Comparing dialect and accented pronunciations on the basis of transcriptions and articulography

Martijn Wieling

University of Groningen, The Netherlands

LDC, September 12, 2014

Overview

- Obtaining sensitive pronunciation distances automatically
  - The need for sensitive sound segment distances
  - Obtaining sensitive sound segment distances
  - Evaluation at sound level and aggregate level
  - Applications
- Comparing pronunciations using articulography
  - Articulography introduction
  - Study 1: Native vs. non-native English
  - Study 2: Comparing Dutch dialects

Collaborators












The need for sensitive segment distances

- In our research on language variation, we employ pronunciation distances (on the basis of alignments)
- We would like to improve alignment quality and the distances
- There is no widely accepted procedure to determine phonetic similarity (Laver, 1994)
- Here we use the distribution of pronunciation variation to determine similarity
- In line with language as *un systême oû tout se tient* (focus on relations between items, not items themselves; Meillet, 1903)

Our starting point: the Levenshtein distance

(restriction: vowels are not aligned with consonants)

- The Levenshtein distance measures the minimum nr. of insertions, deletions and substitutions to transform one string into another

mɔəlkə	delete ɔ	1
məlkə	subst. ə/ɛ	1
mɛlkə	delete ə	1
mɛlk	insert ə	1
mɛlək		
		4

m	ɔ	ə	l		k	ə
m		ɛ	l	ə	k	
	1	1		1		1

- Note the implicit identification of sound correspondences

Our starting point: the Levenshtein distance

(restriction: vowels are not aligned with consonants)

- The Levenshtein distance measures the minimum nr. of insertions, deletions and substitutions to transform one string into another

mɔəlkə	delete ɔ	1
məlkə	subst. ə/ɛ	1
mɛlkə	delete ə	1
mɛlk	insert ə	1
mɛlək		
		4

m	ɔ	ə	l		k	ə
m		ɛ	l	ə	k	
	1	1		1		1

- Note the implicit identification of sound correspondences

Counting sound segment correspondences

- Counting the frequency of sound segments (in the Levenshtein alignments)

p	b	...	ʊ	u	Total
5×10^5	2×10^5	...	90,000	9×10^5	10^8

- Counting the frequency of the aligned sound segments (in the Levenshtein alignments)

	p	b	...	ʊ	u	
p	2×10^5	10,650	...	0	0	
b		88,000	...	0	0	
⋮			⋮	⋮	⋮	
ʊ				65,400	5,500	
u					4×10^5	
						Total: 5×10^7

- Probability of observing [p]: 5×10^5 / 10^8 = 0.005 (0.5%)
- Probability of observing [b]: 2×10^5 / 10^8 = 0.002 (0.2%)
- Probability of observing [p]:[b]: 10,650 / 5×10^7 = 0.0002 (0.02%)

Association strength between segment pairs

- Pointwise Mutual Information (PMI): assesses degree of statistical dependence between aligned segments (x and y)

$$\mathrm{PMI}(x, y) = \log_2 \left(\frac{p(x, y)}{p(x)\, p(y)} \right)$$

- $p(x, y)$: relative occurrence of the aligned segments x and y in the whole dataset
- $p(x)$ and $p(y)$: relative occurrence of x and y in the whole dataset
- The greater the PMI value, the more segments tend to cooccur in correspondences

Association strength between segment pairs

- Probability of observing [p]:[b]: 10,650 / 5×10^7 = 0.0002
- Probability of observing [p]: 5×10^5 / 10^8 = 0.005
- Probability of observing [b]: 2×10^5 / 10^8 = 0.002

$$\mathrm{PMI}(x, y) = \log_2\left(\frac{p(x, y)}{p(x)\, p(y)}\right) \Rightarrow$$

$$\mathrm{PMI}([\mathrm{p}], [\mathrm{b}]) = \log_2\left(\frac{0.0002}{0.005 \times 0.002}\right)$$

$$\mathrm{PMI}([\mathrm{p}], [\mathrm{b}]) \approx 4.3$$

Using PMI values with the Levenshtein algorithm

- Idea: use association strength to weight edit operations
- PMI is large for strong associations, so invert it (0 - PMI)
  - Strongly associated segments will have a low distance
- PMI range varies, so normalize it between 0 and 1
- Use PMI-induced weights as costs in Levenshtein algorithm
  - Cost of substituting identical sound segments is always set to 0

The PMI-based Levenshtein algorithm

- We use the standard Levenshtein algorithm to calculate the initial PMI weights and convert these to costs (i.e. sound distances)
- These sensitive sound distances are then used as edit operation costs in the Levenshtein algorithm to obtain new alignments, new counts, and new PMI sound distances
- This process is repeated until alignments and PMI sound distances stabilize
- Besides improved alignments (Wieling et al., 2009), this procedure automatically yields sensitive sound segment distances

m	ɔ	ə	l		k	ə
m		ɛ	l	ə	k	
	0.20	0.15		0.12		0.12

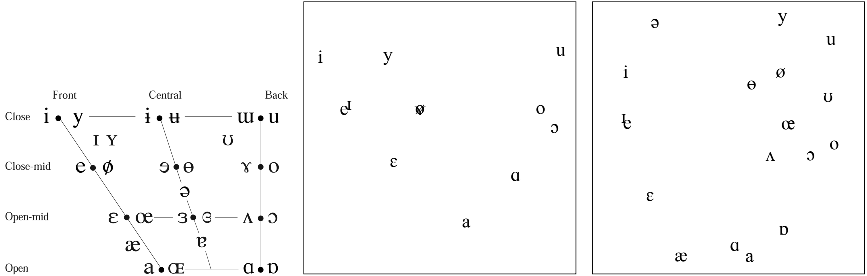
The PMI-based Levenshtein algorithm

- We use the standard Levenshtein algorithm to calculate the initial PMI weights and convert these to costs (i.e. sound distances)
- These sensitive sound distances are then used as edit operation costs in the Levenshtein algorithm to obtain new alignments, new counts, and new PMI sound distances
- This process is repeated until alignments and PMI sound distances stabilize
- Besides improved alignments (Wieling et al., 2009), this procedure automatically yields sensitive sound segment distances

