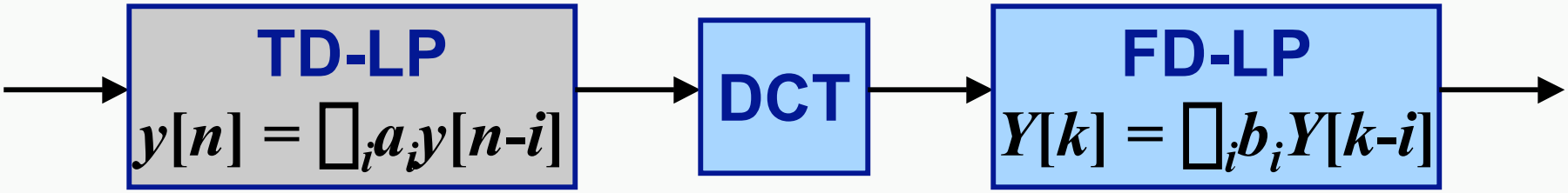
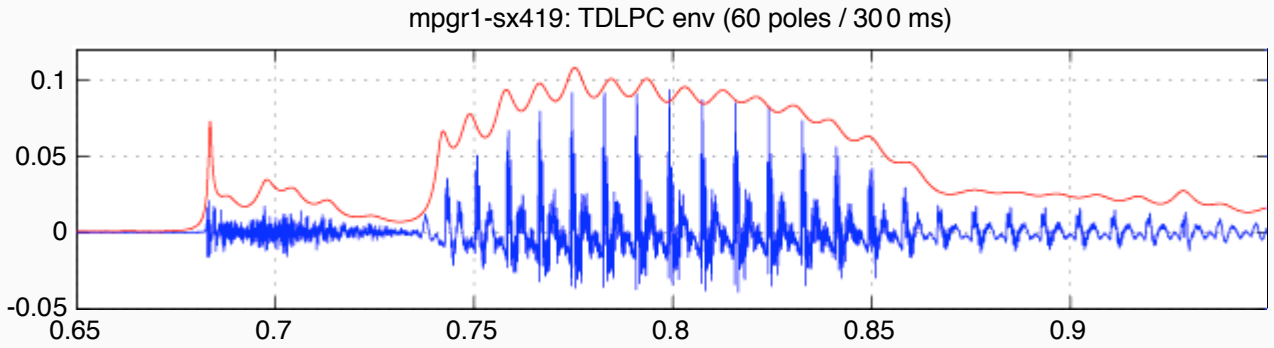
Frequency-Domain Linear Prediction (FDLP) Features

- Idea: Parametric model of sub-10ms temporal envelope structure



- 'Poles' represent temporal peaks:



FDLP features: Issues (1)

- Time window for analysis: ~ 300ms

  Poles per window: ~ 20

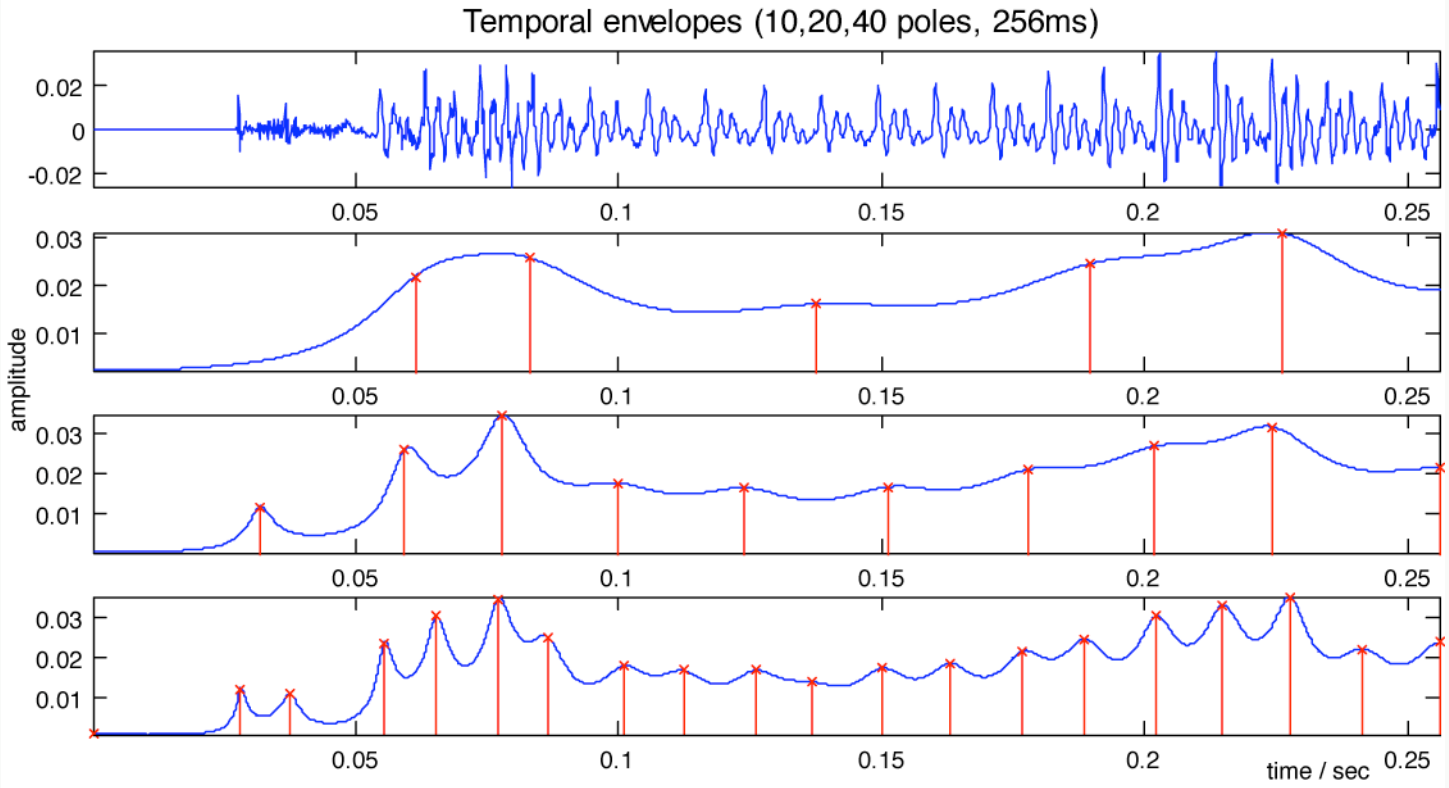
  - model where to put poles within window

