# Analyzing vowel distances

Therese Leinonen

university of groningen

LOT winter school, Groningen 21th January, 2009

# Overview

- Segment distances in dialectometric research
- Acoustic properties of vowels
- Measuring vowel distances
- Perception of vowels
- Inducing sound segment distances using Pair HMM's

# Using segment distances in the Levenshtein algorithm

- theorem: given segment distances, Levenshtein algorithm finds optimal alignment
- what are good segment distances?
- various feature systems: Vieregge-Cucchiarini, Almeida-Braun (Heeringa 2004)
- "acoustic" distance (Heeringa 2004)
- very limited improvement over binary segmental table

# Phonetic Puzzle

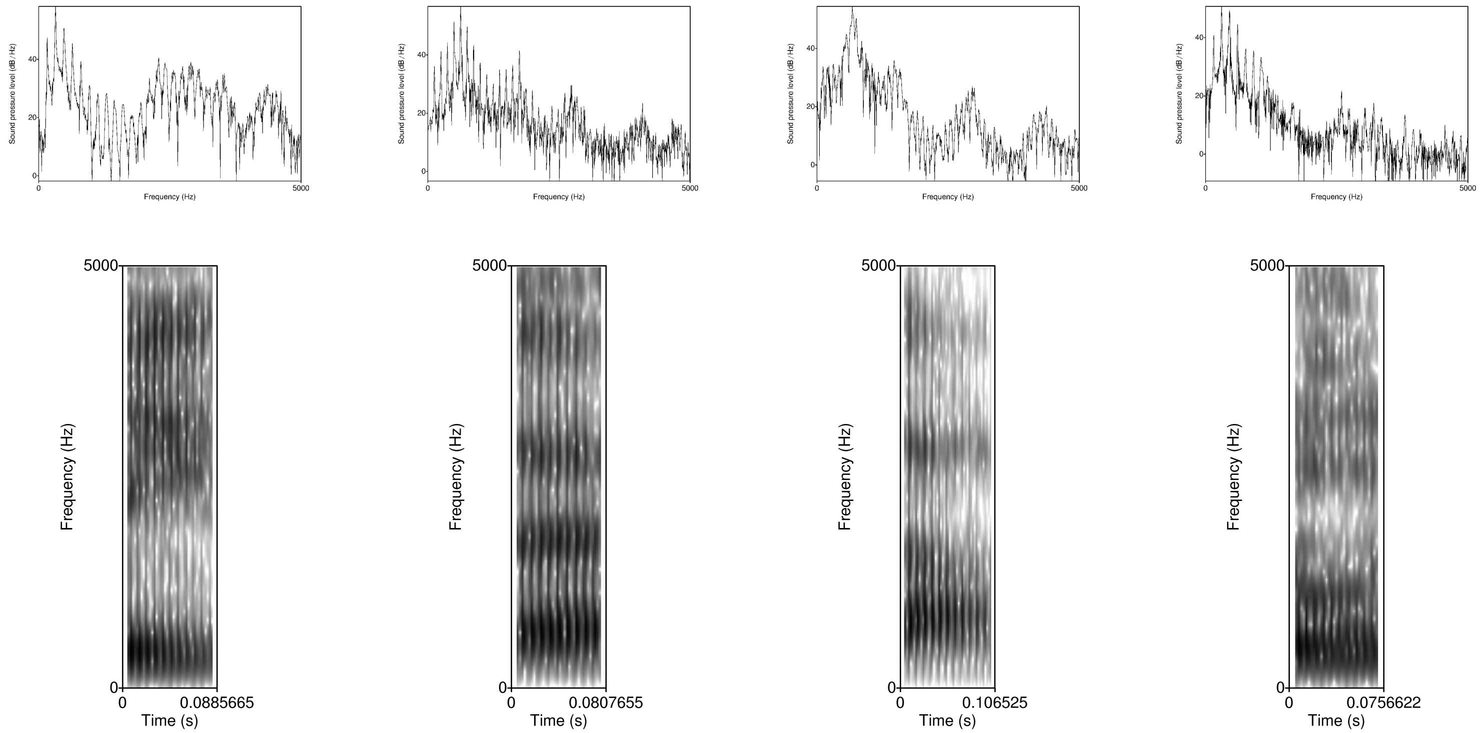
Why is detailed phonetic information not helping?