m	ɔ	ə	l		k	ə
m		ɛ	l	ə	k	
	0.20	0.15		0.12		0.12

Evaluation at the sound segment level: Dutch

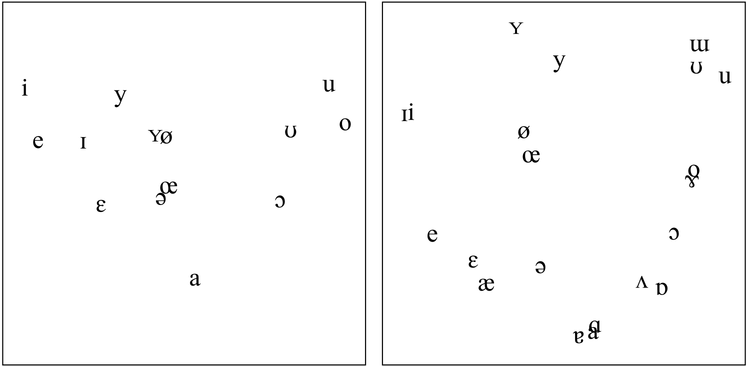
(MDS visualization of PMI distances captures 76% of the variation)



(a) IPA (b) Acoustics (F1 and F2) (c) PMI distances

Evaluation at the sound segment level: German

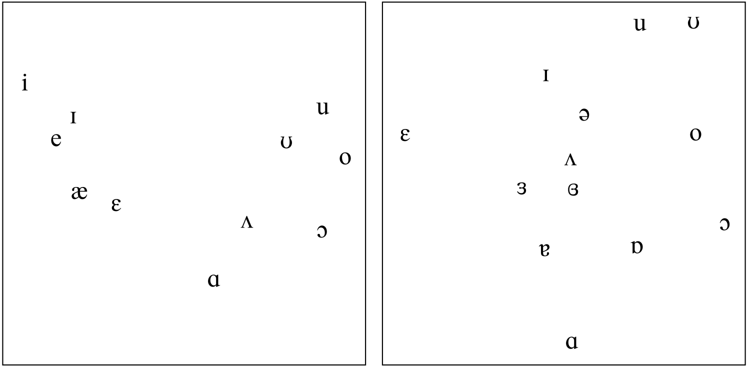
(MDS visualization of PMI distances captures 70% of the variation)



(a) Acoustics (b) PMI distances

Evaluation at the sound segment level: U.S. English

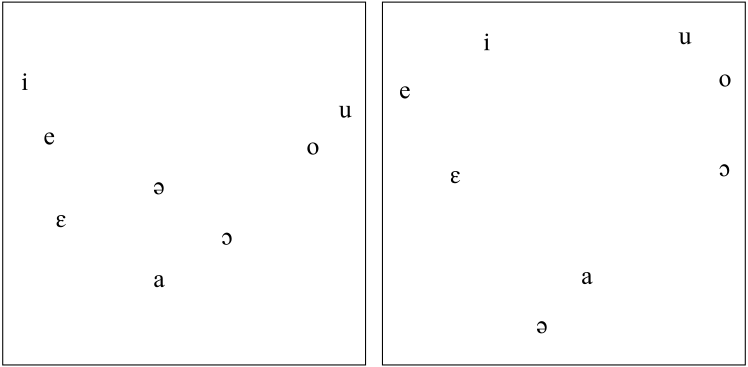
(MDS visualization of PMI distances captures 65% of the variation)



(a) Acoustics (b) PMI distances

Evaluation at the sound segment level: Bantu

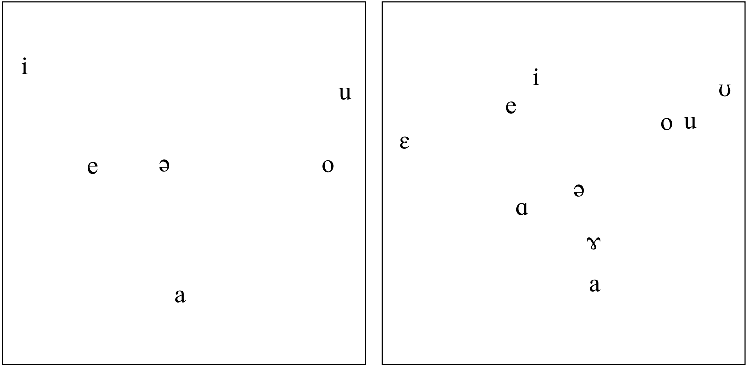
(MDS visualization of PMI distances captures 90% of the variation)



(a) Acoustics (b) PMI distances

Evaluation at the sound segment level: Bulgarian

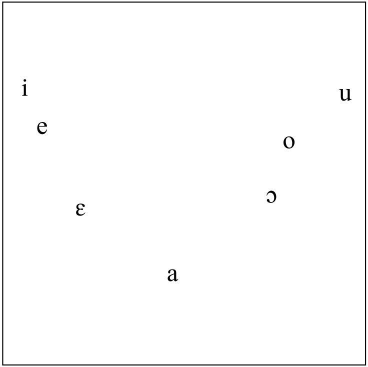
(MDS visualization of PMI distances captures 86% of the variation)



