Functional data analysis of tongue articulation in Gothenburg and Malmöhus Swedish /i:, y:, u-:/

Schötz, Susanne; Frid, Johan; Löfqvist, Anders; Gustafsson, Lars

Published in:
[Host publication title missing]

2013

Functional data analysis of tongue articulation in Gothenburg and Malmöhus Swedish /iː, yː, ʉ̟ː/

Susanne Schötz[1], Johan Frid[1], Lars Gustafsson[1] & Anders Löfqvist[2]
[1] Humanities Lab, Centre for Languages & Literature, Lund University, Sweden
[2] Dept. of Logopedics, Phoniatrics and Audiology, Lund University, Sweden

Abstract

Articulatory data collected from nine speakers each of Gothenburg and Malmöhus Swedish were used in a Functional Data Analysis (FDA) to study tongue articulation dynamics, more specifically the height and frontness of the tongue body and tip in the palatal vowels /iː, yː, ʉ̟ː/. Standard z-score transformations were used for speaker normalisation. Results showed that the tongue articulation for /iː/ and /yː/ is generally similar, and significantly different from /ʉ̟ː/ in both Malmöhus and Gothenburg Swedish. We also found a subdivision of Gothenburg Swedish into two subtypes, where type 1 resembled Malmöhus Swedish more. Significant differences in tongue body height were found between all varieties for all of the vowels, except for /yː/ between Gothenburg type 1 and Malmöhus Swedish.

Introduction

The Swedish vowel system is fairly rich, and Swedish vowels have some particularly unusual and distinctive features. One such feature is that there are three contrastive long front, close vowels /iː, yː, ʉ̟ː/, characterised by a relatively small acoustic and perceptual distance. The magnitude of the lip opening is regarded as the major distinctive feature: unrounded /iː/, outrounded /yː/, and inrounded /ʉ̟ː/ (Fant, 1959; Ladefoged & Maddieson, 1996). Specifically the contrast between /yː/ and /ʉ̟ː/ is considered highly unusual among the world's languages. The tongue articulation is assumed to be basically identical, but the documentation of this is incomplete, especially for the articulatory dynamics (Ladefoged & Maddieson, 1996:295–296).

In many varieties of Swedish, /iː, yː, ʉ̟ː/ are also characterised by a slight diphthongisation or consonantal offglide at the end. For /iː/ and /yː/, this is typically made with the tongue dorsum as a [j] sound, while for /ʉ̟ː/ the gesture is achieved by the lips approaching each other as a [β] sound (McAllister et al., 1974; Hadding et al., 1976). The different diphthongisations at the end of these vowels contribute to maintaining the distinctions between them. The articulatory dynamics of vowels in Swedish, specifically of palatal vowels, has not been subjected to any systematic phonetic production study. Spectral changes have been claimed to be more important for vowel perception than static cues, see e.g. (Nearey, 1989; Strange, 1989).

Another rare feature is the nowadays fairly wide-spread realisation of /iː/ and /yː/ in Swedish with Viby-colouring, i.e. with a "damped" quality /ɨː/ and /ʉ̟ː/ (Ladefoged & Maddiesson, 1996; Bruce, 2010). There is disagreement in the Swedish phonetics literature if the major constriction for the damped /iː/ and /yː/ is further front compared to their regular counterparts, and basically alveolar, or instead further back and rather central (Björsten et al, 1999; Engstrand et al., 2000). However, as adequate articulatory data seem to be lacking, these views are at best intelligent speculations.

The purpose of this study was to use Functional Data Analysis to examine tongue articulation of Swedish vowels. We focus on the vowels /iː, yː, ʉ̟ː/ in two regional varieties of Swedish; Gothenburg Swedish (GS) and Malmöhus Swedish (MS), spoken in and near Gothenburgh and Malmö, respectively. The aim was to find out if the tongue positions are similar for these vowels as previously assumed, and if there are any regional differences. An additional aim was to learn more about the articulatory dynamics of palatal vowels in Swedish. We expected the tongue positions in the dimensions open–close and front–back to be similar in for /iː/, /yː/ and /ʉ̟ː/ in both GS and MS. Furthermore, we expected to find regional differences in the articulation of /iː/ and /yː/, as Viby-colouring is more common in GS than in MS (Bruce, 2010).

