Published in: The Scottish Consortium for ICPhS 2015 (ed.):
Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: University of Glasgow, 2015.
at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0231.pdf
(downloaded on: 07.03.2017/11:25).

Investigating dialectal differences using articulography

Martijn Wieling[1], Fabian Tomaschek[2], Denis Arnold[2], Mark Tiede[3] and R. Harald Baayen[2,4]

[1]University of Groningen, [2]University of Tübingen, [3]Haskins Laboratories, [4]University of Alberta
m.b.wieling@rug.nl, {fabian.tomaschek, denis.arnold, harald.baayen}@uni-tuebingen.de, tiede@haskins.yale.edu

ABSTRACT

The present study introduces articulography, the measurement of the position of tongue and lips during speech, as a promising method to the study of dialect variation. By using generalized additive modeling to analyze articulatory trajectories, we are able to reliably detect aggregate group differences, while simultaneously taking into account the individual variation across dozens of speakers. Our results on the basis of Dutch dialect data show clear differences between the southern and the northern dialect with respect to tongue position, with a more frontal tongue position in the dialect from Ubbergen (in the southern half of the Netherlands) than in the dialect of Ter Apel (in the northern half of the Netherlands). Thus articulography appears to be a suitable tool to investigate structural differences in pronunciation at the dialect level.

Keywords: Articulography, Generalized additive modeling, Dialectology.

1. INTRODUCTION

Currently, many studies in sociolinguistics and dialectology investigating pronunciation variation focus on the acoustic properties of vowels (e.g., [2, 5, 13, 18, 26]). Since the seminal study of Peterson & Barney [20], formant measurements have been the classical way to measure vowel quality.

Labov, Yaeger and Steiner [15] initiated the formant-based approach in sociolinguistics by studying English formant-based vowel variation in the United States of America. Since then many other studies assessing dialect variation have used formant-based methods, such as [2] investigating regional Dutch dialect variation, and [6] and [14] studying American English regional variation. While formant-based measures provide a convenient quantification of the acoustic signal, automatic formant detection is far from perfect and requires manual correction in about 17-25% of cases [1, 8, 26]. Furthermore, formant-based methods are not well suited for investigating variation in the pronunciation of consonants.

In contrast to concentrating on the acoustic signal, it is also possible to use the underlying articulatory gestures (i.e. the movement of lips and tongue, etc. needed for the production of speech [4]). As ease of articulation is one of the known factors driving linguistic change [23], this also makes sense from a diachronic perspective. Only a few studies have investigated dialect and sociolinguistic variation by focusing on the movement of the articulators. Corneau [7] applied electropalatography to compare the palatalization gestures in the production of /t/ and /d/ between Belgium French and Quebec French, while Recasens and Espinosa [21] used the same method to investigate differences in the pronunciation of fricatives and affricates in two variants of Catalan. While electropalatography only contains information about the tongue's position when it is touching the palate, ultrasound tongue imaging is able to track (most of) the shape of the tongue during the whole utterance. The sociolinguistic relevance of tracking the shape of the tongue was clearly shown by Lawson and colleagues [17], who showed that the /r/ pronunciation in Scottish English was socially stratified (with middle-class speakers generally using bunched articulations, whereas working-class speakers more frequently used tongue-tip raised variants). As a consequence, they suggest that "articulatory data are an essential component in an integrated account of socially-stratified variation".

A third method to obtain information about the tongue during speaking is electromagnetic articulography (EMA; [11, 12, 22]). EMA allows the trajectories of small sensors attached to several points in and near the mouth (i.e. on the tongue and lips) to be measured in three-dimensional space and over time. As a point-tracking technique, it is excellently suited for quantitative analysis. To our knowledge, however, this method has not yet been applied to investigate dialect variation.

In this study, we assess articulatory dialect differences between Dutch dialects using electromagnetic articulography. As there is much speaker-related variation in articulatory trajectories [32], we are studying a large group of speakers. Furthermore, we are taking an aggregate perspective by including dozens of words simultaneously, in order to investigate if there are high-level differences between the two dialects.

To analyze the articulatory data, we propose a flexible statistical approach, generalized additive modeling (GAM; [9, 30]). The advantage of using

this approach (explained in more detail below) is that it is able to model the non-linear trajectories of the tongue sensors in multiple dimensions over time, while also taking into account individual variation. Given that generalized additive modeling is a regression approach, it is excellently suited to assess the influence of the predictors of interest (in our case the contrast between the two groups) on the articulatory trajectories. Furthermore, this method has been applied successfully to analyze articulatory data in previous studies [24, 25]. In the following, we discuss the methods and results obtained in this study.

