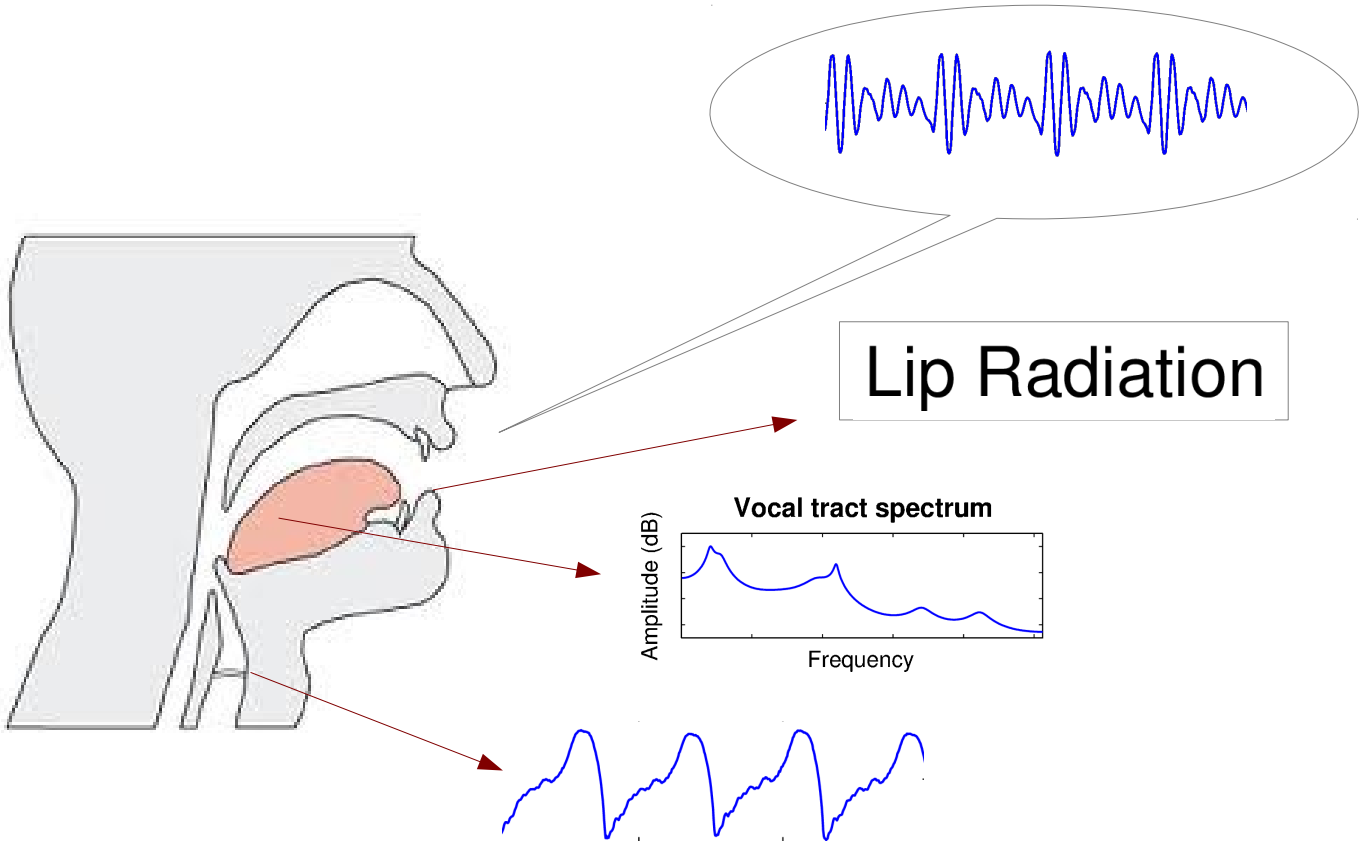
Using phonetic feature extraction to determine optimal speech regions for maximising the effectiveness of glottal source analysis