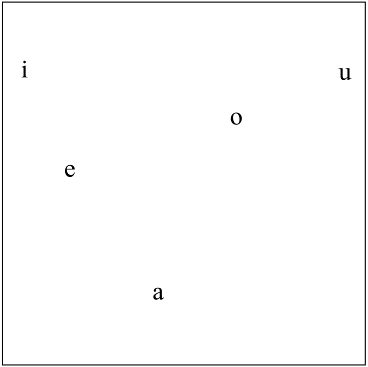
(a) Acoustics (b) PMI distances

Evaluation at the sound segment level: Tuscan

(MDS visualization of PMI distances captures 97% of the variation)



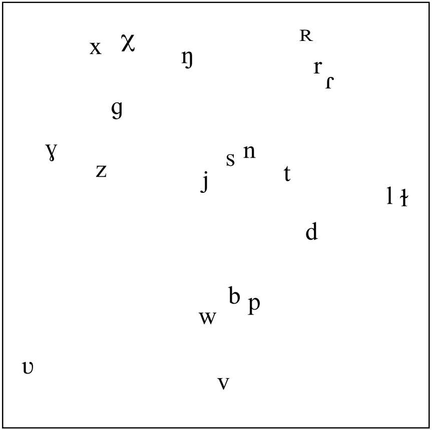
(a) Acoustics



(b) PMI distances

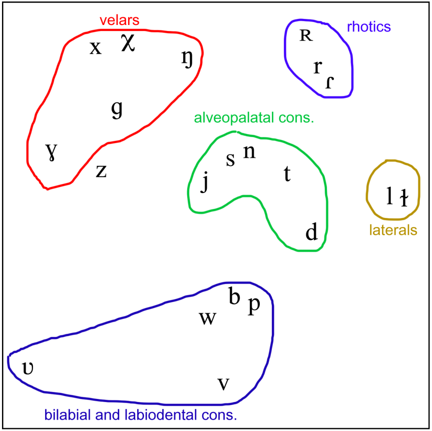
MDS visualization of Dutch consonants

(MDS visualization of PMI distances captures 50% of the variation)



MDS visualization of Dutch consonants

(place (3 groups) dominates over manner (2 groups) and voicing (no groups))



Quantitative evaluation of vowel distances

(Wieling, Margaretha & Nerbonne, 2012, *Journal of Phonetics*)

	Pearson's r	Explained variance (r^2)
Dutch	0.672	45.2%
Dutch w/o Frisian	0.686	47.1%
German	0.633	40.1%
German w/o ə	0.785	61.6%
US English	0.608	37.0%
Bantu	0.642	41.2%
Bulgarian	0.677	45.8%
Tuscan	0.758	57.5%

- In the following, the PMI-based method is used to obtain pronunciation distances on the basis of English accented speech

Quantitative evaluation of vowel distances

(Wieling, Margaretha & Nerbonne, 2012, *Journal of Phonetics*)

	Pearson's r	Explained variance (r^2)
Dutch	0.672	45.2%
Dutch w/o Frisian	0.686	47.1%
German	0.633	40.1%
German w/o ə	0.785	61.6%
US English	0.608	37.0%
Bantu	0.642	41.2%
Bulgarian	0.677	45.8%
Tuscan	0.758	57.5%

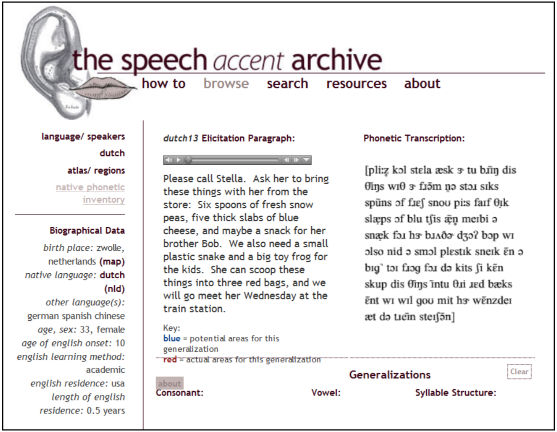
- In the following, the PMI-based method is used to obtain pronunciation distances on the basis of English accented speech

The Speech Accent Archive

(available online at http://accent.gmu.edu)

George Mason University

Audio example



Listen to an example

Validating the PMI-based Levenshtein distances

- There is only a single study investigating the relation between Levenshtein pronunciation distances and perceptual distances
  - Gooskens and Heeringa (*LVC*, 2004)
  - Focusing on Norwegian dialects
  - The reported correlation strength was $r \approx 0.7$
- We conducted a new study based on the Speech Accent Archive, investigating the relation between perceptual and Levenshtein pronunciation distances

Outline of the perception experiment

- We provided six sets of 50 audio samples (286 distinct speakers; 272 non-native) and asked native American English speakers to rate their native-likeness on a scale from 1 (very foreign sounding) to 7 (native American English)
- More than 1100 (!) people filled in the questionnaire

Results of the perception experiment

(Wieling et al., forthcoming, *Language Dynamics and Change*)

