Laboratório de Fonologia
Universidade Federal de Minas Gerais
Faculdade de Letras • sala 3104

CEFALA

POSLIN·UFMG FALE FACULDADE DE LETRAS UFMG

Articulatory and vocal speaker variability in connected speech

Maria Mendes Cantoni (Universidade Federal de Minas Gerais, mmcantoni@gmail.com)
Adelino Pinheiro Silva (Police Academy of Minas Gerais, adelinocpp@gmail.com)

1 INTRODUCTION

Sources of variation in speech [1,2]

Linguistic variation: phonetic-phonological, coarticulatory

Speaker-related variation: sociolinguistic, personal

Between speaker variability: anatomical or physiological (difference on vocal tracts or difference on motor routines used by different speakers)

Within-speaker variability: biomechanical (differences on how speech movements are actually implemented by the same individual.)

Problems

1) The role of different components of the vocal tract in speaker identification is not clear [3].

2) Only a few studies on speaker variability have used connected speech [4].

In this study

We address the two types of speaker variability, with the aim to untangle the role of articulatory structures and voice in speaker identification in connected speech.

We intend to answer the following questions:

- How much speaker variation is due to articulation differences and how much is due to voice differences?
- Which acoustic measures are more robust for speaker identification in connected speech?

2 METHODS

Materials

- Vowels in spontaneous speech
- Recordings from 18 speakers from CEFALA-1, a Brazilian Portuguese database [5]
- Manual segmentation and labelling in Praat [6]

Data Coding

- Linguistic variables relevant to the language sound system: preceding and following context, vowel quality, nasality, stress degree, number of syllables, syllable structure

Measurements

- Acoustic measurements after [3], divided into articulatory and vocal and estimated in mean and variation coefficient: duration, formants (F1, F2, F3, F4 and dispersion), intensity, f0, spectral slope at four spectral regions, SNR, CPP

Modelling [cf. FIG 1]

A - Selection of vowels common to all speakers

B - Feature extraction (measurements)

C - Bootstrapping

D - Data split into training (70%) and test sets

E - Generalized linear regression model (GLM) fit, with linguistic variables as factors $X_n \approx V_{C|n} + (V_{S|n} + V_{NC|n+\varepsilon n}) \rightarrow X_n \approx V_{C|n} + \varepsilon_{S,NC|n}$

Using the normalized Euclidean-distance of residuals from GLM (GLM_RES):

F - Logistic regression model fit (training set only)

G - Logistic inference for same (SS) and different speaker (DS) classification

Model evaluation

- Empirical cross-entropy (ECE) calculation for model fit evaluation
- Multi-dimensional scaling (MDS) for comparison with raw data and also PCA from raw data

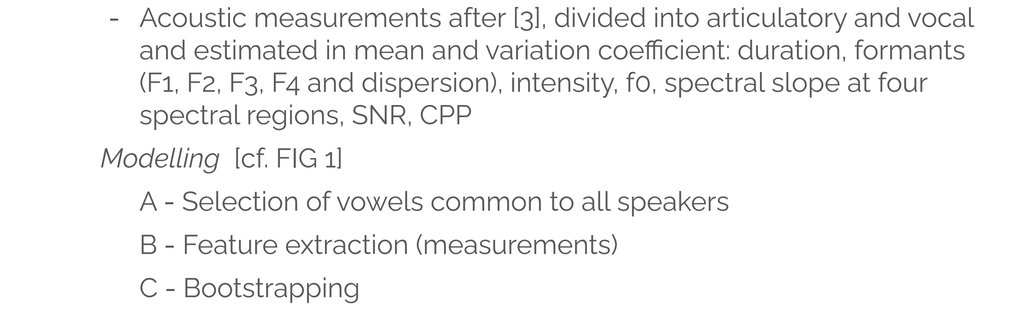


FIG 1 - Modelling procedures

3 RESULTS

-

PCA from raw data:

- Clear separation between female and male voices
- 9 PCs to account for 95% variance