John Kane, Irena Yanushevskaya, John Dalton, Christer Gobl, Ailbhe Ní Chasaide

Monday August 26th, 2013
Interspeech
Lyon, France

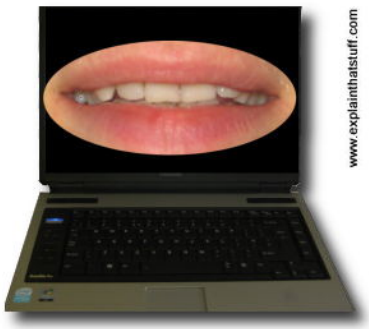
Glottal source analysis

Speech production / Glottal inverse filtering



Glottal source in speech technology

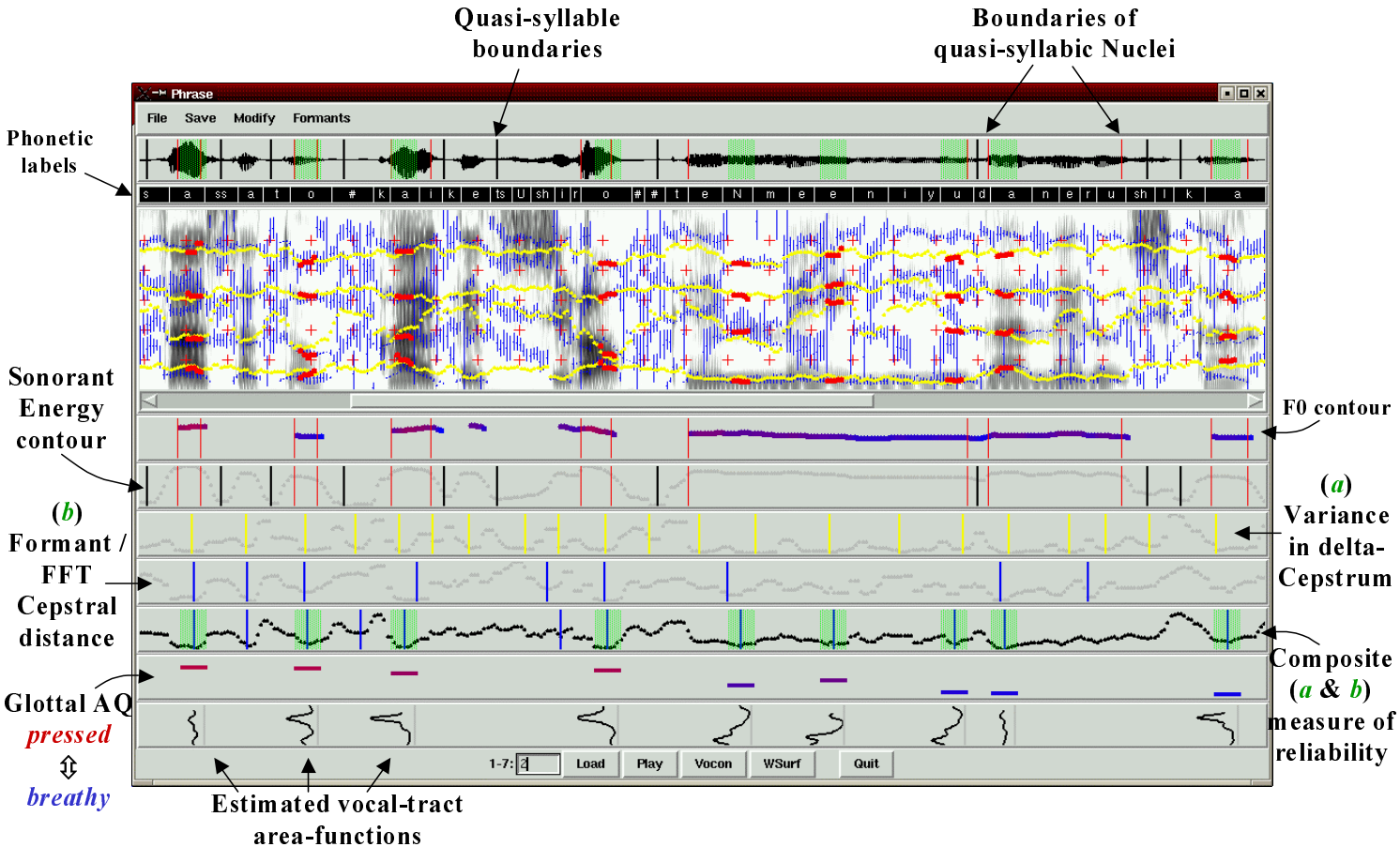
Speech synthesis



Speech recognition

Speaker verification

Previous work - Centres of reliability (Mokhtari et al.)