- We calculated the average native-likeness score for every speaker
- We used the PMI-based Levenshtein algorithm to obtain the automatic pronunciation distances for each of these speakers and the average U.S. speaker (based on 115 samples of native U.S. English speakers born in the U.S.)
  - The distance between two speakers is calculated by averaging the word pronunciation distances between the two speakers
- The correlation between the log-transformed PMI-based Levenshtein distance and the average native-likeness was r = **-0.81**
  - Average agreement of individual raters with average ratings: $r = 0.84$
- We may conclude that the Levenshtein distance is a valid measure to determine perceptually sound pronunciation distances

Results of the perception experiment

(Wieling et al., forthcoming, *Language Dynamics and Change*)

- We calculated the average native-likeness score for every speaker
- We used the PMI-based Levenshtein algorithm to obtain the automatic pronunciation distances for each of these speakers and the average U.S. speaker (based on 115 samples of native U.S. English speakers born in the U.S.)
  - The distance between two speakers is calculated by averaging the word pronunciation distances between the two speakers
- The correlation between the log-transformed PMI-based Levenshtein distance and the average native-likeness was ***r* = -0.81**
  - Average agreement of individual raters with average ratings: $r = 0.84$
- We may conclude that the Levenshtein distance is a valid measure to determine perceptually sound pronunciation distances

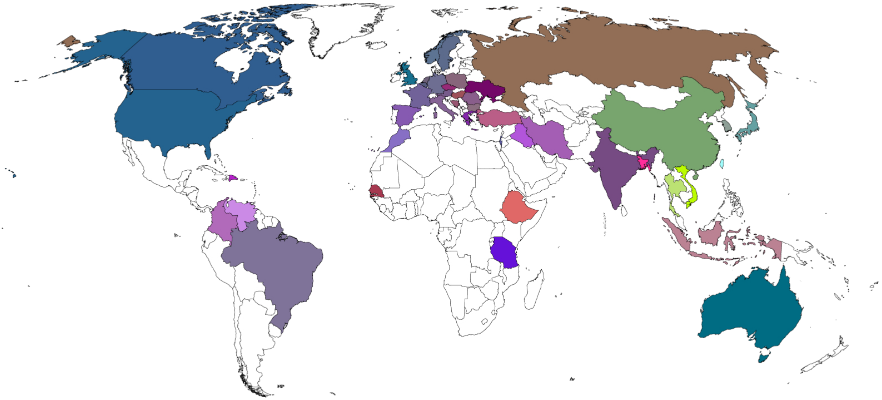
MDS visualization of accent distances: recipe

- We used 989 phonetically transcribed samples from the SAA
- We grouped the transcriptions (i.e. speakers) per country
- For non-English speaking countries, we excluded speakers who moved to an English-speaking country before age 13
- We only included countries with at least 5 speakers

- Pronunciation distances between countries were calculated using the PMI-based Levenshtein algorithm and are visualized using MDS

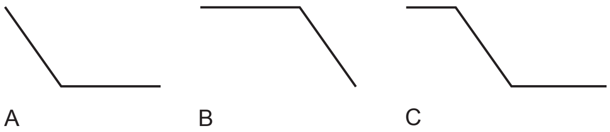
MDS visualization of accent distances

(88% of the variation in the original pronunciation distances is captured)



Determinants of English accent strength: background

- The importance of age of English onset (AEO) for second language (L2) proficiency is undisputed (Ellis and Bogart, 2007)
- But, much debate about existence of critical period (CP): AEO before which learning may still result in native-like performance (Johnson and Newport, 1989)
- Presence of CP operationalized as non-linear influence of AEO (i.e. abrupt changing of slope) on L2 proficiency (Birdsong, 2006)



Determinants of English accent strength: statistics

- Earlier (incorrect) approaches to assess CP (see Vanhove, 2013):
  - Subjectively binning groups of speakers based on AEO and assessing significant L2 performance differences (e.g., Johnson and Newport, 1989)
  - Selecting a subjective AEO breakpoint and comparing the correlation between L2 performance and AEO in both groups (e.g., DeKeyser, 2012)
- Better to use piecewise regression (Baayen, 2008)
  - Here applied to the PMI-based LD from the average American English speaker for 806 non-native SAA speakers + speaker and country info

Piecewise (mixed-effects) regression

1. Find best potential breakpoint

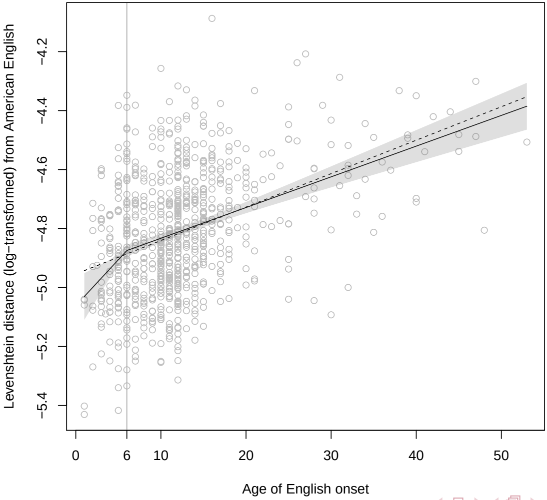
```
for (breakpoint in 1:30) {
    dat$ShiftedAEO = dat$AEO - breakpoint
    dat$PastBreakPoint = as.factor(dat$ShiftedAEO > 0)
    m = lmer(LD.log ~ ShiftedAEO:PastBreakPoint + (1|Country), ...)
    deviances[i] = deviance(m)
}

breakpoint = which(deviances == min(deviances))
dat$ShiftedAEO = dat$AEO - breakpoint
dat$PastBreakPoint = as.factor(dat$ShiftedAEO > 0)
```