Material and method

Nine speakers of GS (5 females, 4 males, 20–47 years) and 9 speakers of MS (4 females and 5 males, 23–62 years) were recorded by means of electromagnetic articulography (Carstens AG 500). Twelve sensors were placed on the lips, jaw and tongue, and also on the nose ridge and behind the ear to correct for head movements. *Figure 1* shows the sensor positions and one subject with sensors attached. In this study, our focus was on the tongue tip and body (sensors 1 and 2). The speech material consisted of 20 repetitions from each speaker of /iː, yː, ʉ̟ː/ in carrier sentences *De va inte hVt utan hVt ja sa* (It was not hVt, but hVt I said), where the target words were stressed. The sentences were displayed on a computer screen one at a time in random order, and the speakers were instructed to read them in their own dialect at a comfortable speech rate.

Figure 1: The twelve sensor positions recorded, and one speaker with the sensors attached.

Error detection and speaker normalisation

Noise and measurement errors in articulatory data may come from a) quick movements by the speaker, b) sensors moving too close to each other, c) sensors breaking or falling off, and d) calculation errors. In order to detect and exclude such errors, we used a two-step process. All /iː, yː, ʉ̟ː/ vowels were segmented manually in Praat (Boersma & Weenink, 2013) and used as acoustic landmarks to trim the data set. Plots for sensors traces 1–3 were used to visually identify and exclude vowels with errors. The remaining errors and outliers were removed with the package 'robustbase' (Rousseeuw et al., 2012) in the R statistical environment (R Development Core Team, 2013), using a method that calculates location (mu) and scale (tau) from articulatory data using robust methods (Maronna & Zamar, 2002). In our case, all the position data in all repetitions of each vowel /iː, yː, ʉ̟ː/, each of the sensors (1–3), and each spatial dimension (x, y, z), for each speaker were used to calculate the mean value of all the individual repetitions of each vowel. If the mean value of a repetition was above or below mu +/- tau, it was marked as an outlier and excluded.

In order to compensate for differences in oral anatomy between speakers, data was normalized using *z*-score transformation.

FDA smoothing and aligning

Functional Data Analysis (FDA) is a technique for timewarping and aligning a set of signals to examine differences between them. FDA techniques and applications to speech analysis were first introduced by Ramsay et al. (1996), and further developed by Lucero et al. (1997), Lucero and Löfqvist (2005) and Gubian et al. (2011). In FDA, a function or function system is fitted to the data, and the fitting coefficients are examined instead of the original data. A commonly used function form are B-spline functions (Ramsey et al. 2009), which are flexible building blocks for fitting curves to approximate a large number of different shapes. In essence, spline functions are placed at overlapping, equidistant intervals throughout a sensor trace. By selecting weights for each spline, the overall shape becomes similar to the actual sensor trace. The degree of similarity may be controlled so that it does not overfit. It is possible to select: a) the number of spline functions ('knots'), b) the order (how well higher-order derivatives are preserved) and c) the amount of roughness ('lambda'). In this study, FDA was used to smooth the sensor traces, and to standardise the time to facilitate comparisons between repetitions. All FDA processing was done using the R package 'fda' and the following parameters for creating the B-spline basis: knots=20, order=6, lambda=$1e^{-2}$.

Analysis of tongue height and frontness

Sensors 1 and 2 were selected to represent the tongue tip and body (see *Figure 1*). We plotted the FDA processed contours for the tongue dynamics in height and frontness for the tongue body and tongue tip, and compared the positions and dynamics within each regional group as well as across the two regional varieties. Statistical analysis was done with functional *t*-tests, an extension of the classical *t*-test where the *t*-statistic is a function of time, using the function *tperm.fd* in the 'fda' package. Functional *t*-tests are described in detail in Ramsey et al. (2009).

