Investigating dialectal differences using articulography

Martijn Wieling[a,], Fabian Tomaschek[b], Denis Arnold[b], Mark Tiede[c], Franziska Bröker[b], Samuel Thiele[b], Simon N. Wood[d], and R. Harald Baayen[b,e]*

[a]Department of Humanities Computing, University of Groningen, [b]Department of Quantitative Linguistics, University of Tübingen, [c]Haskins Laboratories, [d]Department of Statistics, University of Bath, [e]Department of Linguistics, University of Alberta

**Corresponding author: Martijn Wieling, Oude Kijk in 't Jatstraat, 9712 EK Groningen, Netherlands, +31503635979, wieling@gmail.com*

Abstract
The present study uses electromagnetic articulography, by which the position of tongue and lips during speech is measured, for the study of dialect variation. By using generalized additive modeling to analyze the articulatory trajectories, we are able to reliably detect aggregate group differences, while simultaneously taking into account the individual variation of dozens of speakers. Our results show that two Dutch dialects show clear differences in their articulatory settings, with generally a more anterior 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. A comparison with formant-based acoustic measurements further reveals that articulography is able to reveal interesting structural articulatory differences between dialects which are not visible when only focusing on the acoustic signal.

Keywords: Articulography; Dialectology; Generalized additive modeling; Articulatory settings

Introduction

At present, most studies in dialectology and sociolinguistics investigating pronunciation variation focus on the acoustic properties of vowels (e.g., Clopper & Pisoni, 2004; Labov, 1980; Leinonen, 2010; Recasens & Espinosa, 2005; Adank et al., 2007; Van der Harst et al., 2014). Since the seminal study of Peterson & Barney (1952), formant measurements have been the method of choice for measuring vowel quality. While the first and second formant are generally assumed to model height and frontness of the tongue body, this relationship is far from perfect (Rosner and Pickering, 1994). For example, an increase in F2 can be caused by a more anterior tongue position, but also by a decrease in lip rounding or tongue body shape (Lindblom & Sundberg, 1971; Harrington et al., 2011).

Labov et al. (1972) have spearheaded the formant-based approach in sociolinguistics by studying English formant-based vowel variation for a large number of speakers from various areas in the United States of America. Since then many other studies assessing dialect variation have used formant-based methods. For example Adank et al. (2007) investigated regional Dutch dialect variation, and both Clopper and Paolillo (2006) and Labov et al. (2005) studied American English regional variation. While formant-based measures provide a convenient quantification of the acoustic signal, the approach is not without its problems. First, since the shape of the vocal tract influences the formant frequencies (e.g., women generally have higher formant frequencies than men), some kind of normalization is required (see Adank et al., 2004 for an overview of various approaches) and choosing one method over another introduces a degree of subjectivity into the analysis. Furthermore, automatic formant detection is imperfect and requires manual correction in about 17-25% of the cases (Adank et al., 2004; Eklund & Traunmüller, 1997; Van der Harst et al., 2014). Especially when using multiple formant measurement points per vowel (which is arguably better than using only the mid-point of the vowel; see Van der Harst et al., 2014), this becomes very time-consuming. For this reason whole-spectrum methods (obtained by band-pass filtering the complete acoustic signal) have also been used in language variation research. In her dissertation, Leinonen (2010) studied Swedish dialect variation based on the automatic whole-spectrum analysis of Swedish vowel pronunciations. A drawback of this type of analysis, however, is that it is highly sensitive to the amount of noise in the acoustic recordings (Leinonen, 2010, p. 152). Furthermore, both formant-based and whole-spectrum-based methods are not suitable for investigating variation in the pronunciation of consonants.

Another approach to investigating pronunciation variation is the use of transcriptions to describe the pronunciation of a speaker. By using transcriptions, a representative encoding of the

impression of the acoustic signal is obtained which can be used to assess pronunciation differences between groups of speakers. Even though "[t]ranscription is a messy thing" (Kerswill & Wright, 1990, p. 273), transcriptions are frequently used in dialectometry where aggregate analyses based on a large set of linguistic items are instrumental for obtaining an objective view of dialectal variation and its social, geographical and lexical determinants (see Wieling and Nerbonne, 2015 for an overview). A clear advantage of using transcriptions is that they are excellently suited for a quantitative analysis (see, e.g., Wieling et al., 2012). A drawback of using transcriptions is that the speech signal is segmented into discrete units, which means that fine-grained subphonemic (phonetic) differences, such as co-articulation effects, are frequently ignored (as these are less reliably transcribed; Goeman, 1999, p.35). In addition, reduced word forms may be reconstructed automatically by human listeners, effectively interpolating sounds which are not present in the acoustic signal (Kemps et al., 2004), and this may affect transcription quality as well. Of course, for a careful phonetic analysis, a narrow transcription is necessary. For example, Sebregts (2015) distinguished many different pronunciations of /r/ by several hundred Dutch speakers through a careful phonetic analysis.

Instead of focusing on transcriptions based on the acoustic signal, it is also possible to examine the articulatory gestures underlying speech (i.e. the movement of lips and tongue, etc. involved in its production; Browman and Goldstein, 1992). Given that ease of articulation is important for linguistic change (Sweet, 1888; see also Sebregts, 2015, Ch. 7.3.3), this also makes sense from a diachronic perspective. Furthermore, focusing on the articulatory gestures will provide more details about the pronunciations than can be identified on the basis of the (discrete) transcriptions. Only a limited number of studies have investigated dialect and sociolinguistic variation by focusing on the movement of the speech articulators. Most of these studies have employed either electropalatography (EPG) or ultrasound tongue imaging. With EPG, the contact between the tongue and the hard palate is monitored with a custom-made speaker-specific artificial palate containing several electrodes. Corneau (2000) applied this method to compare the palatalization gestures in the production of /t/ and /d/ between Belgium French and Québec French, and Recasens and Espinosa (2007) used it to investigate differences in the pronunciation of fricatives and affricates in two variants of Catalan. While EPG only contains information about the tongue's position when it is touching the palate, ultrasound tongue imaging is able to track most of the tongue surface as it moves during the whole utterance. The sociolinguistic relevance of tracking the shape of the tongue was clearly shown by Lawson et al. (2011), who demonstrated that /r/ pronunciation in Scottish English was socially stratified, with middle-class speakers generally using bunched articulations, while working-class speakers more frequently used tongue-tip raised variants. Consequently, Lawson et al. (2011, p. 257) suggest that "articulatory data are an essential component in an integrated account of socially-stratified variation".

There are some drawbacks associated with the two articulatory observational methods described above. The clear drawback of EPG is that it is very costly, as a custom-made artificial palate needs to be constructed for each participant. In addition, EPG does not yield information about the tongue position when it is not touching the palate. While ultrasound tongue imaging does provide this information, it is not always complete as interposed sublingual air pockets are introduced when the tongue is raised or extended, and shadowing from the mandible and hyoid bones may cause the tongue tip and the tongue root to become invisible (Tabain, 2013). Furthermore, analysis of resulting tongue shapes can be impressionistic, as tracking a single flesh point on the tongue is not possible (Lawson et al., 2011; but see Davidson, 2006). Moreover, unless otherwise corrected (cf. Whalen et al. 2005), the imaged tongue shape is relative to the position of the probe and jaw, not to palatal hard structure, and thus evaluation of tongue height across vowels is problematic.

Electromagnetic articulography (EMA; Hoole and Nguyen, 1999; Perkell et al., 1992; Schönle et al., 1987) is a point-tracking approach and therefore distinct from the two methods above. An EMA device tracks as a function of time small sensors attached with dental adhesive to various flesh points associated with the speech articulators. Radio-frequency transmitters induce voltages in the sensor coils positioned within the field of the device, and sensor position and orientation are subsequently reconstructed by comparing these voltages to known reference values. With good spatial (< 0.5 mm) and temporal (100 Hz) tracking resolution, it is well suited for quantitative analysis because the resulting trajectories are amenable to established statistical approaches. Of course, EMA has drawbacks as well. Because the sensors are monitored through wires, attachment is possible only in the anterior third of the vocal tract. Although speakers readily adapt to speech with attached sensors

they nonetheless constitute a potential perturbation of normal speech, and in particular to minimize such perturbation the tongue tip is tracked indirectly, through sensor placement behind the true apex. Tongue sensor placement introduces variability, as the relative placement of each sensor will not be the same for each speaker given individual differences in speaker morphology. And while current EMA systems support spatial tracking in 3D and can thus in principle track parasagittal movement, in practice sensors are typically placed only midsagittaly. In sum, all approaches have their own advantages and disadvantages. In this study we opted to use EMA in order to track the position of three sensors attached midsagittaly to the tongue.

Until recently, EMA dialectal studies have been conducted with a relatively small number of speakers (e.g., Recasens and Espinosa, 2009: three speakers). Because there is much speaker-related variation in articulatory trajectories (Yunusova et al., 2012), it is fortunate that due to technical advancements including a larger number of participants is becoming increasingly common (e.g., Yunusova et al., 2012: 19 speakers; Koos et al., 2013: 25 speakers). In our study, we continue this development by including 34 speakers. To our knowledge, this is the largest sample size used in an articulography study to date.

In this study, we focus on Dutch pronunciation variation from an aggregate articulatory perspective. Only very few published studies have investigated variation in the Dutch language from an articulatory perspective. Scobbie and Sebregts (2010) focused on investigating a single feature, namely allophonic Dutch variation in the pronunciation of /r/ using ultrasound recordings. However, due to the low number of speakers (five) and the ultrasound approach, the description of the results remained rather impressionistic. Ooijevaar (2015) investigates variation in Dutch liquids using ultrasound tongue imaging, while Strycharczuk & Sebregts (2015) use the same technique to investigate /r/-allophony. Another study (Chan et al., 1995) collected laryngograph recordings for a total of nine Dutch speakers, but did not quantify the results as it was part of a large data collection project (EUROM.1). Finally, one clinical study has used EMA to investigate Dutch speaking children with developmental apraxia of speech (Nijland et al., 2004) in a sample of three children (plus three healthy controls).[1]

Of course, many studies have investigated pronunciation variation in Dutch dialects from various other perspectives. For example, as mentioned above, Adank et al. (2007) investigated the acoustic properties of vowels in several regional varieties of Dutch spoken in the Netherlands and Flanders. They observed clear regional variation in the formant-based measurements. Another type of study focusing on Dutch dialects is exemplified by Goeman (1999), who investigated a specific feature in Dutch dialects, namely the loss of [t] in final word pronunciation (i.e. t-deletion). He identified several (geographical constrained) groups within the Netherlands exhibiting specific t-deletion patterns. Following Nerbonne et al. (1996), Heeringa (2004) took an aggregate dialectometric perspective and quantified pronunciation differences by focusing on the transcriptions and comparing those using the edit distance measure. On the basis of comparing hundreds of words between hundreds of locations in the Dutch-speaking language area, he was able to identify the major dialect areas of the Netherlands. In his dissertation (Figure 9.7, p. 234), he identified the four main dialect areas as the Frisian dialect area (in the northwest of the Netherlands), the Limburg dialect area (in the southeast of the Netherlands), the Low-Saxon dialect area (in the northeast of the Netherlands) and the Central Dutch dialect area. Similarly, Wieling et al. (2007, 2011) identified relatively comparable dialect areas using a different dataset of Dutch dialect transcriptions.

As articulatory data (in the sense of lingual instrumentation) is not readily available for Dutch dialects, we collected dialect (and standard Dutch) pronunciations at two different sites. To ensure the

[1] Additionally, there is one conference proceedings paper investigating Dutch pronunciation variation from an aggregate articulatory perspective (Wieling et al., 2015). However, the present study is an extended version of that study, and offers a more detailed report of the methods and results presented by Wieling et al. (2015). In addition, this study does not only focus on dialect variation, but also on variation in standard Dutch. Note that the results presented here are slightly different from those discussed by Wieling et al. (2015), as in the present study a subset of the data (i.e. only young speakers) was analyzed using an improved version of the generalized additive modeling software. Furthermore, in this study we also controlled for the non-speech resting position of the sensors.

dialects were not too similar, we collected our data at one site in the Low-Saxon dialect area (i.e. the village of Ter Apel), and at another site in the Central Dutch dialect area (i.e. the village of Ubbergen).

Given that the goal of this study is to assess articulatory (dialect) pronunciation differences from an aggregate perspective, we include many participants and items. In addition, we propose a flexible statistical approach, generalized additive modeling (GAM; Hastie and Tibshirani, 1990; Wood, 2006) for analyzing articulography data. The advantage of using this approach (explained in more detail below) is that it is able to model the nonlinear trajectories of the tongue sensors in multiple dimensions over time, while also taking into account individual variation. As 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.

Given that the generalized additive modeling technique is relatively new, we also provide a more frequently used approach to analyze this type of data, namely linear discriminant analysis. Furthermore, we will contrast the articulatory results to those on the basis of traditional formant analysis. This will allow us to investigate the potential differences between the two perspectives. While we certainly expect articulatory differences between the two groups of speakers due to their different dialect background, we do not have a clear hypothesis about the specific characteristics of these differences. In that sense, our study is exploratory. In the following, we will discuss the methods and results obtained in this study.

Articulatory data collection

Our study was conducted on-site in 2013 at two high schools in the Netherlands. The first school "RSG Ter Apel" was located in Ter Apel (in the northern half of the Netherlands, i.e. in the Low Saxon dialect area), while the second school "HAVO Notre Dame des Anges" was located in Ubbergen (in the southern half of the Netherlands, at a distance of about 150 kilometers from Ter Apel, i.e. in the Central Dutch dialect area). Figure 1 shows the location of both data collection sites. The approximate location of the dialect border distinguishing the Low Saxon Dialect area from the Central Dutch dialect area is indicated by a red dashed line. At each school data were collected onsite during a single week by two researchers of the University of Tübingen (MW and DA in Ter Apel and MW and FT in Ubbergen; at both sites, MW attached all sensors). In Ter Apel, 23 speakers participated, but the data of two speakers was excluded as it contained tracking inconsistencies due to a malfunction of the reference sensor. Furthermore, we excluded the data of six adult participants (born between 1939 and 1967) in Ter Apel, as the remaining participants in both locations were children born between 1994 and 2000 (no adults participated in Ubbergen). Of the remaining 15 speakers six were female and nine male with an average year of birth of 1996;6 (average age 16;6). In Ubbergen, 25 high school students participated, but the data of six speakers was excluded (five speakers did not speak the regional dialect, and the reference sensor malfunctioned for one speaker). The remaining 19 participants (17 male, two female[2]) were born between 1994 and 2000 with an average year of birth of 1996;6 (average age 16;6). Before participating, participants were informed about the nature of the experiment and required to sign the informed consent form (if participants were under 18, their parents had to sign an informed consent form as well). Each data collection session lasted a total of 50 minutes for which the participants were financially compensated.

