Validating Sound Segment Distances Induced by Pair Hidden Markov Models by Acoustic Distances

Therese Leinonen

Seminar in Statistics and Methodology
18th of April 2007

Outline

- About the experiment
- Transforming the data
- Regression
- Conclusions

The experiment

- how to incorporate information about sound segment distances to improve sequence distance measures for use in dialect comparison?
- Pair Hidden Markov Models (PHMM) 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

The experiment

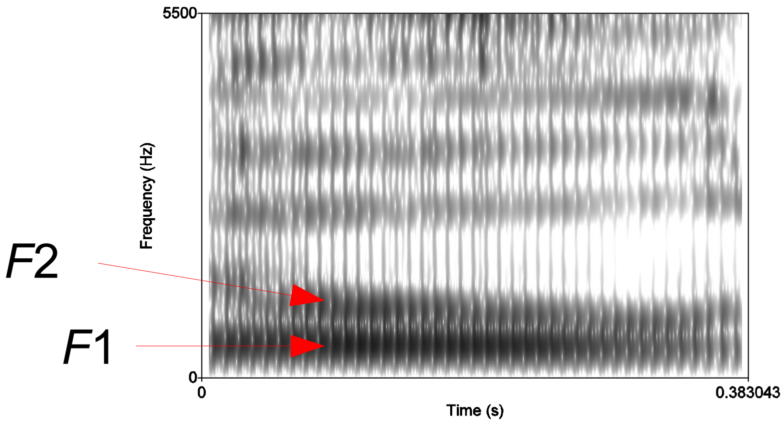
- we validate the substitution probabilities by acoustic measures
- acoustic data: pronunciation of Standard Dutch monophthongs by 50 male (Pols et al., 1973) and 25 female speakers 25 female (Nierop et al., 1973) speakers
- Euclidean distances of *F*1 and *F*2:

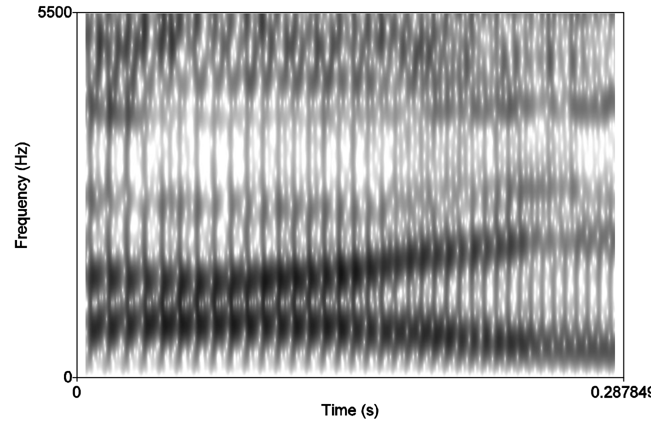
$$\sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2}$$

Formants?

- the sound we produce with our vocal chords consists of a base tone and its harmonics
- 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
- 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 (*F*1 and *F*2), corresponding well with vowel height and advancement, are usually enough to distinguish vowels from each other

Formants?