- hypothesis 1: transcriptions are phonetically unreliable
- hypothesis 2: previous attempts were too ambitious, trying to characterize *all* distinctions
- hypothesis 3: we are past the size where fine discrimination matters
- others?

# Acoustic properties of vowels

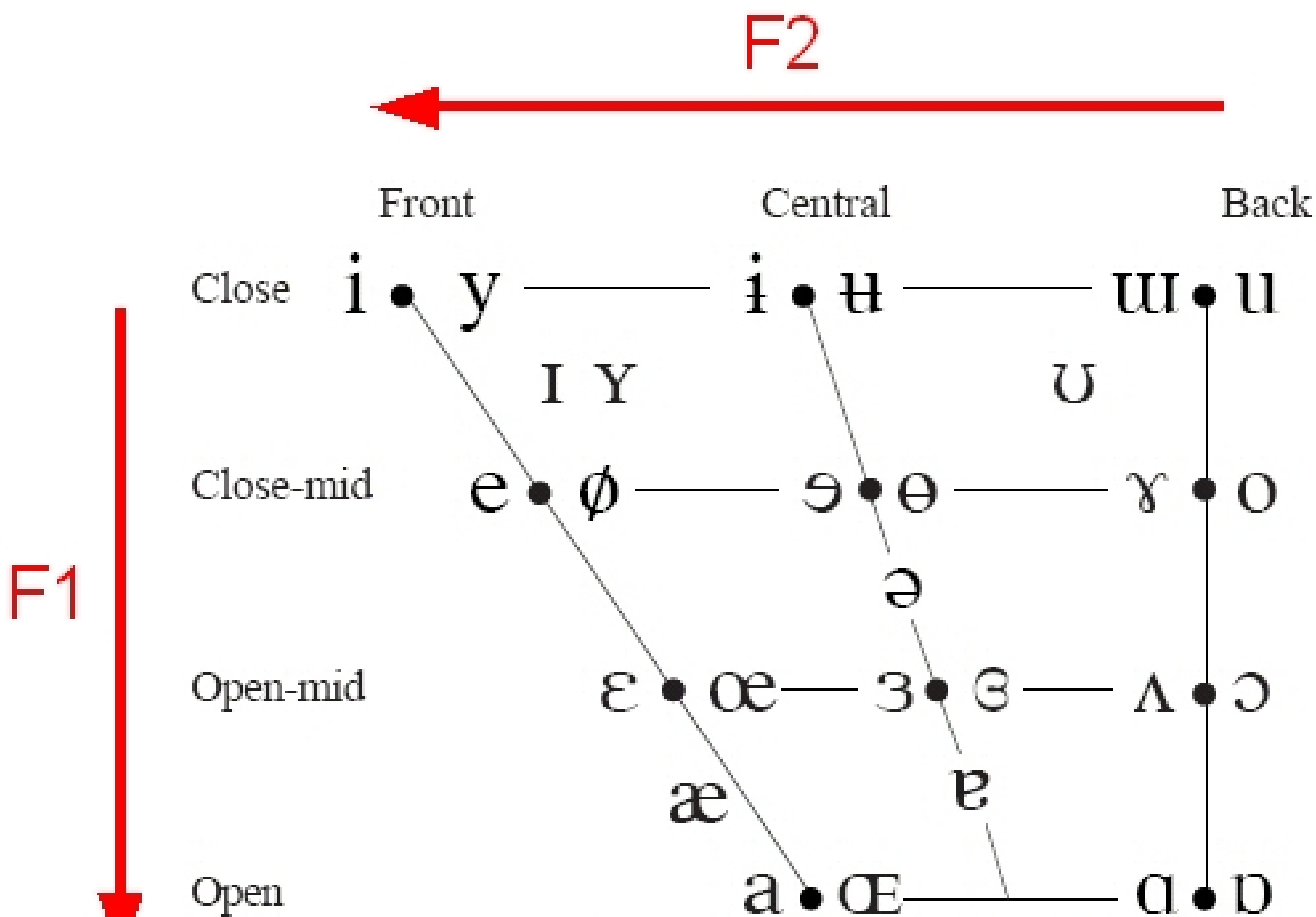
- the vocal tract is a resonator that resonates on given frequencies; by changing the size and shape of the tract (by moving the position of tongue, lips, yaw) we can adjust the resonant frequencies
- the sound we produce with our vocal chords consists of a base tone and its harmonics
- when some harmonic of the sound from the vocal chords matches or is close by a resonant frequency it will cause resonance
- formants = peaks in the frequency spectrum resulting from resonance in the vocal tract
- our perception of vowels is based on recognizing the formant frequencies characterizing each vowel
- the first two formants (F1 and F2), corresponding well with vowel height and backness, are usually enough to distinguish vowels from each other

