Comparing L1 and L2 speakers using articulography

Martijn Wieling[1], Pauline Veenstra[1], Patti Adank[2], Andrea Weber[3] and Mark Tiede[4]

[1]University of Groningen, [2]University College London, [3]University of Tübingen, [4]Haskins Laboratories
m.b.wieling@rug.nl, pauline.veenstra@gmail.com, p.adank@ucl.ac.uk, andrea.weber@uni-tuebingen.de, tiede@haskins.yale.edu

ABSTRACT

This study uses articulography, the measurement of the position of tongue and lips during speech, as a tool to quantitatively assess the differences between pronunciations of native and non-native (Dutch) speakers of English. In our study, we focus on two pairs of English sound contrasts: /s/-/ʃ/ and /t/-/θ/. Our analysis focuses on the anterior-posterior position of the tongue tip during the pronunciation of minimal pairs containing the contrasting sounds. Our results indicate that the contrast between /s/ and /ʃ/ made by the Dutch L2 speakers is slightly reduced compared to the contrast produced by the English L1 speakers. For the contrast /t/-/θ/, our findings show that while native English speakers clearly produce this contrast, Dutch speakers do not. Our results line up with earlier studies on the basis of acoustic data, and also illustrate that articulography is a suitable method of investigating pronunciation differences between first and second language speakers.

Keywords: Articulography, Second language acquisition, Mixed-effects regression

1. INTRODUCTION

Especially when learning begins at a later age, second language (L2) learners typically have a noticeable first language (L1) accent in their pronunciation of the target L2 [10]. Current speech learning models such as Flege's Speech Learning Model (SLM; [9]) or Best's Perceptual Assimilation Model (PAM; [5]) capture difficulties in L2 pronunciation by taking into account the phonetic similarity of (or contrasts between) sound segments in the L1 and L2 language inventories. For example, the SLM predicts that segments in the L2 which are sufficiently different from the segments in the native language are easier to learn than those which are relatively similar (i.e. the former map to a new sound segment category, while the latter merge with an existing category).

Instead of focusing on acoustic and perceptual differences, it is also possible to focus on the differences between the underlying articulatory gestures (i.e. the movement of lips and tongue needed for the production of speech [7]). While many studies have investigated the effect of providing visual feedback based on movements of the articulators on second language learning [4, 11, 17, 19], only few studies have studied L2 pronunciations from the perspective of the movements of the speech articulators. Nissen and colleagues [21] studied differences in tongue movements in Spanish and Korean bilingual speakers, while Chakraborty and Goffman [8] looked at kinematic measures (i.e. lip and jaw movement) of stress in non-native (Bengali) speakers of American English. Furthermore, while there are an increasing number of corpora which contain articulatory data (e.g., the Edinburgh DoubleTalk corpus [23]), only a single corpus containing L2 speakers exists to date (i.e. the EMA-MAE corpus containing Mandarin Accented English [15]).

The goal of the present study is to fill this gap by investigating differences in the articulation between native and non-native (Dutch) speakers of English. Given the suitability of electromagnetic articulography (EMA; [13, 14, 22]) to track and quantify the movement of several sensors attached to tongue and lips, we will employ this method here.

In this study, we focus on two different native English sounds: /ʃ/ and /θ/, both of which are not included in the phonemic inventory of Dutch [6]. When speaking English, Dutch speakers tend to substitute /θ/ (acoustically and perceptually) with [t] [12, 27]. For this reason, we will contrast the articulation of words containing /θ/ to similar words containing /t/ instead (i.e. minimal pairs). The fricative /ʃ/ can be seen as an allophone of /s/ in the Dutch language (though it does occur in loan words from English, such as 'match') [16]. Indeed, Johnson and Babel [16] showed that Dutch L2 speakers of English perceived a smaller contrast between the two sounds than native English speakers. Consequently, we will contrast the articulation of words containing /ʃ/ with (minimal pair) words containing /s/.

In the following, we discuss the methods and results obtained in this study.

