Quantitative identification of dialect-specific articulatory settings

Martijn Wieling[1] and Mark Tiede[2]

[1]*University of Groningen, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, The Netherlands,*

[2]*Haskins Laboratories*

Abstract

The purpose of this study was to quantitatively contrast the articulatory settings of two Dutch dialects. Tongue movement data during speech was collected on site at two high schools (34 speakers) in the Netherlands using a portable electromagnetic articulography device. Comparing the tongue positions during pauses in speech between the two groups revealed a clear difference in the articulatory settings, with significantly more frontal tongue positions for the speakers from Ubbergen in the Southeast of the Netherlands compared to those from Ter Apel in the North of the Netherlands. These results provide quantitative evidence for differences in articulatory settings at the dialect level.

I. Introduction

Honikman (1964: 1) defined articulatory settings as “the overall arrangement and manoeuvring of the speech organs necessary for the facile accomplishment of natural utterance”. In her article, she noted characteristic articulatory setting differences between the English and French languages, such as the tongue being anchored laterally to the roof for the English speakers, versus anchored centrally to the floor for the French speakers. Even though Honikman (1964) gave this phenomenon the label we use today, much earlier reports of language-specific articulatory settings have been given. For example, Sweet (1890: 74) noted that “[e]very language has certain general tendencies which control its organic movements and positions, constituting its organic basis or basis of articulation”. But as Laver (1978) notes in a historical overview of the concept of articulatory settings, even as early as the 7th century, general language-specific distinctions with respect to articulation have been discerned.

The characterizations of Honikman (1964) and others before her have been qualitative in nature, that is, by describing the observed general movements of the articulators. More recently various attempts have been made to identify differences in articulatory settings quantitatively by means of acoustic analysis (see Gick et al., 2004 for an overview). Unfortunately, such an approach is complicated by the inability to separate differences in the articulatory settings from differences in segmental targets. As Laver (1978: 11) notes "no articulatory setting normally applies to every single segment a speaker utters". As a consequence, various researchers have focused on investigating the existence of language-specific resting positions during pauses in speech utterances (dubbed the "pre-speech posture" by Perkell, 1969) in order to characterize articulatory settings. As it is not possible to accomplish this through acoustic analysis, these studies necessarily investigate the position of the articulators. For this purpose, techniques such as X-ray, ultrasound, electromagnetic articulography (EMA), and real-time magnetic resonance imaging can be used (Mennen et al., 2010; Ramanarayanan et al., 2013).

Gick et al. (2004) used X-ray data (with a sample of ten speakers) to show that there were language-specific articulatory settings for English versus French speakers. They found that compared to English, French was characterized by a greater pharynx width, a lower tongue body, a lower tongue tip, a less protruded upper lip, and a more protruded lower lip (although see Wilson, 2013 for a different pattern), but that velum and jaw positions did not differ significantly between the two groups.

Using ultrasound imaging, Wilson and Gick (2014) showed that bilinguals have distinct articulatory settings for their two languages (French and English) if they are perceived as having native fluency in both languages. In their sample of eight bilingual speakers, four were rated as being native in both of their languages, while for the other four this was not the case. The speakers in the first group generally exhibited articulatory setting differences between their two native languages which were in line with the differences between English native speakers and French native speakers: lower tongue tip height (reported by Gick et al., 2014 and Wilson, 2013) and more lower lip protrusion for French speakers compared to English speakers (reported by Wilson, 2013). The speakers in the second group did not exhibit a similar pattern.

Święciński (2013), in a sample of four speakers, suggested that Polish speakers with a better command of the English language had learned to vary their articulatory settings on the basis of the language they spoke. In his study, the two speakers with the greatest command of the English language showed significant differences in the pre-speech posture (i.e., more frontal and higher tongue position) when speaking English compared to Polish, whereas no significant differences were observed for the two less proficient speakers.

