Hät, här, höt and hör

An Articulatory and Acoustic Study of the Swedish Vowels /ɛː/ and /øː/

Anne Schumacher
anne.schumacher.up@gmail.com

Master Thesis, 15 hp
Language and Linguistics
Specialisation: Phonetics
March 2014

Supervisors:
Susanne Schötz & Johan Frid

Lund University
Centre for Language and Literature

Abstract

This thesis is concerned with the examination of the Swedish vowels /ɛː/ and /øː/ for qualitative differences between their pre-r allophones and the allophones occurring in other contexts. For this purpose, electromagnetic articulography and speech recordings were carried out on Swedish speakers from the two dialectal areas Malmö and Stockholm.

Articulatory and acoustic analyses of the collected data show that the pre-r allophone of /ɛː/ has a significantly more open vowel quality compared to its pre-t allophone in the dialects spoken around Malmö and Stockholm. Tongue position data suggest the same for the pre-r allophone of /øː/, although this could not be supported by the results of the formant analysis. The data further suggest that there is a significantly larger difference in openness in the East Swedish dialects than in the Southern ones.

Contents

List of Figures

1 Introduction

From a phonetician's point of view, the Swedish vowel system is very interesting because it is fairly rich in comparison to many other languages and it comprises some exotic features as for example the existence of three close front vowels. Until recently, phonetic studies of Swedish vowels have been mainly acoustic. Electromagnetic articulography offers the possibility to examine articulatory properties and processes during speech production in much greater detail.

As with any basic research, this project aims at deepening the understanding of fundamental principles, in this case the articulatory principles of the production of Swedish vowels. Having detailed information about the configurations of the tongue, lips, mandible and the other articulators during the production of certain sounds or sound combinations can be very helpful in speech therapy and language teaching. Extending the empirical knowledge of sound production is also indispensable when working with articulatory synthesis, a method of speech synthesis which is based on models of the articulatory processes that occur in the human vocal tract.

The phenomenon that this project focuses on is that of the Swedish vowels /ɛː/ and /øː/ which appear to exhibit differences in quality depending on their phonetic context. Several authors have pointed out that those vowels are produced more openly when they occur before rhotic consonants but so far this observation has not been supported by articulatory data. In this thesis the pronunciation of these vowels is investigated on by means of electromagnetic articulography. Acoustic analysis is carried out to validate and support the findings of the articulatory examination of the vowels.

Section 2 provides a general overview of the phonetic features of the vowels /ɛ/ and /ø/ in Swedish, some background information on electromagnetic articulography as well as a summary on previous studies that are related to this thesis. The hypotheses that are being posed for this study are presented in section 3 and the methods that were used to test them are described in section 4. Section 5 summarises the results of the analysis, which are discussed in section 6. Finally, section 7 provides a brief summary of the thesis.

Three different vowel notations will be used throughout this thesis: the phonetic square bracket notation [ɛ] which attempts to capture the actual accurate pronunciation of a segment and the phonological slash notation /ɛ/ which rather denotes phonological contrasts. In those cases where a vowel is viewed in a very general context or when it occurs within an example word it will be displayed in its orthographical form.

2 Background

2.1 The Vowels /ɛ/ and /ø/ in Swedish

Swedish has nine long vowels, each of which has a short cognate. The phonemes /ɛ/ and /ø/ that correspond to the orthographic forms *ä* and *ö* (resp.) belong to the group of the front mid vowels of the Swedish vowel system (see table 1). The short ö has two different transcriptions in the relevant literature, it can either be represented as [ø̞] or [œ] (Leinonen, 2010, p. 17). Authors that use the latter assume thus that the pronunciation of the short /ø/ is more open. The degree of opening of the long /ɛ/ is assumed to vary largely among the different dialects. In Norrland and southern Skåne the vowel is pronounced very openly while it is rather closed in Stockholm (Malmberg, 1971, p. 61). As indicated in table 1, the short counterparts of /ɛː/ and /eː/ are very similar in quality, and in some Swedish dialects they are even homophonous, like in the varieties spoken in Södermanland and Gotland (Bruce and Engstrand, 2006; Elert, 2000).

	FRONT				CENTRAL		BACK	
	Unrounded		Rounded		Rounded		Rounded	
	Long	Short	Long	Short	Long	Short	Long	Short
HIGH	iː	ɪ	yː	ʏ	ʉː		uː	ʊ
HIGH MID	eː		øː				oː	
		e – ɛ				ɵ		
LOW MID	ɛː			œ				ɔ
LOW		a					ɒː	

Table 1 – The Swedish long and short vowels. Straight lines connect long-short cognates that are marked differently in the height or front-back dimension. (source: Bruce and Engstrand, 2006, p. 20)

2.1.1 Vowel Lowering

Many phoneticians have asserted that the vowels /ɛ/ and /ø/—in their short and long variants—are generally lowered before /r/. One of them is Claes-Christian Elert who pointed out that this lowering occurs in all Swedish varieties (Elert, 2000, p. 32 and 48). The same appears to be true when the vowels occur before the so-called supradental consonants that correspond to the written forms *rt*, *rd*, *rn*, *rl* and *rs*. According to Elert, the degree of opening or vowel height varies relatively much among the Swedish dialects.

The vowels have the highest degree of opening in this context in the dialects spoken at the east coast of Sweden (mostly in Stockholm but also in Finland) and the differences are less apparent towards the west of the country (starting in the western parts of Södermanland and Västmanland). In some varieties of Swedish no allophones of /ɛ/ and /ø/ exist, so that the vowels have the same pronunciation in all positions (Leinonen, 2010, p. 22). Elert (2000, p. 47) also noted that the long ö vowel exhibits a changing degree of openness in its pronunciation across the generations. Older generations in Central Sweden would mostly produce the vowel as [øː] while younger speakers tended to use the more open pronunciations [ø̞ː] or [œː] instead—regardless of the context.

Engstrand (2004, p. 115) confirmed Elert's view and pointed out that the allophonic shift from [ɛ] to [æ] and from [ø] to [œ̞] before /r/ and the above mentioned supradentals can often be observed in Swedish. According to Engstrand, this affects not only the long /ɛː/ and /øː/ variants but also their short equivalents as in the words *Bert* [bæʈː], *bärs* [bæʂː], *mört* [mœ̞ʈː] and *börs* [bœ̞ʂː].

Already in the seventies, the more open pronunciation of /ø/ and /ɛ/ before /r/ was discussed by Bertil Malmberg (1970, p. 83; 1971, p. 20). He pointed out that the tongue position for the more open allophone of /øː/ was often also further back compared to the standard allophone and assumed that the function of this contrasting property might be to indicate that the following phoneme is an /r/. The same author pointed out that in the South Swedish dialects the more open allophone of ä before /r/ only occurred with the short vowel. The difference is apparent in the contrast of *färg* vs. *häst* but not in *bära* vs. *äta* (Malmberg, 1971).

Bolfek-Radovani (2000) noted that the pronunciation of the long ä vowel as an extremely open [æː] before /r/ is seen as a stigmatized variant in Uppland of which the speakers themselves are highly aware.

Riad (2013) has even noticed a tendency of younger people pronouncing the long ä and ö more openly in all positions. Where older speakers of most Swedish dialects usually have a clear distinction between the ö and ä vowels occurring before /r/ and in any other position (e.g. *lära* vs. *läsa* and *höra* vs. *fösa*), younger speakers tended to pronounce the vowels in a lowered manner in all contexts and would often even fail to hear the difference. This tendency of using the more open allophone in all contexts, however, does not seem to be equally common in the entire country. Some speakers only produce one of the vowels more openly, some apply this feature to both the long and short vowels and some only to the long ones. In dialects spoken in Västergötland, the shift in pronunciation even seems to happen in the opposite direction since /ø/ is articulated with a higher tongue position before /r/ which results in a quality comparable to the "original" /ø/ in other contexts pronounced by older speakers.

Riad (2013) regards this shift in vowel pronunciation as a normal process which is typical for Germanic languages. A possible explanation for this change could be the asymmetry in the Swedish vowel system which leaves quite some room for low front vowels. Moreover, a typical generalisation process seems to take place in the case of lowering ä and ö in all positions: the rule that /ɛ/ and /ø/ should be lowered before /r/ is extended to all contexts by younger speaker generations.

2.1.2 Diphthongisation

In many dialects long vowels are often diphthongised. According to Bruce (2010, p. 121), diphthongisation is generally underestimated and is not considered enough in most Swedish vowel studies and in vowel studies of other languages.

In the South Swedish dialects the vowels are generally characterised by an on-gliding or rising diphthongisation in which the tongue moves from a less peripheral articulation to a more peripheral one (Engstrand, 2004; Elert, 2000; Bruce, 2010). The long front vowels like /ɛː/ and /øː/ however do not follow this pattern. Their articulation begins in fact more openly and moves then towards the centre to end with a vowel quality that can be described as the Standard Swedish one. In the southern dialects /ɛː/ and /øː/ therefore become [^{æ}ɛː] and [^{oe}øː] (Elert, 2000, p. 38) or [æ͡ɛ] and [ɶ͡œ] (Bruce, 2010, p. 122).