Results

Tongue body height

Within each variety, the contours for /ʉ̟ː/ are often clearly separated from /iː/ and /yː/, which in turn often overlap, and significant differences in tongue body height (*Figure 2*, column 1) were found between /ʉ̟ː/ and /iː, yː/ (pairwise functional *t*-tests, $p < 0.05$).

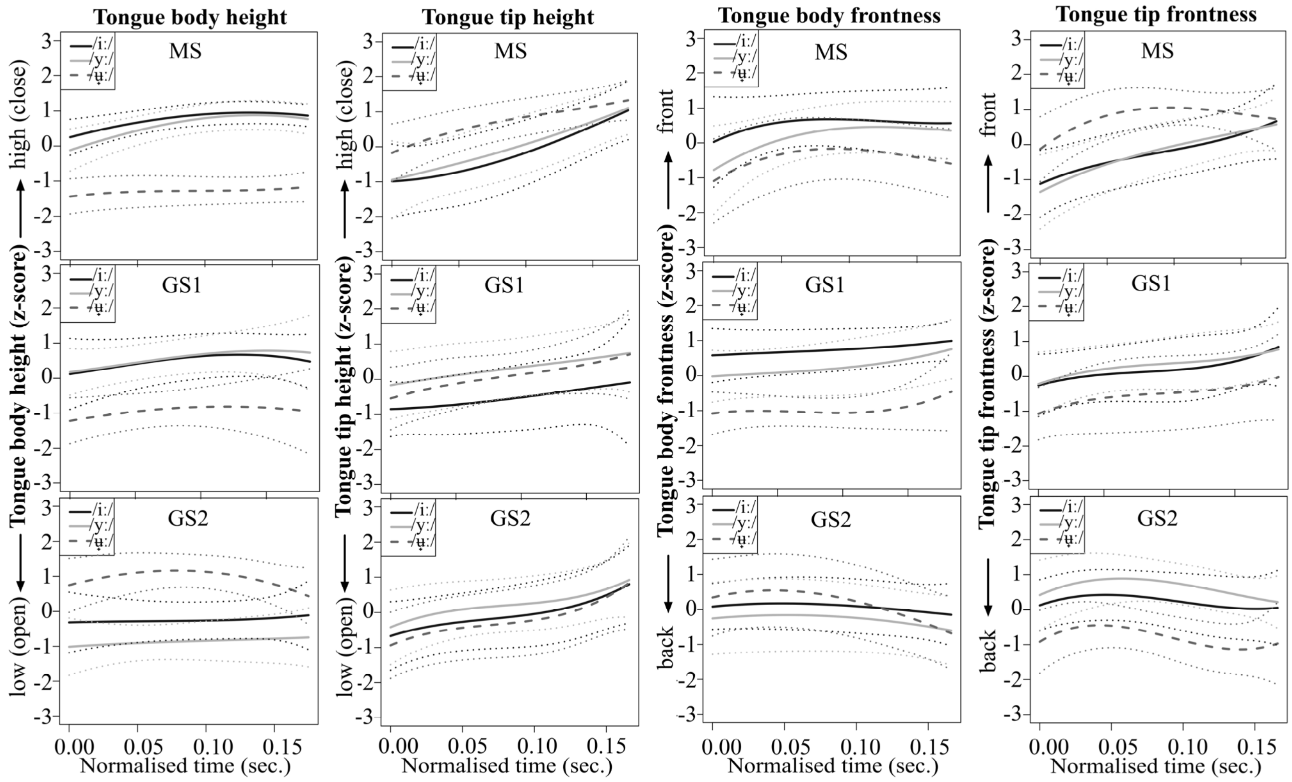

Figure 2. Tongue articulation (z-scores of tongue body and tongue tip height as well as frontness) as a function of normalised time for the vowels /iː, yː, ʉ̟ː/ in Malmöhus Swedish (MS) and two types of Gothenburg Swedish (GS1 and GS2); mean values for each variety (dotted lines: standard deviation).

