LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Institute of Phonetics and Speech Processing

The influence of the place of articulation on the speaker specificity of German phonemes

Carola Schindler & Christoph Draxler

Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University of Munich

{carola.schindler|draxler}@phonetik.uni-muenchen.de

Motivation

- Speaker identification and verification requires features with low intra-speaker and high inter-speaker variability

Previous Work
- Focuses on vowel formants
- Consonants less explored, but often with good results (Kavanagh 2012)

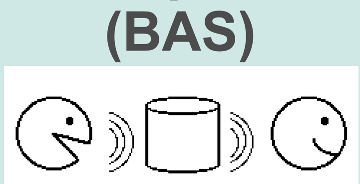
Nasals
- Nasal cavity is complex with a speaker-specific shape (Rose 2002)
- Cannot be modfied willingly

Fricatives
- Crowded articulatory space –> speaker has to produce every sound with great accuracy to make it distinguishable for the listener (Stevens 1971, Gordon et al. 2002, Lorenzen 2004)
- Great diversity in articulation of fricatives between speakers (Narayanan & Haker 1995, Newman et al. 2001, Silbert & De Jong 2008)
- May be affected negatively by phone signal quality
- Nasals and fricatives are common in German

Previous study on speaker-specific information in nasals and fricatives (Mook & Draxler 2012):
- The phonemes /m,n,f,s/ gave the best F-ratio values
- The phonemes /n/ and /s/ always reached higher F-ratios than /m/ and /f/

Method

Verbmobil II – corpus (BAS)



(Hess et al. 1995)

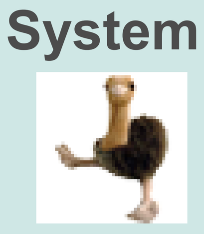
- Spontaneous German dialogues
- 49 male speakers from Munich
- 23 – 185 utterances per speaker

Munich Automatic Segmentation System (MAUS)

(Schiel 2004)

- Automatic segmentation and labelling of speech signal files

EMU Speech database System



- Calculation of the spectra

R Software for Statistical Computing

- Extraction of all required phonemes
- Calculation of the spectral moments at segment midpoint
- Statistical analysis (Analysis of variance)

Spectral moments have beneficial properties:
- Easy to compute
- Can be applied to vowels and consonants
- Directly related to articulation and acoustics (important in forensics and in court)

Statistics

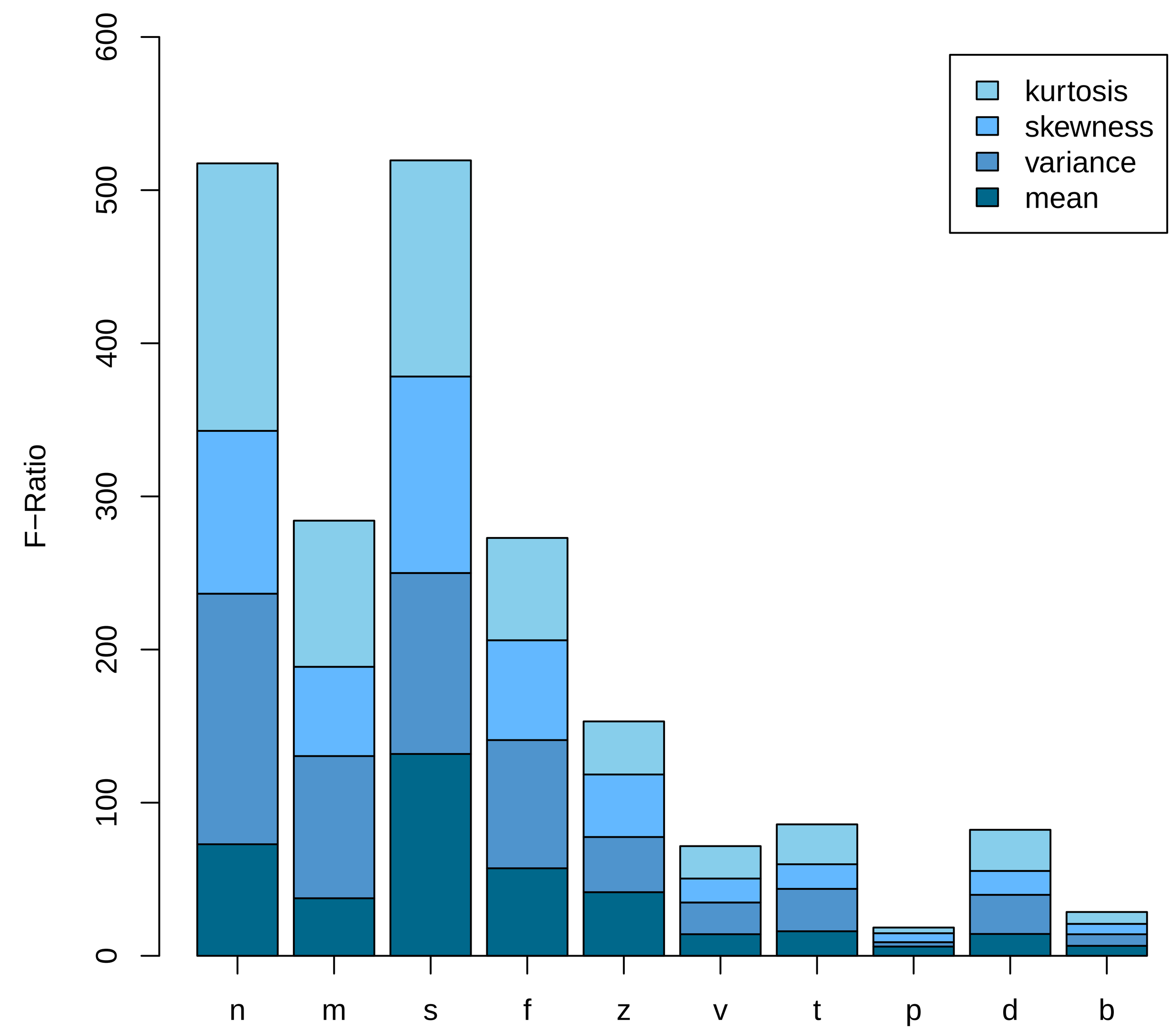
- The F-ratio relates intra- and inter-speaker variabilty
- The larger the F-ratio, the higher the speaker discriminating potential