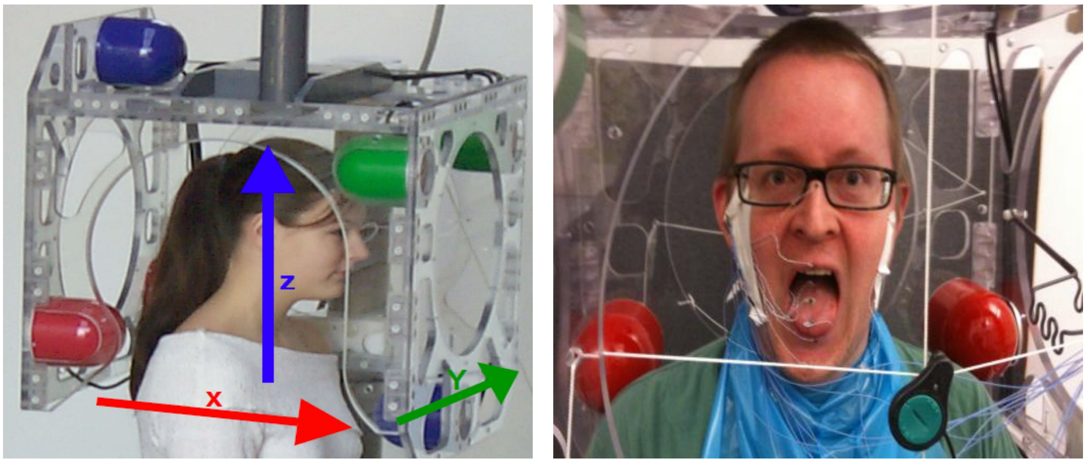
In the East Central Swedish dialects the long vowels rather undergo an off-gliding or falling diphthongisation. The first segment is pronounced as the Standard Swedish one and is followed by a weaker [e]- or [ə]-like vowel element (Leinonen, 2010, p. 19-20). In these dialects /ɛː/ and /øː/ are mostly pronounced as [ɛː^{ə}] and [øː^{ə}] (Elert, 2000, p. 40) or [ɛ͡ə] and [œ͡ə] (Bruce, 2010, p. 127) and the tongue is thus moving from a more peripheral articulation to a less peripheral one.

Diphthongisation also occurs in the dialect spoken on Gotland, although the pattern is not as clear as for the above mentioned varieties. Some diphthongs are on-gliding while others are off-gliding and some vowels like /iː/, /yː/ and /aː/ are usually not diphthongised at all. Because of this irregular pattern, Gotland vowels are sometimes described as vague diphthongs (Elert, 2000, p. 43). According to Elert (2000), /ɛː/ is pronounced as [^{e}ɛː] and /øː/ as [øː^{ʏ}] in the Gotland dialect.

2.2 Electromagnetic Articulography

Electromagnetic articulography (EMA) is a method of measuring speech movements. It is a point-tracking method which means that individual flesh-points are being tracked rather than considering the behaviour of an entire articulator.

For this procedure, small sensors which act as receiver coils are attached to the participant's articulators, and the subject's head is placed inside a magnetic field. When a receiver coil moves within the magnetic field, a voltage is induced in the sensor which is inversely proportional to the distance to the coil transmitting the magnetic field. Because of this relation, the exact position of the sensor within the field can be computed from the signal (Stone, 1969; Hoole and Nguyen, 1997; Schötz et al., 2011). An example of the experimental setup is shown in figure 1.



(a) side view **(b)** front view

Figure 1 – Subjects inside the Carstens Articulograph AG500

The advantage of point-tracking systems like electromagnetic articulographs is that the movements and interactions of several articulators inside and outside the vocal tract can be studied simultaneously at a very high speed and with a high spatial accuracy. The spatial resolution has been calculated to be approximately 0.5 mm for older EMA devices (Perkell et al., 1992) and it is even higher nowadays. This makes the method very applicable and useful for detailed studies of displacements, timing and coordination of articulatory components in speech production. This technique has been used for examining vowel qualities (Schötz et al., 2011, 2012, 2013) and for coarticulation research (Jackson and Singampalli, 2008) but it has also been applied in the context of medical treatment of patients suffering from Broca's aphasia and apraxia of speech. Katz et al. (1999) reported the use of EMA for a regular assessment of a patient's speech motorics in order to improve and monitor the patient's therapy.

In contrast to the older x-ray method, EMA bears no known biological hazards because of the short exposure times and low field strengths (Stone, 1969). Hoole and Nguyen (1997) state that even though there are currently no grounds for disquiet, it is advisable to avoid pregnant subjects and wearers of pace-makers in studies that involve EMA as a simple precaution. Another advantage with this method is that it allows for a high number of utterances to be recorded in a single session. According to Hoole and Nguyen (1997), a session duration of 30 minutes or even more is feasible. Moreover, EMA may be combined with other methods such as electromyography, ultrasound and electropalatography without resulting in unacceptable levels of mutual electromagnetic interference (Hoole and Nguyen, 1997). This allows for the possibility of studying articulation from more perspectives in even more detail.

Unfortunately, EMA devices are quite expensive and their handling requires a lot of training. A further limitation of the method is that it is invasive which means that the sensors are physically touching the articulators. Therefore, inner structures like the pharynx, the velum and the posterior part of the tongue can hardly be measured due to the natural gag reflex. On the other hand, the sensors are very small (about 3 to 4 mm in diameter and 2 to 3 mm high) and they are usually not reported by the subjects to be a source of disturbance when attached to the tongue. Subjects might encounter slight disturbances though when a sensor is placed closer than 1 cm to the tongue tip (Hoole and Nguyen, 1997). No disturbances are generally reported for the thin wires attached to the sensors, as long as they run out of the side of the mouth rather than over the tongue tip.

Using EMA to examine phonetic properties has numerous advantages over using acoustic methods. One of the more traditional ways to study vowel quality is to measure formants. While the first formant (F_1) gives information about the openness of the vowel, the vowel frontness can generally be derived from the second formant (F_2). Measuring formants gives us a pretty good idea about the vowel quality, but it is never precise enough to reveal the exact vowel parameters like the tongue position, the constriction size and the degree of opening. It is also known that there is no perfect correspondence between the formants and the vowels height resp. the vowel advancement, but rather are both F_1 and F_2 influenced by the change in any vowel parameter. Moreover, measurement errors occur quite frequently when formants are calculated automatically from the speech signal, especially when two formants are placed quite close to each other (Leinonen, 2010, p.

23-24).

In this project the focus lies on the analysis of the articulatory data that have been collected. Acoustic data will also be considered and compared to the articulatory analysis. The articulatory data were collected using the electromagnetic articulograph AG500 by Carstens Medizinelektronik GmbH which—unlike many other devices—tracks three-dimensional articulatory movements and has a fairly high sampling frequency of 200 Hz.

2.3 Previous Studies

Swedish vowels have of course been described and analysed by various phoneticians, but so far most analyses that have been carried out are of acoustic nature. Gunnar Fant was one of the first scientists to collect Swedish speech data in order to give an acoustic description and classification of all phonetic units of the language. He has conducted several studies on this matter, some of which can be found in Fant (1973). In 1948, he recorded all 9 Swedish vowels in their long and short versions plus the pre-r allophones of ä and ö, produced by 14 different speakers. He then extracted the formant frequencies and calculated mean values for male and female speakers (Fant, 1973, p. 35ff.). Although he does not specifically mention so, Fant's formant tables indicate higher F_1 values for the pre-r allophones of ä and ö than for the remaining ö and ä allophones that were recorded. Later, Fant carried out articulatory studies of Swedish vowels by means of x-ray tracings (Fant, 1959, 1983). However, these studies only comprise rather general phoneme descriptions and do not examine the vowel lowering before /r/.

More recent articulographic studies on Swedish vowels have been carried out within the VOKART project (Exotic vowels in Swedish – an articulographic study of palatal vowels) at Lund University. Schötz et al. (2011) have conducted a pilot study on the Swedish vowel /iː/. Two speakers were analysed, one South Swedish speaker who pronounced the vowel as a regular [iː] and an East Central Swedish one who produced the vowel as [ɨː], with the typical "damped" or "Viby-coloured" quality. The articulatory data showed that [ɨː] was pronounced further back with an overall lower tongue position but with a higher position of the tongue tip compared to [iː]. This was supported by the acoustic data which revealed a lower F_2 for [ɨː], indicating a more centralised vowel quality compared to [iː].