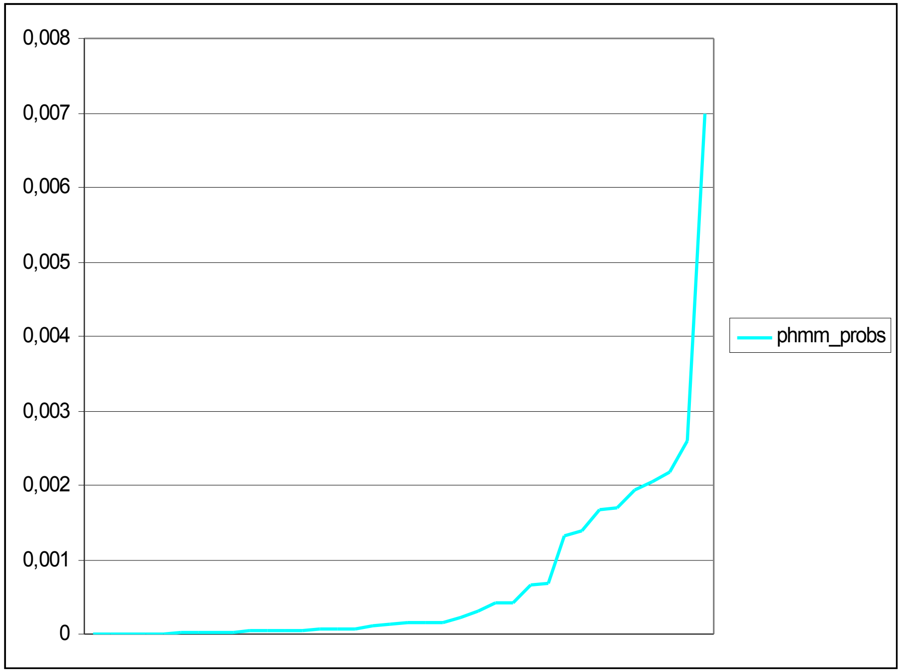
[u:] this vowel has low values for both *F*1 and *F*2 since it is a closed back vowel, it is also slightly diphthongized

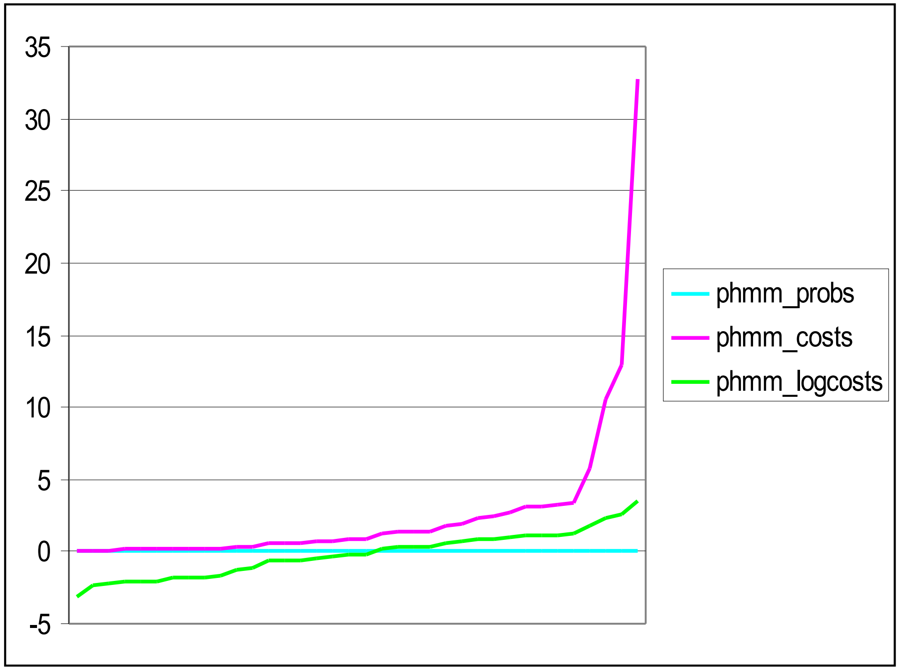
[æi] this diphthong begins as an open front vowel and goes to an even more fronted but closed vowel, accordingly *F*1 starts high and is lowered while *F*2 raises

Transforming the data, substitution probabilities

- 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, substitution probabilities