spectrum

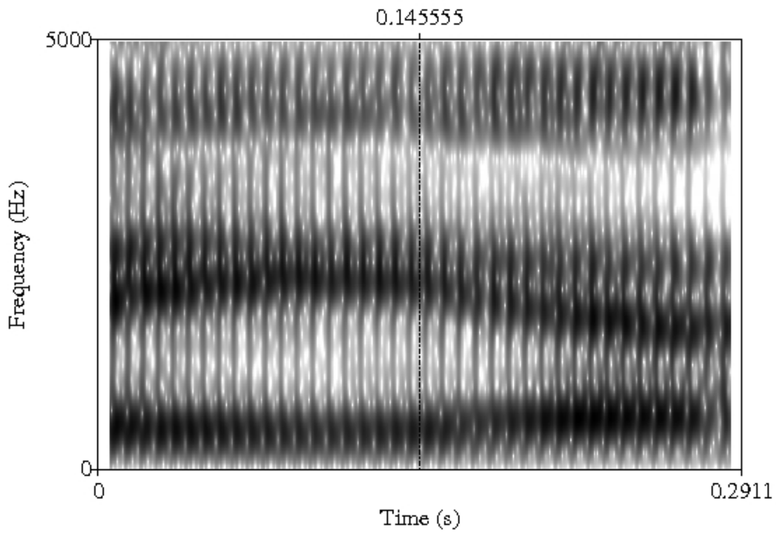


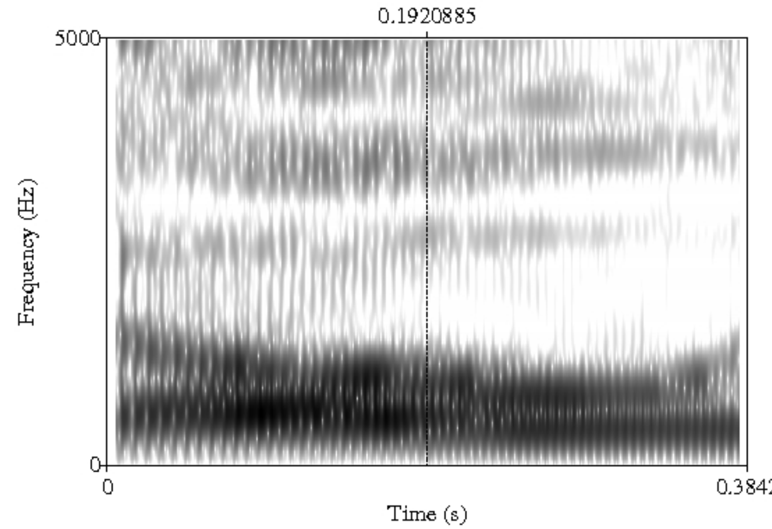

spectrogram

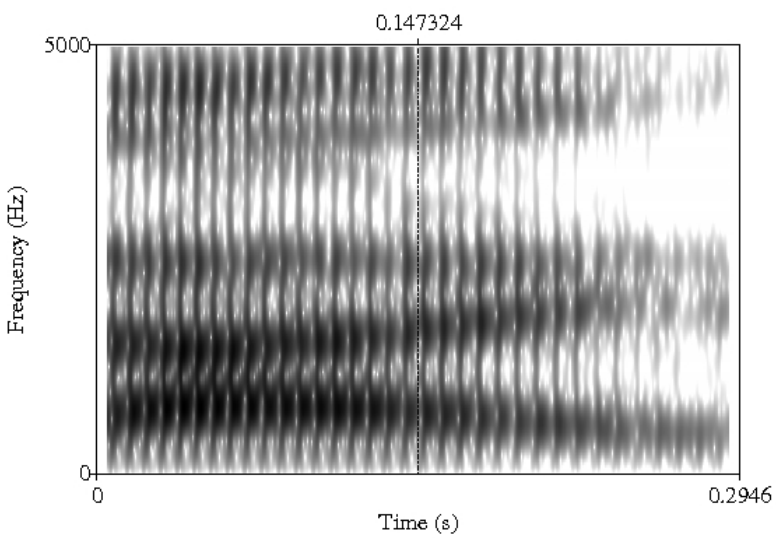