Schötz et al. (2012) analysed the diphthongisation of the vowel /uː/ in South Swedish dialects. Acoustic data revealed that the vowel is realised dynamically, with a considerable F_2 movement. The articulatory data confirmed the dynamical realisation of that vowel by showing significant differences in tongue displacement throughout the diphthong for different speakers of the Malmöhus dialect.

Schötz et al. (2013) examined tongue articulation dynamics in the vowels /iː, yː, ʉ̟ː/ of Gothenburg and Malmöhus Swedish. Functional Data Analysis of the articulatory data showed that the tongue positioning for /iː/ and /yː/ was similar and significantly different from /ʉ̟ː/ in both dialects. The tongue body height was significantly different for all three vowels in both Swedish varieties.

To the best of the author's knowledge, the Swedish vowels /ɛː/ and /øː/ have not been studied by means of EMA before.

3 Hypotheses

As was shown in chapter 2.1, it is assumed that there are differences in the production of the long Swedish vowels /ɛː/ and /øː/ when they occur before the rhotic /r/ and when they precede other phonemes. Furthermore, there are indications for differences in the vowel height among different dialects. The hypotheses that arise from these assumptions are as follows:

1. It is expected that the vowels /ɛː/ and /øː/ have significantly more open qualities (articulatoryly and acoustically) when they precede /r/ than when they precede /t/.
2. This difference is expected to be larger in the Eastern Swedish dialects than in the Southern dialects. Speakers from in and around Stockholm should thus show a larger contextual difference in vowel height than speakers from in and around Malmö.

These hypotheses will be tested by means of EMA and acoustic analysis. The openness of the vowels will be determined by measuring the tongue height and the frequency of the first vowel formant.

4 Method

The material for this study was collected within the VOKART project by Susanne Schötz and Johan Frid (Schötz et al., 2011, 2012, 2013). So far, 29 native Swedish speakers from the three dialectal areas Stockholm, Gothenburg and Malmö have been recruited and recorded in the Humanities Lab at Lund University.

4.1 Participants

Out of the 29 speakers that were recruited for the VOKART project, 19 were considered in this study. One of these subjects was excluded from the analysis due to technical failure. The remaining participants (8 women and 10 men) grew up in the southern part of Sweden (9 subjects with birthplaces in Malmö, Landskrona, Dalby and Skurup) and in East Sweden (9 subjects from Stockholm, Uppsala, Linköping, Norrköping, Norrtälje and Lidingö). The participants were between 21 and 61 years of age (mean=40.1 , sd=12.9, see also table A in the appendix) and each subject received an ice-cream voucher as compensation.

4.2 Recording Procedure

The articulographic data were collected using the Carstens Articulograph AG500 with a sampling rate of 200 Hz. The voice recordings were made using the built-in microphone of the articulograph with a sampling frequency of 16 kHz.

After a short introduction to the experimental procedure, 10 sensors were placed on the participant's tongue, jaw and lips as shown in figure 2 and two additional sensors were placed on the nose ridge and behind the ear to correct for head movements. The front most sensor on the tongue was placed about 1 cm inwards from the tongue tip (sensor number 1 in figure 2) to ensure that the subject was able to articulate without feeling any major disturbances.

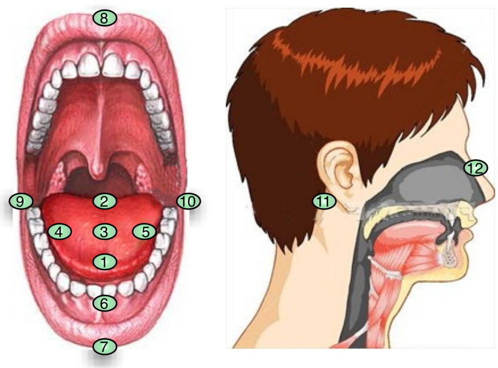


Figure 2 – The sensor positions used in the experiment

The participant was then asked to read out a few sentences in order to habituate the articulators to speaking with the equipment in the vocal tract. Afterwards the subject received instructions to move the tongue tip forth and back on the palate in order to capture the palate shape. For further habituation, the participant was then asked to orally describe a picture before the test sentences were recorded.

The test material consisted of short sentences which were presented one at a time on a computer screen and which were read aloud by the participant. One recording session lasted about one hour including the fixation of the sensors.

4.3 Material

The speech material consisted of the carrier phrases "De va inte hVt utan hVt ja sa." and "De va inte hVr utan hVr ja sa." (*It was not hVt/hVr but hVt/hVr I said*) where V was always replaced by one of the nine long Swedish vowels. The material was presented in 11 different sentences where each vowel occurred twice in the hVt context, once in first and once in second sentence position. Only the vowels ö and ä occurred also in the hVr context. The sentences were presented in random order and each sentence occurred 10 times in the experiment to make sure to have enough back-up data in case of technical failure.

Putting the vowels of interest into the contexts *hVt* and *hVr* had been a conscious choice. Using an alveolar sound has the advantage that the tongue does not need to move very far after producing the vowel as it would have to for an upcoming velar sound. Thus, between /h/ and /t/ the vowels are expected to be least biased since they should not induce any major coarticulatory effects. The very similar *hVd* context has been used in many different studies involving vowel recordings (e.g. Peterson and Barney, 1952; Chia and Sondereggerb, 2007). Using the voiceless equivalent offers an advantage for the extraction of the vowel from the speech signal. A vowel is easily demarcated from a succeeding voiceless stop since the closing phase produces a clear boundary in the waveform and spectrogram. Another reason for choosing this context was the fact that most hVt sound combinations result in actual Swedish words, with the four exceptions of /hoːt/, /hʉːt/, /hɛːt/ and /høːt/.

4.4 Data Preparation

The sensor positions were automatically calculated from the signal that was recorded by the articulograph using the CalPos software which was provided by the manufacturer. Head movement correction was carried out by Carstens' NormPos software which produces a separate position file for every recording. Every position file contained a table of all the sensors and their positions in all three dimensions in 5 ms time intervals. The vowels were manually segmented in Praat (Boersma and Weenink, 2013) by Susanne Schötz through identifying the vowel boundaries in the sound files with the help of the spectogram and waveform. The sound and position files were then cut according to these boundaries, resulting in one position file and one sound file per vowel recording.

The statistical analysis software R (R Development Core Team, 2008) was used to extract the relevant articulatory data and collect them into four tables per speaker. Each of these tables contained the height dimension of sensors 1-3 from all recordings of one of the vowels. The extraction of the formants from the acoustic data was done in Praat using a script that measured the frequencies of F_1, F_2 and F_3 every 5 ms and saved them to a separate table per recording. Although the data of the three first vowel formants were collected, only the F_1 frequencies were considered for the analysis because this formant is known to have the strongest association with the tongue height (Ladefoged, 2000, p. 176).

4.5 Data Analysis

The statistical analysis of the articulatory and acoustic data was carried out in R. Outliers were removed from both data sets using the robust τ-method as proposed by Maronna and Zamar (2002). To compensate for anatomical differences between speakers, the data were normalised using z-score transformation.

Functional Data Analysis (FDA) was then used to compare and analyse the z-scores of the position data and the formant data. This method was first introduced by Ramsay et al. (1996) and is very useful for the alignment of a set of signals with different amounts of data points. In FDA, smoothed functions are fitted to the data which makes it easier to compare different observations with each other because they can be plotted on the same time scale. This method has been used in similar studies by Schötz et al. (2013) and Gubian et al. (2011).

Four different measures were used for the comparison between different data sets: means, standard deviations, pairwise functional t-tests and absolute difference values. Mean values and standard deviations were plotted in order to get a visual impression on the data. Functional t-tests were applied to check for significant differences between the mean values of diverse data sets. The absolute differences were only calculated for the two vowel contexts (hVt vs. hVr) in order to find out which dialect group exhibited a larger difference in average tongue height or F_1 frequency.

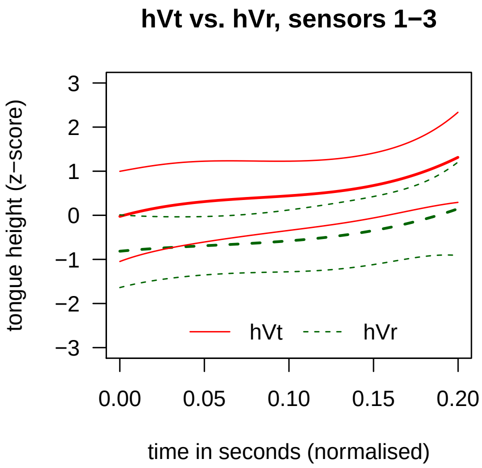
Using FDA, the average tongue heights (mean over all speakers and all instances of /ɛː/ and /øː/) for the hVt context and the hVr context were plotted as time functions, first for each sensor separately. Functional t-tests were applied to each of the graph pairs (see figure A in the appendix). In order to obtain a more general impression on the tongue height, the values for the three sensors were then combined (i.e. the average value across sensors 1-3 was calculated) and the tongue heights were compared between the two contexts by plotting the means and standard deviations, and by applying functional t-tests (see figure 3). For the comparison of the two dialectal groups, the data were then divided into South and East speakers and for a more detailed analysis they were further separated into the two different vowels /ɛː/ and /øː/. The average tongue heights for hVt and hVr were plotted together with the standard deviations and functional t-tests were applied (see figure 4 and figure B in the appendix). The absolute differences between the two context functions (hVt vs. hVr) was calculated for the South and East group and plotted as time functions (see figure 5).

