Predicting Tongue Shapes From A Few Landmark Locations

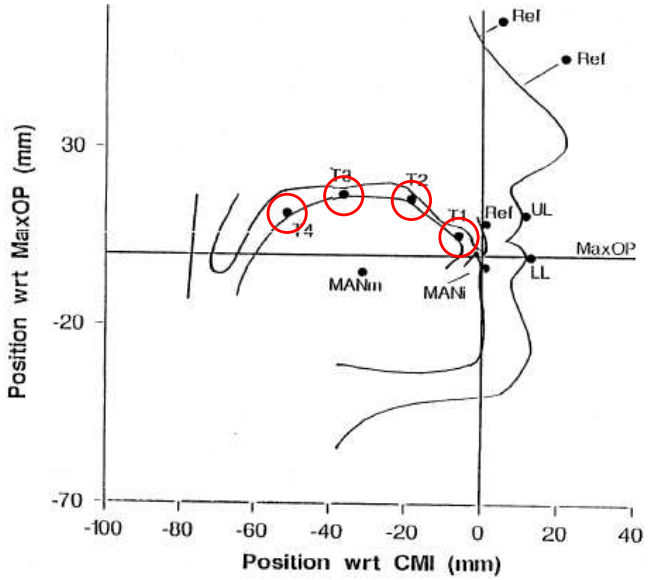
Chao Qin[1], Miguel Á. Carreira-Perpiñán[1],
Korin Richmond[2], Alan Wrench[3], Steve Renals[2]

[1]EECS, School of Engineering, UC Merced, USA
[2]Centre for Speech Technology Research, University of Edinbugh, UK
[3]Queen Margaret University, Edinburgh, UK

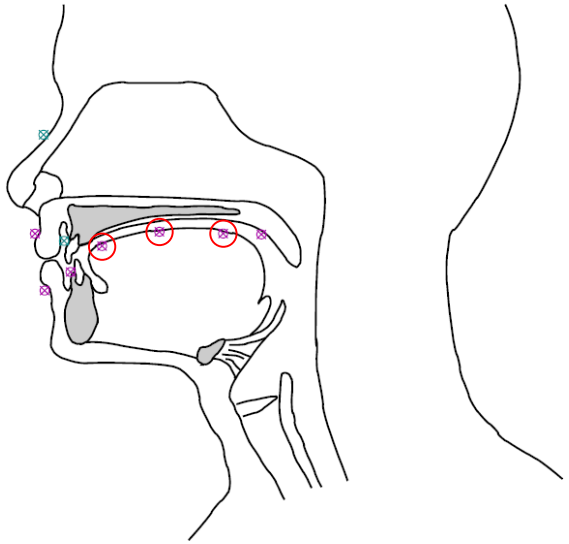
Introduction

- Tongue is the most important speech production articulator
- Articulatory datasets only provide sparse representation of tongue.