The EMA data were collected with a portable 16-channel device (WAVE, Northern Digital Inc.) at a sampling rate of 100 Hz, and automatically synchronized to the audio signal (recorded at 22.05 kHz using an Oktava MK012 microphone) by the controlling software (WaveFront, Northern Digital Inc.). This software also corrected for head movement using a 6DOF reference sensor attached to each participant's forehead. The microphone and EMA device were connected to the controlling laptop via a Roland Quad-Capture USB Audio interface.

We attached three sensors to the midline of each participant's tongue using PeriAcryl 90 HV dental glue. One sensor (T3) was positioned as far backward as possible without causing discomfort for the speaker. Another sensor (T1) was positioned about 0.5 cm behind the tongue tip. The

[2] As the gender distribution across the two groups was unbalanced, we ran an additional analysis focusing only on the male speakers. As this analysis revealed the same pattern which was observed for the whole group, we included all young speakers in the analysis reported in this paper.

remaining tongue sensor (T2) was positioned approximately midway between the other two sensors.[3] The average absolute distance between the front and the back sensor was about 24 mm, and did not differ significantly between the two groups. Attaching all sensors took about 20 minutes. Whenever sensors came off during the course of the experiment, they were reattached at their original location. To align the positional data to axes comparable between speakers, a separate biteplate recording (containing 3 sensors, see Figure 2) was used during processing to rotate the data of each speaker relative to the occlusal plane (Hoole & Zierdt, 2010; Yunusova et al., 2009) and to translate to a common origin on the biteplate ('X' in Figure 2; note that this origin does not influence the normalized sensor positions, due to our preprocessing steps outlined below).

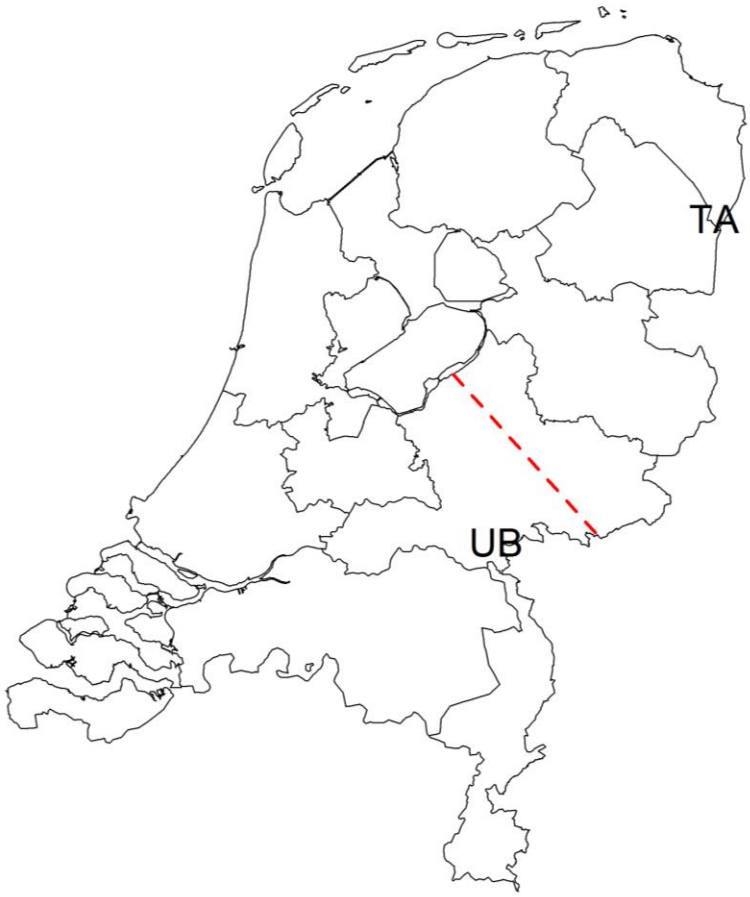

Figure 1. Location of the two data collection sites (TA: Ter Apel, UB: Ubbergen) in the Netherlands. The red dashed line shows the approximate dialect border between the Low Saxon dialect area and the Central Dutch dialect area.

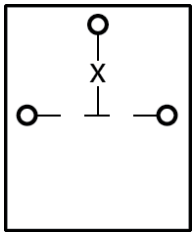

Figure 2. Schematic representation of the biteplate. Circles mark the sensor positions. The 'X' marks the origin.

[3] Besides the three tongue sensors, we also 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.

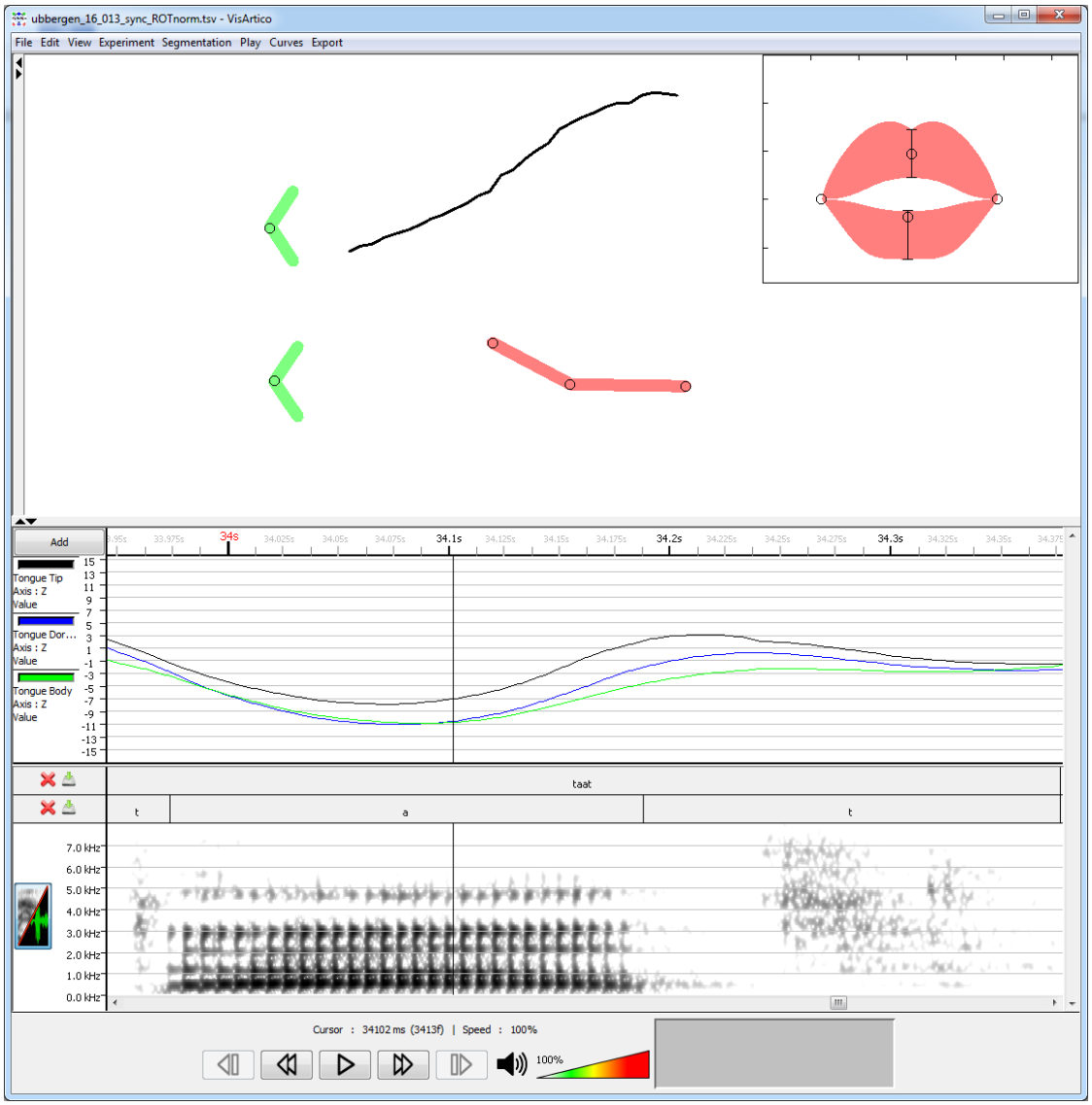

Figure 3. Visualization using VisArtico (Ouni, 2012) of the type of data collected. The top-right inset shows a frontal view of estimated lip posture on the basis of two sensors placed midsagittaly at the vermillion border. The top-left part shows a schematic representation of a midsagittal view of the two lip sensors (in green) and the three tongue sensors (in red). An approximation of the palate of the speaker is also shown in black. Directly below this visualization are the vertical trajectories in the inferior-superior dimension for the three tongue sensors during the pronunciation of the standard Dutch CVC sequence *taat*, [tat]. Below those trajectories, the segmentation and the spectral plot is shown.

The experiment was divided into two parts. In the first part, participants had to name 70 images (e.g., the image of a ball) in their own dialect (repeated twice, in random order), presented on a computer screen. To familiarize the participants with the images 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 corrected by the experimenter. Table 1 shows the 70 words with the transcriptions of the approximate pronunciations in the two dialects. The five words which have the same approximate pronunciation in both dialects are marked in bold face in the table (as these are used in a validation analysis, explained below). In the second part, participants had to read 27 CVC sequences out loud (C: /t,k,p/, V: /a,i,o/, e.g., [tap]) in *standard Dutch* (this was emphasized during the explanation of this part). Importantly, students are familiar with the standard Dutch language, as it is the language in which they are taught at school. Again, each item was pronounced twice and in randomized order. By including both standard Dutch pronunciations and dialect words, we are able to evaluate if common tongue movement trajectories can be observed in both types of speech. A visual impression of the data obtained can be seen in Figure 3.

Word	Ter Apel	Ubbergen	Word	Ter Apel	Ubbergen
bal	baːl	bɑl	molen	møln	møːlə
ballen	baːln	bɑlə	muggen	mœgŋ	mʏjə
been	bʌin	beˑn	negen	negn	neˑjə
beer	bɪːr	bɪˑʀ	ogen	ogŋ	oˑjə
bel	bɛːl	bɛl	oog	oːx	oːç
bellen	bɛːln	bɛlə	oor	ʊːr	ɔːʀ
bier	biːr	biˑʀ	paal	pʊːl	pɒːl
biet	bʌit	biˑt	paarden	pɪːrdn	pɪˑʀdə
bijl	biːl	bil	palen	pʊːln	pɒːlə
bijlen	biːln	bilə	peren	pɪːrn	pɪːʀə
blauw	blaːu	blʌu	riem	rʌim	ʀiˑm
bloemkool	blʌumkoul	blumkoˑl	roos	rous	ʀoːs
bogen	bougŋ	boˑçə	schaap	sxʊːpʰ	sçɒːp
boor	bʊːr	bʊːʀ	schaar	sxɪːr	sÇɪːʀ
boot	boutʰ	boˑt	**speen**	**speˑn**	**speˑn**
bril	brɪl	bʀɪl	speer	spɪːr	spɪːʀ
brillen	brɪln	bʀɪlə	step	stɛbə	stɛp
deuk	døːk	dœːk	ster	stʌrə	stɛʀ
deur	dʏːr	dœˑʀ	stoel	stʌulə	stul
dolfijn	dɔlviˑn	dɔlfɛːn	stoelen	stʌuln	stulə
fruit	frʌytʰ	fʀʌːt	taarten	tʊːtn	tœʀtə
geit	xɛit	çɛːt	**tol**	**tɔl**	**tɔl**
geld	xɛːltʰ	çɛlt	tollen	tɔln	tɔlə
harp	haːp	hɑʀp	tor	tɔrə	tɔʀ
kameel	**kəmeːl**	**kəmeːl**	treinen	trʌinn	tʀɛˑnə
kamelen	kəmeːln	kəmeˑlə	uil	uːlə	yːl
kar	kaːrə	kɛʀ	uilen	uln	yˑlə
kat	kat	kɑt	vingers	vɪŋərs	fɪŋəʀs
kersen	kaːzn	kɛʀsə	violen	vioːln	fioˑlə
kruk	krœgə	kʀɛk	vlaggen	vlagŋ	flɑçə
krukken	krœgŋ	kʀɛkə	vliegtuigen	vlʌixtyːgŋ	fliçtyˑjə
lamp	laːmpʰ	lɑmp	vogels	voɣəls	foˑjəls
leeuw	**leˑu**	**leˑu**	**wiel**	**ʊiˑl**	**ʊiˑl**
lepel	lebəl	leˑpəl	wielen	ʊiˑln	ʊilə
linialen	liniʊːln	liniɑˑlə	zagen	zʊːgŋ	zɒːjə

Table 1. List of all 70 words pronounced in the speakers' local dialect. The expected approximate pronunciations are indicated for each location. The five words which have the same expected pronunciation in both locations are marked in bold face.