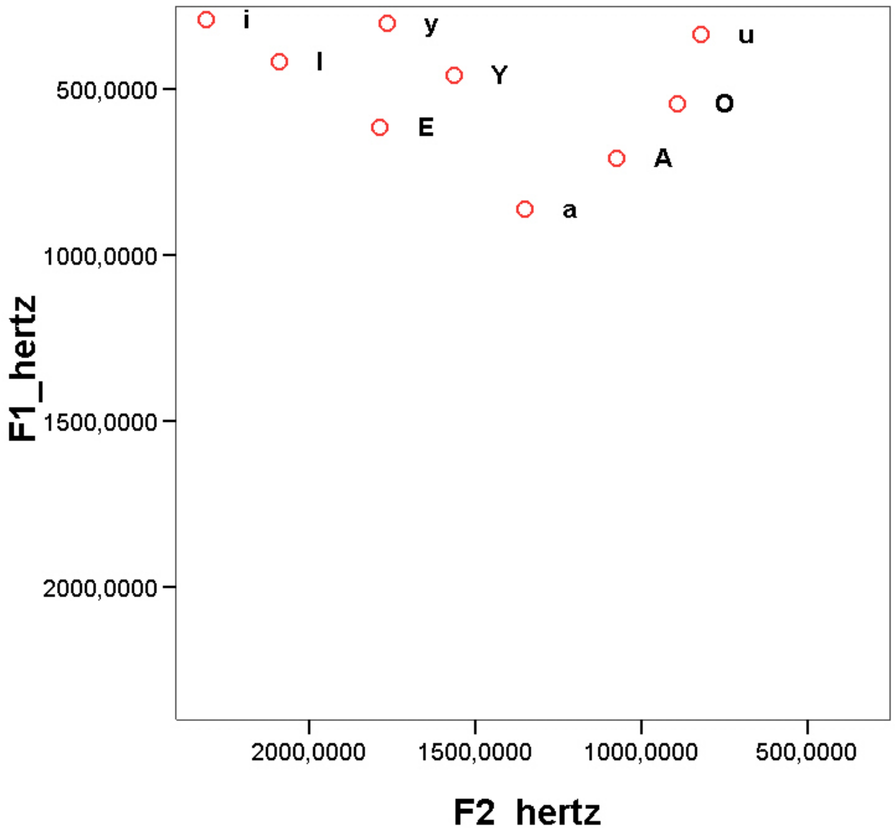




Transforming the data, the acoustic measure

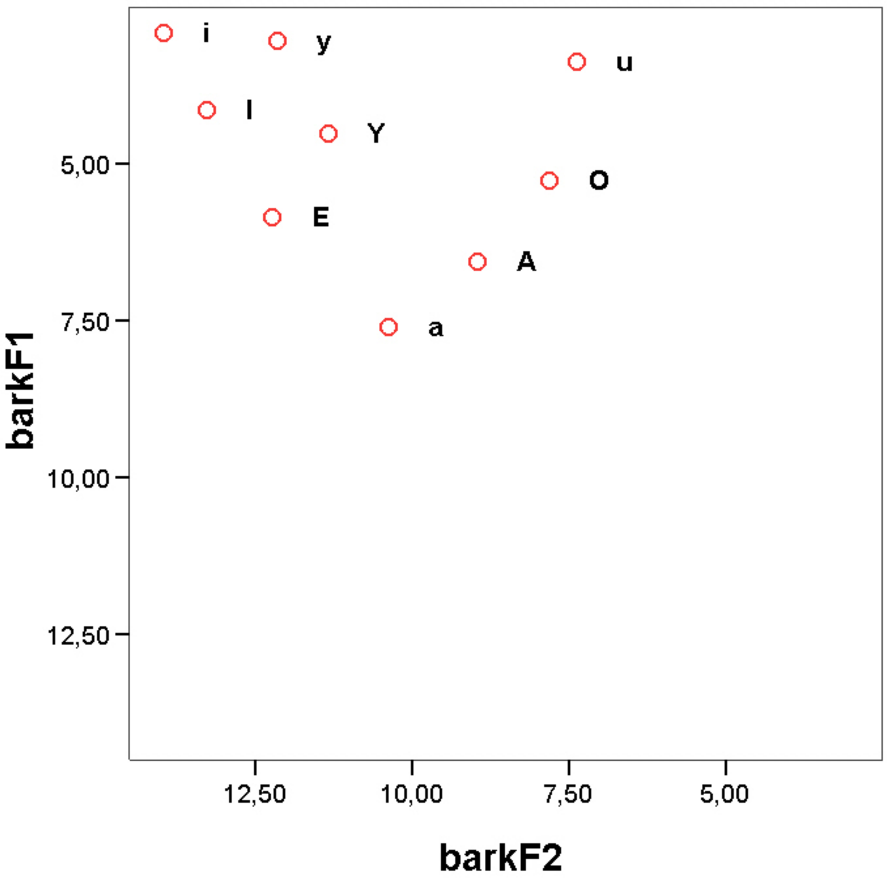
- formants are measured in Hertz
- the Bark scale has a better correspondence to perception, roughly linear below 1000 Hz and roughly logarithmic above 1000 Hz
- formants show variation due to different shapes and sizes of vocal tracts, normalization procedures even out these differences