Wisconsin X-ray microbeam



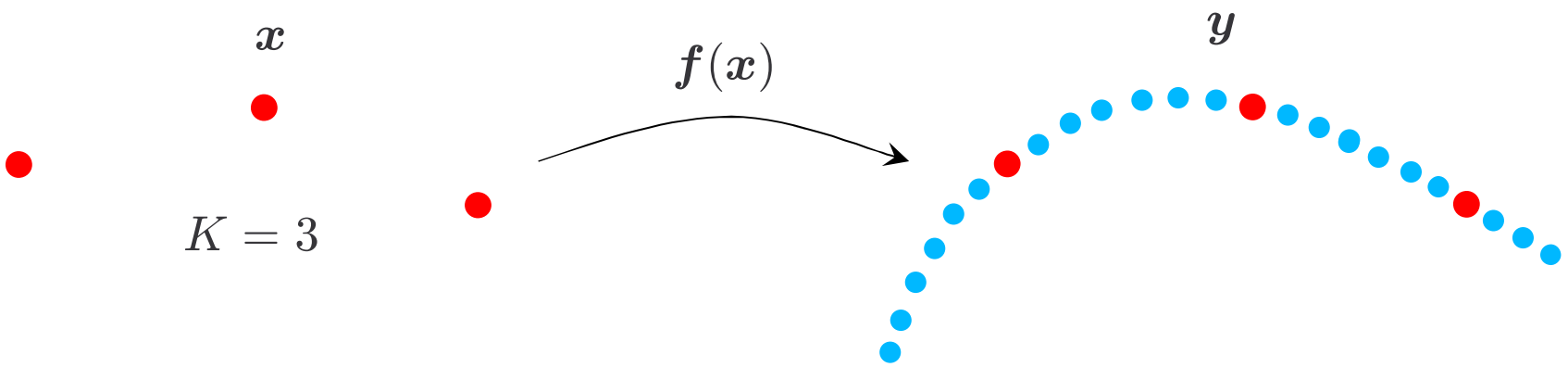
MOCHA



- Questions
  1. Are these 3 or 4 pellets sufficient to reconstruct the tongue shape?
  2. How many are necessary for an accurate reconstruction?
  3. Where to place them optimally?

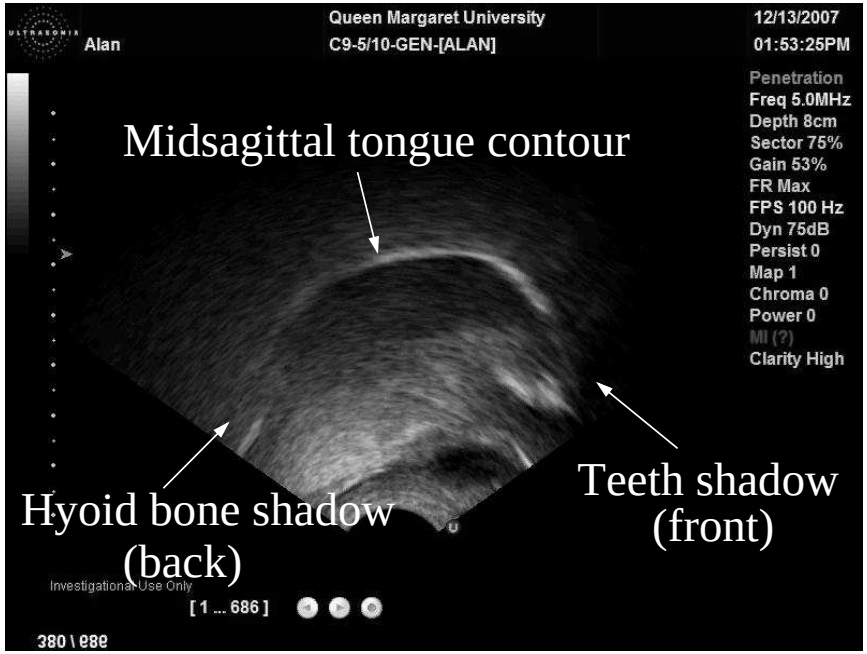
Machine learning approach

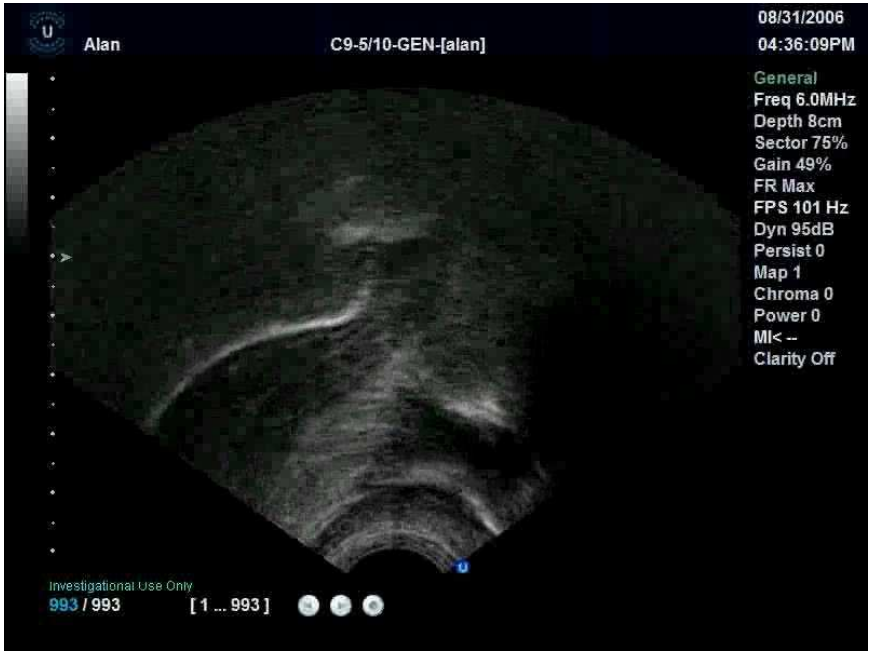
- Assume midsaggital contours
- Collect a training set $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathbb{R}^D$ of tongue contours (ground truth)
- Predict a test contour $\boldsymbol{y}$ from the location $\boldsymbol{x}$ of K pellets using a nonlinear regression: $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$
- Estimate the mapping $\boldsymbol{f}$ from the training set (least-square)