While articulatory settings clearly exist, and differ for different languages, some evidence suggests that differences in articulatory settings might also be observed at the dialect level. Knowles (1973) discusses the urban dialect of Liverpool (Scouse) in terms of articulatory settings, and for example mentions on the basis of a qualitative investigation of his own speech that the Scouse dialect is characterized by more velarized speech than Received Pronunciation (pp. 102-111). Recasens (2010) shows in a sample of fifteen speakers using electropalatography (a technique to monitor contact between the tongue and palate) that distinct tongue position differences can be observed between Eastern Catalan and Valencian (with the latter being characterized by more anterior tongue positions). However, to our knowledge, no study has sought to *quantitatively* investigate the existence of distinct articulatory settings by focusing on the position of the articulators during pauses in speech utterances. While Recasens' (2010) study was certainly quantitative, it did not investigate the position of the articulators during pauses. Two studies by Stuart-Smith (1999a, 1999b) did provide a quantification of articulatory settings for different social groups (male/female, old/young) in Glaswegian (one of the results showed that children showed laxer supralaryngeal articulation than adults), but this was based on transcribing voice quality characteristics on the basis of speech, and did not involve articulatory measurements. However, her approach did enable her to identify differences in (supra)laryngeal settings.

In this study, we will extend the work on investigating differences in articulatory settings at the dialect level by focusing on the pauses during dialectal speech. We will focus on Dutch dialects, as these have been investigated frequently from a quantitative point of view (see e.g., Heeringa, 2004 and Wieling et al., 2007). This study is also distinctive for the large number of speakers included (more than thirty).

II. Articulatory data collection

For our study, articulatory data was collected on site at two high schools in the Netherlands in 2013. The two schools ("RSG Ter Apel" in Ter Apel in the North and "Havo Notre Dame des Anges" in Ubbergen located about 150 kilometers further south) are found on opposite sides of a strong dialect border in the Netherlands, distinguishing the Low Saxon dialects in the North from the Central Dutch dialects to the south of the dialect border (see Wieling et al., 2007). Figure 1 shows a map of the Netherlands in which Ter Apel is marked by a 'T', Ubbergen by a 'U' and the approximate dialect border by a dashed line. The reason these specific locations were chosen was that we had access to the students at the high schools in the two locations. While variability in articulatory movement is greater in adolescents compared to adults (but with no difference between males and females; Walsh & Smith, 2002), testing at high schools gave us access to a very motivated group of participants.

Furthermore, in our analysis (using mixed-effects regression, explained below) we take into account individual speaker variability. Finally, the presence of more variability lowers the probability of discovering differences between groups, and as a consequence our analysis becomes more conservative. At both schools data was collected during a single week. A total of 19 high school students (age at testing between 13 and 18, average year of birth 1996, two females, 17 males) participated in Ubbergen, while 15 high school students participated in Ter Apel (six females, nine males, average year of birth of 1996). [1]

We collected kinematic data from sensors attached to the speech articulators using a portable 16-channel EMA device (Wave, Northern Digital Inc.) at a sampling rate of 100 Hz, automatically synchronized to the audio signal (recorded at 22.05 kHz using an Oktava MK012 microphone). Head-correction was performed using the NDI Wavefront software on the basis of a single 6DOF reference sensor attached to the forehead. Each data collection session lasted about fifty minutes and participants gave consent and received monetary compensation for their participation. Participants were informed beforehand about the nature of the experiment. If they were younger than 18, their parents also had to sign the consent form. Participants were selected only if they spoke the local dialect, which was assessed by the first author before the experiment began. For this purpose, participants had to name images presented on a computer screen in their local dialect. Their response was compared to the expected dialect pronunciations (which were compiled beforehand by an expert on Dutch dialectology, Dr. W.J. Heeringa). If the pronunciation of the participant deviated too much from the expected pronunciation, that speaker was not subsequently included.

For the purpose of this study, we focus on the three tongue sensors which were attached midsagittaly to the tongue of each participant. The sensors were glued to the tongue with PeriAcryl 90 HV dental glue. One sensor (T3) was glued as far back as possible on the tongue without causing the speaker discomfort. The other sensor (T1) was glued approximately 0.5 cm behind the tongue tip. The third sensor (T2) was glued midway between the other two sensors. If sensors came off during the experiment, they were reattached at their original position on the tongue. Due to the purple color of the glue, this position was generally clearly visible. In order to obtain a comparable coordinate system across speakers, a biteplate recording (containing three sensors) was used to rotate the coordinates of each sensor relative to the occlusal plane (Hoole & Zierdt, 2010; Yunusova et al., 2009).