Transforming the data, the acoustic measure



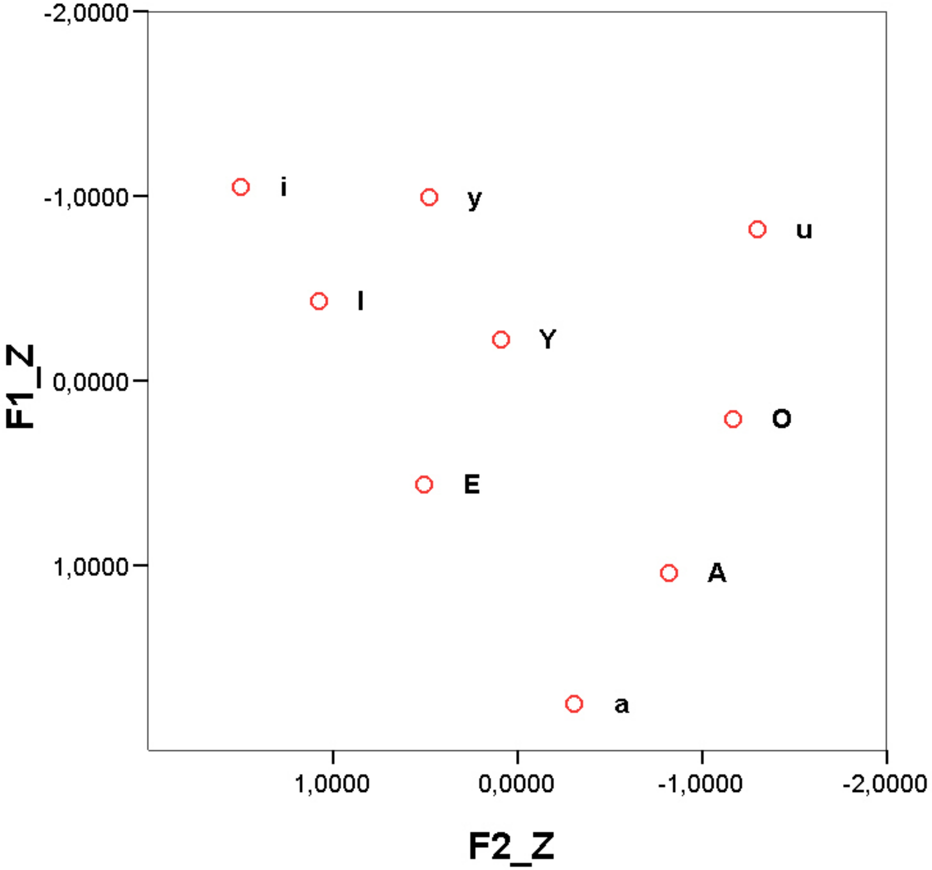
*F*1–*F*2 plot in Hertz

Transforming the data, the acoustic measure



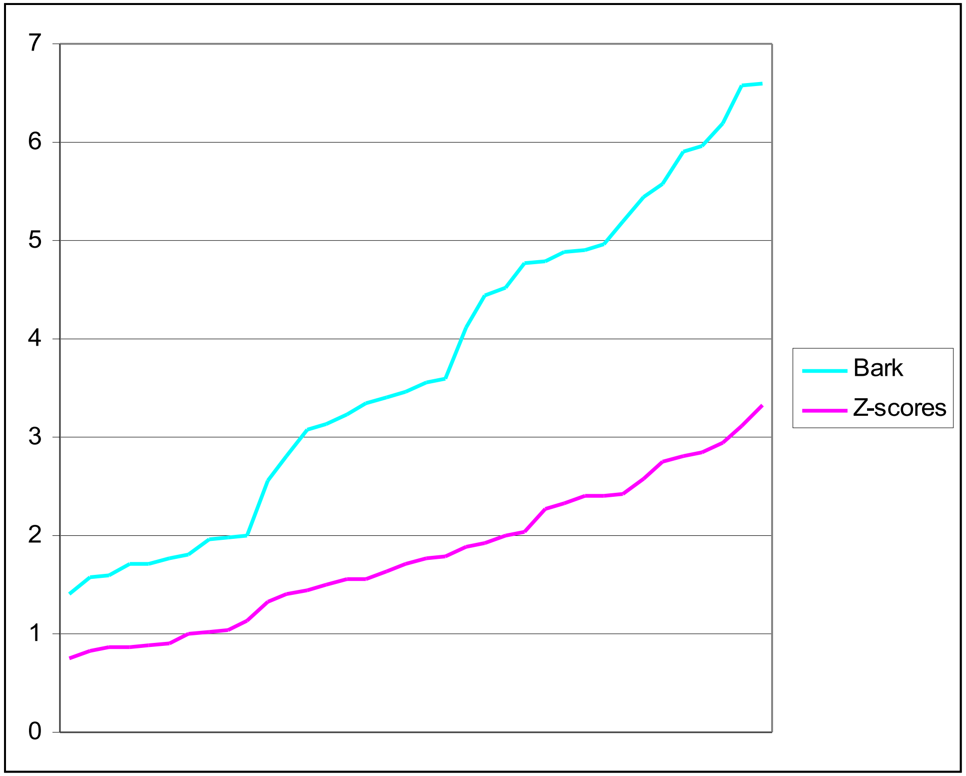
*F*1–*F*2 plot in Bark

Transforming the data, the acoustic measure