Data collection

- Ultrasound data of tongue movement





Data collection

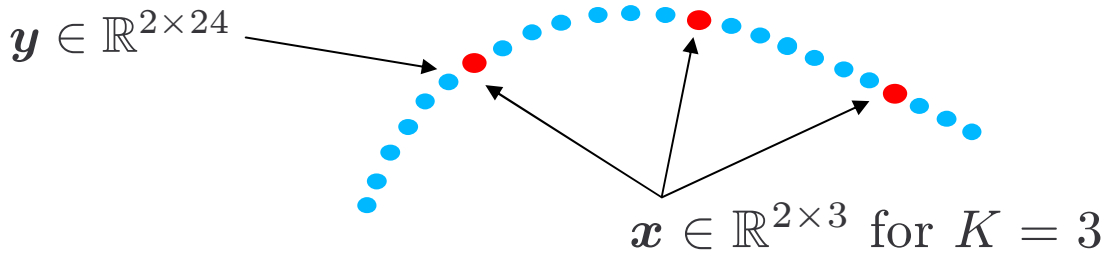
- Ultrasound machine and head stabilization device (QMU)

Data collection

- Tongue contour tracking
  - A difficult task due to noisy ultrasound images
  - Tongue parts are invisible from time to time
  - Our solution: automatic + manual correction
    - Automatic tracking by EdgeTrak (Li et al' 05), based on snake segmentation

- Tongue contour dataset
  - One native English speaker with Scottish accent
  - 20 read TIMIT sentences
  - $N = 6000+$ tongue contours and audio
    - Each contour $\boldsymbol{y} \in \mathbb{R}^{2\times 24}$ = 2D position of 24 points