2. DATA COLLECTION

Our study was conducted in 2014 in the Netherlands and England. A total of 21 native Dutch participants (13 male, 8 female, mean age: 21) were recruited at the Department of Psychology of the University of Groningen in the Netherlands. In England, 22 native (Southern Standard British) English speakers (8 male, 14 female, mean age: 25) were recruited at University College London. Data was collected onsite in Groningen and London. The study was approved by the Ethics Committee Psychology in Groningen and the UCL Ethics Committee in London. Before participating, participants were informed about the nature of the experiment and required to sign an informed consent form. Each data collection session lasted about 90 minutes and participants were compensated with SONA credits or money (£ 15).

The articulography data was collected with a portable NDI Wave 16-channel articulography device with a sampling rate of 100 Hz. Using the NDI WaveFront articulography data capturing software, positional data was automatically synchronized with the audio signal (recorded at 22.05 kHz using an Audio-Technica AT875R microphone). The data was subsequently corrected for head movement via a 6D reference sensor attached to each speaker's forehead. The microphone and the NDI Wave system were connected to the control laptop via a Roland Quad-Capture USB Audio interface. To make the articulatory data comparable across speakers, a separate biteplate recording (containing 3 sensors) was used to rotate the data of each speaker relative to the maxillary occlusal plane [26]. We attached a total of three sensors to the midline of each speaker's tongue using Cyano Veneer Fast dental glue. Before attaching the tongue sensors, we first glued a small (diameter of 0.5 cm), flexible, very thin transparent layer of polyethylene (i.e. plastic) to the bottom of the sensor, which was then glued to the tongue. By adding this layer (with a larger gluing area), sensors did not come off as easily as without the additional layer. One sensor was placed as far backward on the tongue as possible without causing discomfort for the speaker. Another sensor was placed about 0.5 cm behind the tongue tip. The final sensor was placed midway between the other two sensors. Besides the three tongue sensors, additional sensors were placed on the lips and jaw. For this study, however, we only focus on data from the anterior tongue sensor (T1). Attaching all sensors took about 30 minutes. Whenever sensors came off during the course of the experiment, they were reattached.

During the experiment, the English participants had to read aloud various words, non-words, sentences and paragraphs of text in English, while the Dutch participants had to read aloud in English and Dutch. For the present study, we only present data on the English pronunciations of the minimal pairs for /s/-/ʃ/ (11 pairs), and /t/-/θ/ (10 pairs). The complete list of items associated with these sounds is shown in Table 1. Generally, all words were pronounced twice. Each individual word was shown separately on a computer screen, surrounded on both sides by a schwa (ə). Participants were instructed to pronounce these as the corresponding sound (e.g., 'ə crust ə'). This procedure was used to ensure a neutral articulatory starting position when pronouncing the individual words.

Table 1: List of minimal pairs used in this study.

/s/-/ʃ/	/t/-/θ/
crust - crushed	fate - faith
fist - fished	fort - forth
lease - leash	mitt - myth
plus - plush	kit - kith
mess - mesh	tank - thank
rust - rushed	team - theme
save - shave	tent - tenth
seat - sheet	tick - thick
self - shelf	ties - thighs
sign - shine	tongs - thongs
sun - shun	

3. PREPROCESSING AND ANALYSIS

Two male Dutch speakers were excluded from the analysis, as they did not finish the part of the experiment where the minimal pairs shown in Table 1 had to be pronounced (due to sensor attachment problems).

The data for each speaker was manually segmented acoustically at the word level (including the preceding and following schwa). Tongue movement data which was not associated with the pronunciation of the study material was excluded. Given that the sound pairs used in this study mainly contrasted in the anterior-posterior direction, we focused on the movement data for this (x) axis only.

To enable a fair comparison between speakers, the positional information was normalized for each speaker in such a way that 0 in the x-direction indicated the most frontal (anterior) position of the three tongue sensors, while 100 in this direction indicated the position furthest back (posterior) in the mouth. These extremes were based on the pronunciation of all words by the speaker. Clear outliers were removed, and therefore not considered as the maximum or minimum point.

Subsequently, for each word per subject we calculated the average anterior position of the T1 sensor during its pronunciation. We then fitted two separate mixed-effects regression models [3] using the *lme4* package (version 1.1.7) in *R*. The first model focused on /s/-/ʃ/ contrast, while the second model focused on the /t/-/θ/ contrast. The dependent variable of both models was the average anterior position of the T1 sensor (for each word pronunciation). The (fixed-effect) predictors we included were group (English or Dutch) and the word category (/s/ versus /ʃ/ for the first model, and /t/ versus /θ/ for the second model). We also included the interaction between the two predictors, as we are interested in the difference between the two languages with respect to distinguishing the two word categories.