For the analysis of the acoustic data, the average F_1 frequencies (mean over all speakers and all instances of ä and ö) were calculated for the hVt and hVr context and plotted using FDA. Functional t-tests were applied in order to examine the difference between the F1 values in the two contexts (see figure 7). The data were then divided into the two dialect groups and the average F_1 frequencies were plotted (see figure 8). T-tests were applied to the F_1 functions (see figure 9). The same analyses were also carried out for the two vowels separately (see figures C and D in the appendix). Finally, the absolute difference between average F_1 frequencies in hVt and hVr were plotted for the South and East group (see figure 10).

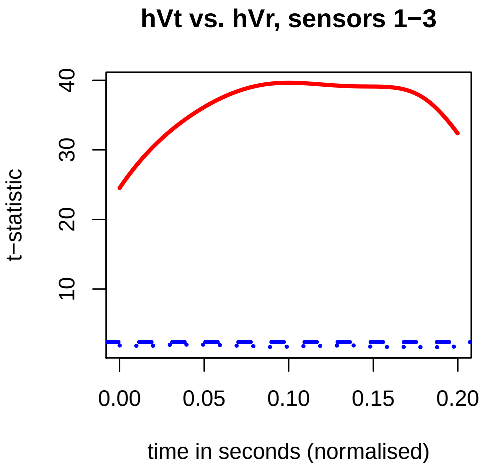
5 Results

5.1 Articulatory Data

Comparing the z-scores of the tongue positions during the articulation of /ɛː/ and /øː/ in the hVt context with the values of the hVr context revealed a significantly lower tongue height for the pre-r allophones. The sensors with the numbers 1-3 that correspond to the positions of the tongue tip, tongue dorsum and tongue blade (see figure 2) were considered for this analysis. The significant difference in the height dimension existed for all three sensors combined as well as for every sensor separately (see figure A in the appendix for separate t-tests for each sensor). Figure 3 shows the mean values and standard deviations for the average tongue height (sensors 1-3 combined) for all 18 speakers and the result of the pairwise functional t-test. The time function plot in figure 3a shows the vertical movement of the tongue during the articulation of the recorded vowels. For both vowel contexts it can be observed that the tongue starts at a lower position within the oral cavity and rises for the entire length of the vowel towards the place of the articulation of the succeeding consonant.



(a) mean values



(b) t-test

Figure 3 – Comparison of average tongue heights between the pre-r and pre-t allophones of /ɛː/ and /øː/ (averaged across all 3 sensors and both vowels); **(a)**: mean values and standard deviations; **(b)**: observed statistic (upper solid line), critical value (0.05) for the pointwise comparison (lower dotted line) and maximum of the t-statistics (middle dashed line)

Separating the speakers into a South and East Swedish dialectal group (see table A in the appendix for details) showed that there was a tendency for a more similar articulation of hVt and hVr in the South group. This was valid for both /ɛː/ and /øː/, as can be seen in figure 4.

Figure 5 shows the absolute difference between the time functions of the tongue articulation of hVt and hVr for the East and South group. For approximately the first 80% of the vowel articulation the East group exhibited a larger absolute difference. Only for the last portion of the vowel the tongue height differed more for the South group. A two-sample t-test revealed that the functions in figure 5 (y-values of the South and East functions) are significantly different from each other ($p<0.003$). This shows that there was a significantly larger difference in average tongue height between the pre-r and pre-t allophones of /ɛː/ and /øː/ for the East group.

Functional t-tests revealed significant differences between average tongue

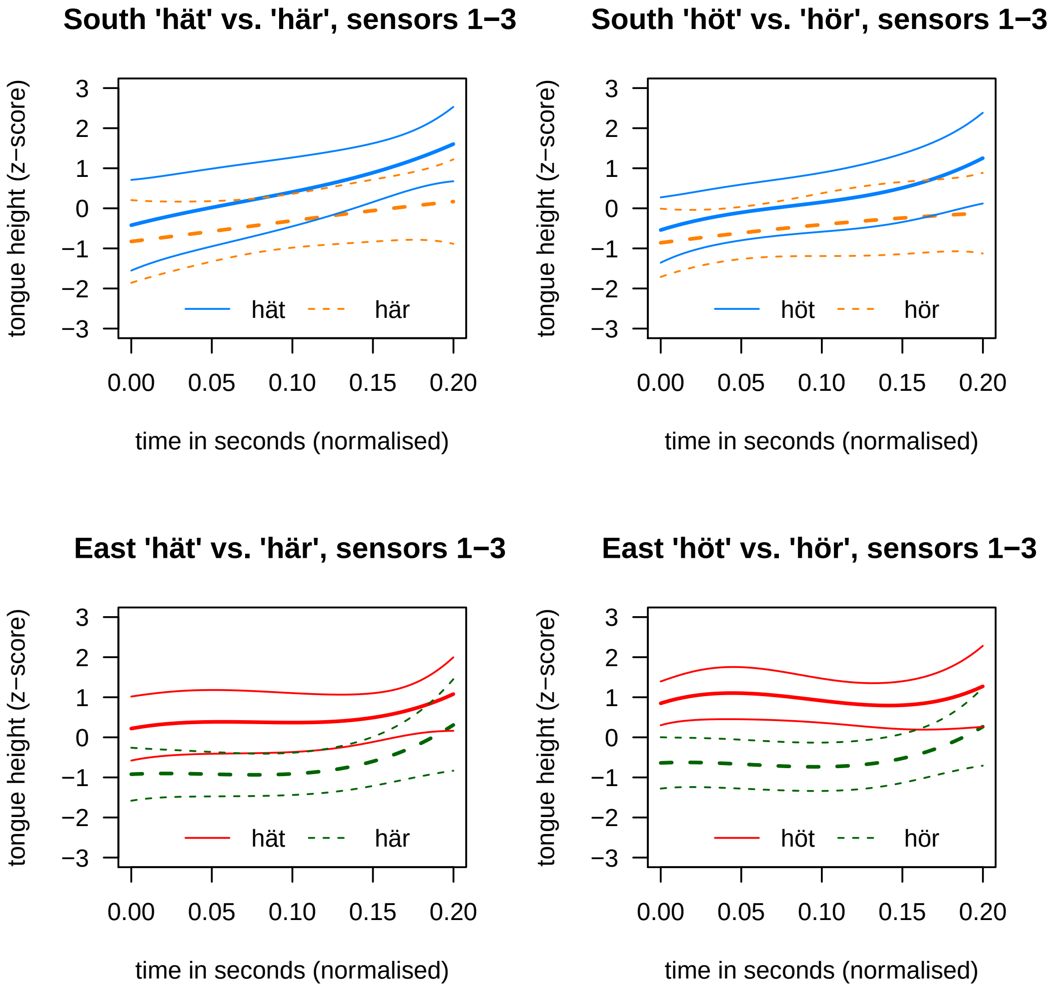


Figure 4 – Average tongue heights (means over all three sensors) measured in the words *hät, här, höt* and *hör* for the South and East group (mean values and standard deviations)

heights for both /ɛː/ and /øː/ and for both dialect groups when sensors 1-3 were combined (see figure B in the appendix). When considered separately, sensors 2 and 3 yielded significantly higher tongue positions for the two pre-r allophones in both groups. Sensor 1, however, exhibited no difference—neither between *här* and *hät*, nor between *hör* and *höt*—for the South group for the predominant portion of the vowel (see figure 6). Only the last vowel portion showed significant differences in tongue height.