Previous work - Phonetic feature extraction

- **Speech synthesis**
- **Speech recognition**

Introduction - Research aims

1. Implement a method for detecting binary phonetic features
2. Quantitatively evaluate phonetic-sensitive glottal source processing

Phonetic feature extraction

Phonetic feature extraction - Speech data & target labels

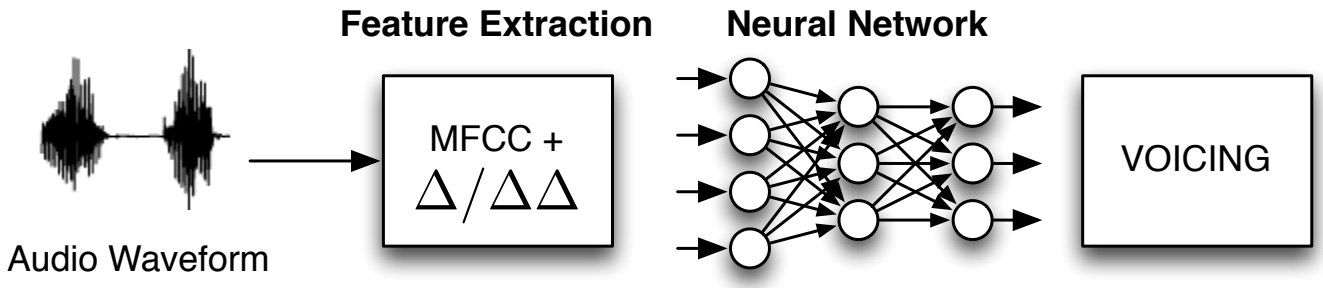
- **ARCTIC:** 9 English speakers, 1000+ sentences each
- **IIIT:** 6 speakers of different Indic languages, 1000 sentences each