2. DATA COLLECTION

Our study was conducted on-site in 2013 at two high schools in the Netherlands. The first school was located in Ter Apel (in the northern half of the Netherlands), while the second school was located in Ubbergen (in the southern half of the Netherlands, at a distance of about 150 kilometres from Ter Apel). At each school data was collected onsite during a single week by two researchers of the University of Tübingen. In Ter Apel, 21 speakers participated (12 male, 9 female). In Ubbergen, 19 speakers (17 male, 2 female) participated. Most speakers were born between 1994 and 2000. Before participating, participants were informed about the nature of the experiment and required to sign an informed consent form. Each data collection session lasted a total of 50 minutes for which the participants were compensated with € 10.

The articulography data was collected with a portable NDI Wave 16-channel articulography device at a sampling rate of 100 Hz. Using the NDI WaveFront articulography data capturing software, the positional data was automatically synchronized with the audio signal (recorded at 22.05 kHz using an Oktava MK012 microphone) and corrected for head movement via a 6D reference sensor attached to each speaker's forehead. The microphone and the NDI Wave articulography device were connected to the control laptop via a Roland Quad-Capture USB Audio interface. To make the positional data comparable across speakers, a separate biteplate recording (containing 3 sensors) was used to rotate the data of each speaker relative to the maxillary occlusal plane [27]. We attached a total of three sensors to the midline of each speaker's tongue using PeriAcryl 90 HV dental glue. One sensor was positioned as far backward as possible without causing discomfort for the speaker. Another sensor was positioned about 0.5 cm. behind the tongue tip. The final sensor was positioned midway between the other two sensors. Besides the three tongue sensors, we glued three sensors to the lips and attached two sensors to the jaw. For the purpose of this study, however, we only focus on data from the three tongue sensors. Attaching all sensors took about 20 minutes. Whenever sensors came off during the course of the experiment, they were reattached.

During the experiment, participants had to name 70 images in their own dialect (repeated twice, in random order). To familiarize the participants with the images (such as a picture of a sheep) and to make sure they knew what each image depicted, they were asked to name each image in their local dialect once before the sensors were attached. In case the participant failed to use the correct word, he or she was helped or corrected by the experimenter.

3. PREPROCESSING AND ANALYSIS

The data for each speaker was manually segmented acoustically at the word and phone level. Tongue movement data which was not associated with the pronunciation of the study material was discarded. The duration of each word's pronunciation was time-normalized between 0 (start of the word) and 1 (end of the word) for each speaker. As the tongue sensors were attached to the midline of the tongue, we only included the position in the *z*-direction (i.e. tongue height: inferior-superior) and *x*-direction (i.e. tongue backness: anterior-posterior) in our analysis. To enable a fair comparison between speakers, the positional information was normalized for each speaker in such a way that 0 in the *z*-direction indicated the lowest (inferior) point of the three tongue sensors and 100 the highest (superior) point. Similarly, 0 in the *x*-direction indicated the most frontal (anterior) position of the three tongue sensors, while 100 in this direction indicated the position furthest back (posterior) in the mouth. These extremes were based on the pronunciation of all words by the speaker. Clear outliers were removed, and therefore not considered as the maximum or minimum point.

Since the articulatory trajectories of the individual tongue sensors are clearly non-linear, we used generalized additive modeling to analyze the data [9, 30]. Generalized additive modeling is a flexible regression approach which allows for non-linear dependencies (via so-called splines) and interactions. In our case, the dependent variable is the (normalized) position of the sensor, which we model as a non-linear pattern over (normalized) time using a thin plate regression spline [29]. To prevent overfitting of the data by the spline, generalized cross-validation is used to determine the parameters of the spline during the model-fitting process [30].

As there clearly is much variation in tongue movement associated with speakers and words, any adequate analysis will need to take this into account. Fortunately, the generalized additive modeling procedure implemented in the *R* package *mgcv* (version 1.8.2) allows for the inclusion of factor smooths to represent non-linear random effects These factor smooths are the non-linear equivalent of the combination of random intercepts and random slopes in a mixed-effects regression model. Just as random intercepts and slopes (which are essential in a model where multiple observations are present per speaker and/or word [3]), factor smooths are essential to take the structural variability associated with individual speakers and words into account and thereby prevent overconfident (i.e. too low) *p*-values in assessing the group differences.