2. Model comparison to see if additional complexity is warranted

```
m0 = lmer(LD.log ~ AEO + (1|Country), ...)
m1 = lmer(LD.log ~ ShiftedAEO:PastBreakPoint + (1|Country), ...)
AIC(m0) - AIC(m1) # if > 2 then m1 should be kept
```

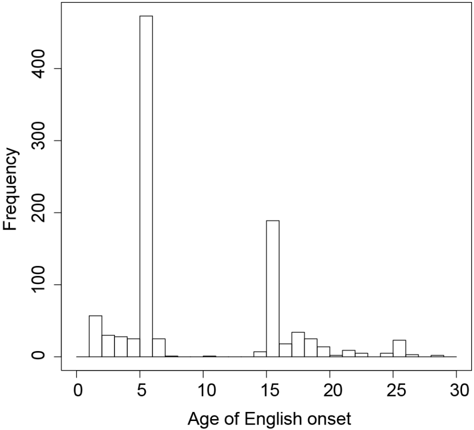
3. Validation of breakpoint via bootstrapping (1000 iterations)

Result on the basis of all data: breakpoint needed

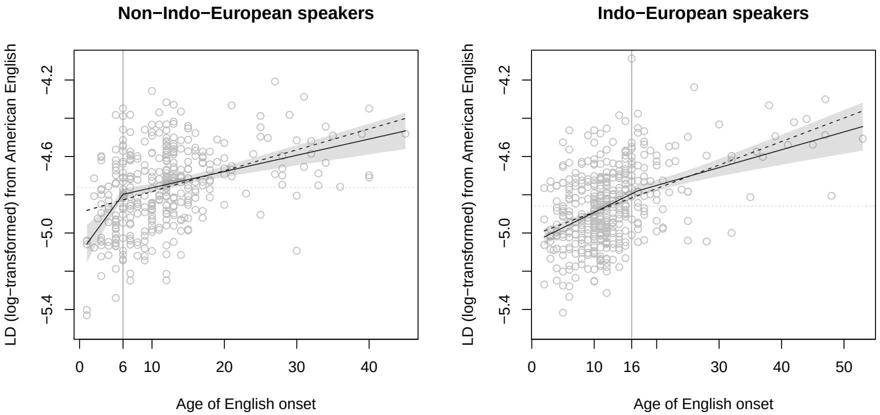


Bootstrap validation: no clear single breakpoint

(adding a breakpoint improved the baseline model in 98% of the bootstrap runs)

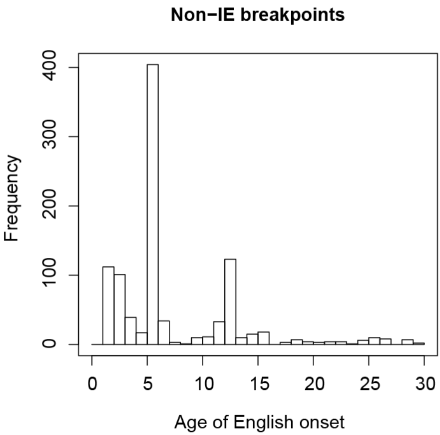


Variation caused by language background?



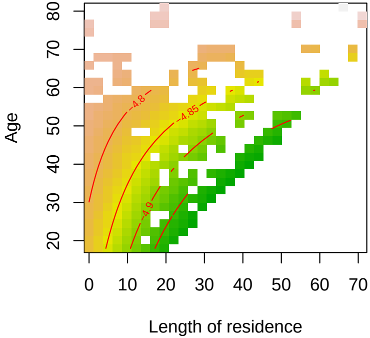
Bootstrap validation: no support for clear breakpoints

(the breakpoint was necessary for 99% [non-IE] and 78% [IE] of the bootstrap runs)



Final model: significantly more vs. less native-like

- IE speakers vs. non-IE speakers
- Nr. of additional languages spoken
- Avg. nr. of years of education per country
- Age of English onset (IE BP at 16, non-IE BP at 6: no CP!)
- Speaker age and length of residence interact:



Recap

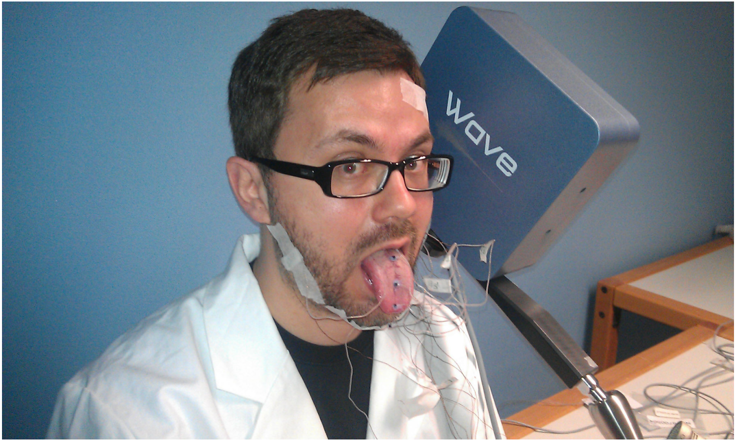
- The PMI approach allows us to induce sensible sound distances
  - The approach is readily applicable to any (dialect) dataset with similar pronunciations
- The PMI-based Levenshtein algorithm yields pronunciation distances which correspond well with perceptual pronunciation distances
- These can adequately be used to investigate determinants of the strength of English accents for many speakers simultaneously: no support for a critical period in L2 acquisition (Wieling et al., submitted)
- "Transcription is a messy thing" (Kerswill and Wright, 1990): Electromagnetic articulography

