Towards articulatory comparison between French and German from MRI data

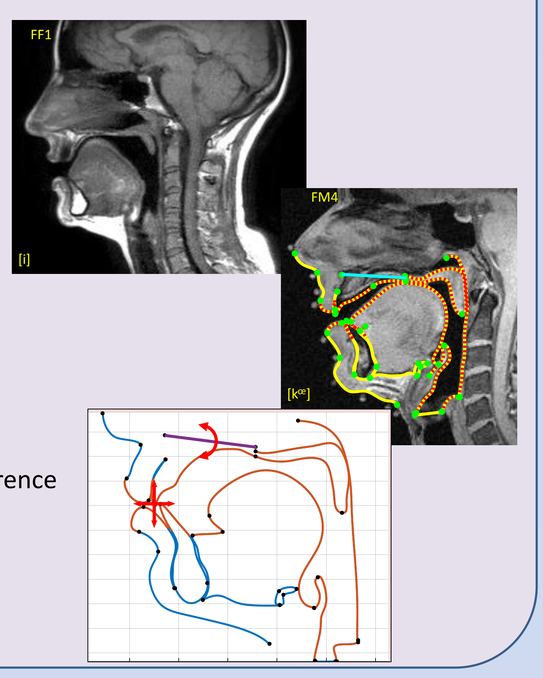
Antoine Serrurier[1,2] & Christiane Neuschaefer-Rube[1]

[1]Department of Phoniatrics, Pedaudiology and Communication Disorders
[2]Department of Medical Informatics
University Hospital and Medical Faculty of the RWTH Aachen University, Germany
aserrurier@ukaachen.de & cneuschaefer@ukaachen.de

1 Context, objectives & approach

- **Context**
  - The vocal tract is used for different **articulatory tasks**: breathing, singing, feeding, speaking different languages, etc.
  - Each task can be associated with a specific **articulatory space**: the range of articulations theoretically producible by making use of the elementary articulatory components corresponding to the task.
  - **Comparing the articulatory spaces**
    - informs about the articulatory gap between the tasks
    - informs about the extent of the required articulatory transfer from one task to another (*e.g.* for second language learning)
  - Challenging comparison due to the large inter-speaker variability
    - Morphology
    - Articulatory strategy
- **Objective**: Compare articulatory spaces of **French (FR)** and **German (DE)**
  - Based on real speaker articulations
  - Taking advantage on an articulatory modelling approach
  - Measuring the discrepancy between the 2 spaces
- **Technical approach**
  - Consider two large datasets of articulatory contours for FR and DE
  - Normalise the contours between speakers to remove the inter-speaker variability related to the morphology
  - Analyse the articulatory variability between the two datasets
  - Compare the articulatory spaces by cross-reconstructing each dataset by models derived from the other one and analyse the errors
  - Compare the articulatory spaces by projecting the two datasets in the same articulatory space

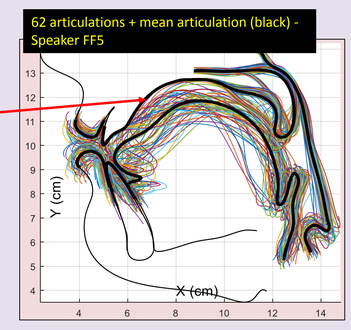
2.1 Speakers & data

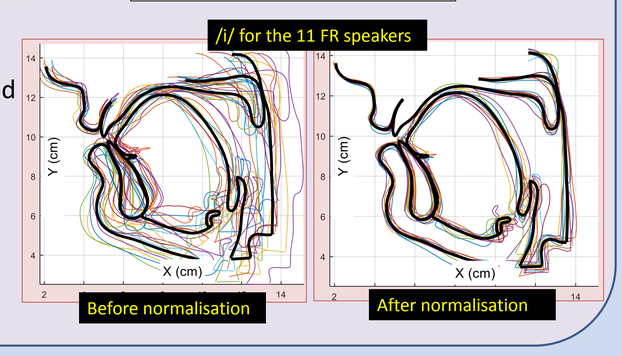


- **Speakers**
  - 11 French speakers
  - 10 German speakers
- **Corpus**: 62 representative sustained articulations for each language
- **Data**: static midsagittal MRI of the vocal tract
- **Processing**
  - Manual organ-based contour segmentation
  - Alignments of the contours on a cranium-based reference coordinate system