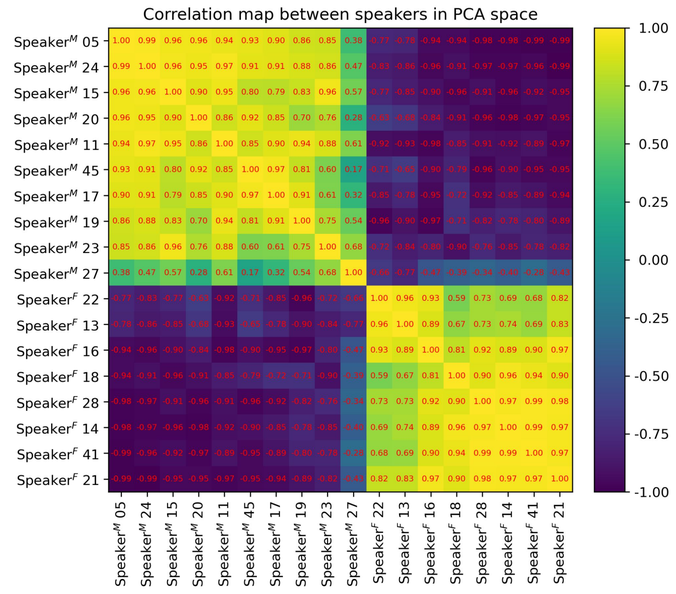


FIG 2 - Speakers correlation map using PCA

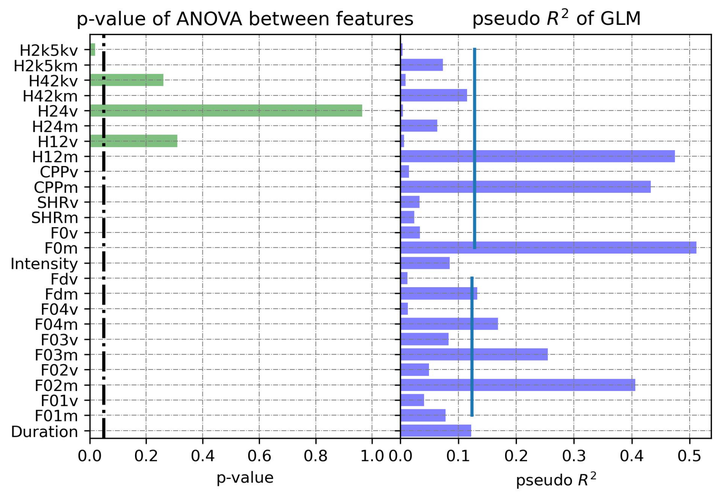
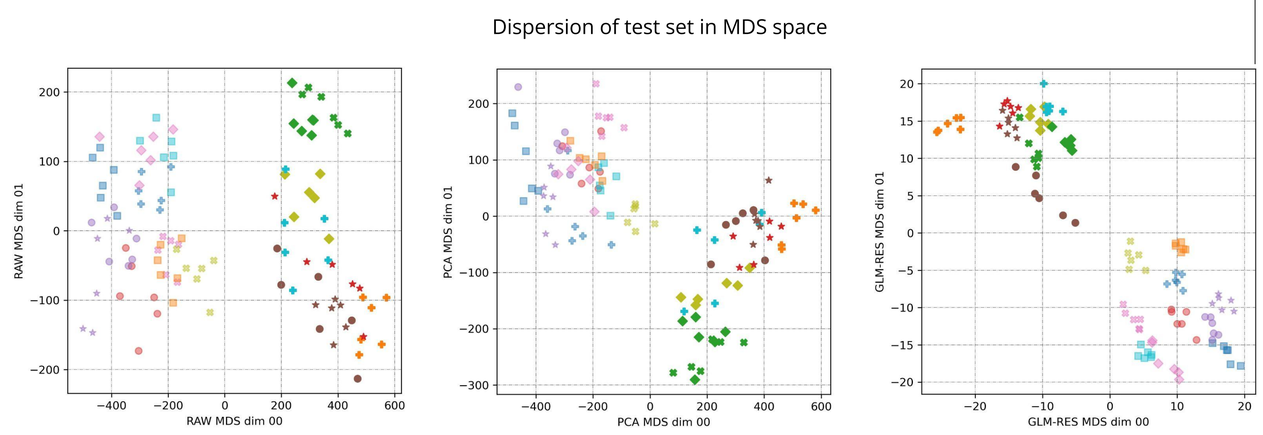
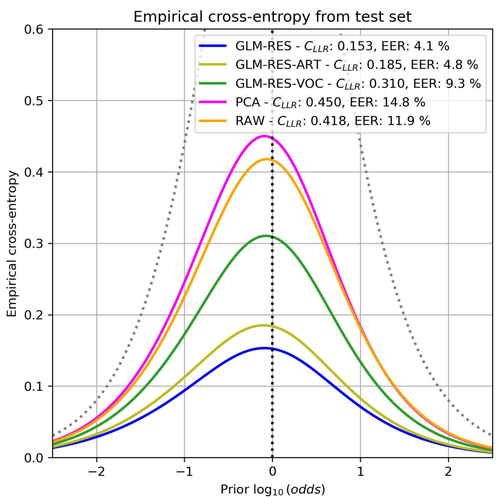


FIG 3 - Significance of each variable in raw data and contribution to model fit in GLM



- XXXXXXXXXXXX

FIG 4 - Multi-dimensional scaling comparative analysis of classsification



- Empirical cross-entropy (ECE, [5]): a measure of classification
- RES (model) performed better than PCA and a classification of the raw data (RAW), with farthest the from classification at chance (dotted line)

FIG 5 - Empirical cross-enropy

4 FINAL REMARKS

-

[1] Ladofoged and Broadbent (1957)
[2] Kilbourn-Ceron and Goldrick (2021)
[3] Lee, Keating and Kreiman (2019)
[4] Lee and Kreiman (2022)
[5] Yehia, Follador and Silva (2019)

[1] Ladofoged and Broadbent (1957)
[2] Kilbourn-Ceron and Goldrick (2021)
[3] Lee, Keating and Kreiman (2019)
[4] Lee and Kreiman (2022)
[5] Yehia, Follador and Silva (2019)