Recap

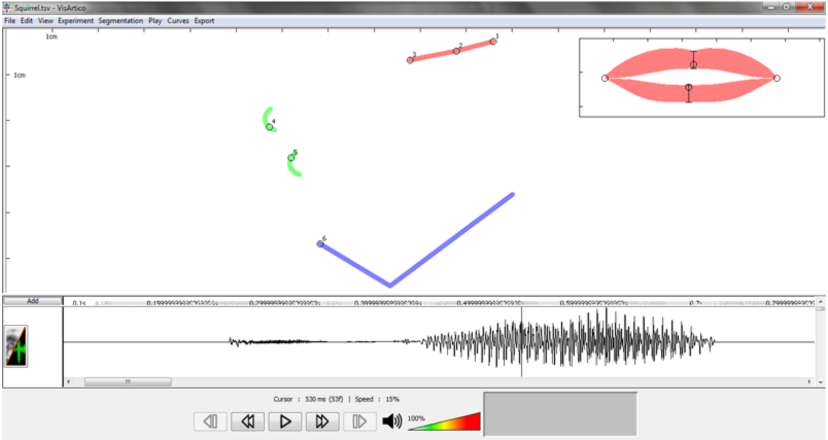
- The PMI approach allows us to induce sensible sound distances
  - The approach is readily applicable to any (dialect) dataset with similar pronunciations
- The PMI-based Levenshtein algorithm yields pronunciation distances which correspond well with perceptual pronunciation distances
- These can adequately be used to investigate determinants of the strength of English accents for many speakers simultaneously: no support for a critical period in L2 acquisition (Wieling et al., submitted)
- “Transcription is a messy thing” (Kerswill and Wright, 1990): Electromagnetic articulography

What is electromagnetic articulography?

Connected to the articulography device



Tongue movement data



Study 1: native vs. non-native English

- 19 native Dutch speakers from Groningen
- 22 native Southern Standard British English speakers from London
- Material:
  - 10 minimal pairs [t]:[θ] repeated twice (e.g., ‘fate’-‘faith’)
    - (the sound [θ] does not exist in the Dutch language)
  - 11 minimal pairs [s]:[ʃ] repeated twice(e.g., ‘lease’-‘leash’)
    - (the sounds [s] and [ʃ] have an allophonic relation in the Dutch language)
- Goal: compare distinctions between both sound contrasts for both groups
- Preprocessing:
  - Focus on frontness of the tongue tip sensor (0.5 cm behind tongue tip)
  - Tongue tip frontness is normalized per speaker and averaged per word

Study 1: results

- Analysis using mixed-effects regression modeling (random-effect factors: word and speaker, including both random intercepts and slopes)
- Contrast [s]:[ʃ]:
  - Both native Dutch and native English speakers make the contrast: more frontal position for [s] ($p = 0.003$)
  - The contrast seems slightly larger ($p = 0.04$, one-tailed) for the native English speakers
- Contrast [t]:[θ]:
  - native English speakers make the contrast ($p < 0.001$) with more frontal positions for the [θ], but Dutch speakers do not clearly make the contrast ($p = 0.25$)

Study 1: discussion

- Results in line with perceptual findings of Johnson and Babel (2010):
  - Dutch speakers rated [s] and [ʃ] as more similar than native English speakers did
- Dutch speakers seem to pronounce /θ/ as /t/ (cf. Hanulikova and Weber, 2012)
- To obtain a better view of the results, however, the remaining material needs to be analyzed:
  - Contrasting /θ/ with /s/ (also w.r.t. German speakers)
  - Comparing the L2 sounds (English) with corresponding L1 sounds (Dutch)

Study 2: articulatory differences in Dutch dialects

- Site visits at two schools (5 days per school)
  - Ter Apel (northern part of the Netherlands): 21 (mostly) young speakers
  - Ubbergen (more southern part of the Netherlands): 19 young speakers
- Procedure:
  - Each participant named 70 images in their local dialect and read 27 four-letter words (e.g., taat: [tat]) in standard Dutch (twice)

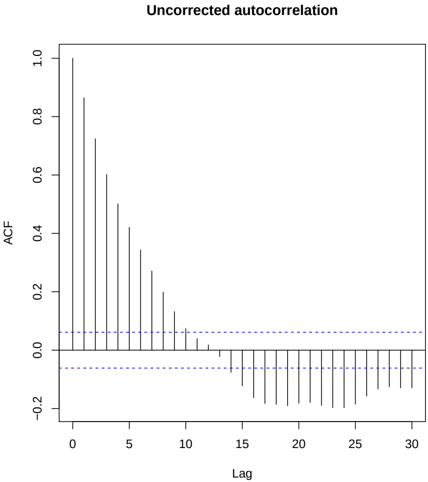
Aggregate analysis of articulatory data

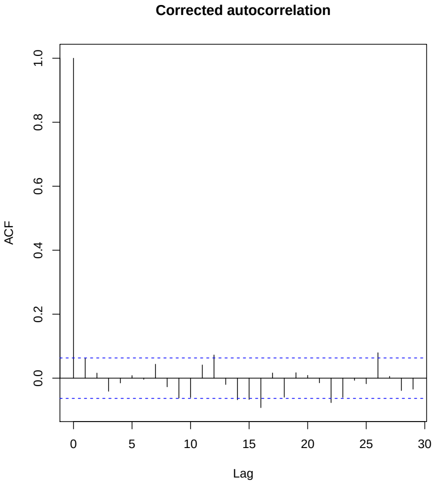
- Generalized additive modeling
  - Models non-linear trajectories (of sensors) over time
  - Takes into account individual variation
  - Is able to assess group differences
  - Covariates can be included
  - Corrects for autocorrelation
  - `R`-package: `mgcv`
- Preprocessing
  - Focus on three tongue sensors
  - Time is normalized from 0 to 1 per word/segment
  - Sensor positions are normalized for each speaker: 0 is most frontal/lowest point in mouth, 1 is most back/highest point in mouth