Reconstructing tongue shape from a few landmarks

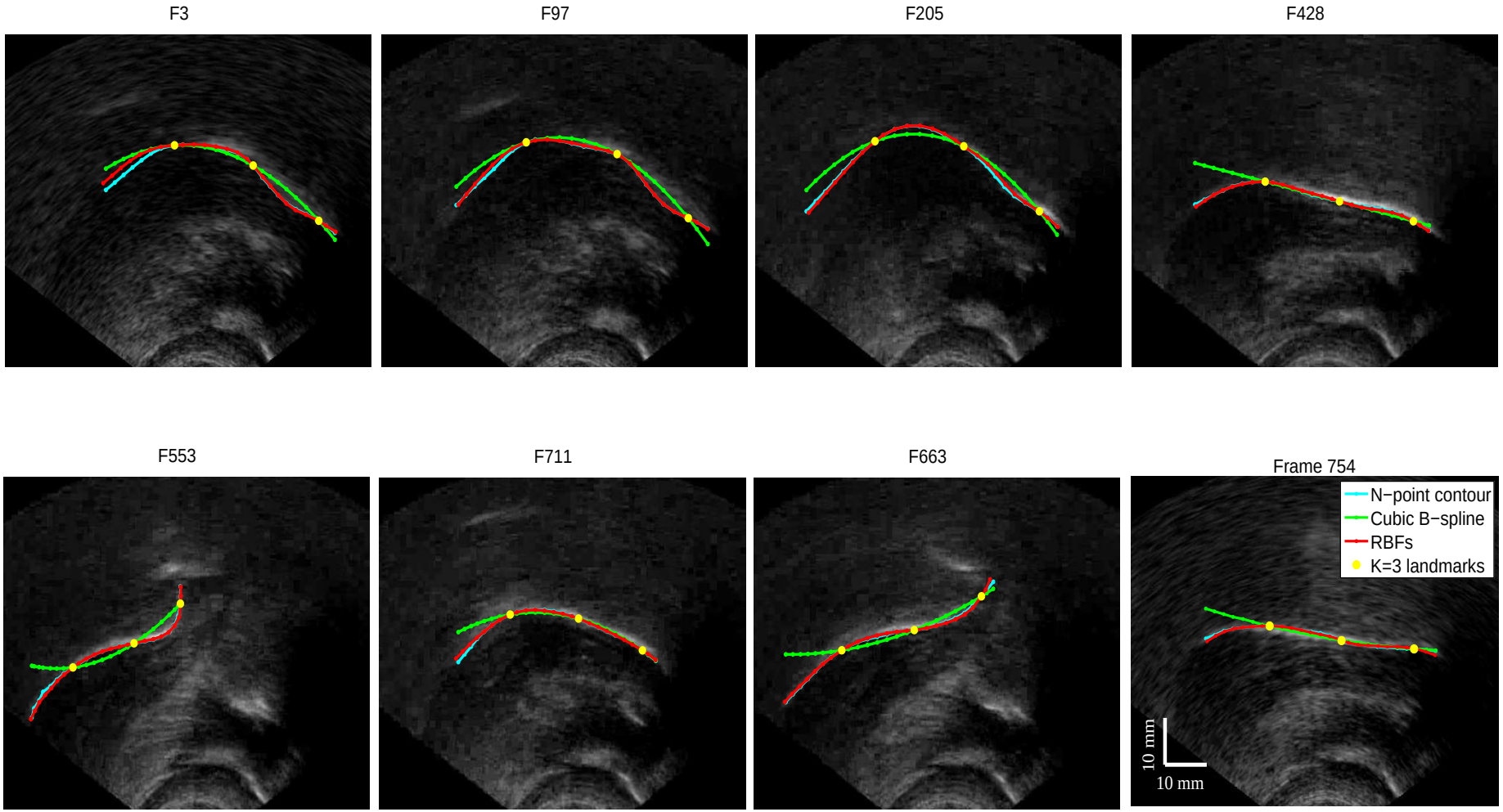


- Unsupervised spline interpolation
  - Uses only information in the K landmarks
  - Smooth but easy to penetrate the palate or teeth, poor extrapolation
- Supervised prediction: learn mapping $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ using a training set
  - Linear prediction
  - Nonlinear prediction

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{\Phi}(\boldsymbol{x}), \phi_i(\boldsymbol{x}) = \exp\left(-\tfrac{1}{2}\|(\boldsymbol{x} - \boldsymbol{\mu}_i)/\sigma\|^2\right)$$

