LUND UNIVERSITY

Functional Data Analysis of Tongue Articulation in Palatal Vowels: Gothenburg and Malmöhus Swedish /i:, y:, u-:/

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

Link to publication

Functional Data Analysis of Tongue Articulation in Palatal Vowels: Gothenburg and Malmöhus Swedish /iː, yː, ʉ̟ː/

Susanne Schötz[1], Johan Frid[1], Lars Gustafsson[1], Anders Löfqvist[2]

[1]Lund University Humanities Lab, Centre for Languages & Literature, Sweden
[2]Dept. of Logopedics, Phoniatrics and Audiology, Lund University, Sweden

susanne.schotz@ling.lu.se, johan.frid@ling.lu.se,
lars.gustafsson@humlab.lu.se, anders.lofqvist@med.lu.se

Abstract

This study used Functional Data Analysis (FDA) to analyse tongue articulation dynamics, more specifically the height and frontness of the tongue body and tip in the palatal vowels /iː, yː, ʉ̟ː/ of two varieties of Swedish. Articulatory data were collected from nine speakers each of Gothenburg and Malmöhus Swedish. 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.

Index Terms: Functional Data Analysis, articulography, palatal vowels, Swedish dialects

1. 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, and exemplified by minimal triplets such as ni, ny, nu, /niː, nyː, nʉ̟ː/ ('you', 'new', 'now'). The magnitude of the lip opening is regarded as the major distinctive feature: unrounded (with spread lips) /iː/, outrounded /yː/, and inrounded /ʉ̟ː/ [1, 2]. Specifically the contrast between /yː/ and /ʉ̟ː/, two similar but still phonemically distinct rounded vowels, is considered highly unusual and exotic 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 (see e.g. [2]:295–296).

1.1. Diphthongisation

In many regional varieties of Swedish /iː, yː, ʉ̟ː/ are also characterised by a slight diphthongisation gesture or consonantal off-glide 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 instead achieved by the lips approaching each other as a [β] sound [3, 4]. The different off-glides (diphthongisations) at the end of these vowels contribute to maintaining the distinctions between them. In addition, they also seem to serve as safeguards to prevent confusion with other neighbouring vowels. 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. [5, 6].

1.2. Viby-colouring

Yet another rare feature among the world's languages is the nowadays fairly wide-spread realisation of /iː/ and /yː/ in Swedish with a "damped" or "buzzing" quality /i̽ː/ and /ʉ̟ː/ [2, 7]. In the Swedish linguistic community this is called Viby-colouring, after the Swedish rural town Viby. Previous acoustic studies of /iː/ [8, 9] indicate that F_1 and F_2 values are similar to the cardinal [ɨ] reported in Catford [10]. In the Swedish linguistics and phonetics literature there is disagreement about whether the point of major constriction for the damped /iː/ and /yː/, is further front, as compared to their regular counterparts (front, close vowels), and basically alveolar, or instead further back and rather central (between front and back) [8]. It should be stressed that these views are at best intelligent speculations, as adequate articulatory data seem to be lacking.

1.3. Purpose, aims and hypotheses

The main purpose of this study was to use Functional Data Analysis and standard speaker normalisation methods to investigate tongue articulation and its dynamics (dipthongisation) of Swedish vowels. Our focus was the three vowels /iː, yː, ʉ̟ː/ in two regional varieties of Swedish; Gothenburg Swedish (GS), spoken in and near the city of Gothenburg in the West of Sweden, and Malmöhus Swedish (MS), spoken in and near the city of Malmö in the South of Sweden. The main 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 (diphthongisation) of palatal vowels in Swedish. We wanted to test two hypotheses. In the first hypothesis we expected the tongue positions in the dimensions open–close and front–back to be similar for /iː/, /yː/ and /ʉ̟ː/ in both Gothenburg and Malmöhus Swedish. The second hypothesis was that we expected to find regional differences in the articulation of /iː/ and /yː/, as Viby-colouring is more common in GS than in MS [7].