Example R code

```
> library(mgcv) # version 1.8.2

> dat = art[art$Sensor == "TB" & art$Axis == "X" & art$Word == "taarten", ]

> model = bam( Pos ~ s(Time,by=Group) + Group + s(Time,Subj,bs="fs",m=1),
             rho = 0.86, ... )

> summary(model)

[...]
Parametric coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      47.86363    2.66743  17.944  < 2e-16 ***
GroupUbbergen    -6.38683    1.80164  -3.545 0.000393 ***

Approximate significance of smooth terms:
                           edf  Ref.df      F  p-value
s(Time):GroupTerApel     6.438   6.701  3.898  0.00041 ***
s(Time):GroupUbbergen    7.992   8.126  9.940 4.72e-14 ***
s(Time,Subj)           641.732 715.000 28.439  < 2e-16 ***
```

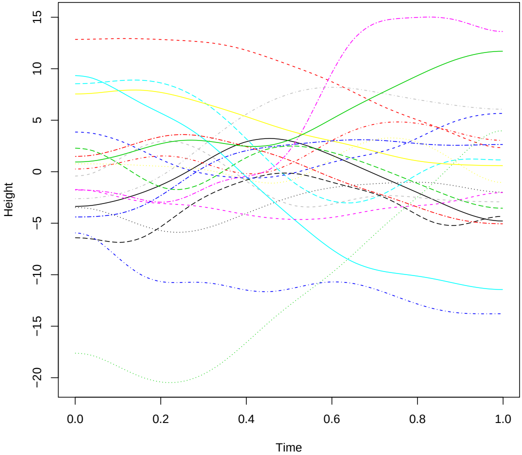
Correcting autocorrelation in the residuals via `rho`





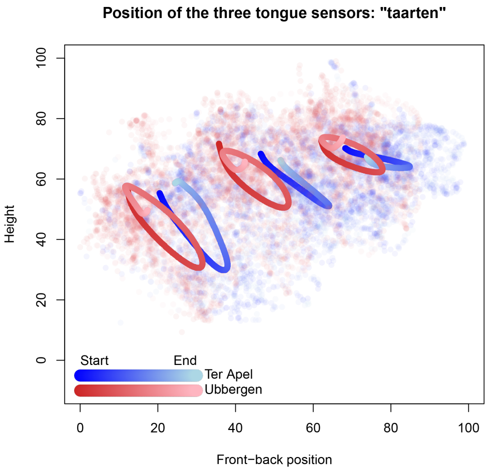
Accounting for individual variation

(via `s(Time,Subj,bs="fs",m=1)`)



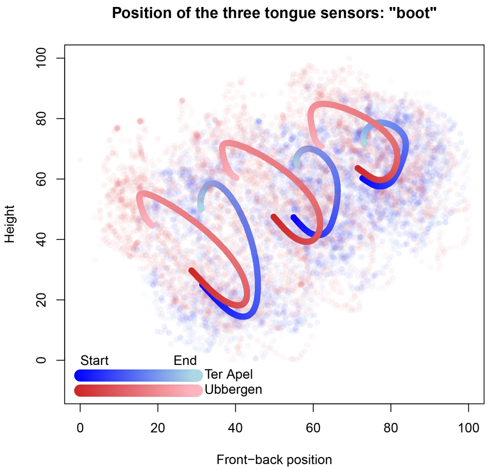
Example of dialect pronunciation differences: "taarten"

Listen to example pronunciations



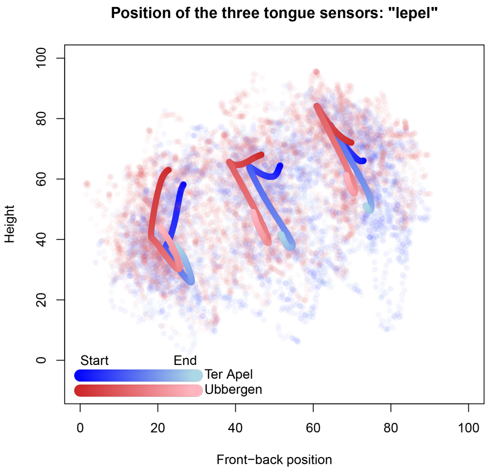
Example of dialect pronunciation differences: "boot"

Listen to example pronunciations



Example of dialect pronunciation differences: "lepel"

Listen to example pronunciations

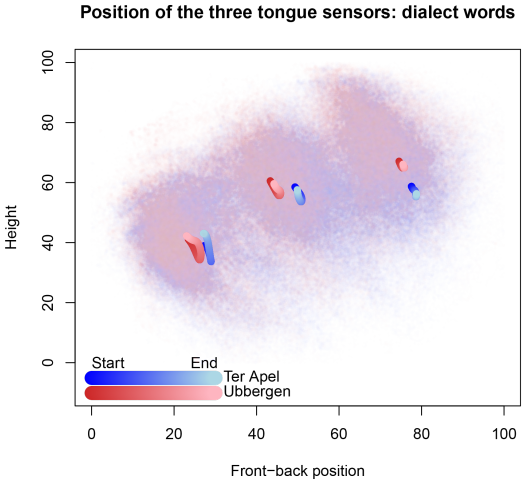


Structural differences between dialects?