$$F = \frac{\text{inter-speaker variance}}{\text{intra-speaker variance}} \quad (1)$$

$$F = \frac{\frac{n}{m-1}\sum_{j=1}^{m}(\mu_j - \bar{\mu})^2}{\frac{1}{m(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \mu_j)^2} \quad (2)$$

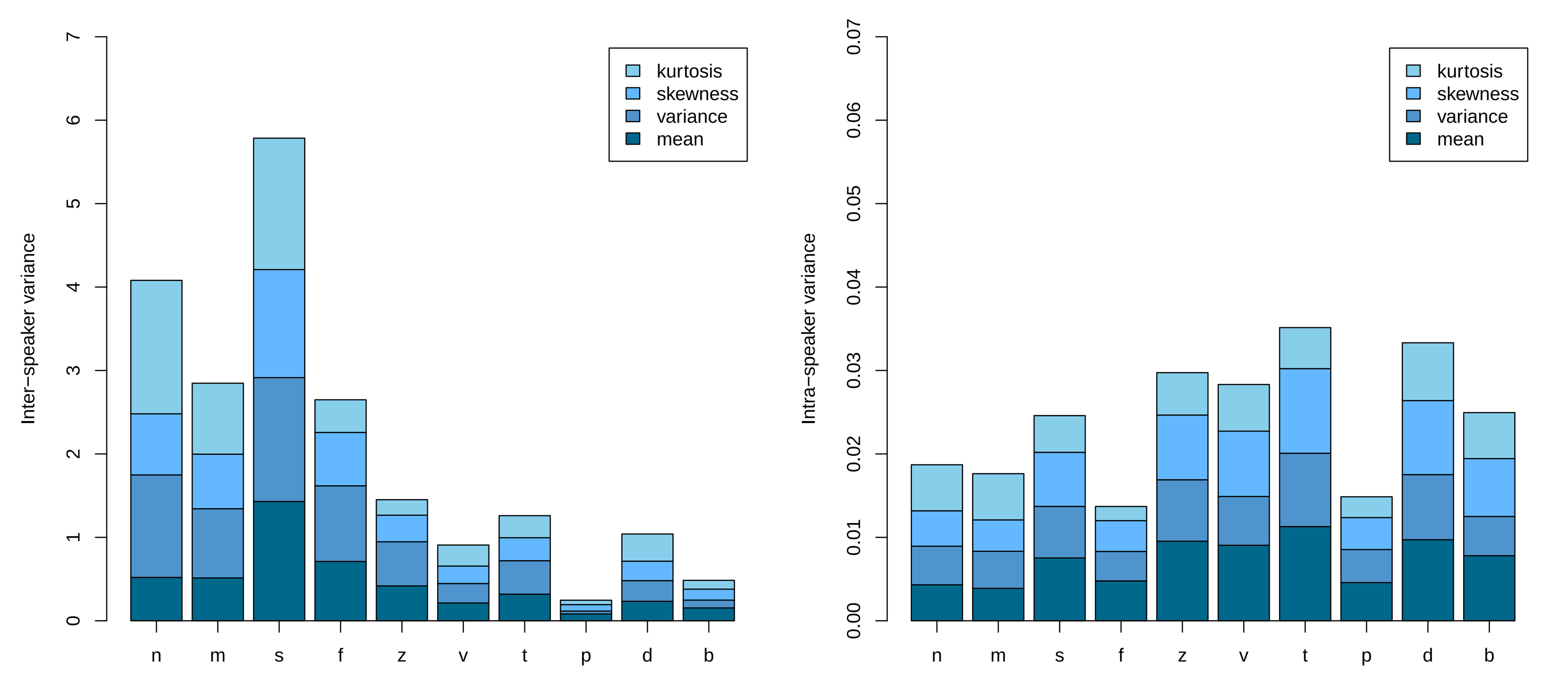
n is the number of utterances, m the number of speakers and x_{ij} the value of the parameter in the ith utterance of the jth speaker.

Results (Summary)



The spectral moments of the alveolar phoneme reached always a higher F-ratio than its labial counterpart.

What causes the differences in the F-ratios?



- Both, the intra-speaker and the inter-speaker variance, are higher for the alveolar than for the labial phonemes,
- but the inter-speaker variance to a much higher degree.

Discussion

The intra-speaker and inter-speaker variance are both higher for alveolar phonemes.

Possible Explanations:

- Different number of resonating cavities: labial phonemes only have one oral resonating cavity, whereas alveolar phonemes have two such cavities –> more room for variance?
- There are more alveolar than labial phonemes. –> Numerical effect? –> No, for instance there are more /n/ than /s/ phonemes, but they still have the same F-ratio.
- Does the context influence the alveolar phonemes more than the labial? –> May be, but that would affect both the inter-speaker and the intra-speaker variance to the same extent.