The experiment consisted of first naming 70 images (e.g., the image of a sheep, pronounced by the participant as an individual word: “sheep”) in their local dialect, and subsequently reading 27 CVC sequences (C: /t,k,p/, V: /a,i,o/) from a computer screen in standard Dutch (see Wieling et al., 2015). Both parts were repeated twice, and the items within each repetition were ordered randomly. The dialectal material was chosen in such a way that a broad overview of Dutch dialect variation was obtained. The CVC sequences contained the /t/, /k/ and /p/ in order to assess movement of the tongue (tip and back) and lips.

Figure 1. Map of the Netherlands indicating the two data collection sites, Ter Apel ('T') and Ubbergen ('U'). The dashed line indicates the approximate location of the dialect border.

III. Articulatory data preprocessing

The (rotated and head-corrected) positions of the tongue sensors were normalized along both the inferior-superior and anterior-posterior axes in such a way that 0 indicated the most inferior (or anterior) position for each sensor, and 1 the most superior (or posterior) position for each sensor.[2] Subsequently, the data for each speaker were manually segmented in PRAAT (Boersma & Weenink, 2015) on the basis of the acoustic signal. Segmentation was conducted both at the segment level as well as at the word level. For the purpose of this study, we only used the word-level segmentation. As the material consisted of the pronunciation of separate, individual words, segmentation at the word-level segmentation was relatively straightforward.

Based on this segmentation, we extracted the articulatory positions associated with the pauses in between the word pronunciations. To be sure that we only extracted positions associated with a true pause, we only considered pauses with a duration of at most 1.5

seconds (longer pauses frequently contained tongue movement associated with swallowing, yawning, or it contained a mispronunciation of a word). There was no lower limit, as due to the setup of the experiment there always was a pause between two succeeding pronounced words. From these pauses, we extracted the articulatory positions over an interval between 0.75 seconds and 0.25 seconds before the start of each pronounced word. If the time between two consecutive words was less than 1 second, the extracted portion of the pause extended from 0.25 seconds after the end of the first word to 0.25 seconds before the start of the second word. If the time between two consecutive words was less than 0.5 seconds, the pause was ignored. The 0.25 second gaps were used as the segmentation was done acoustically and residual articulatory movement can still be present close to the acoustic start or end of a word. Consequently the extracted portion of the pause was at most 0.5 seconds (when the time between two consecutive words ranged between 1 and 1.5 seconds), but could be of shorter duration as well (when the time between two consecutive words was less than 1 second). Note, however, that results remained similar if the extracted portion of the pause was not limited to at most 0.5 seconds, but always ranged from 0.25 seconds after the end of the first word to 0.25 seconds before the start of the second word. The median extracted pause duration was 0.36 seconds (i.e. about 36 measurement points, as the sampling rate was 100 Hz) with an inter-quartile range of 0.24 seconds. About 35% of the pauses had the maximum duration of 0.5 seconds. For each individual pause, the median position for each sensor (T1, T2 and T3) and axis (inferior-superior: z-axis and anterior-posterior: x-axis) over the pause interval was calculated.

IV. Analysis

Our data contains normalized sensor positions in two dimensions for three tongue sensors during approximately 200 pauses per participant, and accordingly we analyzed the data using mixed-effects regression. By using this approach, we are able to take into account the structural variability associated with each individual speaker. For example, the positions of the sensors during the individual pauses were relatively similar for each individual speaker (the average inter-quartile range of the resting positions was 2.6 mm). By using random intercepts (some speakers may have a more frontal pre-speech posture than others) and random slopes (some speakers may show a different pre-speech posture for dialectal vs. standard speech, while others do not), we were able to model the variability associated with each individual, thereby reducing the risk of being overconfident (i.e., reporting *p*-values which are too low). An overview of the merits of mixed-effects regression is given by Baayen et al. (2008). We only included random intercepts and random slopes whenever model comparison using the Akaike Information Criterion (AIC; Akaike, 1974) indicated that the additional complexity was warranted (i.e., resulting in a lower AIC value of at least 2; following the approach of Wieling et al., 2014).