vowel [i] [æ] [a] [u]



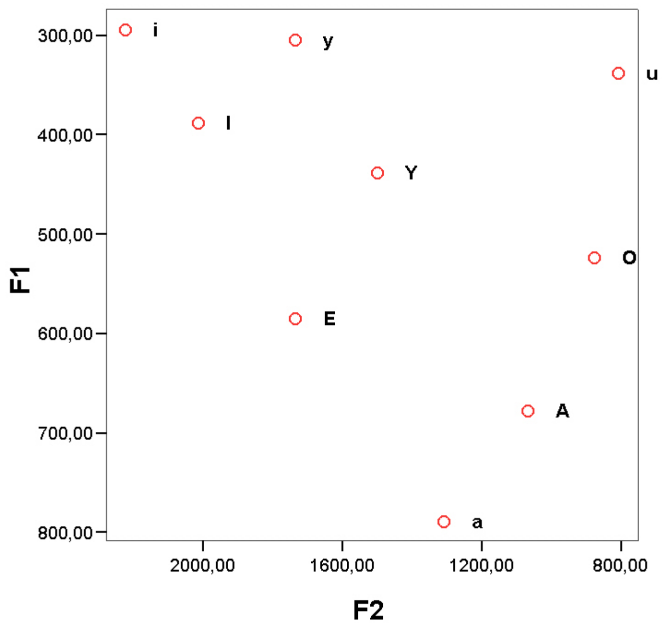

Where symbols appear in pairs, the one to the right represents a rounded vowel.