2.2 Speaker normalisation

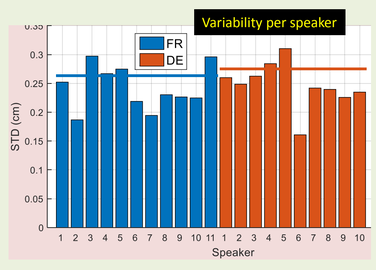
- Remove the variability due to the morphology
- The mean articulation of a speaker is considered to characterise its morphology (large and balanced corpus)
- Remove for each articulation of each speaker the marginal difference between the speaker mean articulation and the overall mean articulation (neutral articulation)
- The remaining inter-speaker variability is considered to be only related to the articulatory strategy
- All further processing done on the normalised articulations

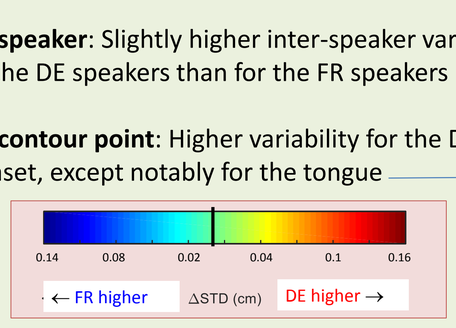


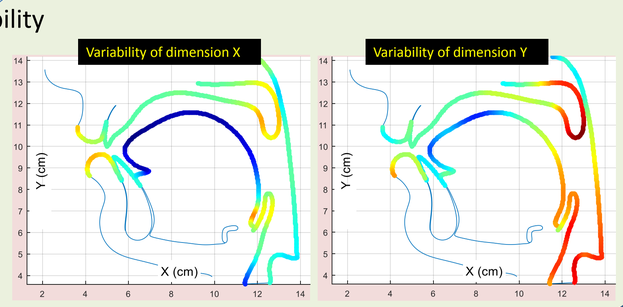


3.1 Variability analysis

- Calculation of the articulatory variability of the two datasets in terms of Standard Deviation (STD)
- **Overall**: Very slightly higher overall variability for the DE dataset (STD = 0.28 cm) than for the FR dataset (STD = 0.26 cm)
- **Per speaker**: Slightly higher inter-speaker variability for the DE speakers than for the FR speakers
- **Per contour point**: Higher variability for the DE dataset, except notably for the tongue →