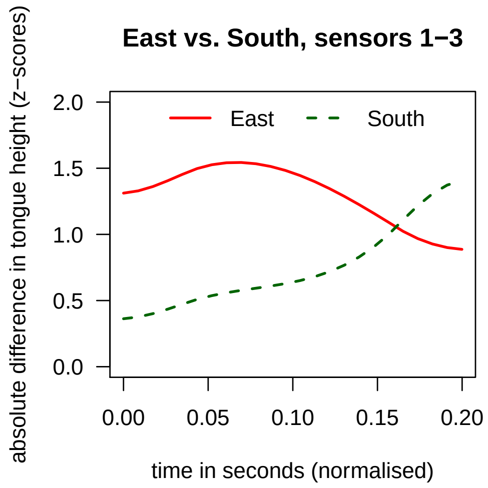


Figure 5 – Absolute difference between the average tongue heights (means over all sensors) in hVt and hVr for the South and East group

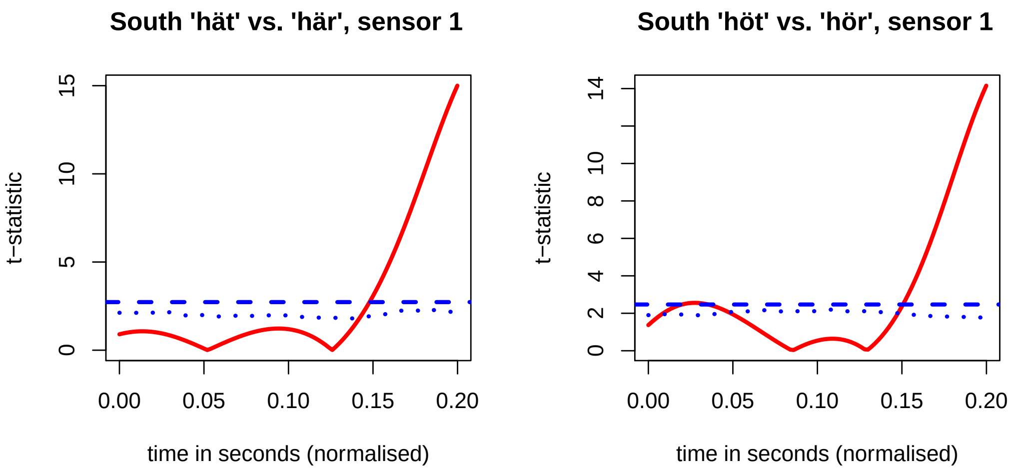
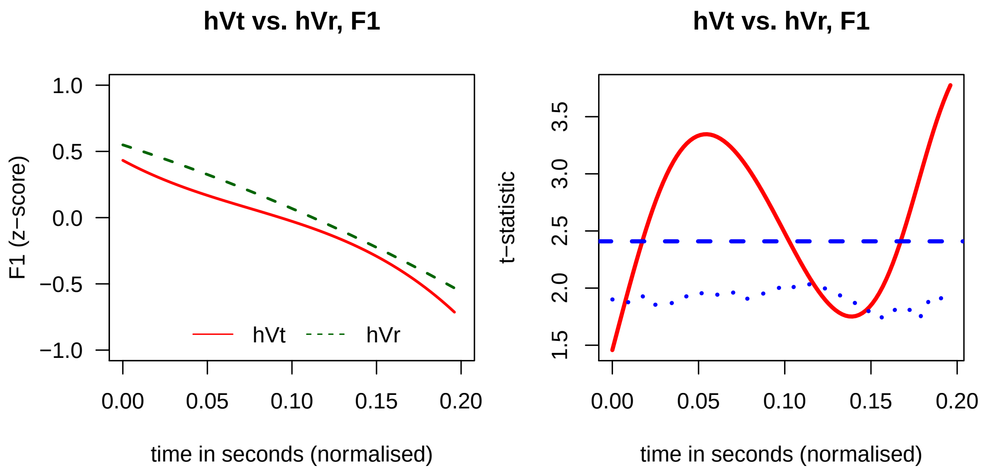


Figure 6 – Functional t-tests for tongue tip height, South group; observed statistic (solid red lines), critical value (0.05) for the pointwise comparison (dotted lines) and maximum of the t-statistics (dashed linse)

5.2 Acoustic Analysis

The F_1 comparison between the hVt and hVr contexts for all 18 speakers yielded only partly significant results. From the mean frequency values in figure 7a it can be seen that the pre-r allophones exhibited somewhat higher F_1 values but functional t-tests showed that this result was only significant for the first and last portion of the vowels (figure 7b).



(a) mean values **(b)** t-test

Figure 7 – Average F_1 frequencies in hVr and hVt; **(a)**: means over all speakers; **(b)**: observed statistic (solid red line), critical value (0.05) for the pointwise comparison (dotted line) and maximum of the t-statistics (dashed line)

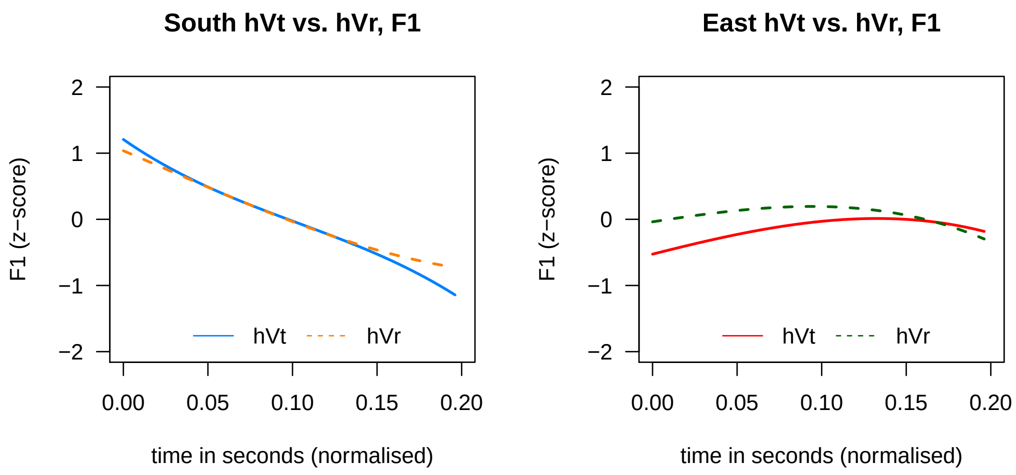


Figure 8 – Average F_1 frequencies in hVr and hVt, South and East group

When the East and South speaker groups were regarded separately, it became visible that the F_1 yielded higher values in the first vowel portion for the pre-r allophones in the East group (figure 8) which was confirmed by t-tests (figure 9). T-tests further revealed a significant difference between hVt and hVr in the very last vowel portion for the South group (see figure 9). More detailed analyses of the four different allophones produced by the two dialectal groups revealed that the F_1 frequency values for *hät* were significantly lower than the values for *här* in both groups (see appendix figure C for mean values and figure D for t-tests). However, for *höt* and *hör* the F_1 values did not differ in the first half of the vowel for the South group and in the final portion for the East group.

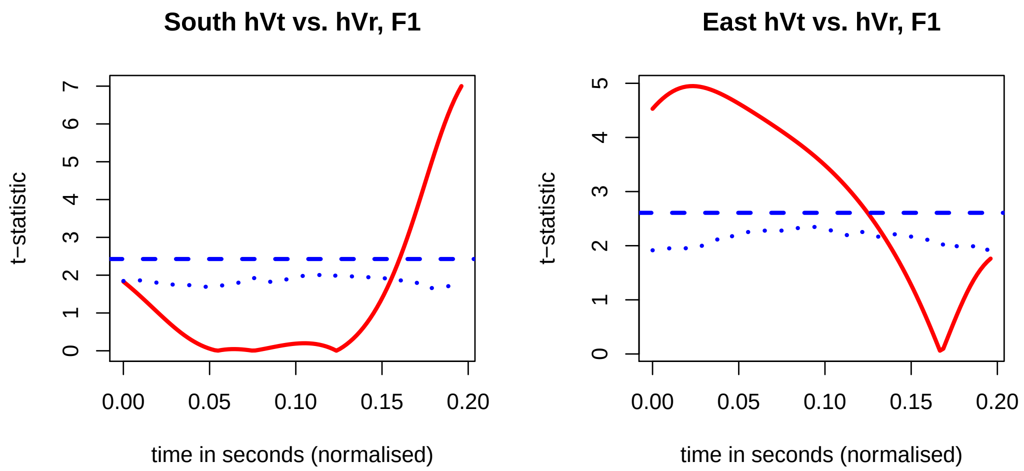


Figure 9 – Pairwise functional t-tests for F_1 frequencies in hVr and hVt for South and East group; observed statistic (solid red lines), critical value (0.05) for the pointwise comparison (dotted lines) and maximum of the t-statistics (dashed lines)

The time functions for the absolute difference between the F_1 mean z-scores of hVt and hVr are plotted for the East and South group in figure 10. For approximately the first 70% of the vowel the East group exhibited a larger difference in the formant values than the South group. A two-sample t-test revealed that the functions in figure 10 (y-values of the South and East time functions) differed significantly ($p<0.0001$) which shows that there was a significantly smaller difference in F_1 values for the South group.