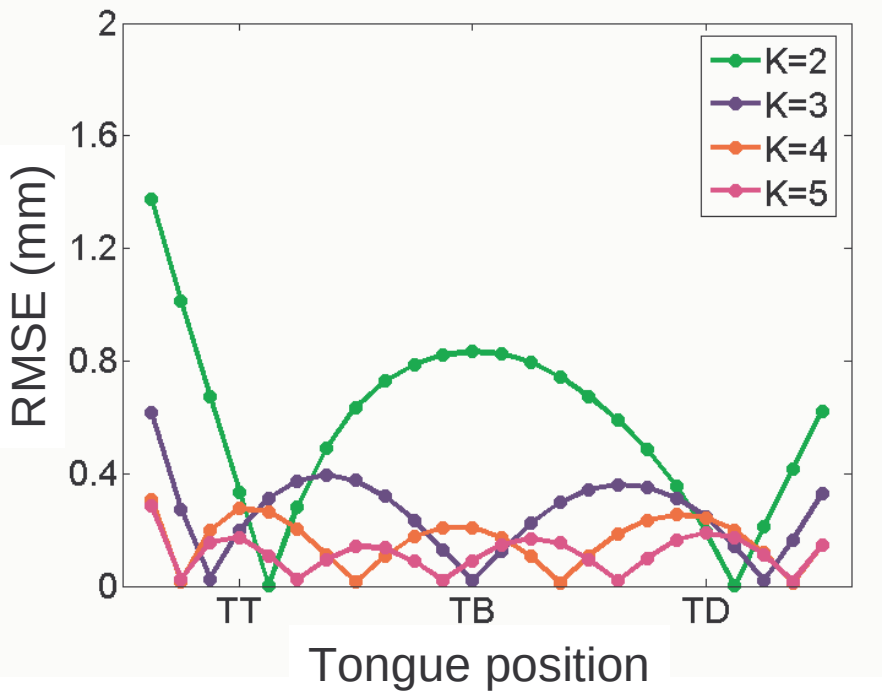
- We use Gaussian Radial Basis Function networks (RBF)
  - Universal mapping approximators
  - Simple and fast training