To account for the (random-effect) variation in tongue position associated with speakers and words, we included random intercepts for speaker and word. As there may be individual variation in how large the difference in average tongue position is for the two categories, we included a by-subject random slope for word category. Finally, to take into account that the difference between Dutch and English speakers in average tongue position might vary per word, we included a by-word random slope for group. Via AIC (Akaike Information Criterion [1]) comparisons (where a reduction in AIC of at least 2 indicates that the higher complexity of the model is warranted compared to the simpler model) we assessed if the inclusion of random intercepts and slopes was necessary.

After determining the best model, we assessed if the predictors remained significant when model criticism was applied (see [2], Ch. 6.2.3). With model criticism, the model is refitted on the data excluding those data points with which the model has trouble fitting, limiting the influence of these problematic outliers.

4. RESULTS

The number of cases in our data set (i.e. the subject-word combinations) for the /s/-/ʃ/ contrast was equal to 1865. Model comparison revealed that the random-effects structure of the /s/-/ʃ/ model required random intercepts for word and participant, as well as a by-participant random slope for word category and a by-word random slope for group. Table 2 shows the fixed-effects structure of the /s/-/ʃ/ model. As model criticism did not change the significance of the predictors substantially, the results shown in Table 2 are based on all 1865 cases.

The interpretation of Table 2 is as follows. The intercept indicates the average position (of the T1 sensor in the *x* direction) for the English speakers for the words in the /s/ category. The second line, Category /ʃ/, indicates that for the English speakers, the words including /ʃ/ as opposed to /s/ are pronounced significantly further back in the mouth (higher values indicate a more posterior value). The third line, Group NL, indicates that the T1 position of the Dutch speakers is significantly more frontal than that of the English speakers for the /s/-words. The final line, Category /ʃ/ : Group NL, indicates how the position difference between the /s/ and /ʃ/-words for the English speakers needs to be changed to fit the Dutch speakers. The negative estimate indicates that the distinction between the two sounds is smaller (marginally significant) than for the English speakers. Figure 1 visualizes this relationship. Note that while the English speakers have a more posterior tongue position, this is likely related to the other phonemes present in the words and not to the sound contrast /s/-/ʃ/.

The number of cases in our data set (i.e. the subject-word combinations) for the /t/-/θ/ contrast was equal to 1575. As model criticism did not change the significance of the predictors substantially, the results shown in Table 3 are based on all 1575 cases. Random intercepts were included for word and participant, as well as a by-participant random slope for word category (a by-word random slope for group was not needed).

Table 2: Fixed-effects structure of the /s/-/ʃ/ model.

	Estimate	Std. Error	*p*-value
Intercept	27.05	1.16	< .001
Category /ʃ/	3.65	1.11	.001
Group NL	-3.85	1.39	.005
Category /ʃ/ : Group NL	-1.47	0.84	.08

Figure 1: Posterior position of T1 sensor during pronunciation of words containing either /s/ or /ʃ/. *p*-values are based on the mixed-effects regression models fitted with the appropriate reference levels.

