F3C0 ($F(2,14) = 10.25$, $p = .002$), and F1C1 ($F(2,11) = 4.36$, $p = .039$), with non-significant results for the other measures. The pairwise tests indicated that significant differences between dialects are manifested as a lower F2C0 value in FS than in CS and ES, higher F2C0 and F3C0 values in CS than in ES, and a higher F1C1 value in FS than in ES (indicative of more formant movement towards lower F1 frequencies in FS than in ES.

There was a significant main effect of dialect in /uː/ on F2C0 ($F(2,15) = 7.04$, $p = .007$), F3C0 ($F(2,15) = 6.42$, $p = .009$), F1C1 ($F(2,84) = 8.89$, $p < .001$), F2C1 ($F(2,13) = 8.67$, $p = .004$), and F3C1 ($F(2,17) = 4.17$, $p = .033$). The main effect of dialect on F1C0 was found to be non-significant. Subsequent pairwise comparisons showed significantly higher F2C0 and F1C1 values in CS compared to ES as well as a lower F2C1 in ES than in CS, where the difference in F1C1 reflects more formant movement towards lower F1 frequencies in CS, and the difference in F2C1 reflects formant movement towards higher F2 frequencies in ES but towards lower F2 frequencies in CS. The pairwise comparisons also showed a significantly higher F2C0 value in CS compared to FS, as well as lower F3C0 and F2C1 values in ES than in FS, where again ES shows formant movement towards higher rather than lower F2 frequencies. There was, however, no indication of a significant dialectal variation in F3C1 in the pairwise tests.

Table 2: *Results of the pairwise comparisons of dialect pairs in each vowel category for the first two DCT coefficients of F1-F3. Results significant at the α-level of 0.05 are highlighted with grey background.*

		CS-ES		**CS-FS**		**ES-FS**	
		C0	**C1**	**C0**	**C1**	**C0**	**C1**
/iː/	**F1**	*t*(13)= 0.68	*t*(13)= 1.35	***t*(15)= 2.90**	*t*(16)= 0.62	*t*(13)= 2.43	*t*(13)= -0.74
	F2	*t*(13)= 1.57	*t*(17)= -1.11	*t*(14)= -0.01	*t*(30)= -0.94	*t*(13)= -1.66	*t*(17)= 0.00
	F3	*t*(13)= 1.94	***t*(25)= -2.56**	***t*(14)= 2.98**	***t*(50)= -3.81**	*t*(13)= 1.13	*t*(23)= -2.17
/yː/	**F1**	-	-	***t*(8)= 2.36**	***t*(9)= 2.83**	-	-
	F2	-	-	*t*(9)= 1.52	*t*(9)= 0.32	-	-
	F3	-	-	***t*(9)= 4.51**	*t*(8)= -0.61	-	-
/ʉː/	**F1**	*t*(14)= -0.53	*t*(15)= 0.37	*t*(15)= 1.29	*t*(18)= -2.19	*t*(14)= 1.85	***t*(12)= -2.80**
	F2	***t*(15)= 2.95**	*t*(14)= -1.64	***t*(15)= 9.28**	*t*(18)= -1.54	***t*(14)= 6.48**	*t*(13)= -0.01
	F3	***t*(15)= 4.52**	*t*(14)= -0.44	*t*(15)= 2.42	*t*(16)= 1.76	*t*(14)= -2.10	*t*(14)= 2.27
/uː/	**F1**	*t*(14)= -2.23	***t*(11)= 4.12**	*t*(15)= -1.09	*t*(15)= 2.44	*t*(14)= 1.12	*t*(13)= -1.59
	F2	***t*(14)= 2.95**	***t*(13)= 3.83**	***t*(15)= 3.48**	*t*(15)= 0.73	*t*(14)= 0.58	***t*(13)= -3.18**
	F3	*t*(14)= 1.91	*t*(13)= 2.47	*t*(15)= -1.67	*t*(15)= 2.53	***t*(14)= -3.58**	*t*(14)= 0.14

4. Discussion

The main purpose of the current study was to compare formant characteristics of long close vowels in three varieties of Swedish. As expected, differences between the varieties were both manifested in the general location of the vowels in the vowel space and in the direction and extent of the formant movement.

The analysis revealed similar cross-dialectal patterns for /iː/ and /yː/, where the main differences between CS and FS occurred in F1 and F3. In FS, /iː/ and /yː/ are realised as more close than in CS with an accompanying decrease in F3, which is, however, not in line with Kuronen [3] who describes /iː/ and /yː/ as more open in FS than CS. Such varying results might be explained by differences in e.g. speaker age and gender (elderly women vs. young men). Furthermore, contrary to our expectations based on [9], we did not find any significant evidence of cross-dialectal variation in the F2 in

/iː/, although Figure 1 suggests that in CS /iː/ might be slightly more front than in ES. The current results also show that /iː/ and /yː/ in CS display more formant movement along the F3 and F1 dimensions respectively than their FS counterparts. A similar significant difference was not found for /iː/ in ES and FS. Only in the F3C0 measure was the ES /iː/ found to be significantly different from its CS counterpart, while no significant differences were found between /iː/ in ES and FS.