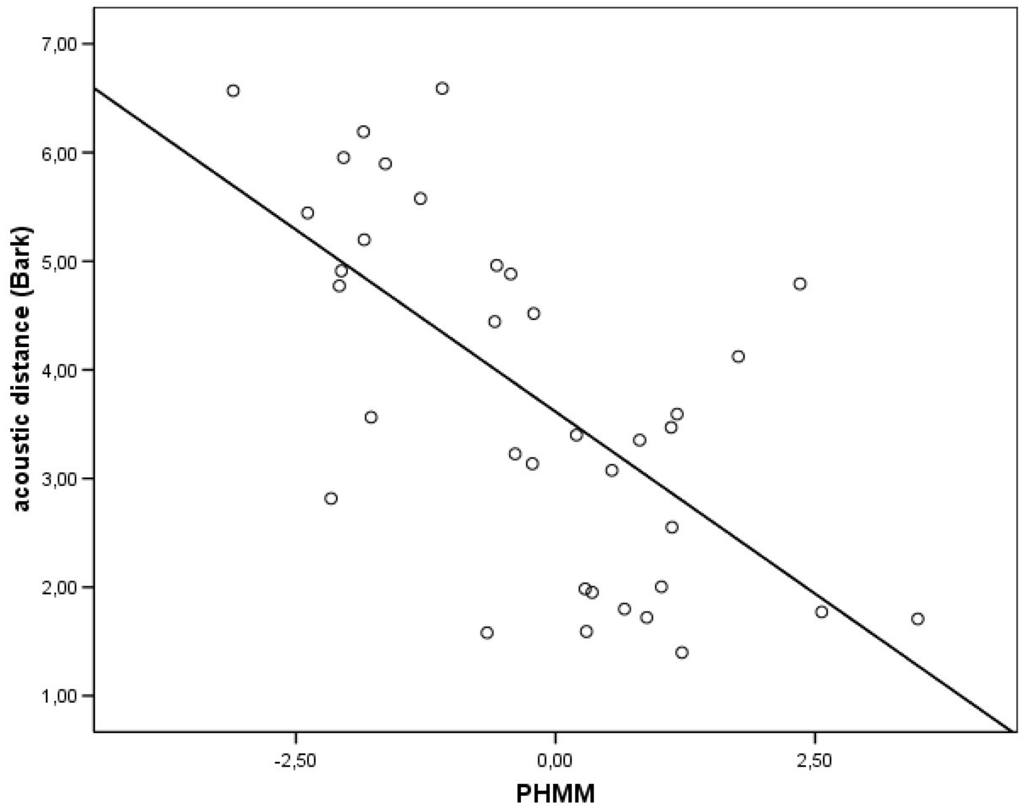


*F*1–*F*2 plot with speaker normalized *z*-values

Transforming the data, the acoustic measure



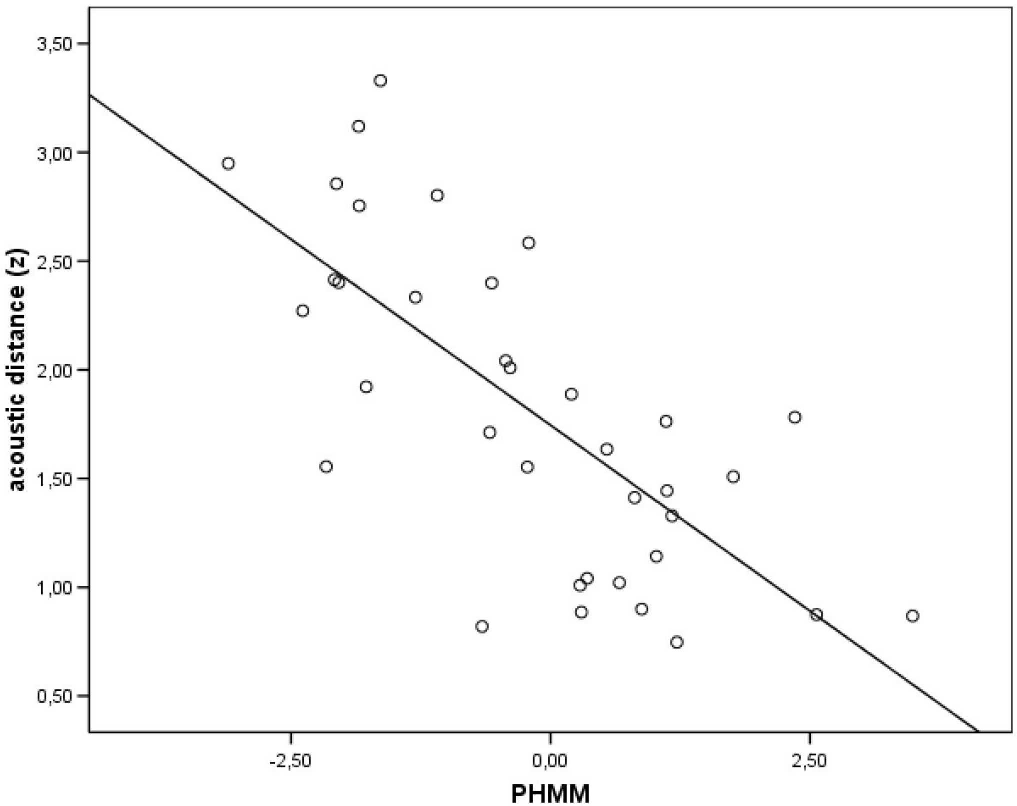
Regression



acoustic distance (Bark) = 3.61 - 0.67 x PHMM