Articulatory data preprocessing

After collecting all articulatory data, the data for each speaker were manually segmented (acoustically) at the phone level. Tongue movement data which were not associated with a pronunciation of one of the words included in our study were discarded. The duration of each word's pronunciation was time-normalized between 0 (acoustic start of the word) and 1 (acoustic 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 inferior-superior direction (i.e. tongue height) and the anterior-posterior direction (i.e. posterior position of the tongue) in our analysis. To enable an appropriate comparison between speakers, we normalized the positions of each sensor separately per speaker. In order to abstract away from differences in where the sensors were placed on the tongue, we determined the position relative to the non-speech resting position of each sensor. Consequently, negative values represented positions below or in front of the non-speech resting position of that sensor (in that direction), whereas positive values represented positions above and behind the non-speech resting position. Higher values thus indicated more superior and posterior positions with respect to the non-speech resting position, whereas lower

values indicated the opposite. The non-speech resting position (i.e. the neutral position of the tongue) was recorded during a separate session of about ten seconds in which the participants were asked to refrain from speaking or swallowing. As the amount of tongue movement may vary per speaker, we normalized the positions by dividing by the total range of movement in each direction. In this way, the difference between the most anterior (or inferior) position and the most posterior (or superior) position was always exactly 1 for each subject. The sign of the difference still indicated the position with respect to the non-speech resting position. For example, for one speaker the normalized posterior positions of T1 could range between -0.4 and 0.6, while for another it could range from -0.8 to 0.2.[4]

Formant extraction

We automatically extracted the first (F1) and second formant (F2) frequencies of the acoustic recording of the vowels in our dataset using the *findformants* function of the *phonTools* R package (Barreda, 2015). This function extracts formants on the basis of the formulas provided in Snell (1993). We extracted the formants for each word separately using a time step of 5 ms (i.e. at 200 Hz). Per time point for which we had articulatory data, we averaged the corresponding formant measurement points (generally about two time points, as the articulatory data was obtained at 100 Hz and the formant data at 200 Hz). As a rough correction of the automatically extracted formants frequencies, we discarded F1 measurements outside of the range 200 – 1000 Hz, and did the same for F2 measurements outside of the range 500 – 3000 Hz. After this step, we normalized the formant frequencies using Lobanov's (1971) z-transformation, as this normalization method was reported by Adank et al. (2004), who also investigated Dutch data, to be an adequate normalization procedure retaining sociolinguistic variation. As automatic formant measurements may be incorrect, we also obtained a set of formant measurements for a subset of the data (for each speaker: 27 CVC sequences and 20 randomly selected dialect words) which were manually corrected (after being automatically generated in PRAAT).

Data analysis: generalized additive modeling

Since the articulatory trajectories of the individual tongue sensors are clearly nonlinear, we use generalized additive modeling to analyze the data (Hastie & Tibshirani, 1990; Wood, 2006; see Baayen, 2013 for a non-technical introduction). Generalized additive modeling is a flexible regression approach which not only supports linear relationships between the dependent variable and the independent variables, but also nonlinear dependencies and interactions.

Generalized additive modeling has been used in articulography before (Tomaschek et al., 2013, 2014; Wieling et al., 2015). Furthermore, the method has been applied to language variation research (Wieling et al., 2011 and Wieling et al., 2014), and to model nonlinear patterns of brain signals across time (e.g., Tremblay & Baayen, 2010; Meulman et al., 2015) and likewise for gaze data (Van Rij et al., 2016a).

In this case our dependent variable is the normalized position of each sensor, which we model as a smooth (i.e. nonlinear) function (SF) over normalized time. The smooth function is represented using a thin plate regression spline (Wood, 2003) which models the nonlinearity as a combination of several low level functions (such as a logarithmic function, a linear function, a quadratic function, etc.).[5] There are other types of splines possible, such as a cubic regression spline (consisting of a series of third degree polynomials), but a thin plate regression spline has better performance and is computationally efficient (Wood, 2003). To prevent overfitting of the data by the SF, the amount of non-linearity (i.e. the wigglyness) of a spline is penalized. Furthermore, generalized cross-validation is used to determine appropriate parameters of the thin plate regression spline during the model-fitting process (Wood, 2006).[6]

---

[4] Importantly, the resulting patterns were relatively similar when another normalization scheme was used instead. This alternative normalization scheme consisted of setting the most anterior (inferior) position of the T1 tongue sensor to 0 and the most posterior (superior) position of the T3 tongue sensor to 1. Consequently, this normalization scheme can be seen as normalizing the inside of the mouth of each speaker between 0 and 1 in both directions. While the non-speech resting position was not involved in this normalization, we included the non-speech resting position as a control predictor in the models which were fit using this normalization scheme.

[5] Intuitively, a spline may be viewed as a flexible band which follows the general pattern of the points.

[6] With generalized cross validation the data is repeatedly fit on a random subset of the data and then validated on the remaining part of the data.

As there is clearly 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 1.8.12) allows for the inclusion of factor smooths to represent full random effects. These factor smooths (for an example, see Figure 4) are a nonlinear alternative to random intercepts and random slopes in a mixed-effects regression model. Just as random intercepts and slopes (which are required in a model where multiple observations are present per speaker and/or words; Baayen et al., 2008), factor smooths are essential for taking the structural variability associated with individual speakers and words into account and thereby prevent anti-conservative (i.e. too low) *p*-values.

As in a common (Gaussian) 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 and normally distributed. However, when analyzing time series which are relatively smooth and slow moving (such as the movement of the tongue over time), the residuals will generally be autocorrelated. This means that the residuals at time t will be correlated with the residuals at time $t + 1$ (see Figure 5, left). In our case, the autocorrelation present in the residuals is very high at about 0.96 at lag 1. If this autocorrelation is not brought into the model, the *p*-values of the model will be too low. Fortunately, the function *bam* of the *mgcv* package we use to create the GAMs is able to take into account the autocorrelation of the residuals (see Figure 5, right, where after correction the autocorrelation at lag 1 has been reduced to below 0.1), thereby 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 large datasets (Wood et al., 2014), such as the data included in this study (about 1.7 million positions: 34 speakers, three sensors, two axes, 97 words repeated twice, and an average duration of about 0.43 seconds, 43 measurement points, per word).

After model fitting, we followed the model criticism procedure put forward by Baayen (2008; Ch. 6.2.3). This procedure showed that the residuals of the models we fitted exhibited non-normality and heteroscedasticity (i.e. the variance of the residuals was not constant across the fitted values). Consequently, after fitting these models, we excluded the data points for which the absolute standardized residuals were greater than 2.5 (i.e. those data points for which the predicted and actual values differed to a large extent). We then refitted the same model on the smaller dataset (generally containing about 98% of the original data). As this procedure resulted in improved characteristics of the residuals, all results reported in this paper are based on the resulting models *after* model criticism. A clear advantage of this procedure is that it reduces the likelihood of reporting effects as significant, if these effects are carried by data points for which the model is not adequate.

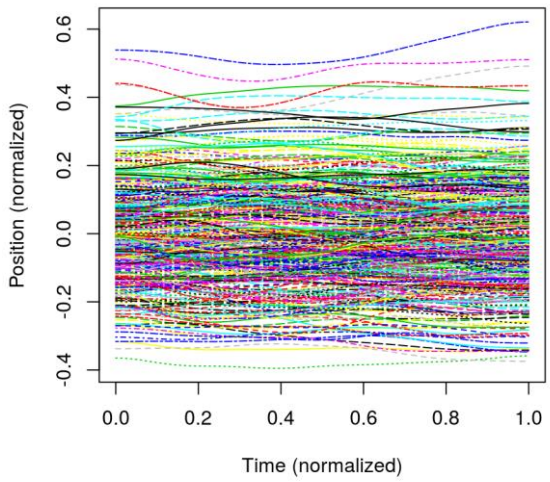

Figure 4. Individual adjustments to the general tongue movement trajectories. As the average of these adjustments is approximately 0 (i.e. centered), both positive and negative adjustments are possible.

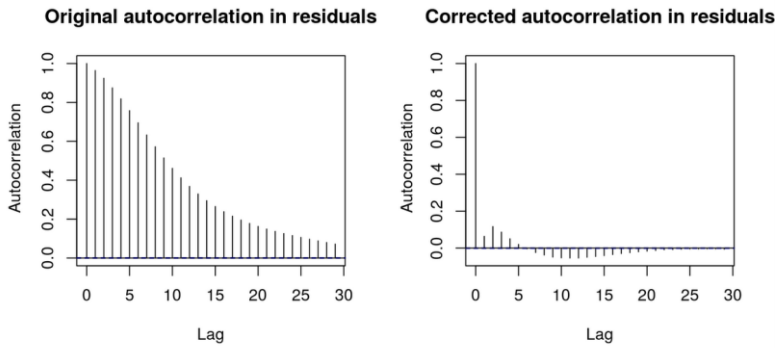

Figure 5. Autocorrelation in the residuals. Left: without correction, right: after correction.

Reproducibility

To facilitate reproducibility and the use of the methods illustrated in this study, the data, methods and results are available as a paper package stored at the Mind Research Repository (http://openscience.uni-leipzig.de) and the first author's website (http://www.martijnwieling.nl).

Results

As an illustration of the generalized additive modeling approach, Figure 6 shows the normalized tongue movement trajectories for each of the three tongue sensors separately during the pronunciation of four dialect words: *taarten*, 'cakes' (generally pronounced [tɑːtn] in Ter Apel and [tœʁtə] in Ubbergen), *bogen*, 'bows' (generally pronounced [bougŋ] in Ter Apel and [boˑçə] in Ubbergen), *tol*, 'top' (spinning toy; pronounced [tɔl] in both dialects), and *kameel*, 'camel' (generally pronounced [kəmel] in both dialects). Similarly, Figure 7 shows the same type of visualization for two CVC sequences in standard Dutch, *taat*, [tat] and *poop*, [pop]. The red and blue dots in the graph indicate the measured tongue positions of both groups. The red (dark) curves indicate the fitted tongue trajectories of the speakers in Ubbergen for word-specific models, whereas the (lighter) blue curves are linked to the speakers in Ter Apel. The relative lightness of each curve visualizes the time course from the beginning of the word (darkest) to the end of the word (lightest). Clearly the articulations for *taarten* and *bogen* are more different (specifically in the shape of the trajectories) than the articulations for *tol* and *kameel* (and also *taat* and *poop* in Figure 7), which only seem to differ with respect to the posterior position (further back in Ter Apel than in Ubbergen). In addition, the pronunciations for *taat* show a greater distinction between the two speaker groups than the pronunciations for *poop*. A general pattern across all six graphs in Figures 6 and 7, however, is that the speakers from Ubbergen appear to have more anterior tongue positions than those from Ter Apel.

The fitted trajectories were obtained by creating a single GAM for each of the six words, simultaneously for all three sensors and both axes. In the GAM specification, a different SF was fitted for each group. The command to fit such a model for a single word (simplified: only for a single sensor in a single dimension) using the function `bam` of the `mgcv` package is:

```
model = bam(Position ~ s(Time,by=Group) + Group +
                       s(Time,Speaker,bs='fs',m=1), rho=0.96)
```

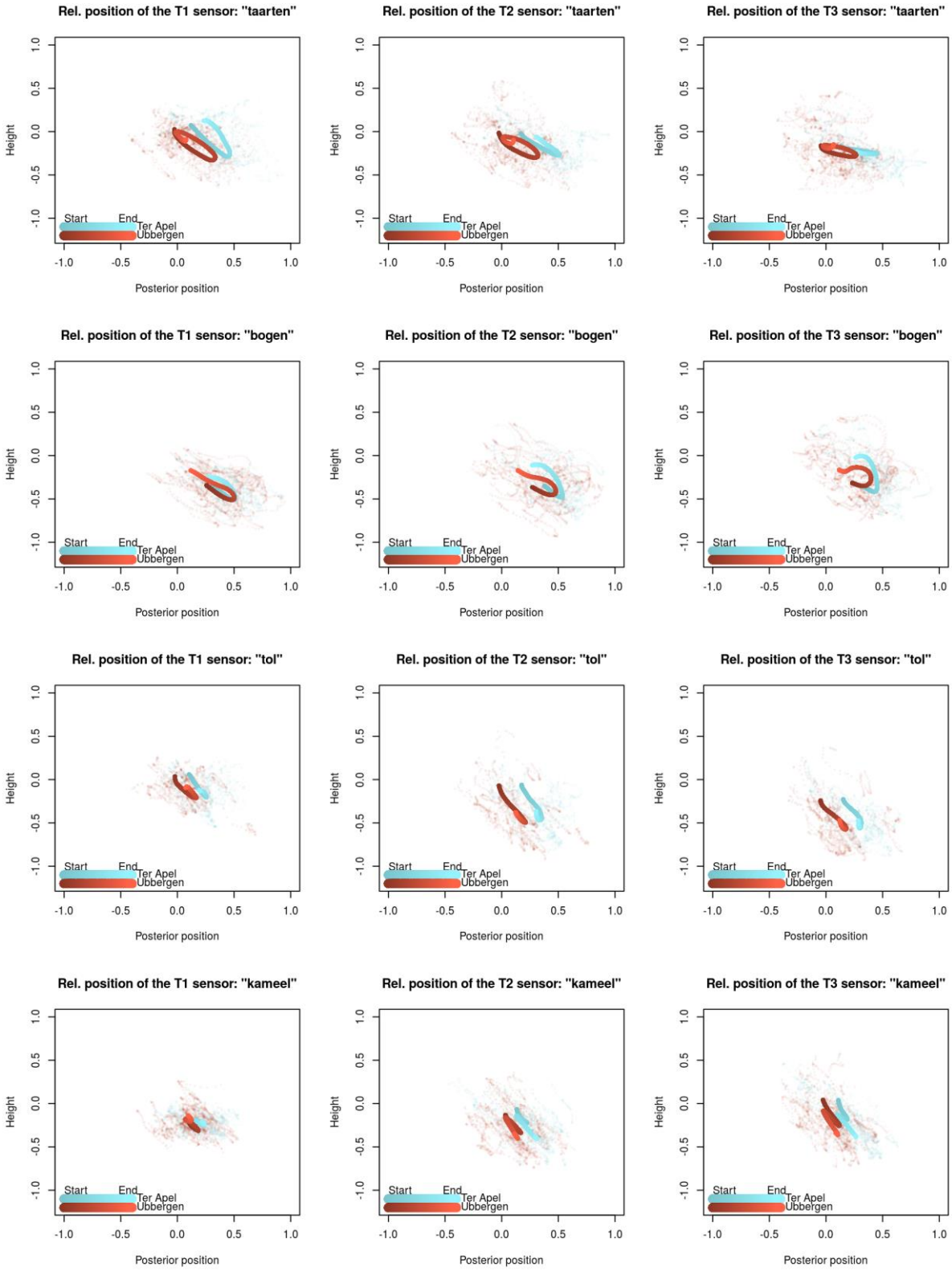

Figure 6. Fitted tongue trajectories (including individual points) of the three tongue sensors (left column: T1, middle column: T2, right column: T3) for the two groups of speakers in two dimensions for four dialect words (one per row). The points represent the normalized position (a range of 1 for each speaker) relative to the non-speech resting position (negative: in front of/below the non-speech resting position; positive: behind/above the non-speech resting position). Higher values on the *x*-axis indicate positions which are further back (posterior). Higher values on the *y*-axis indicate positions which are higher (superior). The darkness of the line indicates the time course of the trajectories (dark: start of the pronunciation, light: end of the pronunciation).