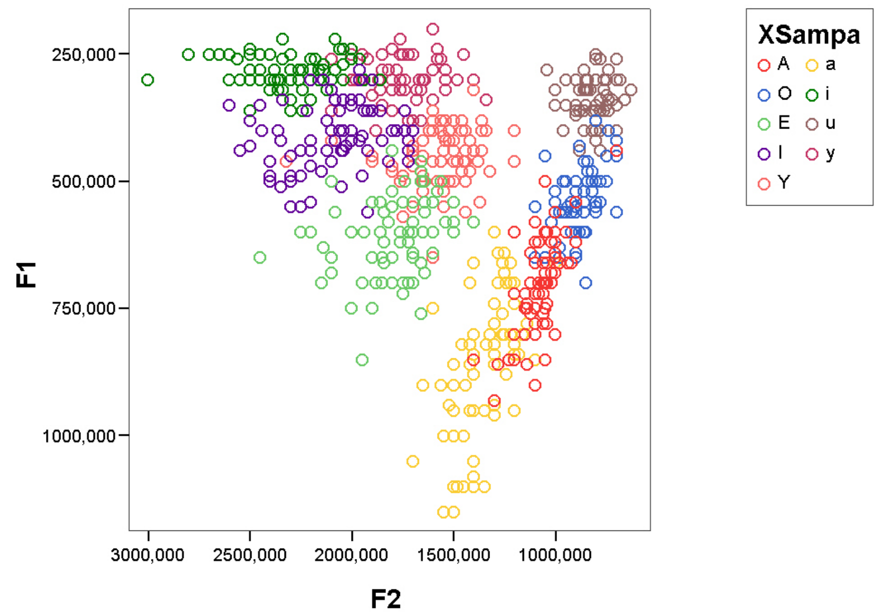

0.145555
5000
0
Frequency (Hz)
0
0.2911
Time (s)


0.1920885
5000
0
Frequency (Hz)
0
0.3842
Time (s)


0.147324
5000
0
Frequency (Hz)
0
0.2946
Time (s)

# Formant frequencies of Dutch vowels



# Measuring the distance between vowels

Euclidean distance:

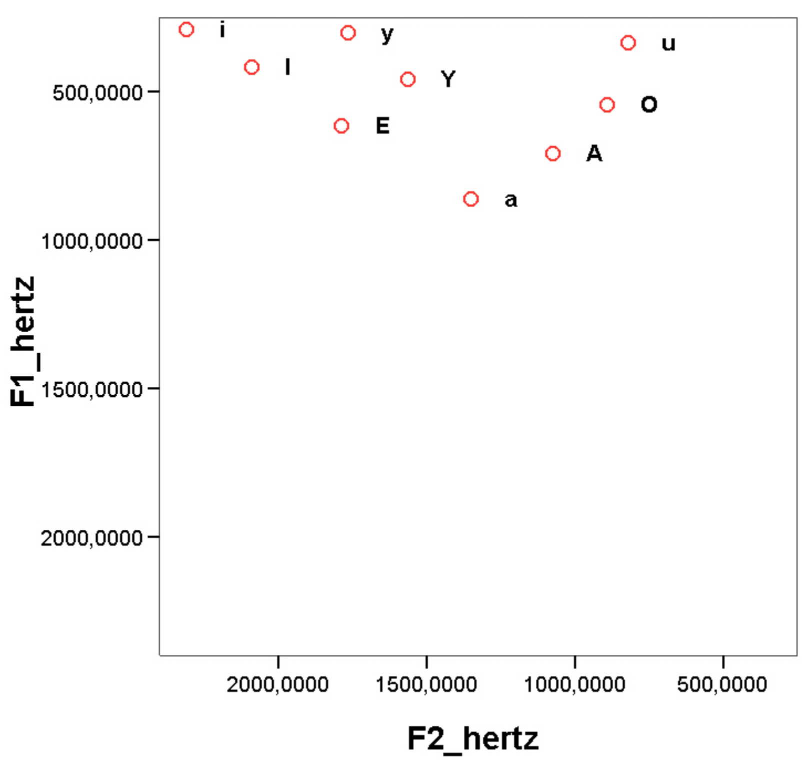
$$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

Euclidean distance of the two first formants:

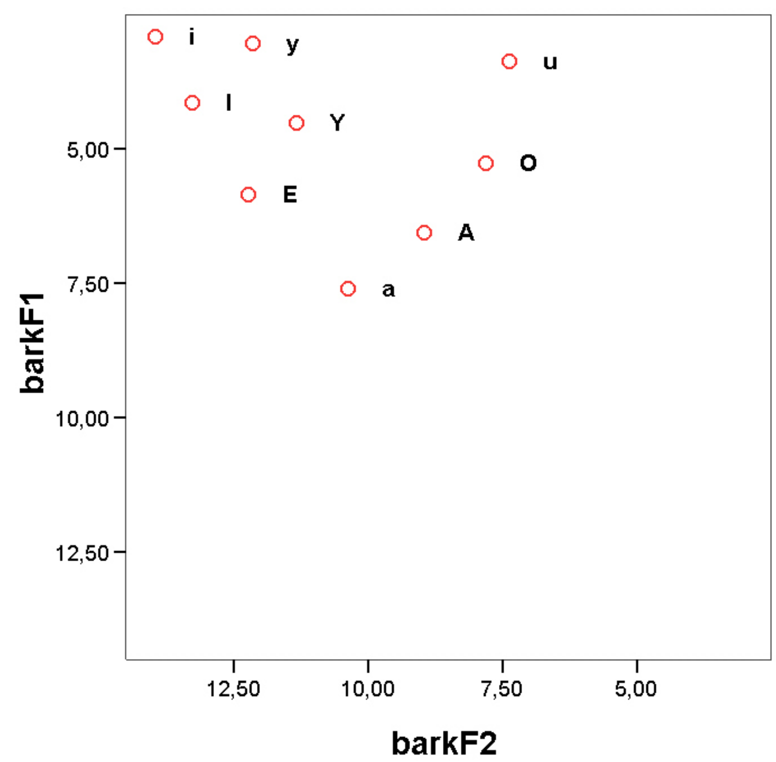
$$d_ij = \sqrt{(F1_i - F1j)^2 + (F2_i - F2j)^2}$$

# Perception of vowels

- frequency is measured in Hertz (cycles per second), however, human perception is roughly linear below 1000 Hz and roughly logarithmic above 1000 Hz
- Bark and Mel are scales that are developed to correspond to perception
- the formant frequencies depend on the size and shape of every speakers vocal tract, as listeners we can normalize for this speaker dependent variation, automatic procedures for doing the same are hard to find
- the inventory of vowel phonemes in a language influences the perception of vowels



i
y
u
I
Y
O
E
A
a
500,0000
1000,0000
1500,0000
2000,0000
F1_hertz
2000,0000
1500,0000
1000,0000
500,0000
F2_hertz

i
y
u
I
Y
O
E
A
a
5,00
7,50
10,00
12,50
barkF1
12,50
10,00
7,50
5,00
barkF2

# Inducing sound segment distances using Pair HMM's

(Wieling, Leinonen and Nerbonne 2007)

- Pair Hidden Markov Models (Pair HMM) were trained to align the pronunciation transcriptions of a large contemporary collection of Dutch dialect data (Goeman & Taeldeman, 1996)

- the PHMM give probabilities of two segments being aligned in the data set – these probabilities can be interpreted as segment distances

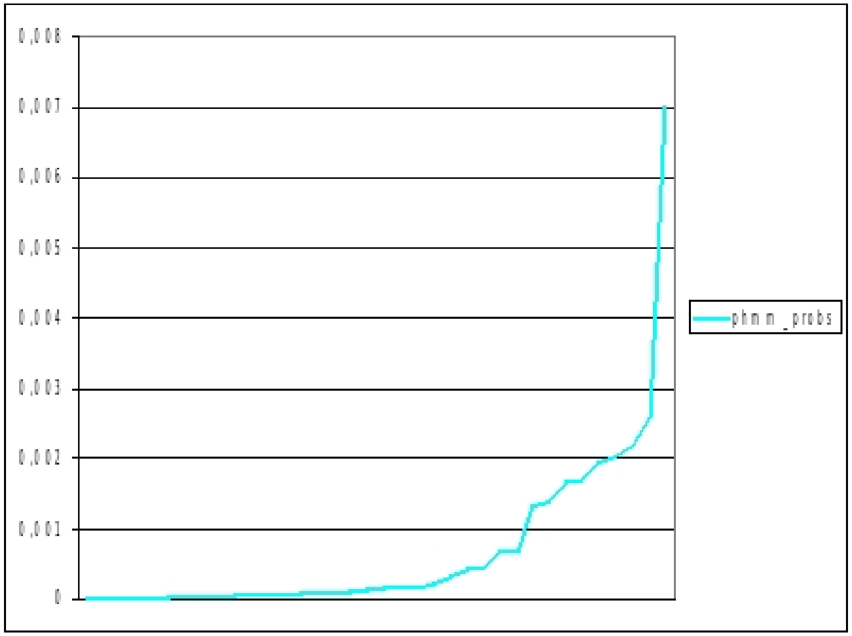
- we validated the substitution probabilities by acoustic measures (Euclidean distance of F1 and F2)