The vowel /ʉː/ displays considerable cross-dialectal variation in vowel backness, as also previously described in [9] and [3]. The current results thus corroborate the observation that the FS /ʉː/ is more back than the CS and ES /ʉː/, but with the addition that this vowel is also significantly more fronted in CS than in ES. The CS /ʉː/ is also realised with a higher mean F3 than the ES /ʉː/, which could be related to its more fronted location. The FS /ʉː/ in addition to its more back location, also features greater F1 movement than the ES /ʉː/, making it more similar to that observed in the CS /uː/. Despite these prominent dialectal differences /ʉː/ displays a similar backing formant movement in all three dialects.

The cross-dialectal variation in /uː/ is just as for /ʉː/ primarily, but not exclusively, confined to the F2 dimension. In CS, /uː/ is slightly more fronted than in ES and FS, a pattern not observed in [9], while in ES /uː/ has a higher F3 value than in FS as well as a slight movement towards higher F2 values rather than towards lower F2 values as seen in CS and FS. While we did not find any significant evidence of dialectal differences in the mean of F1 as reported by Asu et al. [9], our results suggest that the ES /uː/ features less movement along F1 than the CS /uː/, so that the difference between CS and ES increases over the course of the vowel. Moreover, the comparatively straight trajectory of the CS /uː/ contrasts with the curved trajectories of /uː/ in ES and FS in Figure 1, indicating that /uː/ is more monophthongal in ES and FS, as it mainly varies around one point in the vowel space in these dialects (similar curved patterns are observed for monophthongs in [11]). While the number of significant dialectal differences found for /uː/ is surprising, especially considering that Kuronen [3] describes a greater similarity between CS and FS in the realisation of this vowel than in the other close vowels, it is evident that most of these differences are relatively minor.

The current study confirms earlier descriptions of ES lacking the /yː/ phoneme. This is a noteworthy phonological feature of the ES vowel inventory, but the presence or absence of this particular vowel does not seem to have a large effect on the realisation of the other front close vowels in the three dialects. While a shift in the other vowels to ensure the distinction between the front close phonemes would be expected in the vowel systems where /yː/ is present, in CS /iː/ is slightly but not significantly more peripheral than in ES. In CS, /ʉː/ is more fronted than in ES, which decreases rather than increases the contrast between /ʉː/ and /yː/. However, as noted in [9], it does appear that in ES /iː/, /ʉː/, and /uː/ are more evenly spaced than in CS and FS in the F1 and F2 dimensions, as seen in Figure 1.

While the lack of /yː/ clearly sets ES apart from both CS and FS, the CS-ES pair displays the greatest number of statistically significant differences, while the smallest number of such differences was found between ES and FS, if only the results for /iː/, /ʉː/, and /uː/ are taken into consideration. Although these results suggest that ES is less similar to CS than to FS, Figure 1 shows that ES is in many cases intermediate between the other two varieties, and that similarities between ES and CS become less obvious as we move backwards along the F2 dimension, while the opposite holds true for ES-FS.

Overall, the results of the current study imply that cross-dialectal variation is only realised in the F1 and F3 dimensions in the case of /iː/ and /yː/, and primarily but not exclusively in the F2 dimension in /ʉː/ and /uː/. Dialectal differences in F3C1 only appear in the front vowels /iː/ and /yː/, either because the F3 dimension is utilised to a larger degree in high front vowels than in the non-front vowels, or due to more measurement errors in higher formants and coefficients. In general, differences between the dialects were more common in the C0 measures than in the C1 measures. This distribution is not surprising since higher numbered coefficients describe increasingly smaller patterns of the modelled signal, and as vowel steady states or means are usually described as major acoustic vowel features.

5. Conclusions and future work

In the current study, we have investigated the cross-dialectal variation of the long close vowels /iː yː ʉː uː/ in Central Swedish, Estonian Swedish, and Finland Swedish by analysing the first two DCT coefficients of the first three formant frequencies in 368 vowel tokens.

We found cross-dialectal variation in all six acoustic measures as well as in all four vowel categories; with the variation being exclusively limited to the F1 and F3 dimensions in /iː/ and /yː/, while the variation in /ʉː/ and /uː/ is to a large extent manifested in the F2 dimension. The main defining characteristics of FS close vowels were found to be a more close realisation of /iː/ and /yː/, as well as a more back realisation of /ʉː/ which was previously shown in Asu et al. [9] and Kuronen [3]. Prominent features of the long close vowels in CS were found to be a more front realization of /ʉː/ as compared to FS and ES, as well as a higher degree of relatively straight formant movement, particularly in the case of /uː/. As for ES, the most defining characteristic of its long close vowels is the lack of /yː/ and our results suggest that ES in many cases is intermediate between CS and FS.

In future studies, a comparison of the vowel systems of the three varieties of Swedish is planned using similar methods. Further research could potentially answer some questions that were left unanswered here, particularly regarding the surprisingly close realisation of FS /iː/ and /yː/, the more back realisation of FS /ʉː/, and the importance of the observed cross-dialectal differences in the light of an entire vowel system.