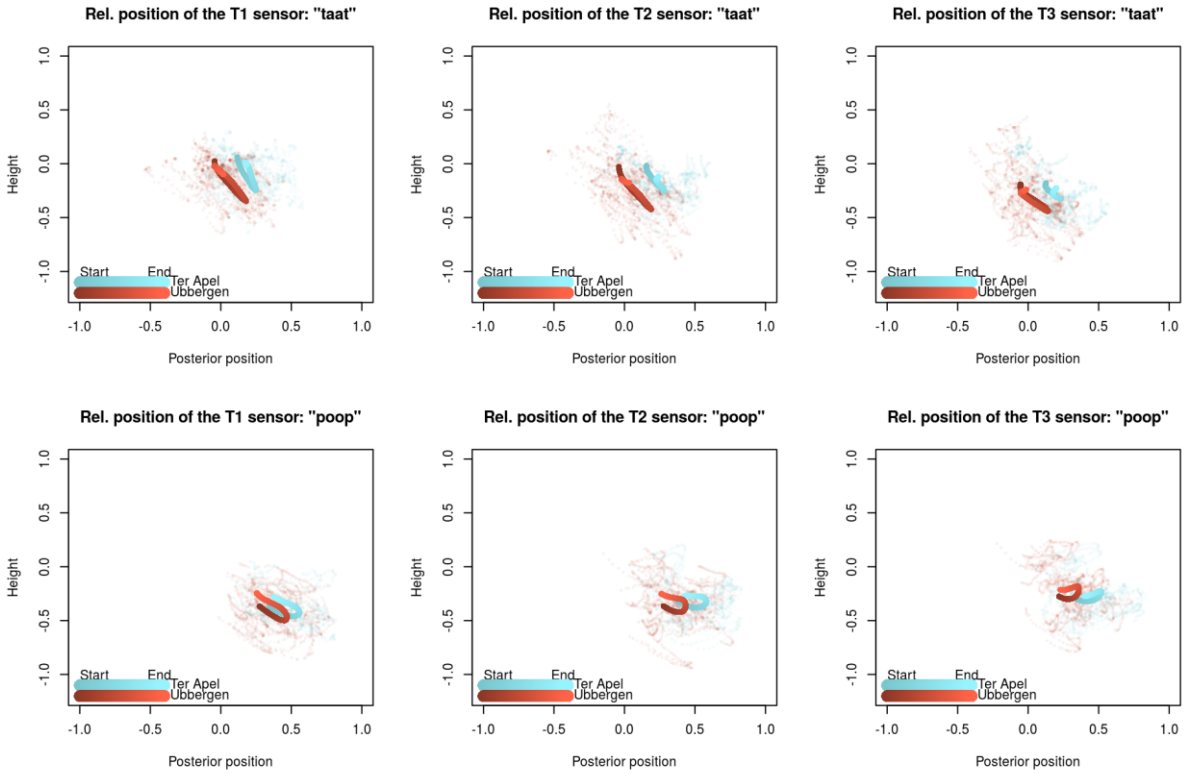

Figure 7. Fitted tongue trajectories (including individual points) of the three tongue sensors (left column: T1, middle column: T2, right column: T3) for the two groups of speakers in two dimensions for two CVC sequences (one per row). The points represent the normalized position (a range of 1 for each speaker) relative to the non-speech resting position (negative: in front of/below the non-speech resting position; positive: behind/above the non-speech resting position). Higher values on the *x*-axis indicate positions which are further back (posterior). Higher values on the *y*-axis indicate positions which are higher (superior). The darkness of the line indicates the time course of the trajectories (dark: start of the pronunciation, light: end of the pronunciation).

The interpretation of this GAM specification is that the sensor position is predicted on the basis of a nonlinear pattern across (normalized) time per group (Ter Apel vs. Ubbergen: `s(Time,by=Group)`), while simultaneously taking into account the speaker-related variation via a factor smooth (the `bs='fs'` block; `m=1` limits the wigglyness of the curve per speaker, which is suitable for these nonlinear random effects). The `rho` value (here fixed at 0.96) indicates the amount of autocorrelation in the residuals which needs to be taken into account (see explanation, above). The linear contrast between the two groups (`Group`) is added to the model as the smooth functions are centered and thus unable to model a constant (intercept) difference between the two groups.

To see at which points the trajectories differ significantly from each other, confidence intervals are needed. These can readily be extracted from the fitted GAM using the R package *itsadug* (version 2.2; van Rij et al., 2016b). Figure 8, visualizing the resulting trajectories and differences for the CVC sequence *taat*, shows that the difference in the posterior position is significant across a large part of the time course, while there is no significant height difference. While this visualization suggests that the distinction between the two groups is necessary (for the posterior position), this should be assessed more formally. There are two approaches for this. The first is fitting a simpler model without the group distinction, and comparing this simpler model to the more complex model having the group distinction to see if the additional complexity is warranted (e.g., by comparing the difference in maximum likelihood scores while taking into account the difference in model complexity). The drawback of this approach is that multiple models need to be fitted, and given that the full model (on the basis of all data) takes a long time to fit (approximately 8 hours using 16 processors simultaneously on a fast server; using a single processor would take about 27 hours) the required amount of time needed for this approach becomes prohibitive.

Consequently, we turn to another approach, which consists of respecifying the model in such a way that it does not fit the SFs for the two groups separately, but rather fits a SF for a single group (i.e. the reference level) and a second smooth function representing the non-linear difference between the two groups (i.e. the difference SF which needs to be added to the SF of the first group to yield the SF of the second group). Additionally, as the SFs are centered, a fixed-effect contrast is included to model the constant (i.e. intercept) difference between the two groups. The associated *p*-values obtained from the model summary for the fixed-effect contrast and the non-linear SF will then directly indicate if the distinction between the two groups is necessary or not, and if the difference consists of an intercept shift and/or a non-linear difference. If the fixed-effect contrast is significant, this indicates that there is a constant (intercept) difference between the two groups (e.g., Ter Apel might show a greater posterior position than Ubbergen). Similarly, if the difference SF is significant, this indicates that the non-linear tongue sensor movement pattern of the two groups differs. The command to fit this type of model (for a single word) is:

```
diff.model = bam(Position ~ s(Time) + s(Time,by=IsTerApelO) +
                 IsTerApelO + s(Time,Speaker,bs='fs',m=1), rho=0.96)
```

In this case `IsTerApelO` is an ordered factor equal to 1 for the speakers from Ter Apel and 0 for those from Ubbergen. The SF containing this predictor, `s(Time,by=IsTerApelO)`, will be equal to 0 when the ordered factor equals 0. This implies that the first smoothing function, `s(Time)`, will be the articulatory trajectory for the Ubbergen group. As the first SF, `s(Time)`, is never equal to 0, this also implies that the second SF, `s(Time,by=IsTerApelO)`, must be equal to the non-linear difference between the Ter Apel and Ubbergen speakers. The fixed-effect predictor `IsTerApelO` models the constant (intercept) difference between the two groups. For the visualization in Figure 8, both the constant difference and the difference SF were significant for the posterior position difference ($p < 0.05$), whereas the height difference was not significant ($p > 0.05$).

While it is useful to focus on the differences in the pronunciation of individual words, an aggregate analysis is able to provide a more general and robust view of tongue trajectory differences. In our aggregate model, we simultaneously analyzed the three tongue sensors and two axes for a large set of words. Rather than using a single `s(Time)` for the reference level (Ubbergen) as in the simple example above, we now need separate patterns over time for each tongue sensor and axis (i.e. height and posterior position for the T1, T2 and T3 sensors). This can be accomplished by adding a by-parameter distinguishing these six levels (i.e. the interaction between sensor and axis, stored in the variable `SensorAxis`). Similarly, rather than a single SF representing the non-linear difference between Ter Apel and Ubbergen (via the use of a by-variable), six difference SFs are needed, one for each combination of sensor and axis. Similarly, six fixed-effect predictors are necessary modeling the constant differences between the two groups. Consequently, six ordered factor predictors are created which are equal to 1 for the group of Ter Apel for a specific sensor and axis. For example, the predictor `IsTA.T1.HO` equals 1 for the positions associated with the inferior-superior axis of the T1 sensor for the Ter Apel group, while `IsTA.T3.PO` is equal to 1 for the positions associated with the anterior-posterior axis of the T3 sensor for the Ter Apel group. The speaker-related variability must also be allowed to vary for each of the six combinations of sensors and axes. This can be achieved by creating a new predictor `SpeakerSensorAxis` representing the interaction between the three predictors `Speaker`, `Sensor`, `Axis` and using this predictor in the factor smooth. Given that we are now aggregating over a large set of words, we also need to take into account the variability per word via a factor smooth. Importantly, as the differences between the two groups might be larger for one word than another, we also need to allow for this variability. Consequently, we construct a new predictor `WordGroupSensorAxis` representing the interaction between the four predictors `Word`, `Group`, `Sensor`, `Axis`. This predictor is used in a separate factor smooth. The specification of this model is as follows:

```
model = bam(Pos ~ s(Time,by=SensorAxis) + SensorAxis +
                  s(Time,by=IsTA.T1.HO) + IsTA.T1.HO +
                  s(Time,by=IsTA.T1.PO) + IsTA.T1.PO +
```

```
s(Time,by=IsTA.T2.HO) + IsTA.T2.HO +
s(Time,by=IsTA.T2.PO) + IsTA.T2.PO +
s(Time,by=IsTA.T3.HO) + IsTA.T3.HO +
s(Time,by=IsTA.T3.PO) + ISTA.T3.PO +
s(Time,SpeakerSensorAxis,bs='fs',m=1) +
s(Time,WordGroupSensorAxis,bs='fs',m=1),
rho=0.96)
```

For example, if `s(Time,by=IsTA.T1.PO)` is found to be significant, this indicates that the non-linear difference between the two groups for the T1 sensor in the anterior-posterior direction is significant, and therefore that it is necessary to distinguish the two groups with respect to the posterior position of the T1 sensor. Similarly, if `IsTA.T1.PO` is found to be significant, this indicates the presence of a significant constant (intercept) difference in the anterior-posterior direction between the two groups.

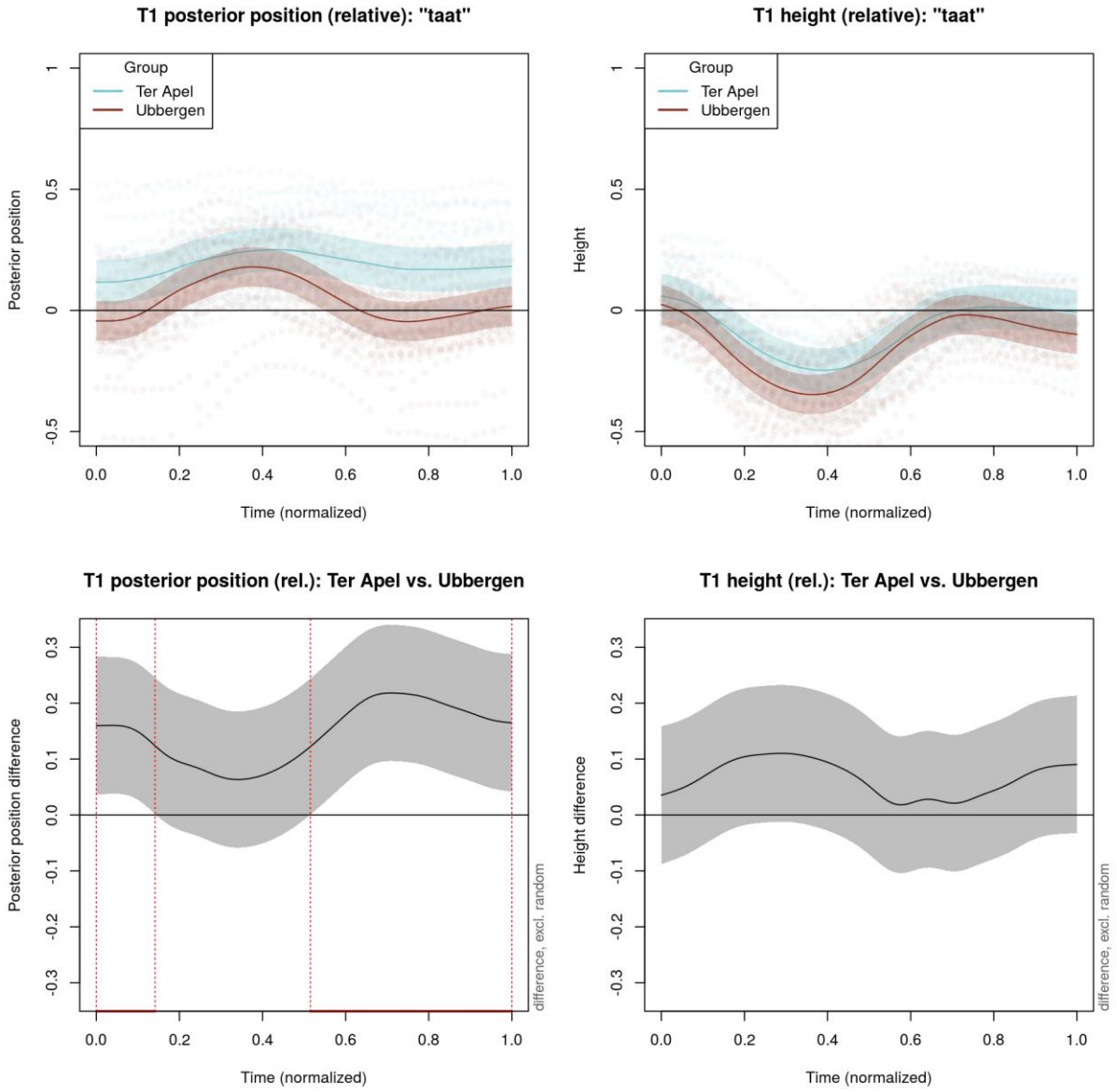

Figure 8. T1 sensor and sensor difference trajectories for the CVC sequence *taat*, [tat], in the anterior-posterior dimension (left) and the height dimension (right) for both groups. The upper graphs show the trajectories per group including 95% confidence bands together with the individual points. The lower graphs show the difference between the two groups including confidence bands (and marked areas where the difference is significantly different from 0: for the posterior position it appears to be significantly different for the pronunciation of the /t/'s but not the /a/ in between) extracted from the fitted GAM (which took the individual variation and autocorrelation in the residuals into account).