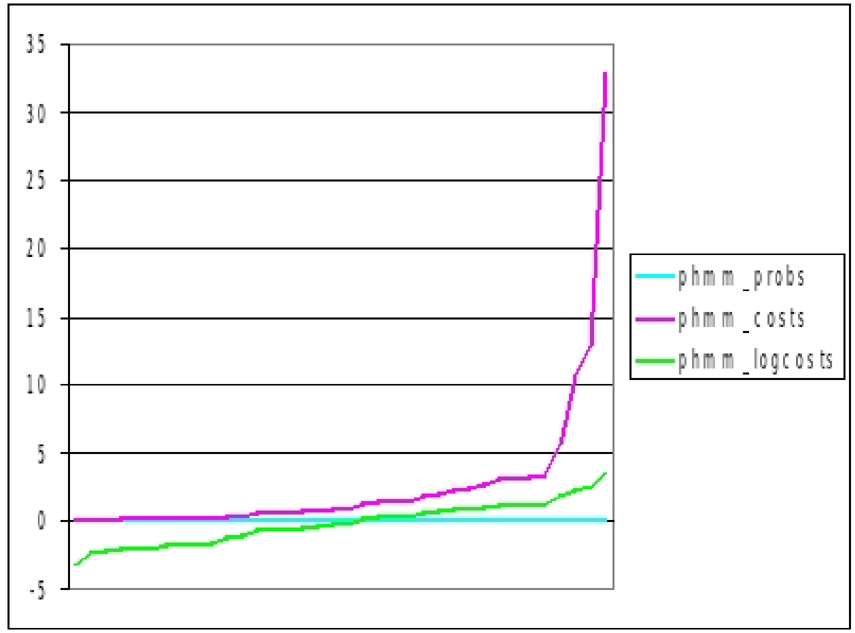
- acoustic data: pronunciation of Standard Dutch monophthongs by 50 male (Pols, Tromp and Plomp 1973) and 25 female speakers 25 female (Van Nierop, Pols and Plomp 1973) speakers