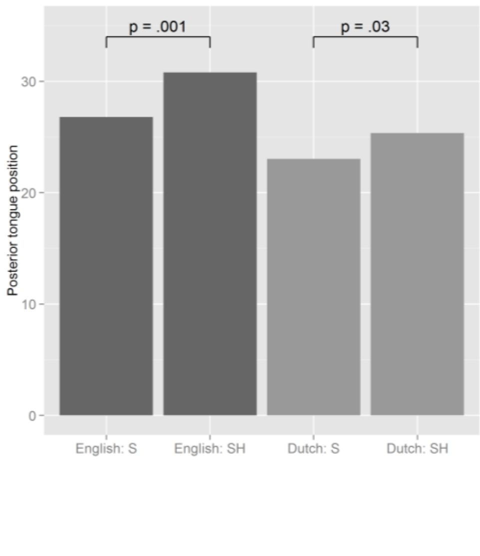
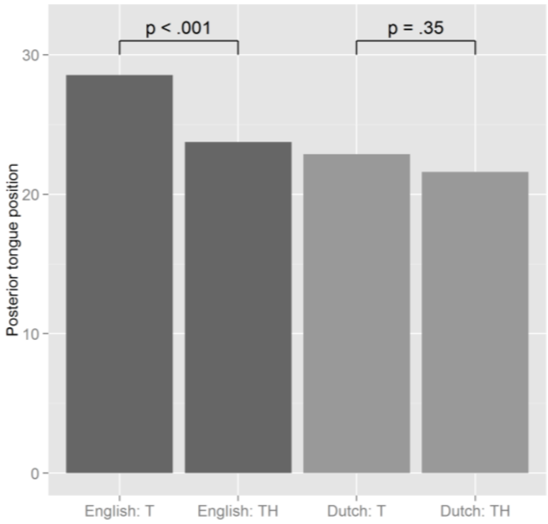


Table 3: Fixed-effects structure of the /t/-/θ/ model.

	Estimate	Std. Error	*p*-value
Intercept	28.52	1.26	< .001
Category /θ/	-4.80	1.26	< .001
Group NL	-5.64	1.51	< .001
Category /θ/ : Group NL	3.58	1.06	< .001

Figure 2: Posterior tongue position of T1 sensor during pronunciation of words containing either /t/ or /θ/. Only the English speakers show a significant difference between /t/ and /θ/. *p*-values are based on the mixed-effects regression models fitted with the appropriate reference levels.



The interpretation of Table 3 is as follows. The intercept indicates the average position (of the T1 sensor in the *x* direction) for the English speakers for the words in the /t/ category. The second line, Category /θ/, indicates that for the English speakers, the words including /θ/ as opposed to /t/ are pronounced significantly more frontal (lower values indicate a more anterior value). The third line, Group NL, indicates that the T1 position of the Dutch speakers is significantly more frontal than that of the English speakers for the /t/-words. The final line, Category /θ/ : Group NL, indicates how the position difference between the /t/ and /θ/-words for the English speakers needs to be changed to fit the Dutch speakers. The estimate is therefore the correction on the (negative) estimate in the second line of Table 3. Consequently, the clear negative difference between /t/ and /θ/ shown by the native English speakers is much less negative (i.e. less strong) for the Dutch speakers. In fact, the difference for the Dutch speakers between the /t/ and /θ/ would be equal to -4.80 + 3.58 = -1.22. A subsequent test showed that this difference is not significant (p = .35). That is, Dutch speakers do not distinguish /t/ from /θ/, whereas native English speakers do distinguish between /t/ and /θ/. Figure 2 visualizes this relationship. Again, while the English speakers have a more posterior tongue position, this is likely related to the other phonemes present in the words and not to the sound contrast /t/-/θ/.

5. DISCUSSION

In this study we have illustrated the use of articulatory data for the purpose of investigating pronunciation differences between L1 and L2 speakers. We found a small, marginally significant difference in how Dutch and English speakers distinguished /s/ from /ʃ/, with Dutch speakers showing a smaller difference in the anterior-posterior position of the T1 tongue (tip) sensor compared to native English speakers. We found a clear, significant difference in how Dutch and English speakers distinguished /t/ from /θ/. English speakers showed a clear contrast, with /θ/ pronounced more anterior than /t/, but Dutch speakers showed no significant difference between the two sounds in the anterior-posterior position of the T1 tongue sensor.

In the context of Flege's Speech Learning Model [9] our results suggest that /θ/ has merged with /t/ for Dutch speakers. (At least, on the basis of the anterior-posterior position of the T1 sensor.) Furthermore, our results line up with the findings reported in [12] and [27] who found that Dutch speakers substituted /θ/ most frequently with /t/. Of course in these studies a categorical distinction was made, whereas in our study a more sensitive and gradual measure of difference was taken into account.

The /s/-/ʃ/ results are in line with the perceptual results reported in [16]. Dutch speakers appear to show a smaller contrast between these two sounds, both in production and perception.

Of course, the analysis we employed here was rather crude. We only focused on a single sensor in a single dimension and obtained an average position across the whole word pronunciation. More advanced methods which take into account the whole (non-linear) trajectory (such as [18, 20, 24, 25]) are likely to reveal additional insights in the specific differences between the native and non-native speakers.