As it might be necessary to distinguish CVC sequences from dialect words (i.e. the difference between the two groups might be larger for the dialect words than for the CVC sequences), we extended the model specification to take this into account (also in the random-effects structure per speaker). The model below shows this extension:

```
model = bam(Pos ~ s(Time,by=SensorAxis) + SensorAxis +
                  s(Time,by=IsTA.T1.HO) + IsTA.T1.HO +
                  s(Time,by=IsTA.T1.PO) + IsTA.T1.PO +
                  s(Time,by=IsTA.T2.HO) + IsTA.T2.HO +
                  s(Time,by=IsTA.T2.PO) + IsTA.T2.PO +
                  s(Time,by=IsTA.T3.HO) + IsTA.T3.HO +
                  s(Time,by=IsTA.T3.PO) + ISTA.T3.PO +
                  s(Time,by=IsCVC.T1.HO) + IsCVC.T1.HO +
                  s(Time,by=IsCVC.T1.PO) + IsCVC.T1.PO +
                  s(Time,by=IsCVC.T2.HO) + IsCVC.T2.HO +
                  s(Time,by=IsCVC.T2.PO) + IsCVC.T2.PO +
                  s(Time,by=IsCVC.T3.HO) + IsCVC.T3.HO +
                  s(Time,by=IsCVC.T3.PO) + IsCVC.T3.PO +
                  s(Time,by=IsTACVC.T1.HO) + IsTACVC.T1.HO +
                  s(Time,by=IsTACVC.T1.PO) + IsTACVC.T1.PO +
                  s(Time,by=IsTACVC.T2.HO) + IsTACVC.T2.HO +
                  s(Time,by=IsTACVC.T2.PO) + IsTACVC.T2.PO +
                  s(Time,by=IsTACVC.T3.HO) + IsTACVC.T3.HO +
                  s(Time,by=IsTACVC.T3.PO) + ISTACVC.T3.PO +
                  s(Time,SpeakerTypeSensorAxis,bs='fs',m=1) +
                  s(Time,WordGroupSensorAxis,bs='fs',m=1),
                  rho=0.96)
```

While much larger, the model resembles the previous model to a large extent. The difference is that there are now two sets of additional ordered factors included. There are six (for three sensors and two axes) new `IsCVC` ordered factors and six new `IsTACVC` ordered factors. The `IsCVC` ordered factors allow the model to represent the non-linear and intercept differences between the CVC sequences and the dialect words. Given that the two types of stimuli differ substantially in their structure, significance of these SFs will not be surprising (nor very informative). However, the `IsTACVC` ordered factors allow the model to represent a potential distinction (both non-linear and in the intercept) between the group difference (i.e. Ter Apel vs. Ubbergen) for the CVC sequences versus the dialect words. For example, the difference between the two groups (for example in the posterior position of the T1 sensor) might be stronger for the dialect words than for the CVC sequences, and this would be reflected in the significance of either `s(Time,by=IsTACVC.T1.PO)` or `IsTACVC.T1.PO`. Of course the model specification above can be made simpler, by excluding non-significant terms.

Following this model specification, we fitted a single large-scale GAM on 1.7 million tongue sensor positions. The model took about 8 hours to fit the model on a high performance server with 16 Intel Xeon E5-2699 v3 processors. As the model fit revealed that only `by=IsCVC` SFs and `IsTA` fixed-effect factors reached significance, we report the results on the basis of the following simpler model:

```
model = bam(Pos ~ s(Time,by=SensorAxis) + SensorAxis +
                  IsTA.T1.HO + IsTA.T1.PO + IsTA.T2.HO +
                  IsTA.T2.PO + IsTA.T3.HO + ISTA.T3.PO +
                  s(Time,by=IsCVC.T1.HO) + s(Time,by=IsCVC.T1.PO) +
                  s(Time,by=IsCVC.T2.HO) + s(Time,by=IsCVC.T2.PO) +
                  s(Time,by=IsCVC.T3.HO) + s(Time,by=IsCVC.T3.PO) +
                  s(Time,SpeakerTypeSensorAxis,bs='fs',m=1) +
                  s(Time,WordGroupSensorAxis,bs='fs',m=1), rho=0.96)
```

The results of the model are shown in Tables 2 (parametric part: fixed effects) and 3 (smooth functions). The explained variance of the model is equal to about 90%, due mainly to the inclusion of the factor smooths per speaker and word. The first line of the parametric part of the model shown in Table 2 simply shows the reference sensor position (i.e. the intercept is equal to the posterior position of the T3 sensor). Furthermore, the next five lines compare the height of the T3 sensor and the height and posterior position of the other sensors to the posterior position of the T3 sensor (i.e. the intercept). While the comparison between height and posterior position is not informative as such, these comparisons are required as the model includes both dimensions simultaneously. Nevertheless, these results show that the normalized posterior positions (with respect to the non-speech resting position) do not differ significantly, whereas the normalized height of the sensors is generally lower than the normalized posterior position.

Lines 7 to 12 of Table 2 are more informative, however. These compare the (constant) posterior position and height difference between the two groups. Clearly, the group differences with respect to the posterior positions of the three sensors are all significant ($p < 0.05$). The positive estimates indicate that the speakers from Ter Apel have a more posterior tongue position than those from Ubbergen. There were no significant height differences between the two groups. As the `IsCVC` and `IsTACVC` fixed-effect factors did not reach significance, this indicates that the pattern is general and holds both for the dialect and the standard Dutch pronunciations.

Table 3 provides information about the SFs in the model and shows (in lines 1 to 6) that there are significant non-linear trajectories associated with the various sensors (for the two axes). Furthermore, lines 7 to 12 of Table 3 show that for various sensors, there are significant non-linear sensor trajectory differences comparing the dialect words to the CVC sequences. However, this is not surprising (or interesting) given that the CVC sequences have a specific structure (a consonant followed by a vowel followed by a consonant), which is not the case for the dialect words (see Table 1). Importantly, note that the difference between the two groups is the same across both CVC sequences and dialect words.

Figure 9 shows a visual impression of the relative position of the three tongue sensors both for the dialect words and the CVC sequences. It is immediately obvious that the position of the tongue sensors is more posterior for the speakers from Ter Apel (reflecting the result shown in Table 2). Figure 10 and 11 provide a visualization of the trajectories over time, as well as their difference, for the dialect words and the CVC sequences, respectively.

Validation

To validate these results, we conducted two additional analyses. In the first analysis, we only analyzed the five dialect words (marked in bold face in Table 1) which had phonologically identical specifications in the two dialects. Results with respect to the fixed effects (i.e. the constant differences between the two groups) are shown in Table 4. As there were no significant non-linear differences between the two dialect groups, the table with the non-linear trajectories is not shown here (but it can be found in Section 6.4.1 of the supplementary material). In the second analysis, we only analyzed the /t/ segments (Table 5). Similar as for the other analysis, the table with the non-linear trajectories is not shown here as there were no significant non-linear differences between the two groups (but see Section 6.5.1 of the supplementary material). Both analyses confirmed the original pattern (shown in Table 2), with the tongue sensors having larger posterior positions in Ter Apel than in Ubbergen (the corresponding lines are marked in italics in Tables 4 and 5), but no height difference between the two groups. Note that despite being in the correct direction and of similar magnitude, the differences were not significant ($0.07 < p < 0.19$; see Table 4) in the first analysis (on the basis of five dialect words). However, this is unsurprising, given that only a small subset of the data was included. The differences were highly significant ($p < 0.001$; see Table 5) in the second analysis.

	Estimate	Std. Error	*t*-value	*p*-value	
Intercept (T3 posterior position)	0.088	0.038	3.1	0.001	**
T2 posterior position vs. T3 posterior position	0.055	0.039	1.4	0.162	
T1 posterior position vs. T3 posterior position	0.059	0.039	1.5	0.131	
T3 height vs. T3 posterior position	-0.267	0.039	-6.8	< 0.001	***
T2 height vs. T3 posterior position	-0.298	0.039	-7.6	< 0.001	***
T1 height vs. T3 posterior position	-0.289	0.039	-7.4	< 0.001	***
T1 posterior position Ter Apel vs. Ubbergen	0.102	0.041	2.5	0.012	*
T1 height Ter Apel vs. Ubbergen	0.035	0.041	0.9	0.386	
T2 posterior position Ter Apel vs. Ubbergen	0.126	0.041	3.1	0.002	**
T2 height Ter Apel vs. Ubbergen	0.017	0.041	0.4	0.669	
T3 posterior position Ter Apel vs. Ubbergen	0.128	0.041	3.1	0.002	**
T3 height Ter Apel vs. Ubbergen	-0.030	0.041	-0.7	0.463	

Table 2. Parametric coefficients of the generalized additive model on the basis of all words (dialect words and CVC sequences), for all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height).

Smooth Functions (SFs)	edf	*F*-value	*p*-value	
s(Time) : T3 posterior position	7.2	2.5	0.130	
s(Time) : T2 posterior position	11.6	8.1	< 0.001	***
s(Time) : T1 posterior position	14.0	11.0	< 0.001	***
s(Time) : T3 height	10.0	11.1	< 0.001	***
s(Time) : T2 height	12.2	24.7	< 0.001	***
s(Time) : T1 height	16.9	18.2	< 0.001	***
s(Time) : T1 posterior position dialect-standard difference SF	13.4	3.8	0.051	
s(Time) : T1 height dialect-standard difference SF	14.2	5.9	< 0.001	***
s(Time) : T2 posterior position dialect-standard difference SF	14.2	4.7	0.013	*
s(Time) : T2 height dialect-standard difference SF	15.6	8.7	< 0.001	***
s(Time) : T3 posterior position dialect-standard difference SF	15.2	6.7	0.219	
s(Time) : T3 height dialect-standard difference SF	15.0	7.5	0.003	**
s(Time, SpeakerTypeSensorAxis) [factor smooth]	3268.2	128.6	< 0.001	***
s(Time, WordGroupSensorAxis) [factor smooth]	10243.4	92.3	< 0.001	***

Table 3. SF terms of the generalized additive model on the basis of all words (dialect words and CVC sequences), for all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height). The first 6 lines show the SFs for the reference level (Ubbergen), whereas lines 7 to 12 represent difference SFs comparing CVC sequences (standard) to the trajectories associated with the dialect words. The edf column indicates the estimated degrees of freedom, which is a measure to reflect SF complexity. The maximum allowed SF complexity was 19 edf (enforced by setting the k-parameter of each SF to 20), and this seems to be sufficiently high as none of the SFs have an edf close to 19. The *p*-value assesses if the SF is significantly different from 0. The final two lines show the factor smooths per speaker and word (i.e. the random effects structure). Non-linear differences between the two groups (whether or not in interaction with the type, dialect or standard) were not found to be significant and not included in the model specification.

	Estimate	Std. Error	*t*-value	*p*-value	
Intercept (T3 posterior position)	0.072	0.054	1.3	0.180	
T2 posterior position vs. T3 posterior position	0.034	0.076	0.4	0.162	
T1 posterior position vs. T3 posterior position	0.037	0.076	0.5	0.131	
T3 height vs. T3 posterior position	-0.258	0.076	-3.4	< 0.001	***
T2 height vs. T3 posterior position	-0.266	0.076	-3.5	< 0.001	***
T1 height vs. T3 posterior position	-0.181	0.076	-2.4	0.017	*
T1 posterior position Ter Apel vs. Ubbergen	*0.102*	*0.077*	*1.3*	*0.189*	
T1 height Ter Apel vs. Ubbergen	-0.007	0.077	-0.1	0.032	
T2 posterior position Ter Apel vs. Ubbergen	*0.141*	*0.077*	*1.8*	*0.069*	
T2 height Ter Apel vs. Ubbergen	-0.003	0.077	-0.0	0.966	
T3 posterior position Ter Apel vs. Ubbergen	*0.123*	*0.077*	*1.6*	*0.112*	
T3 height Ter Apel vs. Ubbergen	-0.039	0.077	-0.5	0.613	

Table 4. Parametric coefficients of the generalized additive model on the basis of five phonologically identical dialect words, for all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height).