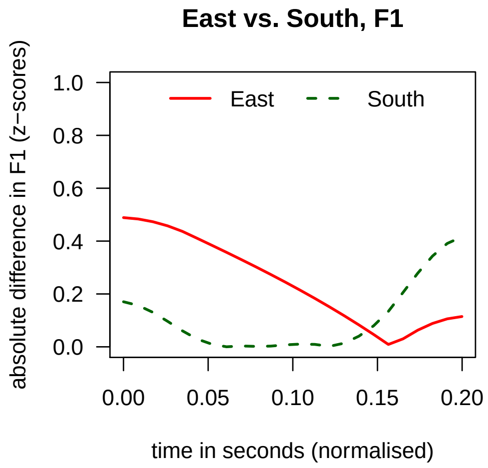


Figure 10 – Absolute difference between average F_1 frequencies in hVt and hVr for the South and East group

6 Discussion

6.1 Hypothesis 1

The first hypothesis stated that the vowels /ɛː/ and /øː/ have significantly more open qualities for the pre-r allophones than when they precede /t/. The results of the articulatory analysis have shown that the average tongue position for the articulation of the pre-r allophones of /ɛː/ and /øː/ was lower than for the pre-t allophones of the same vowels. This difference was significant when the positions of the tongue tip, tongue blade and tongue dorsum were considered separately and combined. The average F_1 values (means over all speakers) were generally higher for the pre-r allophones of /ɛː/ and /øː/, although this result was only significant for the middle vowel portion (from approximately 10 ms to 100 ms on a normalised time scale where the vowel duration is 200 ms) and the vowel offset.

The reason for the similar F_1 frequency values between 100-150 ms of the vowel most likely lies in the fact that the phonemes /ɛː/ and /øː/ were combined for this analysis. /øː/ is considered a close mid vowel while /ɛː/ is an open mid vowel (see table 1), which means that the latter generally has a higher F_1 frequency than the former. Combining the F_1 values of both sounds into one graph makes the contextual difference weaker.

When the acoustic data for the two different vowels were analysed separately, however, it emerged that the recordings of the pre-r allophone of /ɛː/ produced significantly higher F_1 frequencies than its pre-t variant. The first hypothesis which stated that the vowels /ɛː/ and /øː/ have significantly more open qualities when they precede /r/ as when they precede /t/ is thus only partially corroborated by the data that were collected in this study. It can be concluded that in the East and South Swedish dialects the vowel /ɛː/ is lowered when it precedes /r/. Articulatory data further suggests that the same is true for the vowel /øː/, but this was not supported by the acoustic data in this particular study.

6.2 Hypothesis 2

The second hypothesis stated that speakers from in and around Stockholm show a larger difference in vowel height than speakers from in and around Malmö. The results of the separate articulatory analyses of the two dialectal groups have clearly shown that the tongue takes a lower position for the pre-r allophones of both vowels in both regions. The difference in tongue height between hVt and hVr was more apparent in the East group for the first 150 ms of the vowels (on a normalised time scale). During the remaining 50 ms—which represent the vowel offset—the contextual difference was larger in the South group. This result can be explained by slight coarticulatory effects between the vowel and the succeeding consonant. In the South Swedish dialects /r/ is usually pronounced as the uvular fricative [ʁ] or the uvular trill [ʀ] while most other Swedish dialects (including the eastern ones) use the alveolar fricative or approximant [ɹ] or the alveolar trill [r] instead (Elert 2000, p. 65-67, Bruce 2010, p. 150-151). The tongue needs to raise from its vowel position when the following phone is alveolar like [t], but not when preceding a uvular /r/ allophone. Thus, the difference between the articulation of hVt and hVr must be smaller in the vowel offset of the East group, since the tongue undergoes similar movements for both vowel allophones, which is not the case for the South group.

When the position sensors were regarded separately it could be seen that the difference between hVt and hVr was not significant for the height of the tongue tip in the South group. This suggests that the tongue tip is not as essential for the production of different vowel height qualities as other parts of the tongue. Nevertheless, this observation does not dilute the finding that the pre-r allophones are generally produced more openly in both of the examined dialects.

The acoustic analysis yielded similar results as the analysis of the position data. The absolute difference between the F_1 frequencies in hVt and hVr was significantly larger for the East group for the first 150 ms of the vowels. The frequencies in the vowel offset differ more for the South group for the same reasons as stated above. The hVt/hVr comparison even yielded insignificant results for the South group when the /ɛː/ and /øː/ data were combined. A separate vowel formant analysis resulted in significantly higher F_1 frequencies for /ɛː/ in both dialectal groups. Only the first half of the pre-r /øː/ yielded significantly higher F_1 values in the East group. The same allophone was not different from its pre-t version in the South group (except for the vowel offset). This can once again be explained by the fact that the vowel offsets in the words *höt* and *hör* are articulated in a very similar way in the Eastern dialects, but not in the Southern ones. Moreover, /øː/ is naturally produced in a less open manner than /ɛː/ which leads to a sooner articulatory convergence of the tongue and the palate in the word *hör*.

The second hypothesis which stated that the difference between pre-r and pre-t vowel height is larger in the dialects spoken in the East of Sweden than in the dialects spoken in the South could therefore be corroborated by the results of this study. Both articulatory and acoustic analyses of the vowels /ɛː/ and /øː/ indicated that there was a larger difference in vowel height for speakers from in and around Stockholm than for speakers from in and around Malmö.

6.3 Limitations

I would like to point out that the findings from this study do only concern a certain scope and should thus not be overgeneralised. First of all it is not completely accurate to speak of the entirety of the South Swedish dialects when only eight speakers from one area within Scania are considered (and the same goes of course for the East Swedish dialect group). Secondly, the pre-r allophones were merely compared to one phone in another consonant context, namely the pre-t one. There is indeed no indication for the existence of any other /ɛː/ or /øː/ allophones in Swedish (apart from those mentioned in this study) and thus there is no reason to believe that the articulation of those vowels is essentially different in any other contexts. However, for the sake of completeness it is advisable to examine the characteristics of these vowels in other surroundings as well. Finally, diphthongisation has not been further analysed in this study, even though, it was pointed out in section 2.1.2 to be an important characteristic of Swedish vowels which should not be underestimated. Analysing the particular effects that diphthongisation might have on vowel lowering might be appropriate in order to understand this process to an even better extent and it might even help explain some of the outcomes of this study.

In the author's opinion, it would be very interesting to look at differences in vowel production between younger and older speaker generations in order to investigate if there is a trend of lowering /ɛ/ and /ø/ in all contexts as suggested by Riad (2013). However, this could not be examined within this study due to an uneven distribution of ages and genders among the participants (see table A in the appendix). This task is therefore being left to future research. Moreover, it might be a good idea to carry out some perception tests to compare the results of this study to how different speakers perceive the differences in vowel quality.

7 Summary

The aim of this study was to examine the pronunciation of the Swedish vowels /ɛː/ and /øː/ since a number of phoneticians had pointed out that the pre-r allophones of these vowels are realised more openly than the allophones occurring in other contexts. This phenomenon was investigated on by collecting speech samples and tongue position data of Swedish speakers from the two dialectal areas Malmö and Stockholm. The participants were asked to produce sentences containing the words *hVr* and *hVt*, where V was replaced by either /ɛː/ or /øː/. The tongue positions and formant frequencies were recorded and analysed by means of Functional Data Analysis.

Both the articulatory and the acoustic data have indicated a significantly more open pronunciation for the pre-r allophone of /ɛː/ compared to its pre-t allophone in both dialect groups. The analysis of the tongue position data suggests the same for the pre-r allophone of /øː/, although this could not be supported by the results of the formant analysis. The comparison between the two dialectal groups showed that the difference in openness is significantly larger for the dialects spoken around Stockholm which agrees with the general opinion of most phoneticians.

The fact that electromagnetic articulography has so far rarely been used to conduct studies on Swedish vowel production leaves many open questions for future research. One example is diphthongisation which is an important feature of Swedish vowels and should therefore be considered in further studies within this area. Furthermore, it would be interesting to examine vowel production across different age groups in order to gain more insight on the vowel shift as an ongoing process.

References

Paul Boersma and David Weenink. Praat: doing phonetics by computer. [Computer program], 2013. Version 5.3.11, retrieved 2 February 2013 from `http://www.praat.org/`.

Jasmin Bolfek-Radovani. Attityder till svenska dialekter – en sociodialektologisk undersökning bland vuxna svenskar i Tierp, Östersund, Linköping, Trollhättan och Malmö. Sociolingvistiska dokument (SoLiD) nr 13, FUMS Rapport nr 201, 2000.

Gösta Bruce. *Vår fonetiska geografi - Om svenskans accenter, melodi och uttal.* Studentlitteratur, 2010.

Gösta Bruce and Olle Engstrand. The phonetic profile of Swedish. *Sprachtypologie und Universalienforschung (STUF)*, 59(1):12–35, 2006.

Xuemin Chia and Morgan Sondereggerb. Subglottal coupling and its influence on vowel formants. *Journal of the Acoustical Society of America*, 122(3):1735–1745, 2007.

Claes-Christian Elert. *Allmän och svensk fonetik.* Norstedts, 2000.

Olle Engstrand. *Fonetikens grunder.* Studentlitteratur, 2004.