2. Method

2.1. Data aquisition

Nine speakers of Gothenburg Swedish (five females and four males, age: 20–47 years, mean = 29, sd = 10.0) and nine speakers of Malmöhus Swedish (four females and five males, age:

23–62 years, mean = 43, sd = 11.7) were recorded by means of electromagnetic articulography (AG 500, Carstens Medizinelektronik). 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 (sensor 1) and the tongue body (sensor 2). The speech material consisted of 20 repetitions from each speaker of /iː, yː, ʉ̟ː/ in carrier sentences of the type "De va inte hVt utan hVt ja sa" (It was not hVt, but hVt I said), where the target words containing the vowels were stressed and produced with prosodic contrastive focus. The sentences were displayed one by one on a computer screen in random order, and the speakers were instructed to read them in their own dialect at a comfortable speech rate. A contour of the palate was also recorded by asking each speaker to move the tongue tip several times back and forth along the midline of their palate.

Figure 1: *The twelve sensor positions used in the study, and one speaker with the sensors attached.*

2.2. Error detection and speaker normalisation

Noise and measurement errors in articulatory data are not uncommon, and may have several causes, including: a) a quick and sudden movement by the speaker, b) two or more sensors moving too close and interfering with each other, c) sensors breaking or falling off, and d) calculation errors of a sensor's spatial position from its distance to the transmitter coils. In order to detect and exclude any noise and other errors, we trimmed the data set to contain only the three vowels /iː, yː, ʉ̟ː/, and then used a two-step process to detect and remove errors and outliers.

For all speakers, the beginning and end of each vowel was manually identified and segmented in Praat [11], and used as acoustic landmarks. Plots of the traces of sensors 1–3 for each vowel were used to visually identify any major measurement errors, and exclude these vowels from further analysis.

Remaining errors and outliers were removed with a method from the package 'robustbase' [13] in the R statistical environment [12]. This method calculates location (mu) and scale (tau) from articulatory data using robust methods [14] (in robust statistics, 'location' and 'scale' roughly correspond to 'mean' and 'standard deviation', respectively). 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 from subsequent processing. In order to compensate for differences in oral anatomy between speakers, data was normalized using z-score transformation. The risk of removing

perfectly legitimate tokens in this way was minimal since all the material had been visually inspected in the previous step.

2.3. Functional Data Analysis smoothing and aligning

Functional Data Analysis (FDA) is a technique for time-warping 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. [18], and further developed by Lucero et al. [19], Lucero and Löfqvist [20] and Gubian et al. [21]. In FDA, a function or function system is fitted to the data, and the fitting coefficients are examined instead of the original data. The function(s) can take many different forms, but a commonly used type are B-spline basis functions [22]. B-spline functions are very flexible building blocks for fitting curves that can approximate a large number of different shapes. In essence, spline functions are placed at overlapping, equidistant intervals throughout a sensor trace. The spline functions also 'collaborate' to form an overall function. 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. More specifically, 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, the FDA procedure served two purposes: a) smoothing the sensor traces, and b) standardising the time to facilitate direct comparisons between repetitions. All FDA processing was performed using the R package 'fda' using the following parameters for creating the B-spline basis: knots = 20, order = 6, lambda = 0.01. Parameters were selected based on intuition. Cross validation-based methods exist [22, 23], but when comparing the smoothed traces with the original ones, overfitting was judged to be less of a problem since the traces were rather smooth to begin with.

2.4. Analysis of Tongue height and frontness

Sensor 2 was selected to represent the tongue body, and sensor 1 was chosen to represent the tongue tip (see Figure 1). We plotted the FDA processed individual 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 using the R function *tperm.fd*. A functional t-test is an extension of the classical t-test where the t-statistic is a function of time. The resulting statistic is represented as a curve that shows the difference between two groups of functions. In order to test a hypothesis, the t-statistic is compared with 1) a pointwise 0.05 critical value found though a permutation test and 2) a maximum 0.05 critical value. Functional t-tests are described in detail in [22].