Binary phonetic classes: {Voiced, fricative, nasal, high vowel}

Target labelling: e.g., FRICATION

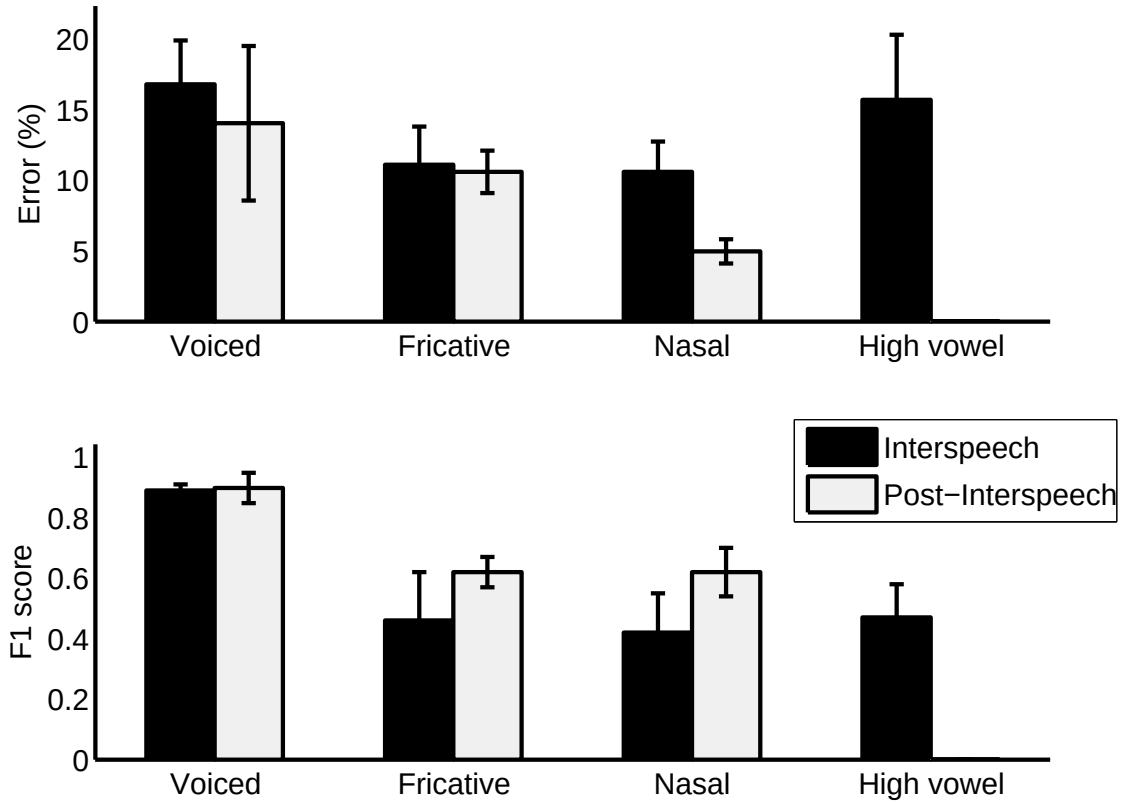
- /a/ => 0
- /f/ => 1
- /t/ => 0

Phonetic feature extraction - Features & learning

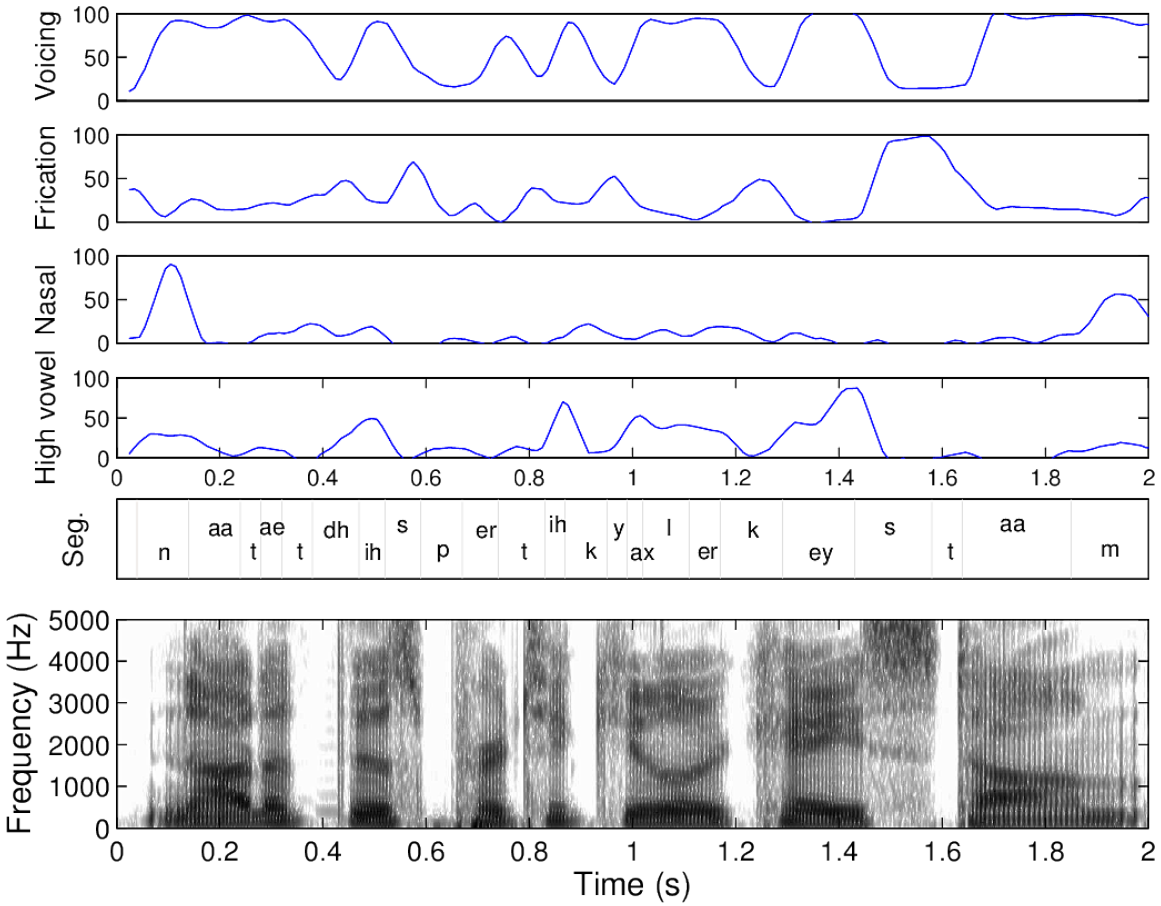


- **Features:** 13 MFCCs with Δ and ΔΔ
- **ANN:** Multi-layer perceptron, one hidden layer, 100 neurons