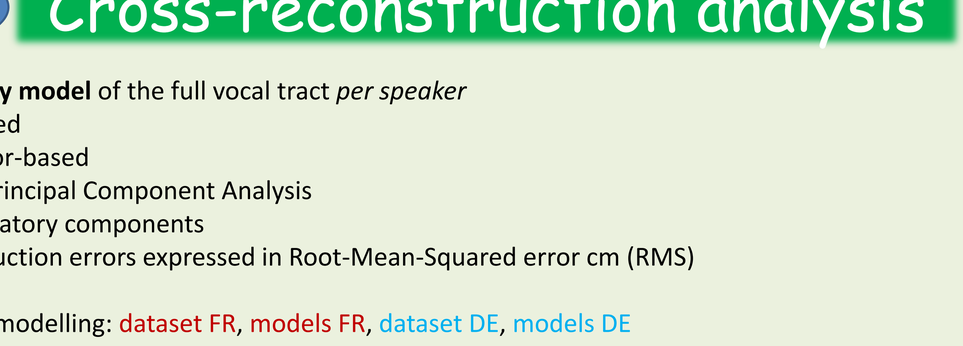




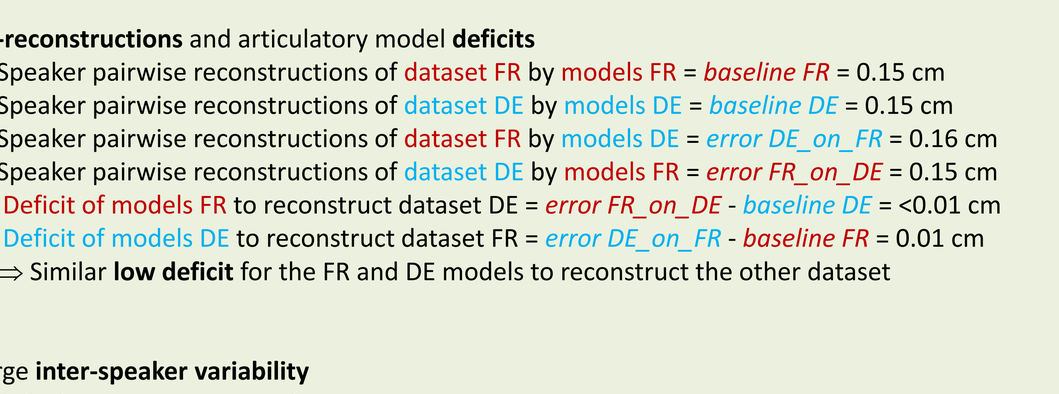


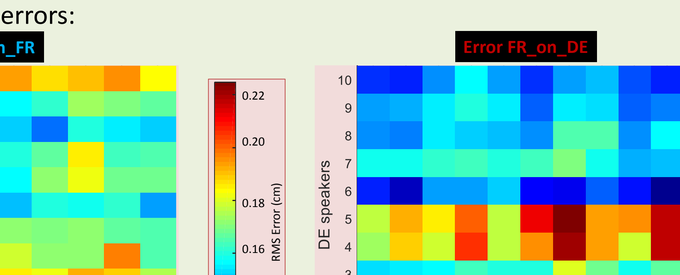
3.2 Cross-reconstruction analysis

- One **articulatory model** of the full vocal tract *per speaker*
  - Data-based
  - Articulator-based
  - *Guided* Principal Component Analysis
  - 14 articulatory components
  - Reconstruction errors expressed in Root-Mean-Squared error cm (RMS)
- Available after modelling: dataset FR, models FR, dataset DE, models DE
- **Cross-reconstructions** and articulatory model **deficits**
  - Speaker pairwise reconstructions of dataset FR by models FR = *baseline FR* = 0.15 cm
  - Speaker pairwise reconstructions of dataset DE by models DE = *baseline DE* = 0.15 cm
  - Speaker pairwise reconstructions of dataset FR by models DE = *error DE_on_FR* = 0.16 cm
  - Speaker pairwise reconstructions of dataset DE by models FR = *error FR_on_DE* = 0.15 cm
  - Deficit of models FR to reconstruct dataset DE = *error FR_on_DE* - *baseline DE* = <0.01 cm
  - Deficit of models DE to reconstruct dataset FR = *error DE_on_FR* - *baseline FR* = 0.01 cm
  - ⇒ Similar **low deficit** for the FR and DE models to reconstruct the other dataset
- But large **inter-speaker variability**
  - Pairwise cross-reconstructions errors:



  - Model deficits per articulation and contour point





3.3 Space projection analysis

- One universal articulatory model of the full vocal build on the whole dataset
- Speaker reconstructions by this universal model = projection of each speaker into the articulatory space of the universal model
- No statistical difference in the range of use of each articulatory component between the FR and DE datasets: FR and DE speakers seem to use the same articulatory components in a similar range

4 Discussion & conclusion

- Comparison of FR and DE articulatory spaces
  - Variability of the DE dataset slightly higher, except for the tongue
  - Similar articulatory spaces
    - Tongue tip of DE models for FR dataset?
    - Velum uvula of FR models for DE dataset?
  - ⇒ The native articulatory degrees of freedom of FR and DE seem sufficient to form articulations of the other dataset
  - Formalisation of an approach to compare the articulatory spaces of two datasets
- Open discussion points
  - Are the tips of the velum and tongue less constrained by the language and more speaker-specific?
  - Large inter-speaker variability: larger than inter-language variability?
  - Observed deficit at the border of the model precisions
  - Number of articulatory component per speaker always the same?