3. Results

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 were found between /ʉ̟ː/ and /iː, yː/ (pairwise functional t-tests, $p < 0.05$). Although the results of the functional t-tests are presented here only for tongue body height, 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
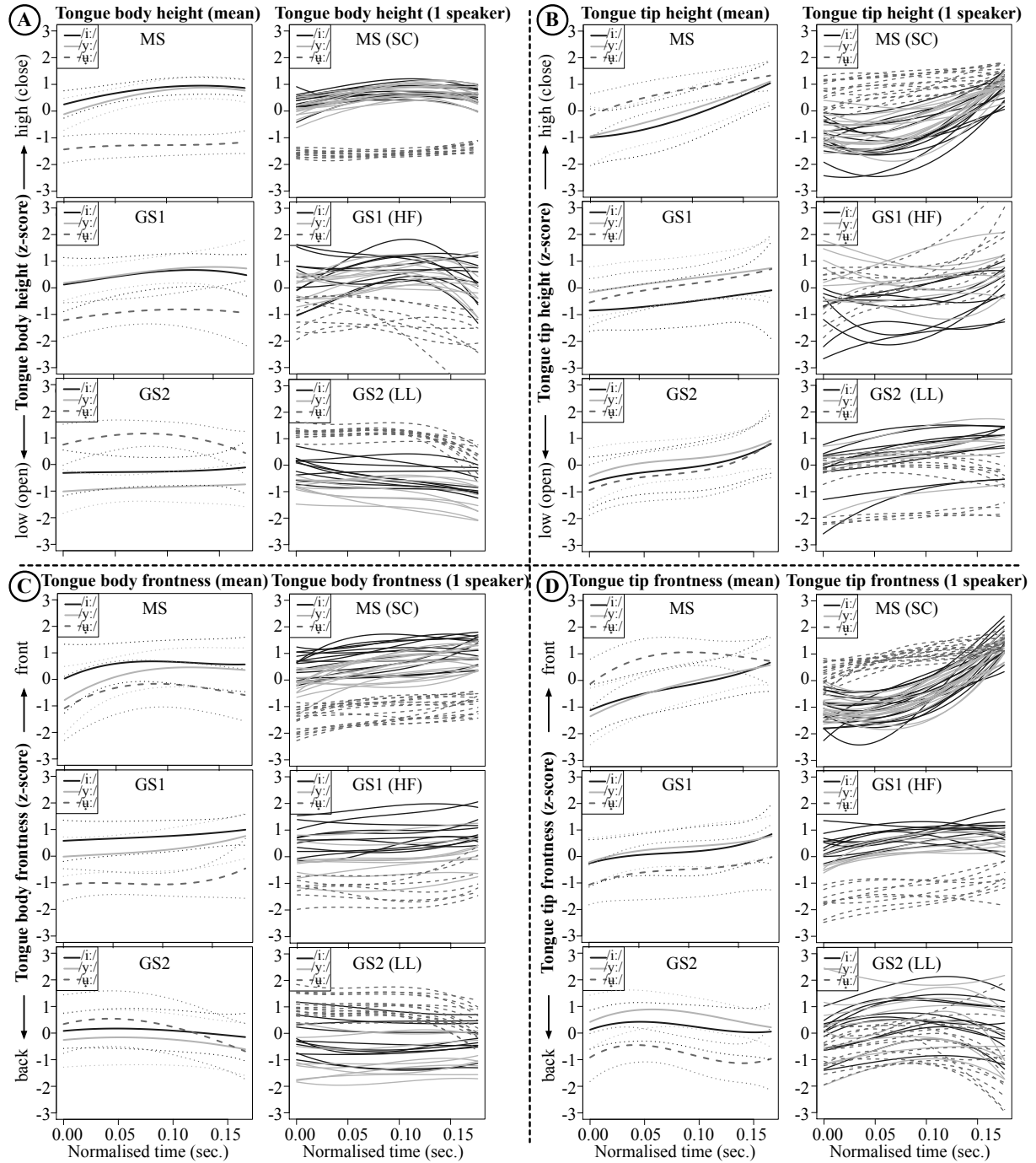
Figure 2: *Tongue articulation (z-scores of tongue body and tongue tip height as well as frontness) as a function of normalised time for the vowels /iː, yː, ʉ̟ː/ in Malmöhus Swedish (MS) and two types of Gothenburg Swedish (GS1 and GS2); mean values for each variety (columns 1 and 3; dotted lines: standard deviation), and individual vowel tokens by one speaker of each variety (columns 2 and 4).*

the other two varieties. In the following sections the results are presented separately for MS and the two GS types.