Phonetic feature extraction - Speaker independent results

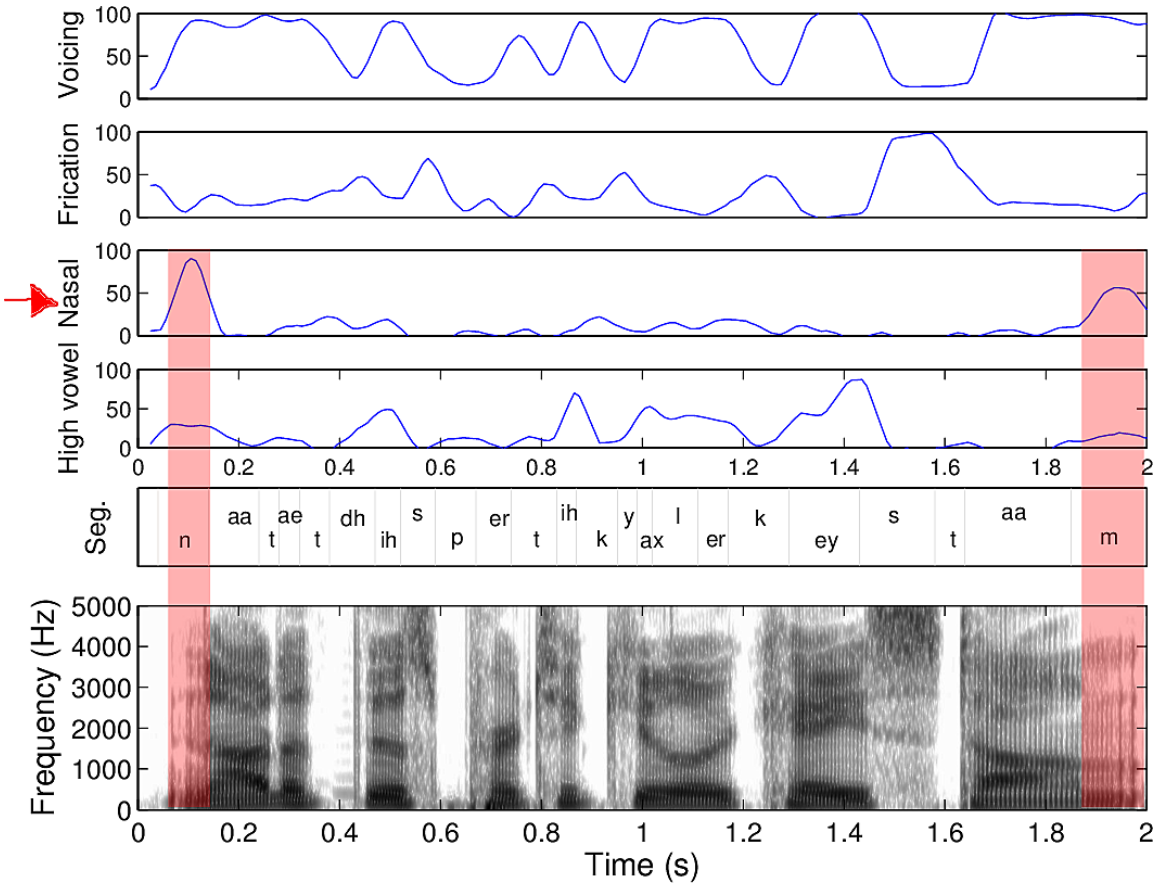


Phonetic feature extraction - Illustration



"Not at this particular case Tom ..."

Phonetic feature extraction - Illustration



"Not at this particular case Tom ..."

Glottal source processing

Glottal source processing

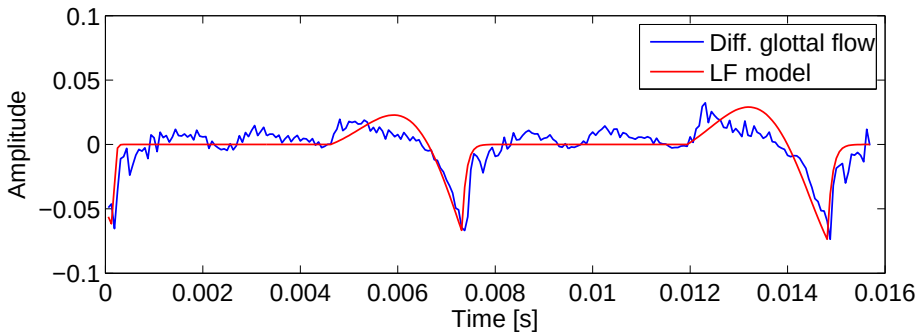
- Glottal source analysis difficult to quantitatively evaluate
- Assessed implicitly here through voice quality classification experiments

Glottal source processing - Speech data

- 6 speakers, 17 TIMIT utterances in 3 phonation types (breathy, modal, tense)