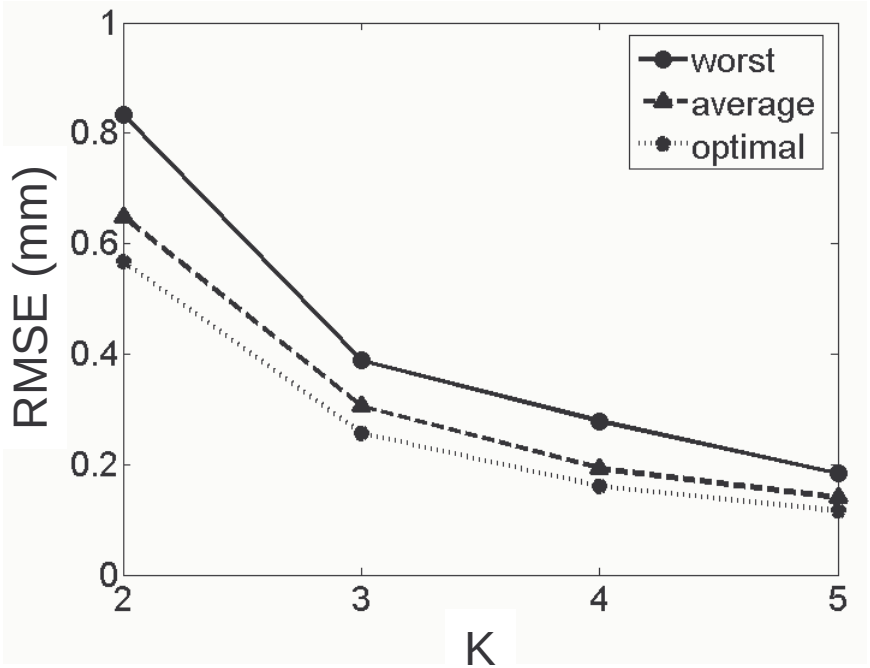
Experimental results



Experimental results by RBF prediction

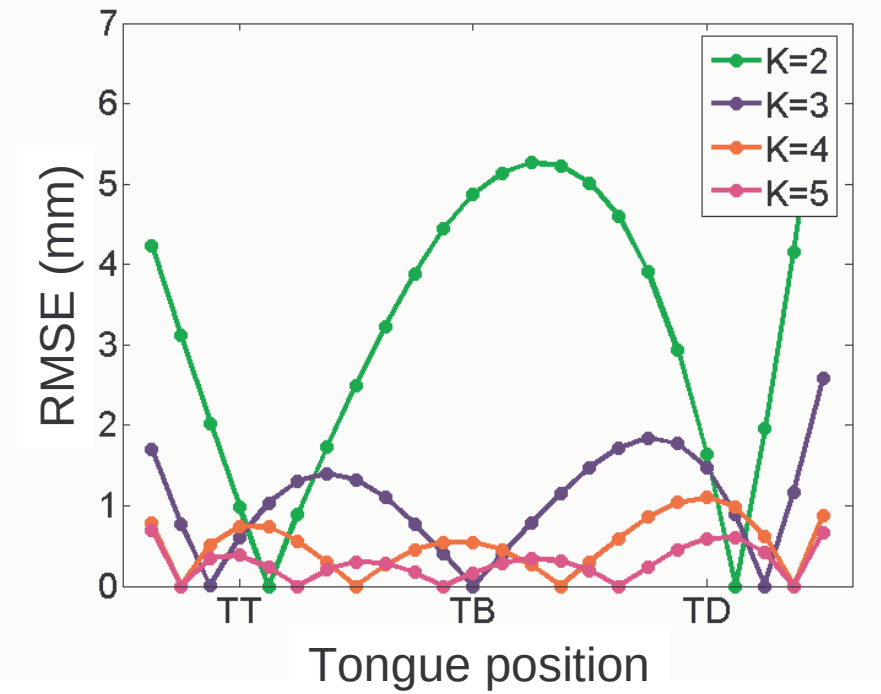
- Landmarks $\boldsymbol{x}$: test each of the $\binom{P}{K}$ combinations, $P = 24, K = 2, 3, 4, 5$
- Ignore unreasonable arrangements of landmarks
  - Divide the contour into K consecutive segments
  - Constrain each landmark to select points from one segment