Just as for a regular linear regression model, the residuals (i.e. the difference between the observed and the estimated values) of a generalized additive model (GAM) have to be independent. However, when analyzing time series data which are relatively smooth and slow moving (such as the movement of the tongue over time), the residuals will generally not be independent (i.e. the residuals at subsequent time points will be correlated). In our case, the autocorrelation present in the residuals is very high at a level of about 0.96. If this autocorrelation is not brought into the model, the confidence bands and *p*-values of the model will be too small. Fortunately, the function *bam* of the *mgcv* package we use is able to control for autocorrelation, enabling a more reliable assessment of the model fit and the associated *p*-values. Another important benefit of the *bam* function is that it is able to work with very large data sets [31]. This is an essential characteristic for our data set as it contains 70 words pronounced by 40 speakers, for 3 tongue sensors in 2 dimensions, with an average of 90 sampling points per word (i.e. 70 x 40 x 3 x 2 x 90 = 1.5 million data points).

4. RESULTS

As an illustration of the generalized additive modeling approach, Figure 1 shows the tongue movement differences in the oral cavity as measured by the three tongue sensors during the pronunciation of two dialect words *taarten*, 'cakes' (generally pronounced [tʊːtn] in Ter Apel and [tœʀtə] in Ubbergen), while Figure 2 shows the same visualization for the word *boor*, 'drill' (generally pronounced [bʊːr] in Ter Apel and [bʊːʀ] in Ubbergen). The red and blue dots in each graph indicate the measured tongue positions of both groups. The red curves indicate the fitted tongue trajectories of the speakers in Ubbergen for both word-specific generalized additive models, whereas the blue curves are linked to the speakers in Ter Apel. The lightness of the curve visualizes the time course from the beginning of the word (dark) to the end of the word (light). Clearly the pronunciations for *taarten* (Figure 1) are markedly different for the two groups, whereas the pronunciations for *boor* (Figure 2) are much more similar. A general pattern across both graphs, however, is that the speakers from Ubbergen appear to have more frontal tongue positions than those from Ter Apel.

Figure 1: Position of the three tongue sensors for both groups for the word *taarten*.

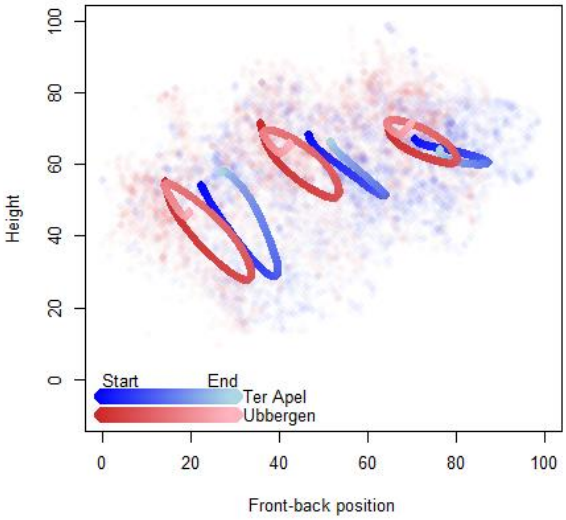
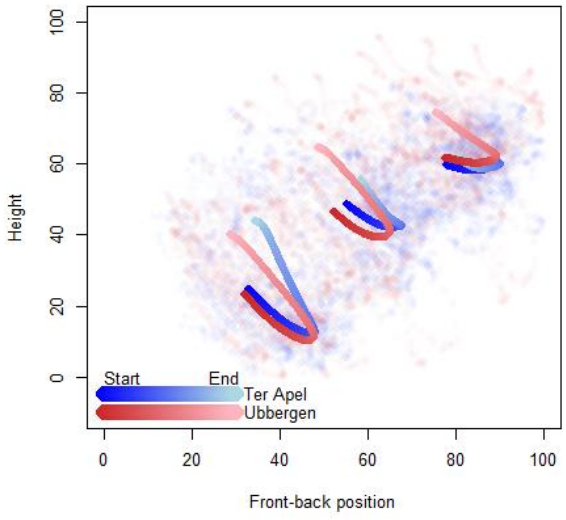


Figure 2: Position of the three tongue sensors for both groups for the word *boor*.



While it is certainly insightful to focus on the differences in the pronunciation of individual words, an aggregate analysis is able to provide a more objective view of tongue trajectory differences. For this purpose we created a large-scale GAM for the three tongue sensors and two axes simultaneously. This GAM assessed the tongue trajectory differences

between the two groups of speakers for all 70 dialect words. Besides including factor smooths to take into account the speaker-related structural variation (per sensor and axis separately), we also included factor smooths per word (per sensor and axis separately) to take into account the word-related variation. In addition, we also corrected for the autocorrelation of the residuals of the model.