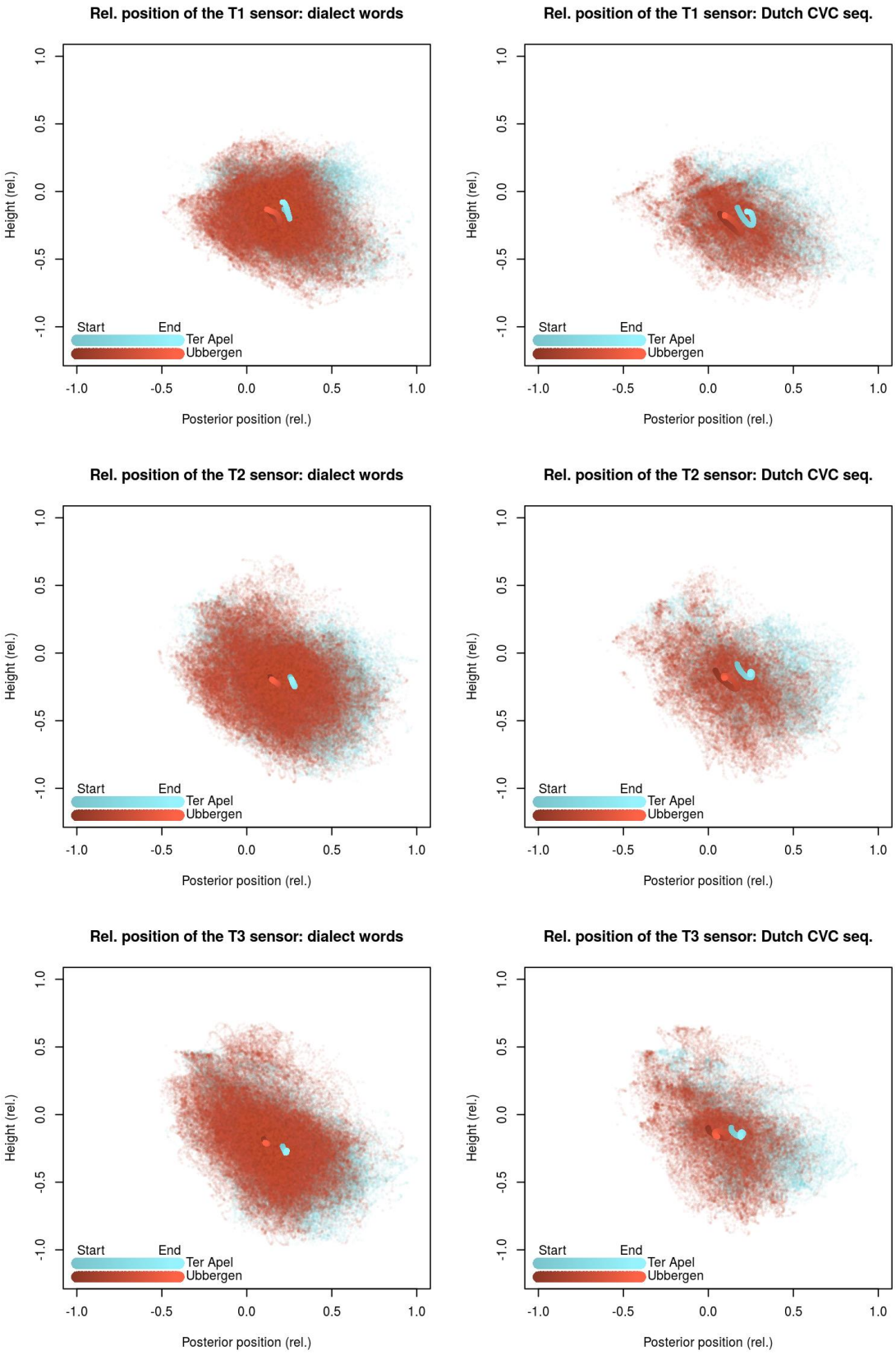

Figure 9. Aggregate fitted tongue trajectories (including individual points) of the three tongue sensors (first row: T1, second row: T2, third row: T3) for the two groups of speakers in two dimensions (posterior position on the *x*-axis, height on the *y*-axis) for all 70 dialect words (first column) and for all 27 CVC sequences (second column). The darkness of the lines indicates the time course of the trajectories (dark: start of the pronunciation, light: end of the pronunciation). The difference in anterior-posterior position is significant in all cases, while the difference in height is not.

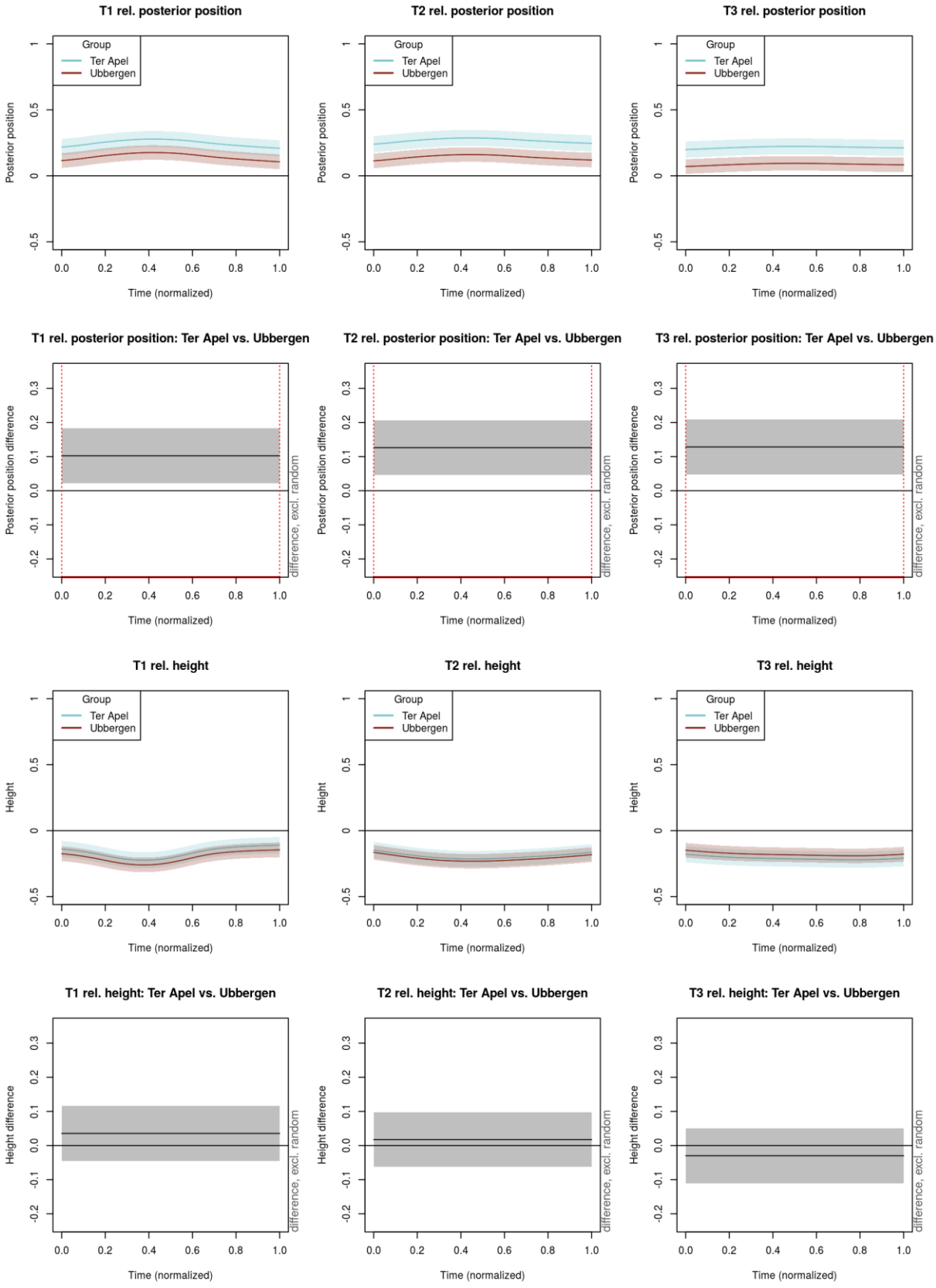

Figure 10. Graphs in row 1: tongue sensor trajectories (T1, T2, T3) aggregated over all 70 dialect words in the anterior-posterior dimension for both groups. Graphs in row 2: differences between the tongue sensor trajectories in the anterior-posterior dimension. The differences are significant across the whole time span ($p < 0.05$; indicated by the red bar) for all tongue sensors (see Table 2). The graphs in row 3 and 4 show the corresponding results for height. None of the height differences are significant ($p < 0.05$).

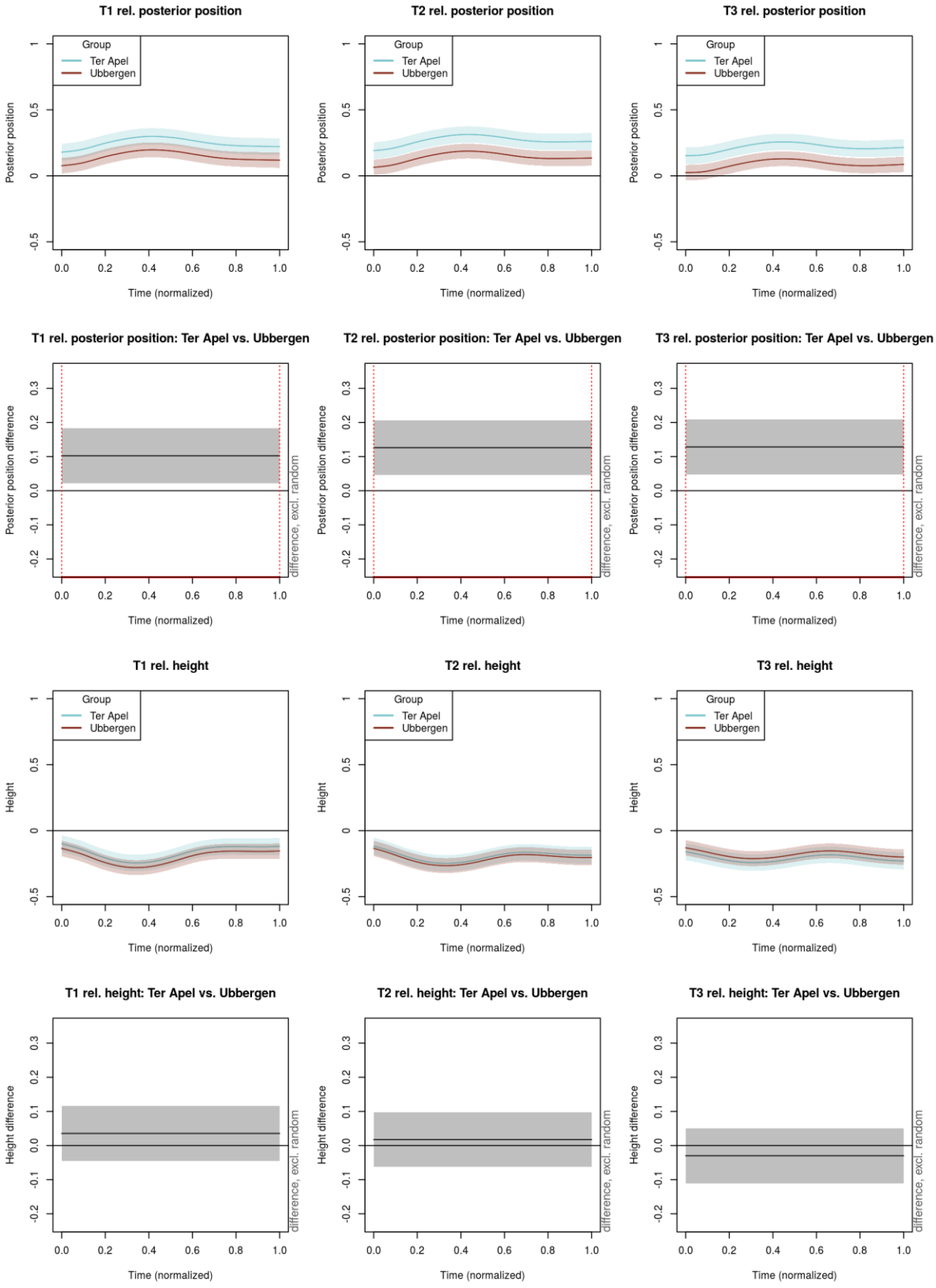

Figure 11. Graphs in row 1: tongue sensor trajectories (T1, T2, T3) aggregated over all 27 CVC sequences in the anterior-posterior dimension for both groups. Graphs in row 2: differences between the tongue sensor trajectories in the anterior-posterior dimension. The differences are significant across the whole time span ($p < 0.05$; indicated by the red bar) for all tongue sensors (see Table 2). The graphs in row 3 and 4 show the corresponding results for height. None of the height differences are significant ($p < 0.05$).

	Estimate	Std. Error	*t*-value	*p*-value	
Intercept (T3 posterior position)	-0.029	0.028	-1.0	0.302	
T2 posterior position vs. T3 posterior position	0.033	0.040	0.8	0.418	
T1 posterior position vs. T3 posterior position	0.018	0.040	0.4	0.664	
T3 height vs. T3 posterior position	-0.138	0.040	-3.4	< 0.001	***
T2 height vs. T3 posterior position	-0.059	0.040	-1.5	0.140	
T1 height vs. T3 posterior position	-0.001	0.040	-0.0	0.976	
T1 posterior position Ter Apel vs. Ubbergen	*0.175*	*0.042*	*4.1*	*< 0.001*	***
T1 height Ter Apel vs. Ubbergen	0.045	0.042	1.1	0.286	
T2 posterior position Ter Apel vs. Ubbergen	*0.215*	*0.042*	*5.1*	*< 0.001*	***
T2 height Ter Apel vs. Ubbergen	0.008	0.042	0.2	0.845	
T3 posterior position Ter Apel vs. Ubbergen	*0.200*	*0.042*	*4.7*	*< 0.001*	***
T3 height Ter Apel vs. Ubbergen	-0.015	0.042	-0.4	0.724	

Table 5. Parametric coefficients of the generalized additive model on the basis of the analysis of the segment /t/ in all words, for all tongue sensors (T1: front, T2: middle, T3: back) and both axes (posterior position and height).

Comparison to linear discriminant analysis

Since generalized additive modeling is a relatively new technique, especially when applied to articulatory data (see Tomaschek et al., 2013 and 2014), we have also analyzed the data using another technique, namely linear discriminant analysis (LDA).[7] In LDA an item's class (in our case the group of the speaker) is predicted on the basis of a set of numerical predictors (in our case the normalized height and posterior position for the three tongue sensors). For both the dialect data and the CVC data, we created five different LDAs using segment-specific positions (i.e. for /a/, /i/, /o/, /k/ and /t/). All ten LDAs showed significant group mean differences (all p's < 0.001) generally in line with the global position differences shown in Table 2. Thus, for both types of data the sensor positions were more posterior for the speakers from Ter Apel than for the speakers from Ubbergen. The probability of correctly classifying the group of the speaker on the basis of the tongue position (on the basis of the three sensors) at a certain time point ranged between 62% and 77% (see Table 6). In sum, the LDA analysis showed that the tongue position (in terms of the posterior position of the three tongue sensors) during the pronunciation of a single segment is useful for predicting from which dialect region a speaker originates. These results are in line with the results on the basis of the generalized additive modeling approach, which also showed clear differences between the two groups.