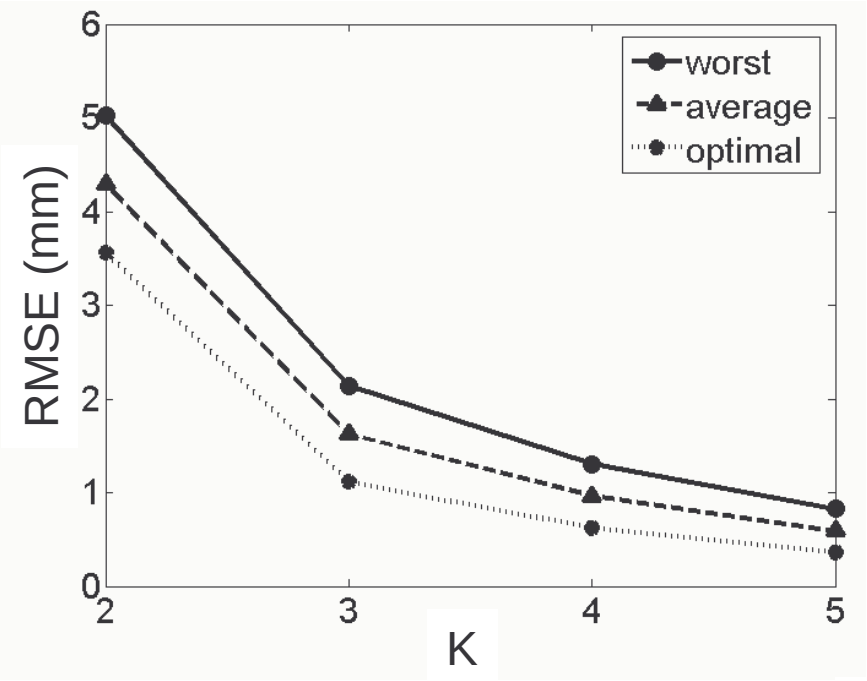




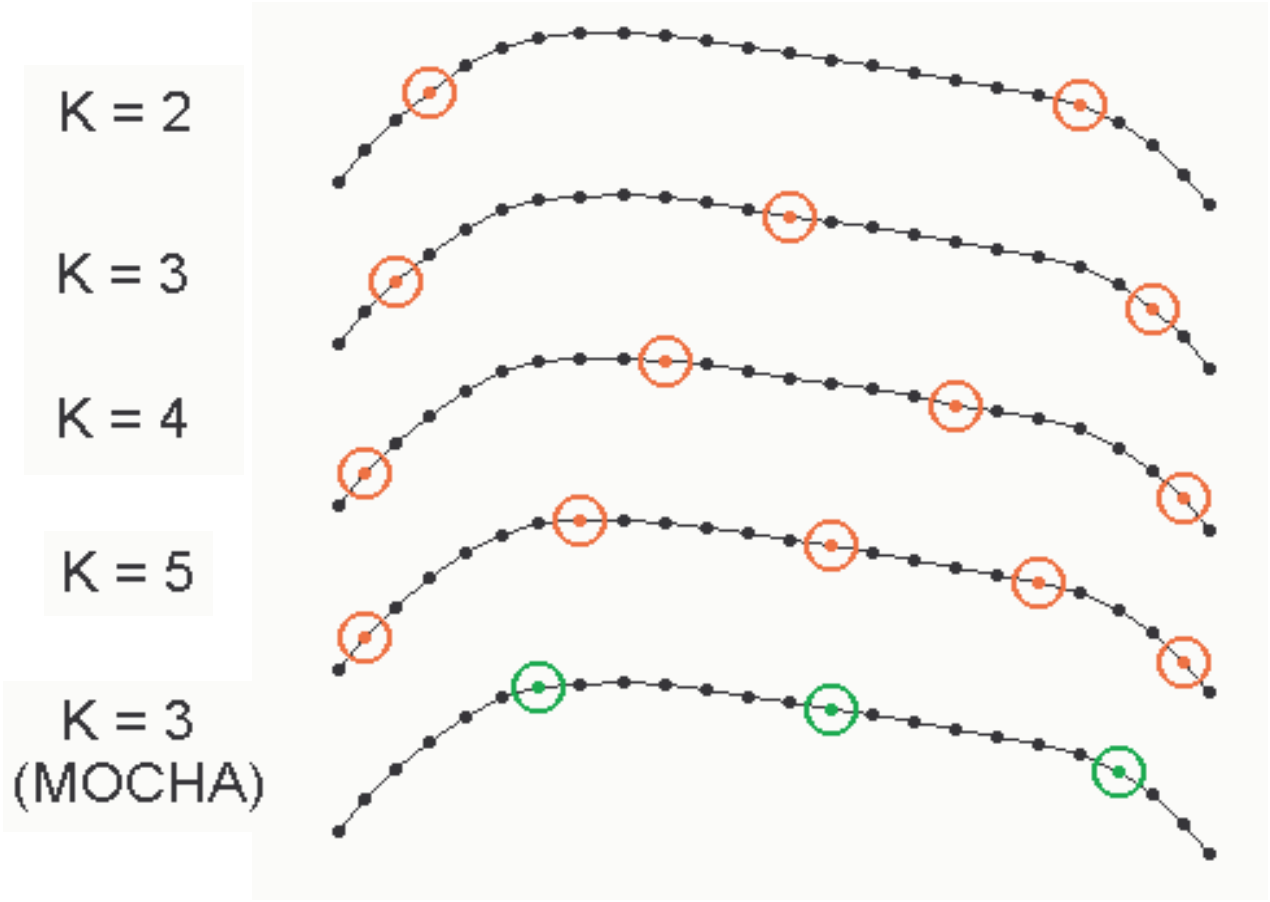
Experimental results by spline interpolation

- Run spline interpolation on the same landmarks' locations as RBF
- Worse than RBF prediction by an order of magnitude





Optimal locations of landmarks



Practical rule: quasi-equidistant placement, more landmarks on the tongue tip

Conclusions

- Using 3 or 4 landmarks is sufficient to predict the tongue shape by a nonliear mapping with RMS error below 0.4mm
- Nonlinear prediction can predict very realistic tongue shapes and is much more reliable than spline interpolation
- Slight difference between the best and the worst landmarks placement
- Useful for determining optimal number and locations of landmarks for EMA and X-ray microbeam techniques
- Approach applicable to reconstruct 3D tongue shapes if 3D data available
- Future work
  - Speaker adaptation
  - Tongue contour animation for vocal tract visualization
  - Augment tongue pellets in MOCHA and X-ray datasets, eg. for articulatory inversion

- Supported by NSF CAREER award IIS-0754089 and Marie Curie Early Stage Training Site EdSST (MESTCT-2005=020568)

Acknowledgement

- Thanks D. Massaro and M. Cohen (UC Santa Cruz) for useful discussions