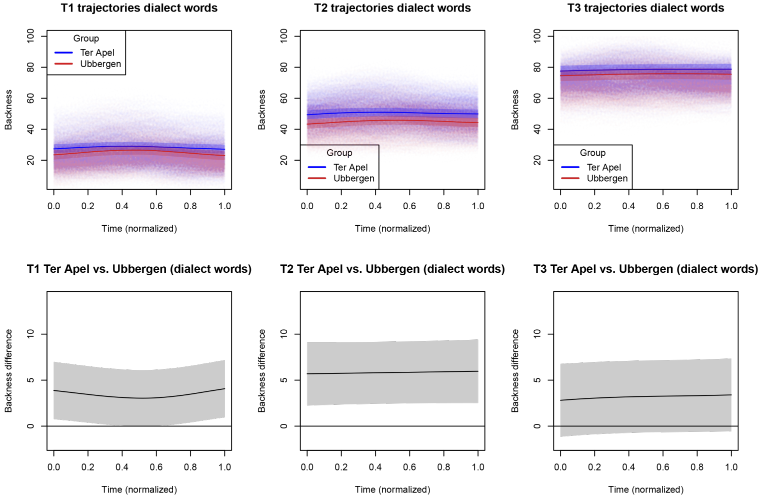
- Rather than focusing on single words, we are interested in the aggregate differences between the two groups of speakers
- In the following, we will report the aggregate results for both the Dutch non-words (accented speech) and the dialectal pronunciations

Aggregate pronunciation differences: dialect words

(significant correlation between F2 and frontness: $r = 0.3$)

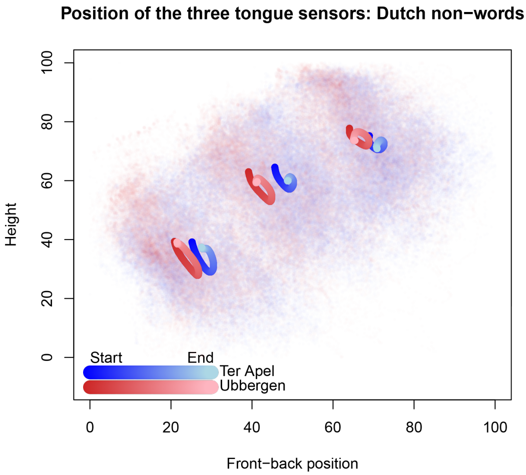


Results of statistical analysis: dialect words

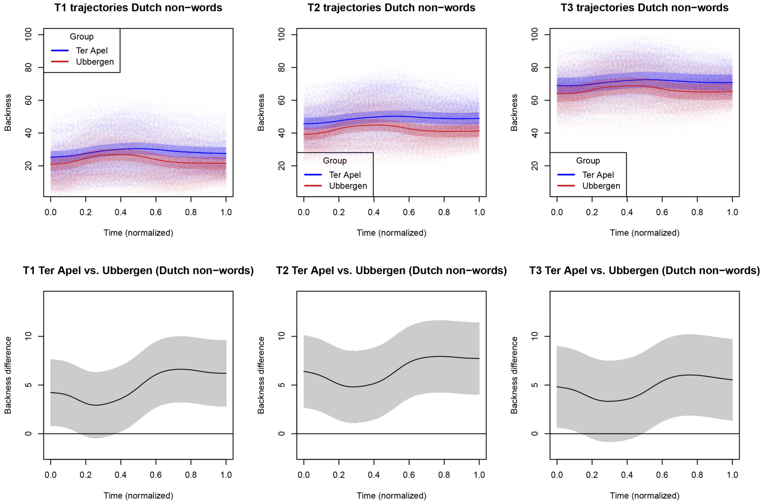


Aggregate pronunciation differences: Dutch non-words

(significant correlation between F2 and frontness: $r = 0.4$)



Results of statistical analysis: Dutch non-words



Study 2: discussion

- Clear dialectal differences: tongue further back for northern speakers
- Previous formant-based study (Adank et al., 2007) did not find F2 differences between the two dialect areas
- Results suggest differences in the articulatory settings (Laver, 1978) at the dialect level

Conclusion

- The articulatory perspective offers information which is not (easily) observable in the acoustic signal
- It is a useful research tool for variationist linguists as well as L2 researchers
- Compared to transcriptions, obtaining articulatory trajectories is more precise, as it takes (e.g.,) co-articulation into account
- Generalized additive modeling allows us to extract the general movement patterns for groups of speakers while taking into account the individual variation
  - Additionally, it allows the simultaneous incorporation of sociolinguistic variables (e.g., age and gender)

Thank you for your attention!

Additional material

PMI-based LD: good alignment performance

(Wieling, Prokić and Nerbonne, 2009)

- Bulgarian dialect data set (152 words in 197 sites)
- Manually corrected Gold Standard set of about 3M alignments

Algorithm	Correct alignments (%)
Hamming distance	79.08
Levenshtein distance	94.48
Damerau-Levenshtein distance	95.34
PairHMM Viterbi	95.37
PMI-based Levenshtein distance	95.50

- N.B. PairHMM emission probabilities similar to PMI-based sound distances: Spearman ρ = -0.97 (subst.) and ρ = -0.74 (indels)

By-country random intercepts of non-native English

(country- and language-related predictors were excluded to generate this plot)