Glottal source processing - Features

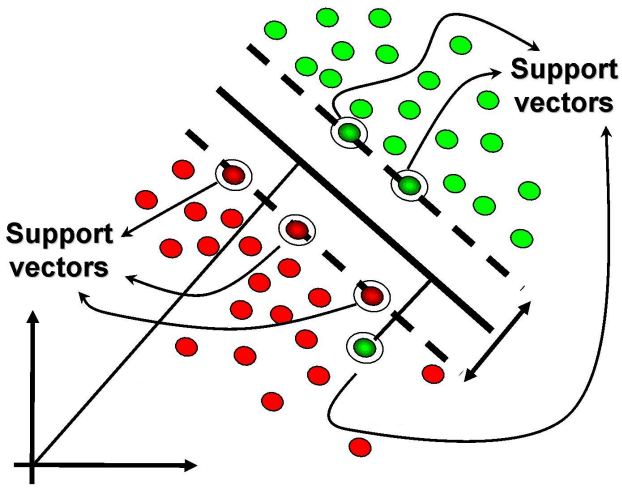
- **Model parameters:** Liljencrants-Fant (LF) model fit using dyProg-LF algorithm => {Ra, Rk, Rg}



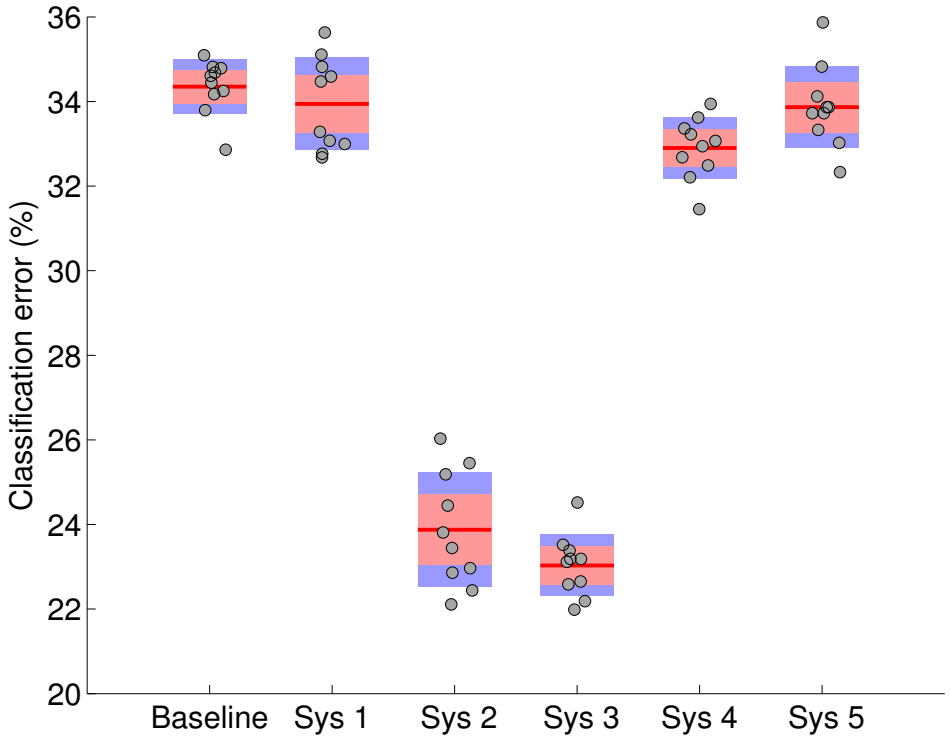
- **Direct parameters:**
  - **NAQ:** Normalised Amplitude Quotient
  - **QOQ:** Quasi-Open Quotient
  - **H1-H2:** Difference in amplitude of first two glottal harmonics

Glottal source processing - Classification

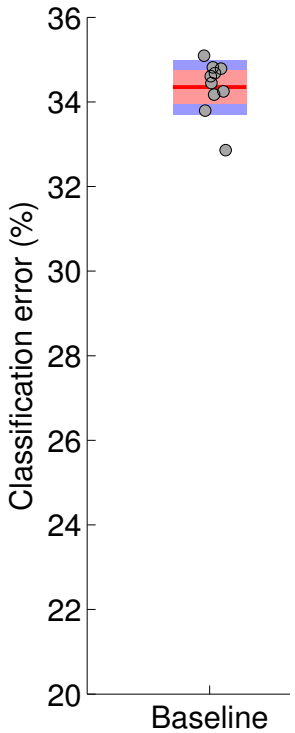
- **Support Vector Machines (SVMs):**
  - One-against-one multi-class architecture
  - Radial Basis Function (RBF) kernel
  - 10-fold cross-validation experiments (incrementally removing feature data from certain phonetic regions)