Gunnar Fant. The acoustics of speech. In *Third International Contress on Acoustics*, pages 188–201, 1959.

Gunnar Fant. *Speech Sounds and Features.* The MIT Press, 1973.

Gunnar Fant. Feature analysis of Swedish vowels - a revisit. 24(2-3):1–19, 1983.

Michele Gubian, Francesco Cangemi, and Lou Boves. Joint analysis of F0 and speech rate with Functional Data Analysis. In *Acoustics, Speech and Signal Processing (ICASSP)*, page 4972–4975, 2011.

Philip Hoole and Noel Nguyen. Electromagnetic articulography in coarticulation research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 35:177–184, 1997.

Philip JB Jackson and Veena D Singampalli. Coarticulatory constraints determined by automatic identification from articulograph data. In *8th International Seminar on Speech Production (ISSP)*, pages 377–380, 2008.

William F. Katz, Sneha V. Bharadwaj, and Burkhard Carstens. Electromagnetic articulography treatment for an adult with Broca's aphasia and apraxia of speech. *Journal of speech, language, and hearing research (JSLHR)*, 42(6): 1355–1366, 1999.

Peter Ladefoged. *A Course in Phonetics*. Thomson Wadsworth, fourth edition, 2000.

Therese Leinonen. *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. PhD thesis, University of Groningen, 2010.

Bertil Malmberg. *Lärobok i fonetik*. Gleerups, 1970.

Bertil Malmberg. *Svensk fonetik*. Gleerups, 1971.

Ricardo A. Maronna and Ruben H. Zamar. Robust estimates of location and dispersion of high-dimensional datasets. *Technometrics*, 44(4):307–317, 2002.

Joseph S. Perkell, Marc H. Cohen, Mario A. Svirsky, Melanie L. Matthies, Inaki Garabieta, and Michel T. T. Jackson. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92(6):3078–3096, 1992.

Gordon E. Peterson and Harold L. Barney. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2):175–184, 1952.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

J.O. Ramsay, K.G. Munhall, V.L. Gracco, and D.J. Ostry. Functional Data Analysis of lip motion. *Journal of the Acoustical Society of America*, 99:3718–3727, 1996.

Tomas Riad. Ungdomarna står för förändringen. Och de äldre knorrar. Språktidningen, March 2013.

Susanne Schötz, Johan Frid, and Anders Löfqvist. Exotic vowels in Swedish: a project description and an articulographic and acoustic pilot study of /iː/. In *Proceedings of Fonetik 2011, Speech, Music and Hearing*, 2011.

Susanne Schötz, Johan Frid, and Anders Löfqvist. South Swedish diphthongisation: an articulographic and acoustic study of /uː/ in the Malmöhus dialect. In *Proceedings of Fonetik 2012*, page 85–88, 2012.

Susanne Schötz, Johan Frid, Lars Gustafsson, and Anders Löfqvist. Functional Data Analysis of tongue articulation in palatal vowels: Gothenburg and Malmöhus Swedish /iː, yː, ʉ̟ː/. In *Proceedings of Interspeech 2013*, 2013.

Maureen Stone. *Principles of experimental phonetics*, chapter 14: Instrumentation for the study of speech physiology, pages 495–524. St. Louis: Mosby, 1969.

Appendix

ID	SEX	AGE	DIALECT
1	f	46	South
2	m	41	South
3	m	50	South
5	f	23	South
7	f	46	South
8	m	52	South
9	f	23	South
11	f	35	South
14	m	32	South
4	m	43	East
6	m	34	East
16	f	45	East
20	m	36	East
22	m	50	East
25	f	21	East
27	m	63	East
28	m	61	East
29	f	21	East

Table A – Summary of subject information

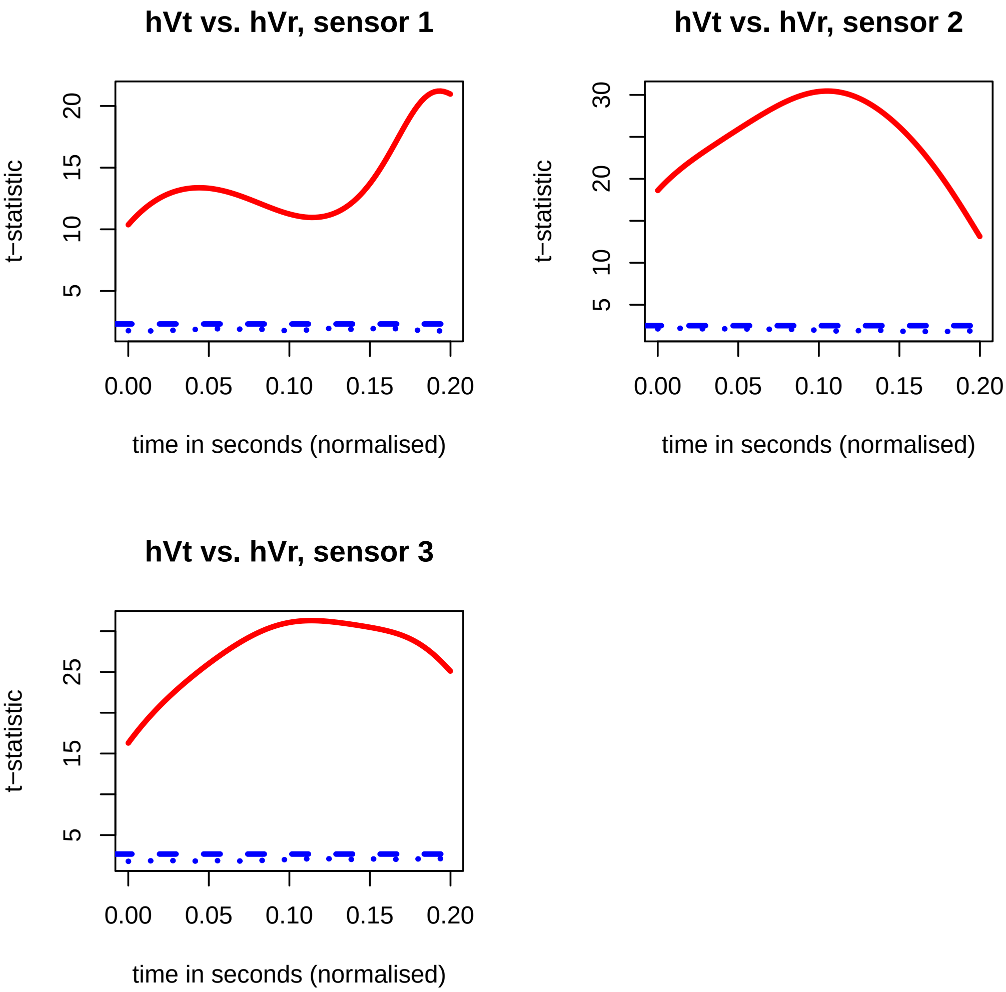


Figure A – Pairwise functional t-tests for average tongue height in hVr and hVt for sensors 1-3; observed statistic (upper solid lines), critical value (0.05) for the pointwise comparison (lower dotted lines) and maximum of the t-statistics (middle dashed lines)

