Journal of Phonetics 59 (2016) 122–143

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

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], R. Harald Baayen[b,e]

[a] Department of Humanities Computing, University of Groningen, Oude Kijk in't Jatstraat, 9712 EK Groningen, The Netherlands
[b] Department of Quantitative Linguistics, University of Tübingen, Germany
[c] Haskins Laboratories, Yale University, United States
[d] Department of Statistics, University of Bristol, United Kingdom
[e] Department of Linguistics, University of Alberta, Canada

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form
21 September 2016
Accepted 22 September 2016
Available online 4 November 2016

Keywords:
Articulography
Dialectology
Generalized additive modeling
Articulatory settings

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.

1. 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, van Hout, & Van de Velde, 2007; Van der Harst, Van de Velde, & Van Hout, 2014). Since the seminal study of Peterson and 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 & 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, Kleber, & Reubold, 2011).

Labov, Yaeger, and Steiner (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, Ash, and Boberg (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, Smits and Van Hout (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

* Corresponding author.
E-mail address: wieling@gmail.com (M. Wieling).

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2016.09.004

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 for 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, Margaretha, & Nerbonne, 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, Ernestus, Schreuder, & Baayen, 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 & 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, Scobbie, and Stuart-Smith (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 & 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 & 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.,