Glottal source processing - Results

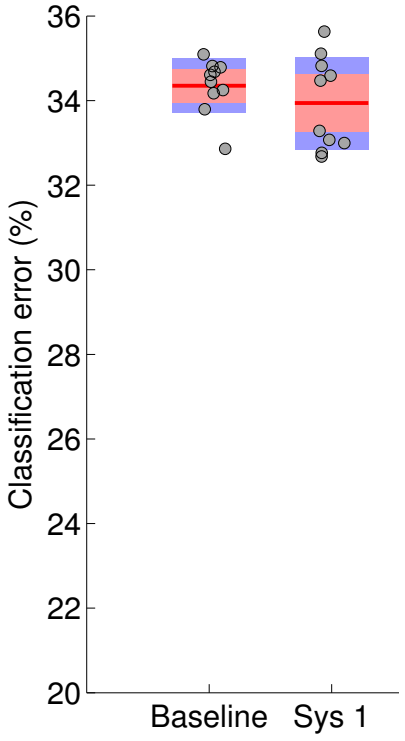


Glottal source processing - Results



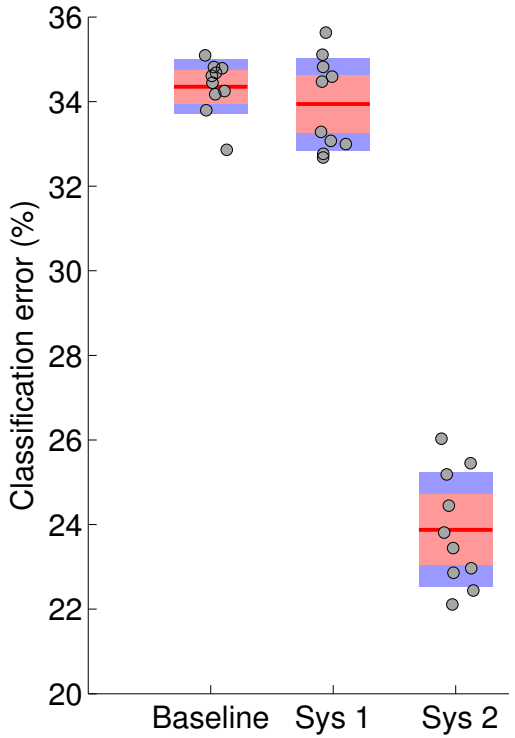
BASELINE: Using all glottal feature data