# Transforming the data

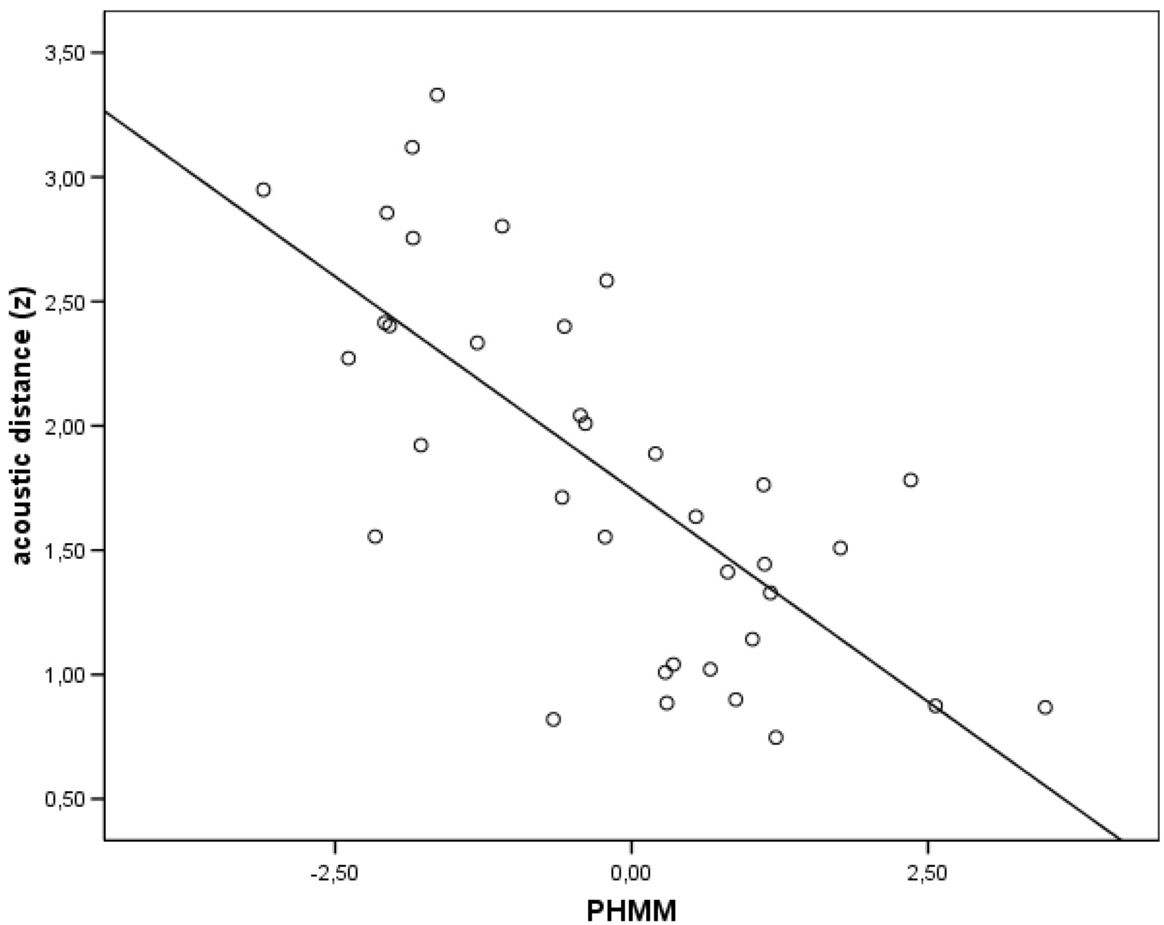
- the occurrence frequency of the phonetic symbols influences substitution probability
- the substitution probabilities are divided by the product of the relative frequencies of the two phonetic symbols used in the substitution
- substitutions involving similar infrequent segments now get a much higher score than substitutions involving similar, but frequent segments – the logarithm of the score is used to bring the scores into a comparable scale

# Transforming the data



# Regression




$$acousticdistance = 1.75 - 0.32 * PHMM$$

$$r = -0.72$$

## Dialectological results:

- dialect classification based on Pair HMM's show only small differences to analyses based on the Levenshtein algorithm ($r$ = 0.89)
- it would be valuable to test the method on dialect data for which perceptual distances are also available

## Conclusions:

- alignments created by the Pair HMM are linguistically responsible
- the linguistic structure influences the range of linguistic variation
- similarity is a satisfying basis of comparison at local levels

# References


Ashby, M. and Maidment, J.(2005), *Introducing Phonetic Science*, Cambridge University Press, Cambridge.

Heeringa, W.(2004), *Measuring Dialect Pronunciation Differences using Levenshtein Distance*, PhD thesis, Rijksuniversiteit Groningen, Groningen.

Pols, L. C. W., Tromp, H. R. C. and Plomp, R.(1973), Frequency analysis of Dutch vowels from 50 male speakers, *Journal of the Acoustical Society of America* **53**, 1093–1101.

Van Nierop, D. J. P. J., Pols, L. C. W. and Plomp, R.(1973), Frequency analysis of Dutch vowels from 25 female speakers, *Acoustica* **29**, 110–118.

Wieling, M., Leinonen, T. and Nerbonne, J.(2007), Inducing sound segment differences using pair hidden markov models, *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, Association for Computational Linguistics, Prague, Czech Republic, pp. 48–56.