Figure 3 provides a visualization comparable to Figures 1 and 2. Obviously the trajectories are less pronounced as they are the average trajectories across all words. Also at this aggregate level, however, there is a clear difference between the two groups: the trajectories of the speakers from Ubbergen are much more frontal (i.e. anterior) than those of Ter Apel.

Figure 3: Position of the three tongue sensors for both groups for all Dutch dialect words.

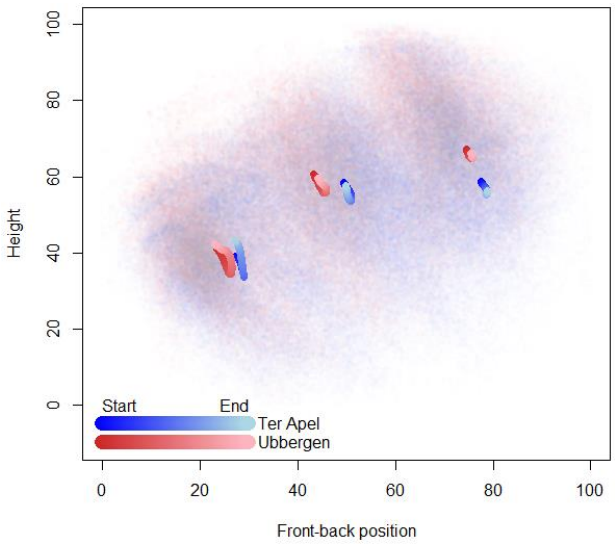
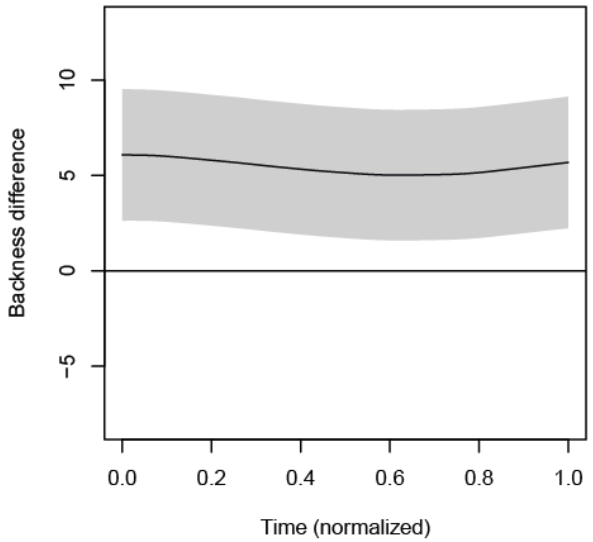


Figure 4: Aggregate backness difference between Ter Apel and Ubbergen for the T2 sensor. The difference is significant ($p < .05$).



In order to evaluate whether the signal we are seeing in the noisy articulatory data is really there, we fitted a GAM to the data with regression curves (one per sensor) for the speakers from Ter Apel, and three difference curves (one per sensor) contrasting the speakers from Ubbergen from those from Ter Apel. Such a difference curve is shown in Figure 4 and visualizes the difference and associated confidence bands for the T2 sensor. Since zero (i.e. the *x*-axis) lies outside the confidence interval of the difference curve for the full time interval, this analysis provides strong support for a measurable anterior-posterior difference between the two dialects. The pattern observed for the T2 sensor replicated for the T1 sensor, but not for the T3 (posterior) sensor, where the confidence bands overlapped with zero for the full time interval (not shown). Consequently, there appears to be a clear difference between the speakers from Ubbergen and Ter Apel with respect to the front of the tongue.

5. DISCUSSION

In this study we have illustrated the use of articulatory data for the purpose of variationist studies. We identified a structural difference in the position of the tongue during speech between the two groups of speakers, with more anterior positions of the (front of the) tongue for the speakers from Ubbergen in the southern half of the Netherlands compared to the speakers from Ter Apel in the northern half of the Netherlands. Due to the high-level analysis we employed, these results might be suggestive of a difference in articulatory settings [10] at the dialect level. However, this needs to be confirmed by focusing on the tongue position during the speakers' interspeech posture (e.g., [28]).

The generalized additive modeling approach proposed here complements other approaches used to analyze articulatory data over time, such as functional data analysis (e.g., [19]) or cross-recurrence analysis [16]. Those methods generally separate amplitude variability from phase variability when comparing articulatory trajectories. The method we propose, however, is especially suitable when articulatory trajectories need to be compared at a higher level of aggregation.

6. REFERENCES

[1] Adank, P., van Hout, R., Smits, R. 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *J. Acoust. Soc. Am.* 116, 1729-1738.

[2] Adank, P., van Hout, R., Velde, H. V. D. 2007. An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *J. Acoust. Soc. Am.* 121, 1130-1141.