Comparison with formant-based patterns

We first assessed whether the automatically obtained formant measurements matched the manual formant measurements by correlating the two measures. The correlations between the automatic and manual formant measurements were reasonably high at $r = 0.87$ ($r^2 = 0.75$, $p < 0.001$) for F1 and $r = 0.83$ ($r^2 = 0.69$, $p < 0.001$) for F2. When investigating the relationship between F1 and height, and F2 and posterior position, however, it was clear that the correlations were stronger when using the manual formant measurements. Consequently, we opted to use the manual formant measurements in the remainder of this section.

Dataset	/a/	/i/	/o/	/k/	/t/
Dialect words	62%	68%	77%	69%	75%
CVC sequences	67%	70%	67%	67%	72%

Table 6. Speaker group classification accuracy on the basis of the height and posterior position of the three tongue sensors.

[7] Note that LDA is not entirely appropriate for data with repeated measures (Lix and Sajobi, 2010). In addition, LDA requires observations to be independent, which assumption is violated in this dataset where each individual speaker contributes many tongue positions. Consequently, the LDA approach may be anti-conservative when applied to this dataset. A repeated-measures LDA approach would be more appropriate, but to our knowledge no such procedure is implemented in R.

While there certainly is no direct relationship between F1 and height and F2 and posterior position, it is generally assumed that the two correlate (Stevens, 1998). In line with this, the average correlation between the normalized height (taking into account the non-speech resting position) of the tongue sensors during the pronunciation of the vowels and the normalized F1 frequency for the vowels in the dialect words was r = -0.22 (r^2 = 0.05; all p's < 0.001), for the CVC sequences (containing only three different vowels) the correlation increased to r = -0.43 (r^2 = 0.18; all p's < 0.001). The correlations are negative as a higher F1 is related to a lower (anterior) tongue position. When looking at the normalized posterior position of the tongue sensors, the average correlation with the normalized F2 frequency was r = -0.44 (r^2 = 0.19; all p's < 0.001) for the dialect words and r = -0.63 (r^2 = 0.39; all p's < 0.001) for the CVC sequences. Again, the correlations are negative as a higher F2 is related to a less posterior (anterior) tongue position. Figures 12 and 13 visualize the scatter plots between the formants and the tongue sensor positions (for all three tongue sensors) for the dialect words and the CVC sequences, respectively. It is clear that the correlations are not very strong, likely because the characterization of tongue position (height and backness) is based on the sparse representation of just three midsagittal points on the anterior tongue. Furthermore, it is also possible that the articulatory pattern is stronger for the consonants than for the vowels. A valid question then becomes whether a formant-based comparison would show the same pattern as the articulatory comparison discussed above.

In the previous section, we observed a more posterior tongue position for both the dialect words as well as the CVC sequences for the speakers from Ter Apel versus those from Ubbergen. Consequently, we would expect lower F2 values for the speakers from Ter Apel compared to those from Ubbergen. However, this is not what we observe. In fact, on the basis of F2, we do not see a significant difference between the two groups for the dialect words (t = -1.46, p = 0.14), nor for the CVC sequences (t = 0.35, p = 0.73). Clearly, this contrasts with the articulatory differences. With respect to the F1 values, and in line with the articulatory results, we do not find significant differences between the two groups for the dialect words (t = -0.11, p = 0.91), nor for the CVC sequences (t = 1.82, p = 0.07).

Of course, the lower sensitivity of the formant measurements might be caused by the different amount of data involved in the respective analyses (all data for the articulatory analysis, versus only vowels with manual formant measurements for the formant-based analysis). Consequently, we also analyzed the same subset of data used for the formant-based analysis from an articulatory perspective. For the dialect words, the difference in posterior position between Ter Apel and Ubbergen was in the same direction as before (with more posterior positions of the tongue sensors for Ter Apel), but it did not reach significance (t = 1.0, p = 0.33). For the CVC sequences, the difference was significantly more posterior for the speakers from Ter Apel compared to the speakers from Ubbergen (t = 2.1, p < 0.05). As before, no significant height differences (all $|t|$'s < 1.0, p's > 0.33) were observed.

In sum, on the basis of our data, articulatory measurements appear to be capturing a dialect-driven difference that corresponding formant measurements do not. Also note that while the (non-significant) F2 difference for the dialect words was in the expected direction, this was not the case for the CVC sequences (with non-significant higher F2 values for the speakers from Ter Apel). Consequently, on the basis of these results, we caution those unfamiliar with articulatory analysis against interpreting formant-based patterns simply with respect to height and backness of the tongue generally, and as characterized by the anterior location of EMA flesh-points specifically.

Note that all of the aforementioned formant-based analyses consisted of mixed-effect regression models where we included type (standard CVC sequences vs. dialect words) as a fixed-effect and assessed if there were group differences (Ter Apel vs. Ubbergen) for both types separately. We included the full random-effects structure (which was also supported by the model, as determined via model comparison) consisting of random intercepts for speaker, word and segment, a by-speaker random slope for type, a by-word random slope for group, and by-segment random slopes for group, type and their interaction. For details, please refer to the supplementary materials.

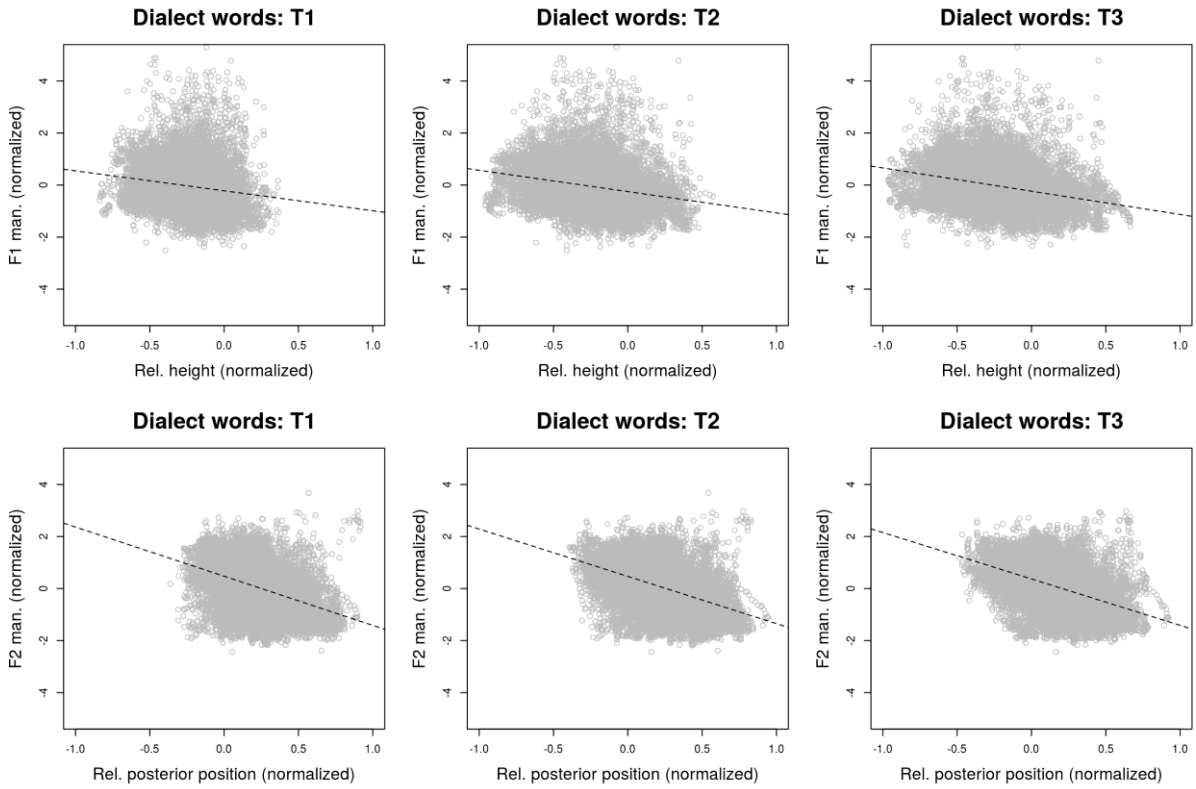

Figure 12. The three graphs in row 1 visualize the relation between normalized relative (compared to the non-speech resting position) height and normalized manually corrected F1 for the three tongue sensors for the dialect words. Those in row 2 show the same for the normalized relative posterior position and normalized manually corrected F2.

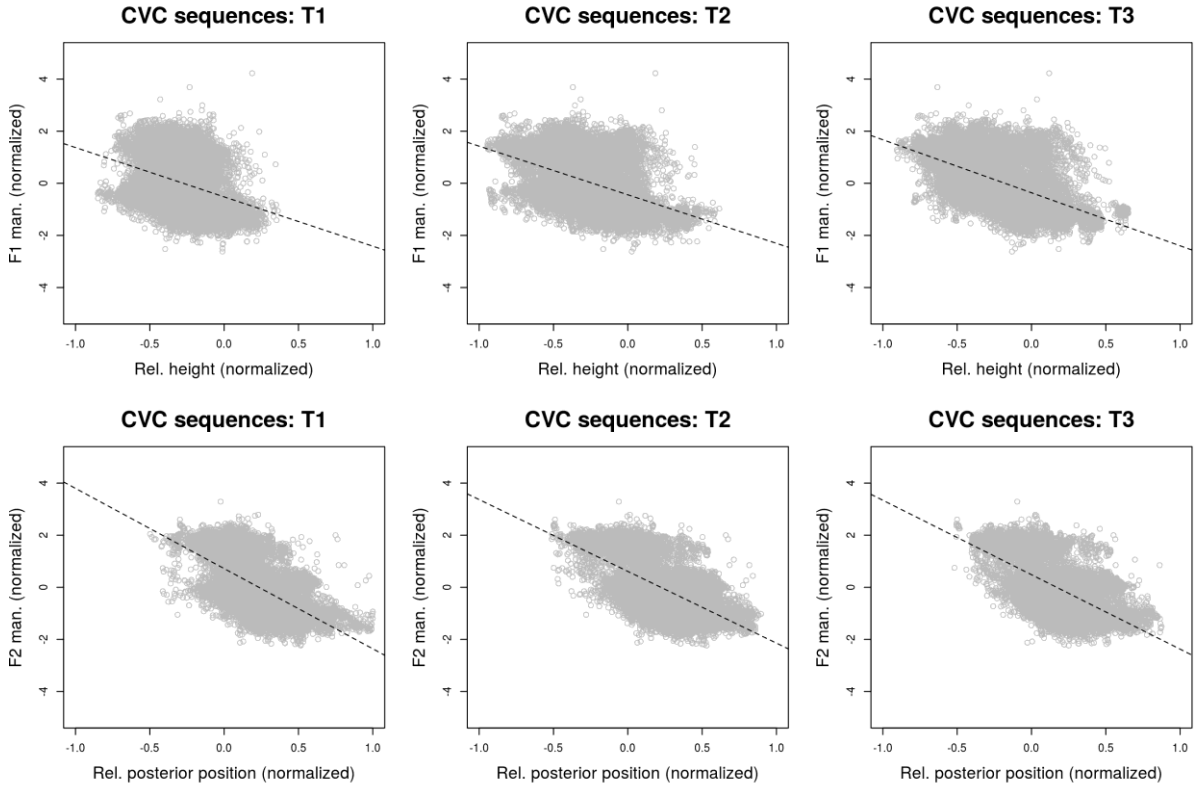

Figure 13. The three graphs in row 1 visualize the relation between normalized relative (compared to the non-speech resting position) height and normalized manually corrected F1 for the three tongue sensors for the CVC sequences. Those in row 2 show the same for normalized relative posterior position and normalized manually corrected F2.

Discussion

In this study we have illustrated the use of articulatory data for the purpose of studying dialect variation. We identified a structural difference in the position of the tongue between the two groups of speakers, with more anterior positions 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.

This difference (specifically with respect to the dialect words) may have been caused by differences in the sound segments of both dialects. For example, the Ter Apel speakers may use more segments which are pronounced further back in the mouth or fewer segments which are pronounced more in the front of the mouth than the Ubbergen speakers. When making a rough categorization from Table 1 (i.e. a dialect-by-dialect assumption) based on primary place of articulation, the proportion of front consonants /t, d, n, s, z, l/ and front variants of /r/, and vowels /i, ɪ, e, ɛ, a, y, ʏ, ø, œ/ was 0.44 for Ubbergen (0.27 for the consonants and 0.16 for the vowels), whereas it was 0.56 for Ter Apel (0.39 for the consonants and 0.17 for the vowels). Furthermore, the proportion of back consonants /k, g, ŋ, x, ɣ, ʁ/ and vowels /u, ʊ, o, ʌ, ɔ, ɑ, ɒ/ was 0.25 for Ubbergen (0.13 for the consonants and 0.12 for the vowels), while it was 0.26 for Ter Apel (0.11 for the consonants and 0.14 for the vowels). Finally, the proportion of the central vowel /ə/ and the central consonants (including the /h/ and labial consonants, as they presumably have a neutral tongue position) /p, b, m, f, v, ç, j, h/ was 0.32 for Ubbergen (0.20 for the consonants and 0.11 for the vowel) and 0.18 for Ter Apel (0.14 for the consonants and 0.04 for the vowel). Figure 14 visualizes these proportions. Ter Apel seems to be characterized by more frontal consonants than Ubbergen, while the difference with respect to front vowels and back vowels and consonants seems only limited. Consequently, this cannot explain our results, given that we observed more posterior positions for the speakers from Ter Apel compared to those from Ubbergen.