The GS speakers generally displayed more variation than the MS speakers in all tongue positions and all vowels. Among the GS speakers we found a subdivision between four speakers (type GS1) who articulated the vowels with similar tongue positions as the MS speakers, and five speakers (type GS2) who generally had a different tongue positions than slightly higher tongue body for /iː/ than for /yː/. The dynamics, represented by the mean contour shapes, are fairly level for /iː/ and /yː/ in GS1 and GS2, but there seems to be some individual variation. In MS, the contours for /iː/ and /yː/ are slightly rising, suggesting a mild closing diphthongisation. /ʉ̟ː/ is relatively level in MS and GS1, but is somewhat arch-shaped in GS2.

Tongue tip height

The tongue tip height contours for /iː/ and /yː/ in MS are similar and somewhat lower than for /ʉ̟ː/ (*Figure 2*, column 2). In GS1, /iː/ seems to be produced with a lower tongue tip than /yː/ and /ʉ̟ː/, while GS2 has similar contours for all three vowels. The dynamics for all the vowels in all three varieties is represented by slightly rising contours, suggesting closing diphthongisations, although some individual variation was observed.

Tongue body frontness

While the tongue body in MS and GS1 is more front for /iː/ and /yː/ than for /ʉ̟ː/ (*Figure 2*, column 3), the opposite pattern is shown in GS2, except towards the final part of the vowels, where there is more variation. In addition, the tongue body seems to be slightly more front for /iː/ than for /yː/ in all varieties. All MS vowel contours rise initially, indicating a forward motion. /iː/ and /yː/ are fairly level in GS1 and GS2, while the tongue body seems to move slightly forward (GS1) or backward (GS2) in the final part of /ʉ̟ː/.

Tongue tip frontness

In MS the tongue tip is further back in /iː/ and /yː/ compared to /ʉ̟ː/, while the opposite pattern is found for GS1 and GS2 (*Figure 2*, column 4). The contours for /iː/ and /yː/ are similar and overlapping in MS and GS1, while /yː/ tends to be a bit more front than /iː/ in GS2. In MS the /iː/ and /yː/ contours are rising, indicating height-harmonic diphthongisations towards more peripheral vowels, while the GS1 contours are moving slightly forward. The GS2 contours are slightly arch-shaped.

Discussion and future work

The results of this study indicate that the tongue articulation for /ʉ̟ː/ is significantly different from /iː/ and /yː/ in both MS and GS. Our hypothesis of similar tongue articulation for the three vowels was thus rejected. In addition, we found more intra-regional variation in GS than in MS, which led to the subdivision into the two types GS1 and GS2. A closer look showed that the GS1 speakers were more often from the outskirts of the Gothenburg area than the GS2 speakers. Furthermore, most GS2 speakers had clear Viby-coloured /iː/ and /yː/, which was not the case for all GS1 speakers. No MS speakers used Viby-colouring. The Viby-colouring may offer one explanation for the differences in tongue articulation. However, a few GS1 speakers did use some kind of Viby-colouring, and we need to investigate further how the speakers articulated both general and Viby-coloured vowels. We will also compare this data to acoustic data, e.g. formant frequencies.

Considerable regional variation was found in this study, not only for each vowel in the front–back and open–close dimensions, but also in the vowel dynamics (diphthongisation). Our hypothesis of different articulation strategies in different regional varieties was thus supported.

In this study we analysed only two discrete points and two dimensions of the tongue: tongue tip and body height and frontness, and used a standard *z*-score transformation for speaker normalisation. Although we did not look at lip rounding, traditionally regarded as the main difference between /iː/, /yː/ and /ʉ̟ː/, our results clearly show differences between these vowels in tongue body height as well. In future studies, tongue body height will be compared to other tongue articulation dimensions as well as to lip rounding. Moreover, we will include other palatal vowels, and compare tongue articulation in MS and GS to that of Stockholm Swedish. We will also investigate more sophisticated speaker normalisation methods.