Glottal source processing - Results



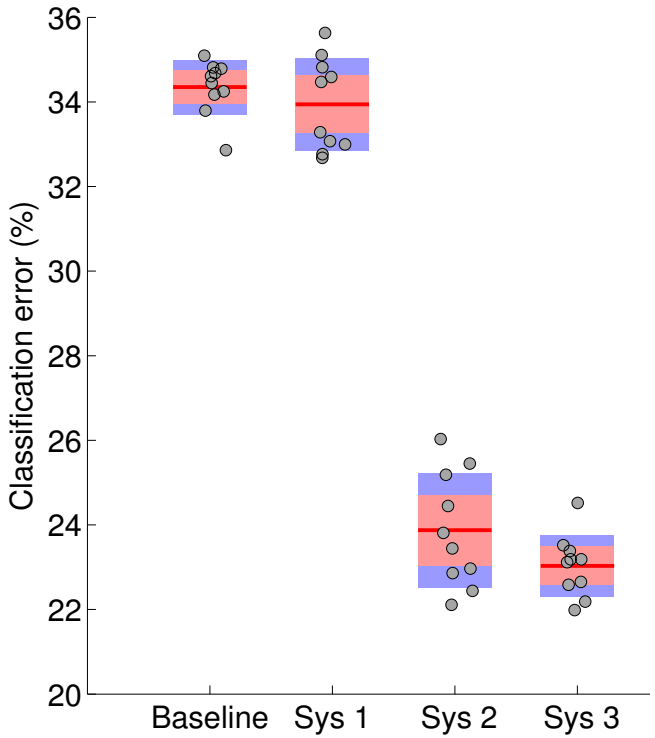
... excluding high vowel regions => :(

Glottal source processing - Results



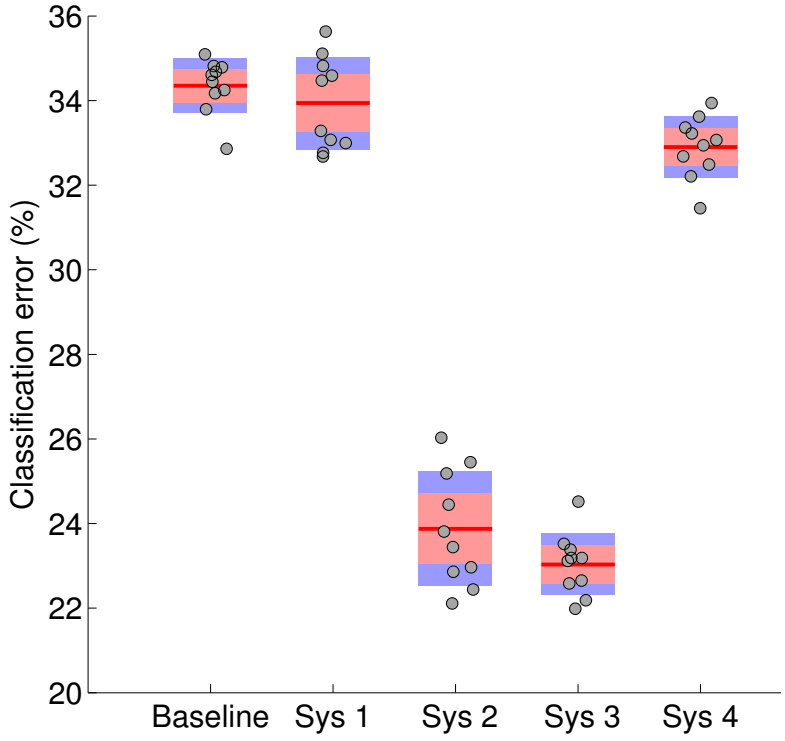
... additionally excluding fricative regions*** => :)

Glottal source processing - Results



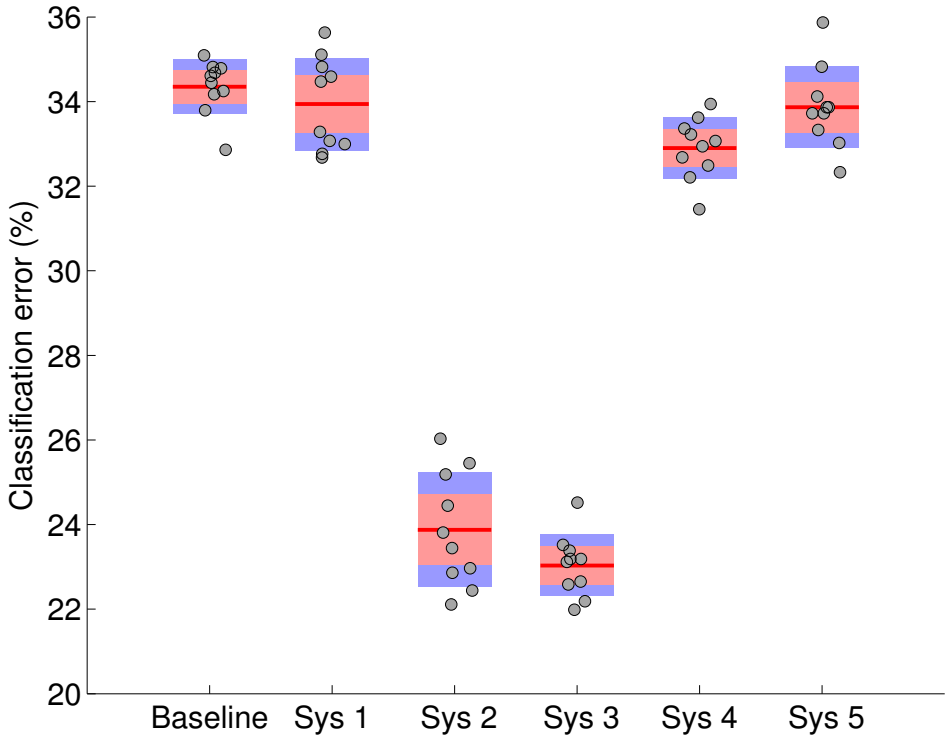
... additionally excluding nasal regions

Glottal source processing - Results



JUST excluding nasal regions* => :)

Glottal source processing - Results



Using phonetic features as input features in the classifier => :(

What did we find?

- Implementation of phonetic feature extraction based on ANNs
- Using information from this (i.e. removing feature data from fricative and nasal regions) significantly improved voice quality classification

Future ...

- Optimise phonetic feature extraction
- Increase set of phonetic features
- Investigate other context-sensitive glottal source processing methods (e.g., adaptive vocal tract model)
- Application in other areas of speech processing

Website: `http://covarep.github.io/covarep`
GitHub: `https://github.com/covarep/covarep`

Thank you!

Website: `http://covarep.github.io/covarep`
GitHub: `https://github.com/covarep/covarep`