Our result also contrasts with previous findings on Dutch dialects of Adank et al. (2007) who did not find a difference in F2 for two (corresponding) groups using only a single formant measurement for monophthongs. However, Van der Harst et al. (2014) show that a dynamic approach using acoustic vowel information (F1 and F2) measured across multiple time points does help in uncovering regional differences. Using a dynamic approach, we still did not discover regional differences on the basis of formants measured at multiple time points. While this may have been caused by noise in the (unobserved) parasagittal position of the tongue, there is also no one-to-one correspondence between F1 or F2 and height or posterior position of the tongue (despite these values being frequently interpreted as such since Bell, 1867). Our results furthermore illustrate that formant-based patterns should not simply be interpreted in terms of height and backness. Additionally, these results highlight the need and use for articulatory data in studies investigating language variation.

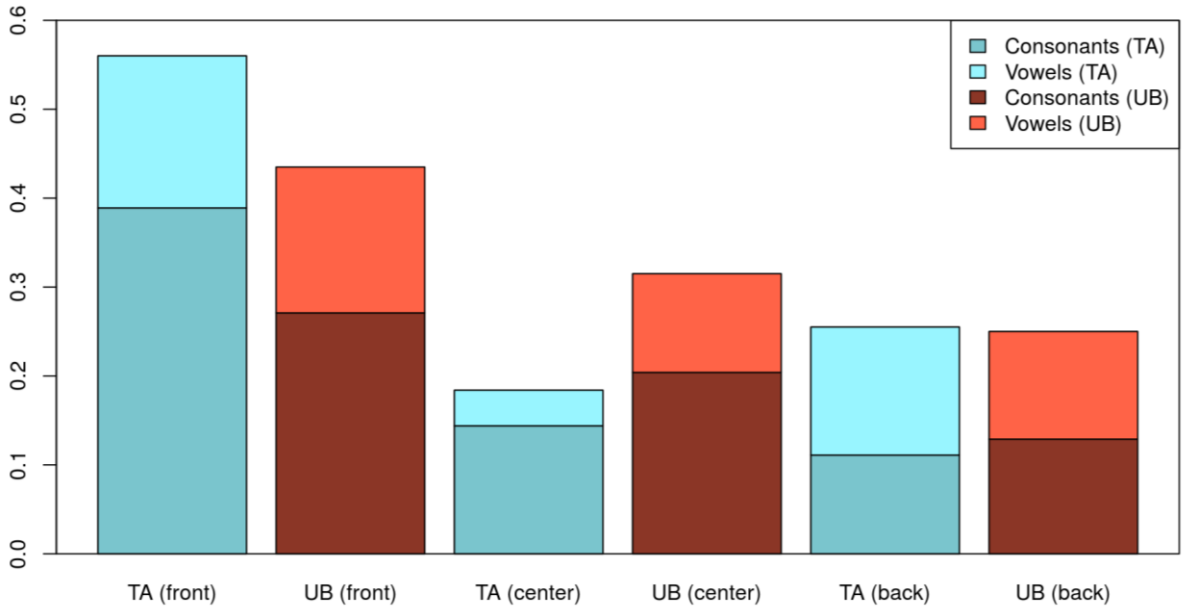

Figure 14. Proportion of front and back vowels and consonants for Ter Apel (TA) and Ubbergen (UB).

Our findings are well interpretable in the context of articulatory settings (Honikman, 1964; Laver, 1978). Honikman (1964, p. 73) defined the articulatory settings as "the overall arrangement and manoeuvring of the speech organs necessary for the facile accomplishment of natural utterance". Given that the speakers from Ter Apel showed a tongue position which was more posterior than the speakers from Ubbergen, both when pronouncing words in their own dialect and in standard Dutch, this suggests that there are distinct articulatory settings for the two dialects, also causing distinguishable accents when pronouncing standard Dutch. Whereas distinct articulatory settings have been identified for individual languages, such as English and French (Honikman, 1964; Gick et al., 2004; Wilson, 2013), no quantitative study of articulation on this scale has identified articulatory setting differences at the dialectal level before.

The generalized additive modeling approach proposed here results in a model of tongue movement over time. While we have not employed this here, the generalized additive model may also be used to determine speed and acceleration of the fitted trajectories. The approach complements other approaches used to analyze articulatory data over time, such as dynamic time warping (Sakoe & Chiba, 1978), functional data analysis (e.g., Lucero, Munhall, Gracco & Ramsay, 1997), or cross-recurrence analysis (Lancia, Fuchs and Tiede, 2014). These methods generally separate amplitude variability from phase variability when comparing articulatory trajectories. The method we propose, however, is particularly suitable when articulatory trajectories need to be compared at a higher level of aggregation for a large number of speakers. Furthermore, our approach is able to take into account individual variation, and correct for autocorrelation in the residuals.

We have shown that the results using generalized additive modeling are in line with those using linear discriminant analysis. However, there are clear benefits of generalized additive modeling over linear discriminant analysis. First, it is not necessary to create separate analyses for each individual segment, and second, the GAM analysis takes into account individual variation by treating speaker as a random-effect factor.

While we only collected data in this study for a small number of older speakers in Ter Apel, future research will focus on collecting articulatory data for more older speakers in order to adequately investigate sound change (comparing older to younger speakers) from an articulatory perspective. Furthermore, it would be very insightful to conduct a vocal profile analysis (e.g., Stuart-Smith, 1999) of the two dialects and compare that to our quantitative instrumental articulatory measurements.

Whereas the two dialects studied here show clear pronunciation differences which can usefully be studied from an acoustic or transcription-based dialectometric perspective, the aggregate articulatory perspective put forward in this study revealed interesting results, which were not identified when taking an acoustic perspective. This further shows that articulatory data is not only an essential component in an integrated account of socially-stratified variation (Lawson et al., 2011), but also for regionally-stratified variation.

Acknowledgements

This work is part of the research program *Investigating language variation physically*, which is financed by the Netherlands Organisation for Scientific Research (NWO) via a Rubicon grant awarded to Martijn Wieling (grant number 446-11-030). Furthermore, this work has benefitted from funding of the Alexander von Humboldt Professorship awarded to R. Harald Baayen (AvH grant number 1141527). First and foremost, we thank the directors and students of the "RSG Ter Apel" and "HAVO Notre Dame des Anges" for their facilitation and participation. We also thank Dankmar Enke, Matthias Villing, Lea Hofmaier, and Amber Nota for their help in segmenting the acoustic data, and Melika Oladazimi for the manual formant correction. Finally, we thank the editor and two anonymous reviewers for their comments and insights which helped to improve the manuscript substantially.

References

Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America, 116*(5), 3099-3107.

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.

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.
Baayen, R. H. (2013). Multivariate Statistics. In Podesva, R. and D. Sharma, D. (eds.), *Research Methods in Linguistics*, pp. 337-372. Cambridge: Cambridge University Press.
Barreda, S. (2015). phonTools: Functions for phonetics in R. R package version 0.2-2.1.
Bell, A. M. (1867). *Visible Speech: The Science of Universal Alphabetics Or, Self-interpreting Physiological Letters, for the Writing of All Languages in One Alphabet*. Simpkin, Marshall & Company.
Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.
Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., & Zeiliger, J. (1995). EUROM- A Spoken Language Resource for the EU. Proceedings of the 4th European Conference on Speech Communication and Speech Technology, pp. 867-870.
Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*(1), 111-140.
Clopper, C. G., & Paolillo, J. C. (2006). North American English vowels: A factor-analytic perspective. *Literary and linguistic computing*, *21*(4), 445-462.
Corneau, C. (2000). An EPG study of palatalization in French: Cross-dialect and inter-subject variation. *Language Variation and Change*, *12*(1), 25-49.
Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variancea). *The Journal of the Acoustical Society of America*, *120*(1), 407-415.
Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, *54*(1), 1-21.
Gick, B., Wilson, I., Koch, K., & Cook, C. (2005). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, *61*(4), 220-233.
Goeman, A. (1999). *T-deletie in Nederlandse dialecten. Kwantitatieve analyse van structurele, ruimtelijke en temporele variatie*. Holland Academic Graphics.
Harrington, J., Kleber, F., & Reubold, U. (2011). The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association*, *41*(02), 137-156.
Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. CRC Press.
Heeringa, W. (2004). Measuring Dialect Pronunciation Differences using Levenshtein Distance. PhD thesis, University of Groningen.
Honikman, B. (1964). Articulatory Settings. In D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, & J.L.M. Trim (Eds.), *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961*, pp. 73-84. London: Longmans, Green & Co. Ltd.
Hoole, P., & Zierdt, A. (2010). Five-dimensional articulography. *Speech motor control: New developments in basic and applied research*, 331-349.
Hoole, P., & Nguyen, N. (1999). Electromagnetic articulography. *Coarticulation–Theory, Data and Techniques, Cambridge Studies in Speech Science and Communication*, 260-269.
Kemps, R., Ernestus, M., Schreuder, R. and Baayen, R. H. (2004) Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*, 117-127.
Kerswill, P., & Wright, S. (1990). The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change*, *2*(03), 255-275.
Koos, B., Horn, H., Schaupp, E., Axmann, D., & Berneburg, M. (2013). Lip and tongue movements during phonetic sequences: analysis and definition of normal values. *The European Journal of Orthodontics*, *35*(1), 51-58.
Labov, William. (1980). The social origins of sound change. In Labov, William (ed.), *Locating language in time and space*. New York: Academic Press.
Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Vol. 1). US Regional Survey.

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. *Journal of Speech, Language, and Hearing Research*, 1-16.

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

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic/r/in Scottish English. *Journal of Sociolinguistics*, *15*(2), 256-268.

Leinonen, T. N. (2010). *An acoustic analysis of vowel pronunciation in Swedish dialects*. PhD thesis, Rijksuniversiteit Groningen.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*, 1166–1179.

Lix, L. M., & Sajobi, T. T. (2010). Discriminant analysis for repeated measures data: a review. *Frontiers in psychology*, *1*.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*, 606–608.

Lucero, J. C., Munhall, K. G., Gracco, V. L., & Ramsay, J. O. (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research*, *40*(5), 1111-1117.

Meulman, N., Wieling, M., Sprenger, S.A., Stowe, L.A., & Schmid, M.S. (2015). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. *PLOS ONE*, *10*(12: e0143328.

Nerbonne, J., Heeringa, W., Van den Hout, E., Van de Kooi, P., Otten, S., & Van de Vis, W. (1996). Phonetic distance between Dutch dialects. In: Durieux, G., Daelemans, W., & Gillis, S. (eds.), CLIN VI: Proceedings of the Sixth CLIN Meeting, Antwerp, pp. 185–202.

Nijland, L., Maassen, B., Hulstijn, W., & Peters, H. (2004). Speech motor coordination in Dutch-speaking children with DAS studied with EMMA. *Journal of Multilingual Communication Disorders*, *2*(1), 50-60.

Ooijevaar, E. (2015). Articulation and acoustics of postvocalic liquids in the Volendam dialect. Proceedings of Ultrafest VII.

Strycharczuk, P., & Sebregts, K. (2015). /r/-allophony and germination: an ultrasound study of gestural blending in Dutch. Proceedings of Ultrafest VII.

Ouni, S., Mangeonjean, L., & Steiner, I. (2012), VisArtico: a visualization tool for articulatory data, Proceedings of Interspeech 2012, September 9-13, 2012, Portland, OR, USA.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, *92*(6), 3078-3096.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175-184.

Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear/l/and dark/l: evidence from two Catalan dialects. *Journal of the International Phonetic Association*, *35*(01), 1-25.

Recasens, D., & Espinosa, A. (2007). An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association*, *37*(02), 143-172.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America*, *125*(4), 2288-2298.

Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production*. Oxford University Press.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43-49.

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 and Language*, *31*(1), 26-35.

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.
Sebregts, Koen (2015). *The Sociophonetics and Phonology of Dutch r*. PhD dissertation, University of Utrecht (LOT 379).
Stevens, K. N. (1998). Acoustic phonetics. Cambridge, MA: MIT Press.
Stuart-Smith, J. (1999b). Voice quality in Glaswegian. *Proceedings of ICPhS99*, pp. 2553-2556.
Sweet, H. (1888). *A history of English sounds from the earliest period: with full word-lists*. Clarendon Press.
Tabain, M. (2013). Research methods in speech production. In: Jones, M. & Knight, R.-A. (eds) *Bloomsbury Companion to Phonetics*, London: Bloomsbury, pp. 39-56.
Tomaschek, F., Tucker, B. V., Wieling, M., & Baayen, R. H. (2014). Vowel articulation affected by word frequency. Proceedings of the 10th ISSP, Cologne, pp. 429-432.
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. Proceedings of the 14th Interspeech, Lyon, pp. 1302-1306.
Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, 151-173.
Van Rij, J., Hollebrandse, B., & Hendriks, P. (2016a). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In: Holler, A., Goeb, C., & Suckow, K. (eds.) *Experimental Perspectives on Anaphora Resolution. Information Structural Evidence in the Race for Salience*. Boston: De Gruyter, pp. 267-293.
Van Rij, J., Wieling, M., Baayen, R.H., van Rijn, H. (2016b). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package, version 2.2.
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. *Language Variation and Change*, *26*(2), 247-272.
Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-Lehouillier, H., Vatikiotis-Bateson. E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech Language and Hearing Research*, *48*, 543-553.
Wieling, M., Heeringa, W., & Nerbonne, J. (2007). An aggregate analysis of pronunciation in the Goeman-Taeldeman-Van Reenen-Project data. *Taal en Tongval*, *59*, 84–116.
Wieling, M., Margaretha, E., & Nerbonne, J. (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics*, *40*(2), 307-314.
Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R.H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669-692.
Wieling, M., & Nerbonne, J. (2011). Bipartite spectral graph partitioning for clustering dialect varieties and detecting their linguistic features. *Computer Speech and Language*, *25*(3), 700-715.
Wieling, M., & Nerbonne, J. (2015). Advances in Dialectometry. *Annual Review of Linguistics*, *1*(1).
Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLOS ONE*, *6*(9), e23613.
Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., & Baayen, R. H.. (2015). Investigating dialectal differences using articulography. Proceedings of ICPhS 2015, Glasgow.
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.
Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95-114.
Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.
Wood, S. N., Goude, Y., & Shaw, S. (2014). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.
Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica*, *64*(2), 94-102.

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