Figure 2 displays the the z-transformed and FDA processed tongue articulation dynamics for /iː, yː, ʉ̟ː/ in Malmöhus Swedish (MS) and two types of Gothenburg Swedish (GS1 and GS2). Mean contours are shown in columns 1 and 3, while columns 2 and 4 show individual vowel contours of a single speaker. Tongue body height is shown to the top left (A), tongue

tip height to the the top right (B), tongue body frontness to the bottom left (C), and tongue tip frontness to the bottom right (D).

3.1. Tongue body height

In MS and GS1, the tongue body is higher for /iː/ and /yː/ than for /ʉ̟ː/. There is also a considerable overlap of /iː/ and /yː/ in these two varieties. GS2 displays the opposite pattern, with a higher tongue body for /ʉ̟ː/ than for /iː/ and /yː/, and with a

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 individual variation, as is shown by the individual vowel contours of GS1 (HF). 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.

3.2. Tongue body frontness

While the tongue body in MS and GS1 is more front for /iː/ and /yː/ than for /ʉ̟ː/, 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 three varieties, although some overlap can be observed in the individual contours in column 2. All vowel contours rise initially in MS, 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/.

3.3. Tongue tip height

The tongue tip height contours for /iː/ and /yː/ in MS are similar and somewhat lower than for /ʉ̟ː/. 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 variation can be seen for the individual speakers in column 4.

3.4. 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. The contours for /iː/ and /yː/ are similar and overlapping in MS and GS1, while /yː/ tends to be somewhat more front than /iː/ in GS2. In MS the contours for /iː/ and /yː/ are rising, indicating height-harmonic dipthongisations towards more peripheral vowels, while the contours in GS1 are moving slightly forward. The contours are slightly arch-shaped in GS2.

3.5. Regional differences in tongue body height

The functional t-tests show full or partial significant differences in most dimensions between the three varieties ($p<0.05$) and the three vowels. However, the nature of the differences is very complex and a full discussion has to be left out due to lack of space. Only the results for tongue body height are presented here. Figure 3 shows pairwise functional t-tests with significantly different tongue body height for all three vowels in GS2 compared to both MS and to GS1. Between MS and GS1, we found significantly different tongue body heights for /ʉ̟ː/, and for the final part of /iː/, but not for /yː/.

4. Discussion and future work

The results of this study indicate that the tongue articulation is significantly different for /ʉ̟ː/ than for /iː/ and /yː/ in both Malmöhus and Gothenburg Swedish. 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 at these two GS types showed that the GS1 speakers were more often from the outskirts of the Gothenburg area than the GS2 speakers. Furthermore, an initial
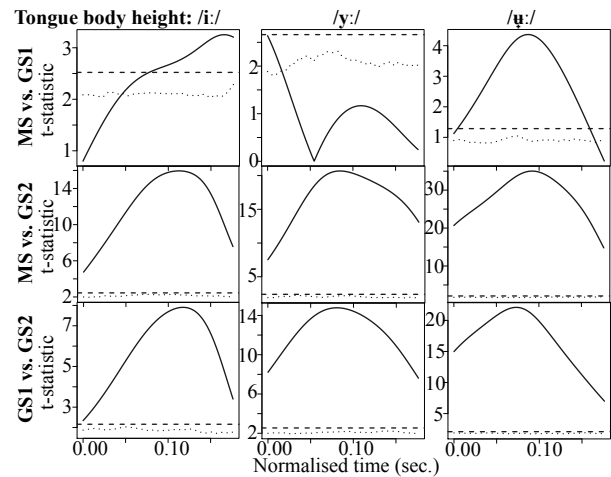