Acknowledgements

This study was carried out within the project Exotic Vowels in Swedish: an articulographic study of palatal vowels. The authors also gratefully acknowledge support from the Linnaeus environment Thinking in Time: Cognition, Communication and Learning, financed by the Swedish Research Council, grant no. 349-2007-8695.

References

Björsten, S., G. Bruce, C-C Elert, O. Engstrand, A. Eriksson, E. Strangert & P. Wretling. 1999. Svensk dialektologi och fonetik – tjänster och gentjänster. *Svenska landsmål och svenskt folkliv*, 7–23.

Boersma, P. & D. Weenink. 2013. *Doing phonetics by computer*. Retrieved from http://www.praat.org/

Bruce, G. 2010. *Vår fonetiska geografi*. Lund: Studentlitteratur.

Fant, G. 1959. Acoustic description and classification of phonetic units. *Ericsson Technics* 1 (Reprinted 1983 in G. Fant, *Speech Sounds and Features*, 32–83. Cambridge, MA: MIT Press)

Engstrand, O., S. Björsten, B. Lindblom, G. Bruce & A. Eriksson. 2000. Hur udda är Viby-i? Experimentella och typologiska observationer. *Folkmålsstudier* 39:83–95. Helsingfors.

Gubian, G., F. Cangemi & L. Boves. 2011. Joint analysis of F0 and speech rate with functional data analysis. *ICASSP*, 4972–4975, Prague.

Hadding, K., H. Hirose & K. S. Harris. 1976. Facial muscle activity in the production of Swedish vowels: An electromyographic study. *Journal of Phonetics* 4:233–245.

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

Lucero, J., K. Munhall, V. Gracco & J. Ramsay. 1997. On the registration of time and the patterning of speech movements, *Journal of Speech, Language and Hearing Research* 40:1111–1117.

Lucero, J. & A. Löfqvist. 2005. Measures of articulatory variability in VCV sequence, *Acoust. Res. Lett. Online* 6:80–84.

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

McAllister, R., J. Lubker & B. Lindblom. 1974. An EMG study of some characteristics of the Swedish rounded vowels. *Journal of Phonetics* 2:267–278.

Nearey, T. 1989. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85: 2088–2113.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*, URL: http://www.R-project.org.

Ramsay, J., G. Hooker & S. Graves. 2009. *Functional Data Analysis with R and MATLAB*. Springer.

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

Rousseeuw, P., C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, M. Maechler. 2012. *robustbase: Basic Robust Statistics*. R package version 0.9-7. URL: http://CRAN.R-project.org/package=robustbase.

Strange, W. 1989. Evolving theories of vowel perception. *Journal of the Acoustical Society of America* 85:2081–2087.

Proceedings of Fonetik 2013

The XXVI[th] Annual Phonetics Meeting 12–13 June 2013, Linköping University Linköping, Sweden

Studies in Language and Culture no. 21

Robert Eklund, editor

Linköping University

Conference website: www.liu.se/ikk/fonetik2013
Proceedings also available at: http://roberteklund.info/conferences/fonetik2013

Cover design and photographs by Robert Eklund
Photo of Claes-Christian Elert taken by Eva Strangert on the occasion of his 80th birthday

Proceedings of Fonetik 2013, the XXVIth Swedish Phonetics Conference
held at Linköping University, 12–13 June 2013
Studies in Language and Culture, no. 21
Editor: Robert Eklund
Department of Culture and Communication
Linköping University
SE-581 83 Linköping, Sweden

ISBN 978-91-7519-582-7
eISBN 978-91-7519-579-7
ISSN 1403-2570

Printed by LiU-Tryck, Linköping, Sweden, 2013