V. Results

We fitted two separate mixed-effects regression models, one for each axis, with as dependent variable the normalized position for each of the three sensors. The model fit on the basis of the inferior-superior position did not show a tongue height difference between the two groups, neither with nor without taking into account (the interaction with) the type of speech (all $|t|$'s < 1.3, p's $> .19$). For completeness, Table 1 shows the fixed effects of this full model including the interaction The random-effects structure consisted of random intercepts for speaker and pause (i.e. linked to the following word), and random slopes for the group differences (Ubbergen versus Ter Apel) per pause and the type of speech (dialect versus standard Dutch CVC sequences per speaker).

The effect size (Ω_0^2, similar to R^2; Xu, 2003) of this full model was 0.25, and the residuals approximately followed a normal distribution. Figure 2 (left) visualizes the non-significant difference between the two groups. In contrast, the model fit on the basis of the anterior-posterior position of the tongue sensors showed a clear significant (fixed effect) difference with respect to the tongue frontness between the two groups ($t = -4.3$, $p < .001$). The speakers from Ubbergen had more frontal tongue sensor positions than the speakers from Ter Apel in the North of the Netherlands (the size of this effect is 9 percent of the total range of the sensors). The random-effects structure of this model was identical to the aforementioned model focusing on the inferior-superior position. The effect size (Ω_0^2) of this optimal model was 0.37, and the residuals approximately followed a normal distribution. There was no significant effect (nor any significant interaction) with the type of speech. Table 2 shows the model summary of the best model (i.e., including only the significant group difference). Figure 2 (right) visualizes the significant difference between the two groups.[3]

Predictor	Estimate	Std. Error	*t*-value	*p*-value
(Intercept)	0.605	0.015	39.47	< .001
Ubbergen vs. Ter Apel	0.026	0.020	1.29	.19
Standard vs. Dialect (Ter Apel)	0.012	0.013	0.91	.36
Standard vs. Dialect (Ubbergen)	0.008	0.012	0.66	.51

Table 1. Mixed-effects regression model for the inferior-superior (*z*) axis. No significant difference between the two dialect groups was observed. Only fixed effects are shown; see text for the random-effects specification.

Predictor	Estimate	Std. Error	*t*-value	*p*-value
(Intercept)	0.438	0.016	27.54	< .001
Ubbergen vs. Ter Apel	-0.090	0.021	-4.26	< .001

Table 2. Mixed-effects regression model for the anterior-posterior (*x*) axis. The difference between the two groups was significant. Only fixed effects are shown; see text for the random-effects specification.

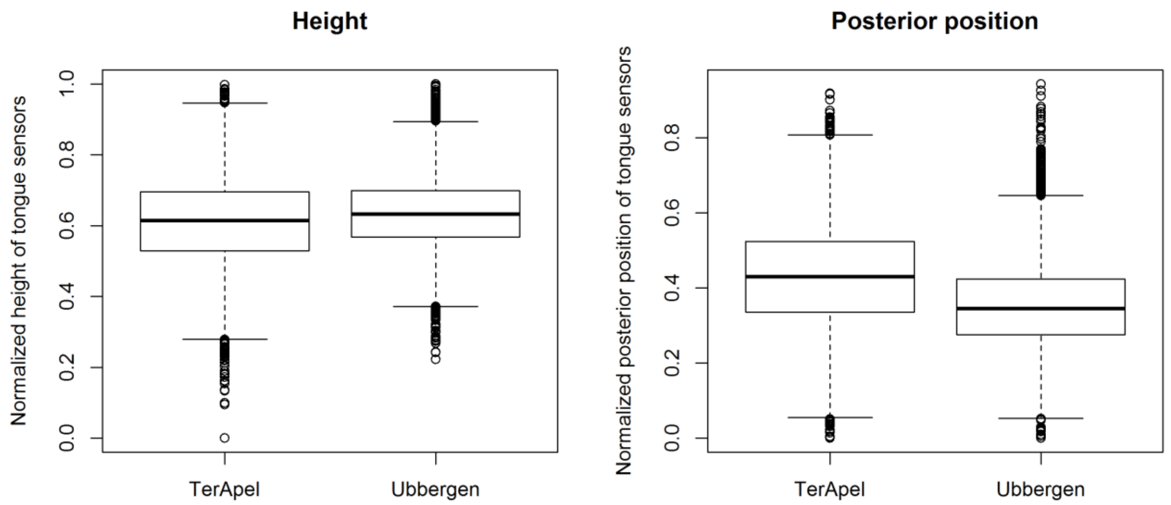

Figure 2. Visualization of the non-significant height difference (left) and the significant posterior position difference between the two groups (right). Larger *y*-values indicate higher (left) or more posterior (right) normalized tongue sensor positions.