Figure 3: *Pairwise functional t-tests for tongue body height of /iː, yː, ʉ̟ː/ in Malmöhus Swedish (MS) and the two Gothenburg Swedish types (GS1 and GS2); solid lines: observed statistic, dashed lines: the permutation 0.05 critical value for the maximum of the t-statistic, dotted lines: the permutation critical value for the pointwise statistic.*

auditive analysis indicated that most GS2 speakers had clear Viby-coloured /iː/ and /yː/, which was not the case for the GS1 speakers. No MS speakers had Viby-coloured /iː/ or /yː/. 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 the articulatory data to acoustic data, including vowel 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 – with focus on tongue body height, and used a standard z-score transformation for speaker normalisation. Our results need to be tested further, and tongue body height should be related to the palate shape and compared to other tongue articulation dimensions as well as to lip rounding. Although this study did not look at lip rounding – traditionally regarded as the main difference between /iː/, /yː/ and /ʉ̟ː/ – our results clearly show that there are differences between these vowels in tongue body height as well. In future studies, we will include other palatal vowels, such as /eː/ and /øː/, and compare tongue articulation in MS and GS to that of Stockholm Swedish. We will also investigate more sophisticated speaker normalisation methods.

5. Acknowledgements

6. References

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

[2] Ladefoged, P. and Maddieson, I., "The Sounds of the World's Languages", Oxford: Blackwells, 1996.

[3] McAllister, R., Lubker J. and Lindblom, B., "An EMG study of some characteristics of the Swedish rounded vowels", Journal of Phonetics 2: 267–278, 1974.

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

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

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

[7] Bruce, G., "Vår fonetiska geografi", Lund: Studentlitteratur, 2010.

[8] Björsten, S. and Engstrand, O., "Swedish 'damped' /i/ and /y/: experimental and typological observations", In Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, August 1999, 1957–1960, 1999.

[9] Schötz, S. Frid, J. and Löfqvist, A.,"Exotic vowels in Swedish – an articulographic and acoustic pilot study of /i:/", In Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong on August 17–21, 2011.

[10] Catford, J. C., "A practical introduction to phonetics", Oxford: Oxford University Press, 1988.

[11] Boersma, P. and Weenink, D., "Doing phonetics by computer" [Computer program], Version 5.3.11, retrieved 2 February 2013 from http://www.praat.org/, 2013.

[12] R Development Core Team, "R: A Language and Environment for Statistical Computing", Webpage: http://www.R-project.org, accessed on 11 February 2013.

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

[14] Maronna, R. A. and Zamar, R. H., "Robust estimates of location and dispersion of high-dimensional datasets", Technometrics 44(4), 307–317, 2002.

[15] Hashi, M., Westbury, J. R. and Honda, K., "Vowel posture normalization", The Journal of the Acoustical Society of America, vol. 104, no. 4, pp. 2426–2437, 1998.

[16] Geng, C. and Mooshammer C., " How to stretch and shrink vowel systems: Results from a speaker normalization", Journal of the Acoustical Society of America 125, 3278–3288, 2009.

[17] Valdés Vargas, J. A., Badin, P. and Lamalle, L., "Articulatory speaker normalisation based on MRI-data using three-way linear decomposition methods", In Interspeech 2012 (13th Annual Conference of the International Speech Communication Association). Portland, Oregon, USA, 2012.

[18] Ramsay J. O., Munhall K. G., Gracco V. L. and Ostry D. J., "Functional data analysis of lip motion", Journal of the Acoustical Society of America 99, 3718–3727, 1996.

[19] Lucero, J., Munhall, K., Gracco, V. and Ramsay, J., "On the registration of time and the patterning of speech movements", J. Speech Lang. Hear Res. 40, 1111–1117, 1997.

[20] Lucero, J. and Löfqvist, A., "Measures of articulatory variability in VCV sequence", Acoust. Res. Lett. Online 6, 80–84, 2005.

[21] Gubian, M., Cangemi, F. and Boves, L., "Joint analysis of F0 and speech rate with functional data analysis", ICASSP, 4972–4975, Prague, Czech Republic, 2011.

[22] Ramsay J., Hooker, G. and Graves, S., "Functional Data Analysis with R and MATLAB", Springer, 2009.

[23] Gubian, M., "A functional PCA Session in Detail", Retrieved from http://lands.let.ru.nl/FDA/tutorials/Interspeech2011_tutorial_FDA_Gubian.zip, 2011.