[3] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390-412.
[4] Browman, C. P., Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica* 49, 155-180.
[5] Clopper, C. G., Pisoni, D. B. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *J. Phon.* 32, 111-140.
[6] Clopper, C. G., Paolillo, J. C. 2006. North American English vowels: A factor-analytic perspective. *Lit. Linguist. Comput.* 21, 445-462.
[7] Corneau, C. 2000. An EPG study of palatalization in French: Cross-dialect and inter-subject variation. *Lang. Var. Chang*. 12, 25-49.
[8] Eklund, I., Traunmüller, H. 1997. Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica* 54, 1-21.
[9] Hastie, T. J., Tibshirani, R. J. (1990). *Generalized Additive Models*. CRC Press.
[10] Honikman, B. 1964. Articulatory settings. In: Abercrombie, D., Fry, D. B., MacCarthy, P. A. D., Scott, N. C., Trim, J. L. M. (eds), *In Honour of Daniel Jones*. London: Longman, 73-84.
[11] Hoole, P., Nguyen, N. 1999. Electromagnetic articulography. In: Hardcastle, W.H., Hewlett, N. (eds), *Coarticulation: Theory, Data and Techniques.* Cambridge University Press, 260-269.
[12] Hoole, P., Zierdt, A. 2010. Five-dimensional articulography. In: Maassen, B., van Lieshout, P. (eds), *Speech Motor Control: New Developments in Basic and Applied Research*. OUP, 331-349.
[13] Labov, W. 1980. The social origins of sound change. In: Labov, W. (ed), *Locating Language in Time and Space*. New York: Academic Press, 251-266.
[14] Labov, W., Ash, S., Boberg, C. 2005. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Walter de Gruyter.
[15] Labov, W., Yaeger, M., Steiner, R. 1972. *A Quantitative Study of Sound Change in Progress.* Philadelphia: U. S. Regional Survey.Lancia, Fuchs and Tiede, 2013
[16] Lancia, L., Fuchs, S., Tiede, M. 2014. Application of concepts from cross-recurrence analysis in speech production: An overview and comparison with other nonlinear methods. *J. Speech Lang. Hear. Res*. 57, 718-733.
[17] Lawson, E., Scobbie, J.M., Stuart-Smith, J. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *J. Sociolinguist.* 15, 256-268.
[18] Leinonen, T. 2010. *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. PhD thesis, University of Groningen.
[19] Lucero, J. C., Munhall, K. G., Gracco, V. L., Ramsay, J. O. 1997. On the registration of time and the patterning of speech movements. *J. Speech Lang. Hear. Res* 40, 1111-1117.
[20] Peterson, G. E., Barney, H. L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am*. 24, 175–184.
[21] Recasens, D., Espinosa, A. 2007. An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *J. Int. Phon. Assoc.* 37, 143-172.
[22] Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., Conrad, B. 1987. Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain Lang.* 31, 26-35.
[23] Sweet, H. 1888. *History of English sounds*. Oxford: Oxford University Press.
[24] Tomaschek, F., Tucker, B. V., Wieling, M., Baayen, R. H. 2014. Vowel articulation affected by word frequency. *Proc. 10th ISSP* Cologne, 429-432.
[25] Tomaschek, F., Wieling, M., Arnold, D., Baayen, R. H. 2013. Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. *Proc. 14th Interspeech* Lyon, 1302-1306.
[26] Van der Harst, S., Van de Velde, H., van Hout, R. 2014. Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Lang. Var. Chang*. 26, 247-272.
[27] Westbury, J. R. 1994. On coordinate systems and the representation of articulatory movements. *J. Acoust. Soc. Am.* 95, 2271–2273.
[28] Wilson, I., Gick, B. 2014. Bilinguals use language-specific articulatory settings. *J. Speech Lang. Hear. Res.* 57, 361-373.
[29] Wood, S. N. 2003. Thin plate regression splines. *J. Royal Stat. Soc: Ser. B (Stat. Methodol.)* 65, 95–114.
[30] Wood, S. N. 2006. *Generalized Additive Models: An Introduction with R.* Chapman & Hall/CRC.
[31] Wood, S. N., Goude, Y., Shaw, S. 2015. Generalized additive models for large data sets. *J. Royal Stat. Soc.: Ser. C (Appl. Stat.)* 64, 139-155.
[32] Yunusova, Y., Rosenthal, J. S., Rudy, K., Baljko, M., Daskalogiannakis, J. 2012. Positional targets for lingual consonants defined using electromagnetic articulography. *J. Acoust. Soc. Am.* 132, 1027-1038.