correlation = -0.65

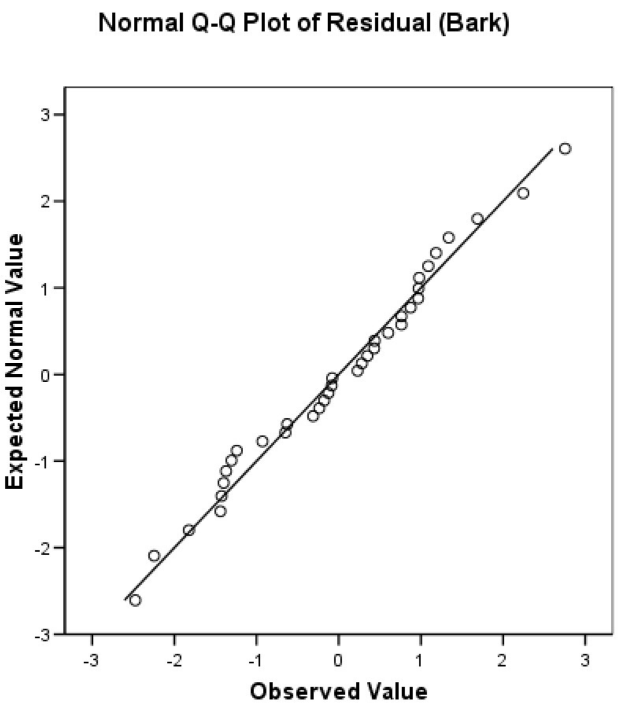
Regression

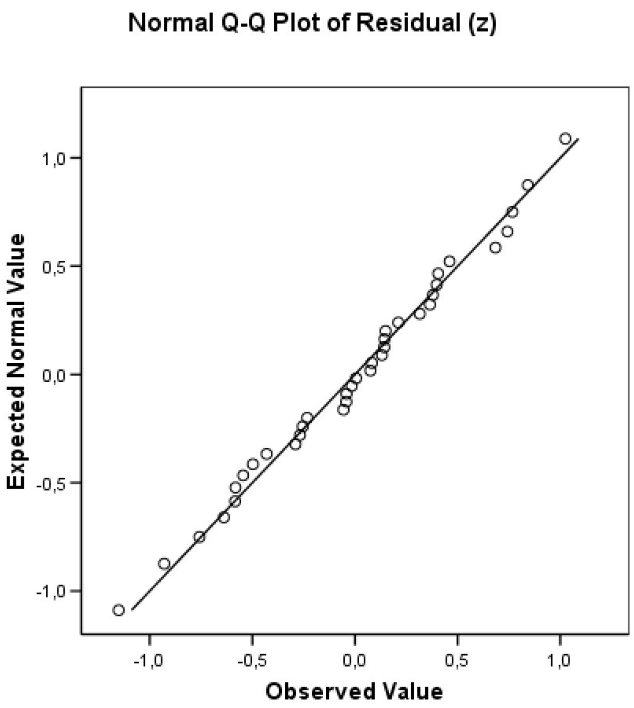


acoustic distance (z) = 1.75 - 0.32 x PHMM

correlation = -0.72

Regression


Normal Q-Q Plot of Residual (Bark)
Expected Normal Value
Observed Value
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


Normal Q-Q Plot of Residual (z)
Expected Normal Value
Observed Value
1,0
0,5
0,0
-0,5
-1,0
-1,0
-0,5
0,0
0,5
1,0

Conclusions

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