6. Acknowledgements

This study was conducted in the framework of the project Estlandssvenskans språkstruktur (ESST), financed by the Swedish Research Council, grant no. 2012-907. The second author was additionally supported by the Estonian Research Council grant IUT2-37. The authors gratefully acknowledge the Lund University Humanities Lab. We would also like to thank the Estonian Swedish speakers who participated in our study, and Joost van de Weijer for advice on the statistical analysis.

7. References

[1] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Oxford, United Kingdom: Blackwell, 1996.

[2] C.-C. Elert, *Allmän och svensk fonetik*. Stockholm: Norstedts, 2000.

[3] M. Kuronen, "Vokaluttalets akustik i sverigesvenska, finlandssvenska och finska," Ph.D. dissertation, Univ. Jyväskylä, Jyväskylä, Finland, 2000.

[4] T. Riad, *Svenskt fonologikompendium*, Stockholm: Univ. Stockholm, 1997.

[5] G. Fant, G. Henningson, and U. Stålhammar, "Formant Frequencies of Swedish Vowels," *STL-QPSR*, vol. 4, pp. 26–31, 1969.

[6] B. Malmberg, *Nyare fonetiska rön och andra uppsatser i allmän och svensk fonetik*. Lund: Gleerups, 1966.

[7] N. Tiberg. Estlandssvenska språkdrag. *Estlandssvenskarnas folkliga kultur*, vol. 6. Uppsala and Copenhagen: Acta Academiae Regiae Gustavi Adolphi 38, 1962.

[8] E. Lagman, *En bok om Estlands svenskar. Estlandssvenskarnas språkförhållanden. 3A.* Stockholm: Kulturföreningen Svenska Odlingens Vänner, 1979.

[9] E. L. Asu, S. Schötz, and F. Kügler, "The acoustics of Estonian Swedish long close vowels as compared to Central Swedish and Finland Swedish," in *Proc. Fonetik 2009*, Stockholm, 2009, pp. 54-59.

[10] T. M. Nearey and P. F. Assmann, "Modeling the role of inherent spectral change in vowel identification," *J. Acoust. Soc. Am.*, vol. 80, pp. 1297–1308, 1986.

[11] R. A. Fox and E. Jacewicz, "Cross-dialectal variation in formant dynamics of American English vowels," *J. Acoust. Soc. Am.*, vol 126, pp. 2603–2618, 2009.

[12] R. Mayr and H. Davies, "A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh," *J. Int. Phon. Assoc.*, vol. 41, no. 1, pp. 1-25, 2011.

[13] O. Ewald, "The Estonian Swedish diphthongs /ai au oi ui/: Acoustic characteristics and cross-dialectal variation," M.S. thesis, Centre for Languages and Literature, Lund Univ., Lund, Sweden, 2016.

[14] C. L. Watson and J. Harrington, "Acoustic evidence for dynamic formant trajectories in Australian English vowels," *J. Acoust. Soc. Am.*, vol. 106, pp. 458–468, 1999.

[15] D. Williams and P. Escudero, "A cross-dialectal acoustic comparison of Northern and Southern British English vowels," *J. Acoust. Soc. Am.*, vol. 136, pp. 2751-2761, 2014.

[16] G. Bruce, C.-C. Elert, O. Engstrand, and A. Eriksson, "Phonetics and phonology of the Swedish dialects – a project presentation and a database demonstrator," in *Proc. ICPhS 99*, San Francisco, 1999, pp. 321–324.

[17] J. M. Hillenbrand, M. J. Clark, and T. M. Nearey, "Effects of consonant environment on vowel formant patterns," *J. acoust. Soc. Am.*, vol. 109, pp. 748–763, 2001.

[18] P. Boersma and D. Weenink. (2017). *Praat: doing phonetics by computer* (Version 6.0.25) [Computer program]. Available: http://www.praat.org/.

[19] R Core Team (2016). *R: A language and environment for statistical computing* [Computer program]. R Foundation for Statistical Computing, Vienna, Austria. Available: https://www.R-project.org/.

[20] P. Escudero, P. Boersma, A. Rauber, and R. Bion, "A cross-dialect description of vowels: Brazilian versus European Portuguese," *J. Acoust. Soc. Am.*, vol. 126, pp. 1379–1393, 2009.

[21] K. Dhana. (2016, April 30). *Identify, describe, plot, and remove outliers from the dataset* [Blog post]. Available: https://www.r-bloggers.com/identify-describe-plot-and-remove-the-outliers-from-the-dataset/.

[22] J. W. Tukey, *Exploratory data analysis*. Reading, PA: Addison-Wesley, 1977.

[23] D. Bates, M. Maechler , B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1-48, 2015.

[24] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-6," 2014, Available: http://CRAN.R-project.org/package=lmerTest.

[25] R. V. Lenth, "Least-Squares Means: The R Package lsmeans," *J. Stat. Softw.*, vol. 69, no. 1, pp. 1-33, 2016.