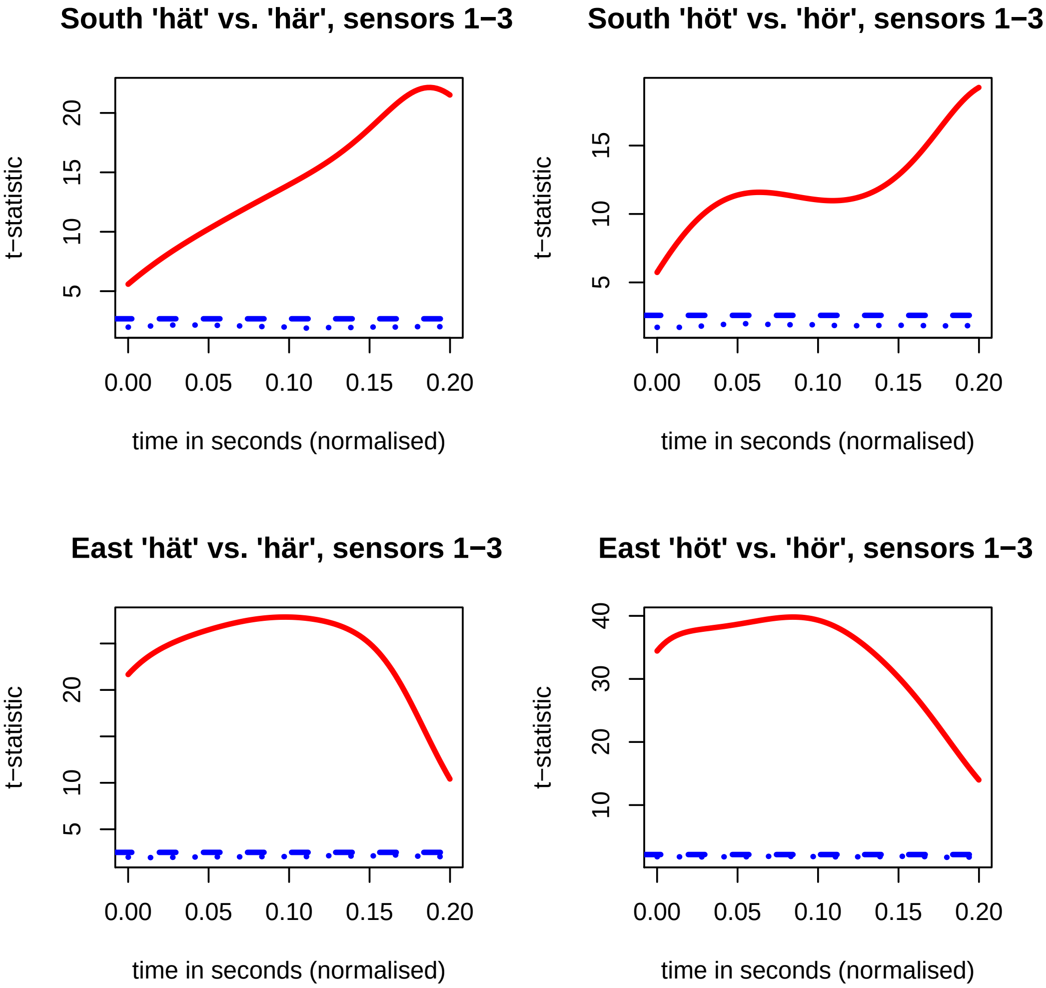


Figure B – Pairwise functional t-tests for the average tongue heights (means over all sensors) in the words *hät, här, höt* and *hör* (South and East group); observed statistic (upper solid lines), critical value (0.05) for the pointwise comparison (lower dotted lines) and maximum of the t-statistics (middle dashed lines)

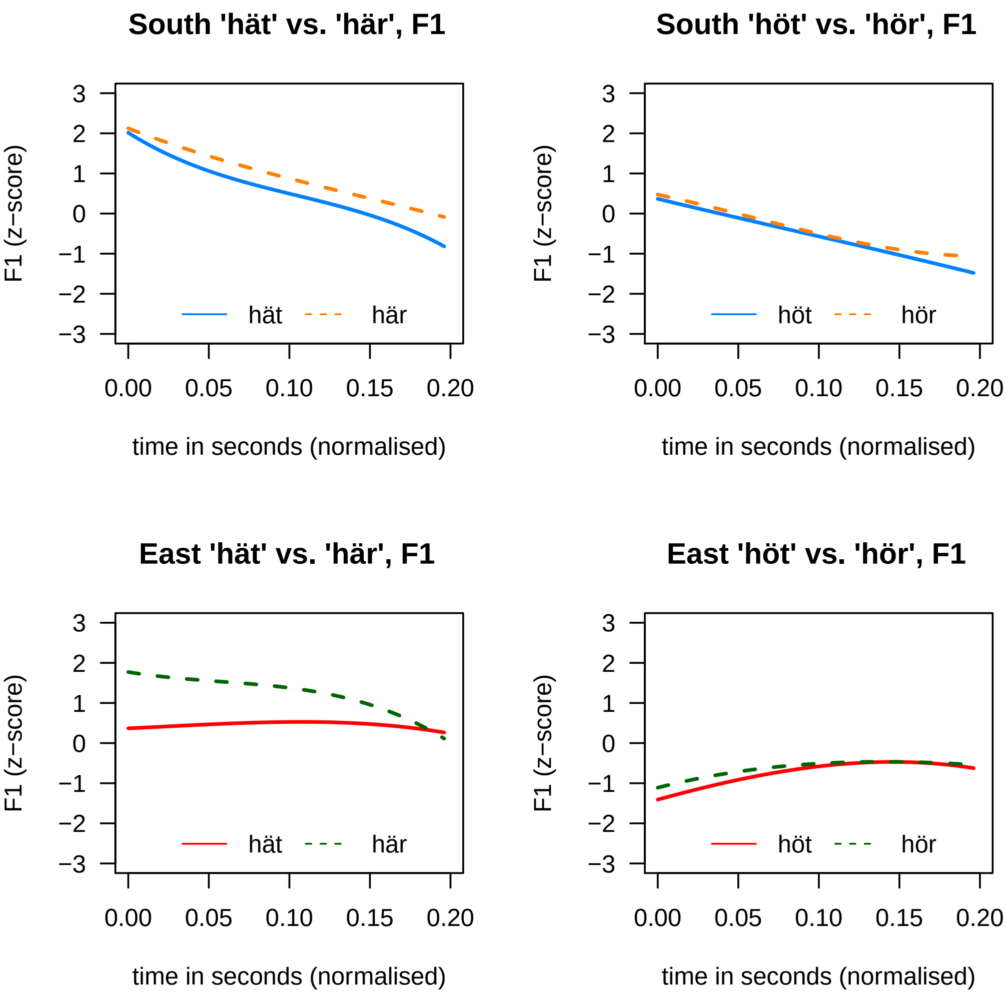


Figure C – Average F_1 frequencies in the words *hät, här, höt* and *hör* (South and East group)

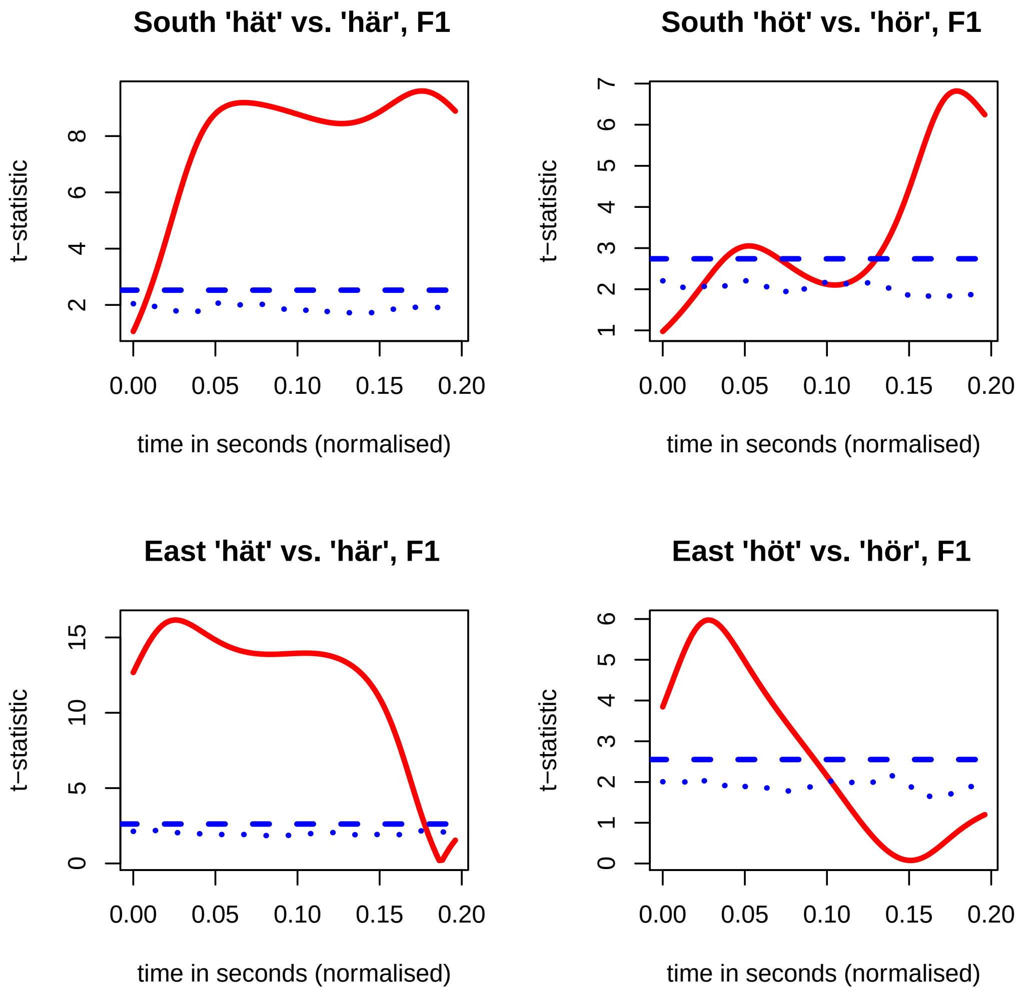


Figure D – Pairwise functional t-tests for average F_1 frequencies in the words *hät, här, höt* and *hör* (South and East group); observed statistic (solid red lines), critical value (0.05) for the pointwise comparison (dotted lines) and maximum of the t-statistics (dashed lines)