6. REFERENCES

[1] Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716-723.

[2] Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

[3] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390-412.
[4] Badin, P., Tarabalka, Y., Elisei, F., Bailly, G. 2010. Can you ‘read’ tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Commun. 52*, 493-503.
[5] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Timonium: York Press, 171-204.
[6] Booij, G. 1995. *The phonology of Dutch*. Oxford: Oxford University Press
[7] Browman, C. P., Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica* 49, 155-180.
[8] Chakraborty, R., Goffman, L. 2011. Production of lexical stress in non-native speakers of American English: Kinematic correlates of stress and transfer. *J. Speech Lang. Hear. Res.* 54, 821-835.
[9] Flege, J. 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Timonium: York Press, 233-277.
[10] Flege, J., Munro, M., MacKay, I. 1995. Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am.* 97, 3125-3134.
[11] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In: Hansen, J., Zampini, M. (eds), *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, 309-322
[12] Hanulikova, A., Weber, A. 2012. Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Atten. Percept. Psychophys*. 74, 613-629.
[13] Hoole, P., Nguyen, N. 1999. Electromagnetic articulography. In: Hardcastle, W.H., Hewlett, N. (eds), *Coarticulation: Theory, Data and Techniques.* Cambridge University Press, 260-269.
[14] Hoole, P., Zierdt, A. 2010. Five-dimensional articulography. In: Maassen, B., van Lieshout, P. (eds), *Speech Motor Control: New Developments in Basic and Applied Research*. OUP, 331-349.
[15] Ji, A., Berry, J. J., Johnson, M. T. 2014. The Electromagnetic Articulography Mandarin Accented English (EMA-MAE) corpus of acoustic and 3D articulatory kinematic data. *Proc. 39th Int. Conf. Acoust. Speech Signal Process.* Florence, 7719-7723.
[16] Johnson, K., Babel, M. 2010. On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *J. Phonetics* 38, 127-136.
[17] Kröger, B., Graf-Borttscheller, V., Lowit, A. 2008. Two and three-dimensional visual articulatory models for pronunciation training and for treatment of speech disorders. *Proc. 9th Interspeech* Brisbane, 22-26.
[18] Lancia, L., Fuchs, S., Tiede, M. 2014. Application of concepts from cross-recurrence analysis in speech production: An overview and comparison with other nonlinear methods. *J. Speech Lang. Hear. Res.* 57, 718-733.
[19] Levitt, J. S., Katz, W. F. 2010. The effects of EMA-based augmented visual feedback on the English speakers’ acquisition of the Japanese flap: A perceptual study. *Proc. 10th Interspeech* Makuhari, 1862-1865.
[20] Lucero, J. C., Munhall, K. G., Gracco, V. L., Ramsay, J. O. 1997. On the registration of time and the patterning of speech movements*. J. Speech Lang. Hear. Res* 40, 1111-1117.
[21] Nissen, S. L., Dromey, C., Wheeler, C. 2008. First and second language tongue movements in Spanish and Korean bilingual speakers. *Phonetica* 64, 201-216.
[22] Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., Conrad, B. 1987. Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain Lang.* 31, 26-35.
[23] Scobbie, J. M., Turk, A., Geng, C., King, S., Lickley, R. J., Richmond, K. 2013. The Edinburgh Speech Production Facility DoubleTalk Corpus. *Proc. 14th Interspeech* Lyon, 764-766.
[24] Tomaschek, F., Tucker, B. V., Wieling, M., Baayen, R. H. 2014. Vowel articulation affected by word frequency. *Proc. 10th ISSP* Cologne, 429-432.
[25] Tomaschek, F., Wieling, M., Arnold, D., Baayen, R. H. 2013. Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. *Proc. 14th Interspeech* Lyon, 1302-1306.
[26] Westbury, J. R. 1994. On coordinate systems and the representation of articulatory movements. *J. Acoust. Soc. Am.* 95, 2271–2273.
[27] Wester, F., Gilbers, D., Lowie, W. 2007. Substitution of dental fricatives in English by Dutch L2 speakers. *Lang. Sci.* 29, 477-491.