VI. Discussion

The quantitative results obtained in this study suggest a distinct pre-speech tongue posture difference between two Dutch dialects, which is present both when the speakers speak in their local dialect and when they speak (accented) standard Dutch. The Low-Saxon dialect from Ter Apel in the North of the Netherlands seems to be characterized by a tongue position which is located further back in the mouth than that of the Central Dutch dialect of Ubbergen. Various studies have quantified differences in articulatory settings between different languages, but this is – to our knowledge – the first study which has done the same for different dialects of the same language on the basis of the tongue position during pauses in speech.

While no previous studies have investigated articulatory settings in Dutch dialects, a few studies have investigated variation in the Dutch language using articulatory measurements. For example, Scobbie and Sebregts (2010) investigated variation in Dutch /r/ using ultrasound recordings. However, as they only included five speakers, their results remained rather qualitative. Furthermore, a single-segment study is unsuitable to shed light on differences in articulatory settings, as differences in articulation of the specific segment and articulatory setting differences cannot be distinguished. Interestingly, Wieling et al. (2015; forthcoming) conducted an articulatory analysis of the tongue movement data associated with the word pronunciations (as opposed to the data associated with the pauses analyzed in this study) of the experiment explained above, and found a similar pattern as reported in the present study, with a more posterior tongue position for the speakers from Ter Apel compared to those from Ubbergen. This suggests that articulatory setting differences may also be observed when analyzing a sizeable amount of variable speech data (i.e., not only focusing on a single segment).

Adank et al. (2007) investigated regional Dutch variation from an acoustic perspective, focusing on formant measurements of the vowels. While there certainly is no one-to-one correspondence between formants and tongue position, tongue positions and formant frequencies do correlate (Lee et al., 2016). As Adank et al. (2007) did not identify a clear first or second formant difference between the speakers from the North versus those from the Central Dutch area, we might have expected the absence of an articulatory difference between the two groups as well. However, using formant measurements is less sensitive than using tongue position information, and also restricts the analysis to vowels. Consequently, it is unclear to what extent our results would be expected to match those of Adank et al. (2007).

Of course our study is not without its limitations. First of all, only two dialects were investigated, and it is not clear in what way the difference in articulatory settings can be generalized to other dialects in the Netherlands. In future work, we will investigate if the pattern indeed holds for other dialects in the same regions in the Netherlands, and if other patterns may be identified as well. Second, while we have not found a clear difference in articulatory settings between the two types of speech (dialect vs. standard), this might have been caused by the characteristics of the speech stimuli (naming images vs. reading specific CVC sequences).

Another limitation is methodological. Even though we have attempted to ensure that we only included real pauses in our data, it is possible that the data we included might have contained tongue movement due to, for example, swallowing or articulatory movements associated with the pronunciation of the preceding or subsequent words. As a consequence, we have attempted to alleviate these potential problems by taking the median of the positions during the pauses, and by including all pauses separately in the analysis, rather than averaging them.

In conclusion, our study has provided quantitative evidence for differences in the articulatory settings between two dialects of the same language. The existence of such differences at the dialect level is in line with characterizations of dialects in terms of articulatory settings by (e.g.,) Knowles (1973) and Stuart-Smith (1999).

---

[1] While the asymmetry in the gender distribution across the two groups might be problematic, the results were similar when only male speakers were included in the analysis.

[2] Note that other normalization choices could have been made. However, two alternative normalization procedures showed a similar pattern of the results, and consequently our results seem relatively independent of the choice of normalization procedure. The first alternative was to normalize all three tongue sensors simultaneously, with 0 indicating the most anterior position of all three tongue sensors and 1 indicating the most

posterior position of all three tongue sensors. The second normalization procedure was similar to the method used currently, but additionally took into account the non-speech resting position (i.e. the position of the tongue recorded when the participants were asked to keep their mouth closed). Using this system, negative values indicated more anterior (or lower) positions of the sensor compared to its non-speech resting position, whereas positive values indicated more posterior (or higher) positions of the sensor compared to its non-speech resting position.

[3] As indicated earlier, the results using a different normalization procedure were generally quite similar, with significantly more posterior positions (and no difference in height) of the tongue sensors for the speakers from Ter Apel compared to those from Ubbergen. The only difference was that when normalizing all three tongue sensors simultaneously, there appeared to be a significant interaction with the type of speech (with the group difference being significantly larger for the CVC sequences). However, this interaction was not observed using the two other normalization procedures.

References

Adank, P., van Hout, R., & Van de Velde, H. (2007). An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *The Journal of the Acoustical Society of America*, 121(2), 1130-1141.

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.

Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 5.3.40, retrieved 5 February 2013 from http://www.fon.hum.uva.nl/praat/.

Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, 61(4), 220-233.

Heeringa, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Dissertation. University of Groningen, 315 p.

Honikman, B. (1964). Articulatory Settings. In: Abercrombie, D. et al. (eds.) *In Honour of Daniel Jones*. London: Longmans, pp. 73-84.

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*. Oxford: OUP, pp. 331-349.

Knowles, G. O. (1973). *Scouse: the urban dialect of Liverpool*. Dissertation. University of Leeds, 406 p.

Laver, J. (1978). The concept of articulatory settings: an historical survey. *Historiographia Linguistica*, 5(1-2), 1-14.

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America*, 139(1), 426-440.

Mennen, I., Scobbie, J. M., de Leeuw, E., Schaeffler, S., & Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research*, 26(1), 13-41.

Perkell, J.S. (1969). *Physiology of speech production: results and implications of a quantitative cineradiographic study*. Cambridge: MIT Press, 120 p.

Ramanarayanan, V., Goldstein, L., Byrd, D., & Narayanan, S. S. (2013). An investigation of articulatory setting using real-time magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 134(1), 510-519.

Recasens, D. (2010). Differences in base of articulation for consonants among Catalan dialects. *Phonetica*, 67(4), 201-218.

Scobbie, J.M. & K. Sebregts (2010). Acoustic, articulatory and phonological perspectives on allophonic variation of /r/ in Dutch. In: Folli, R. & C. Ulbrich (eds.) *Interfaces in Linguistics: New Research Perspectives*. Oxford: Oxford University Press, pp. 257-277.

Stuart-Smith, J. (1999a). Glasgow: Accent and voice quality. In: Foulkes, F. & Docherty, G. (eds.) *Urban voices: Accent studies in the British Isles*. Routledge, pp. 203-222.

Stuart-Smith, J. (1999b). Voice quality in Glaswegian. *Proceedings of ICPhS99*, pp. 2553-2556.

Sweet, H. (1890). *A Primer of Phonetics*. Oxford: Clarendon Press, 113 p.

Święciński, R. (2013). An EMA study of articulatory settings in Polish speakers of English. In: Waniek-Klimczak, E. & Shockey, L.R. (eds.) *Teaching and researching English accents in native and non-native speakers*. Berlin: Springer, pp. 73-82.

Walsh, B., & Smith, A. (2002). Articulatory Movements in Adolescents. Evidence for Protracted Development of Speech Motor Control Processes. *Journal of Speech, Language, and Hearing Research*, 45(6), 1119-1133.

Wieling, M., Heeringa, W., & Nerbonne, J. (2007). An aggregate analysis of pronunciation in the Goeman-Taeldeman-Van Reenen-Project data. *Taal en Tongval*, 59(1), 84-116.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, 90(3), 669-692.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., & Baayen, R. H. (2015). Investigating dialectal differences using articulography. Proceedings of ICPhS 2015, Glasgow, August 10-14. Paper number 0231.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S.N., Baayen, R. H. (forthcoming). Investigating dialectal differences using articulography. *Journal of Phonetics*.

Wilson, I. (2013). Articulatory settings of French and English monolinguals. In: Ooigawa, T. (ed.), *Sophia University Working Papers in Phonetics*. Tokyo, Japan: Sophia University, 39-58.

Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, 57(2), 361-373.

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22), 3527-3541.

Yunusova, Y., Green, J. R., & Mefferd, A. (2009). Accuracy assessment for AG500, Electromagnetic Articulograph. *Journal of Speech Language and Hearing Research*, 52, 547-555.