Waveform

10 poles

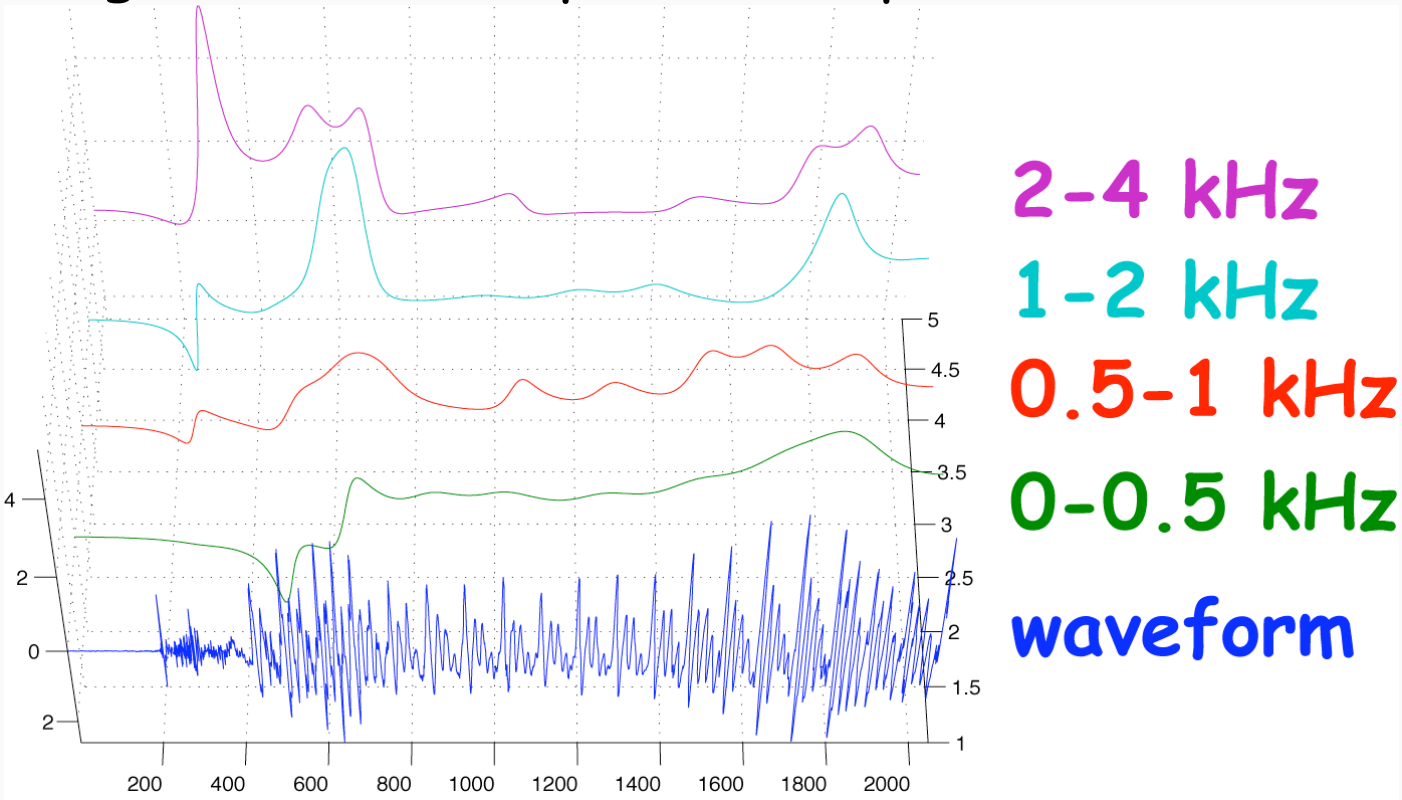
20 poles

40 poles



FDLP features: Issues (2)

- Frequency structure
  - single band envelope or multiple subbands?

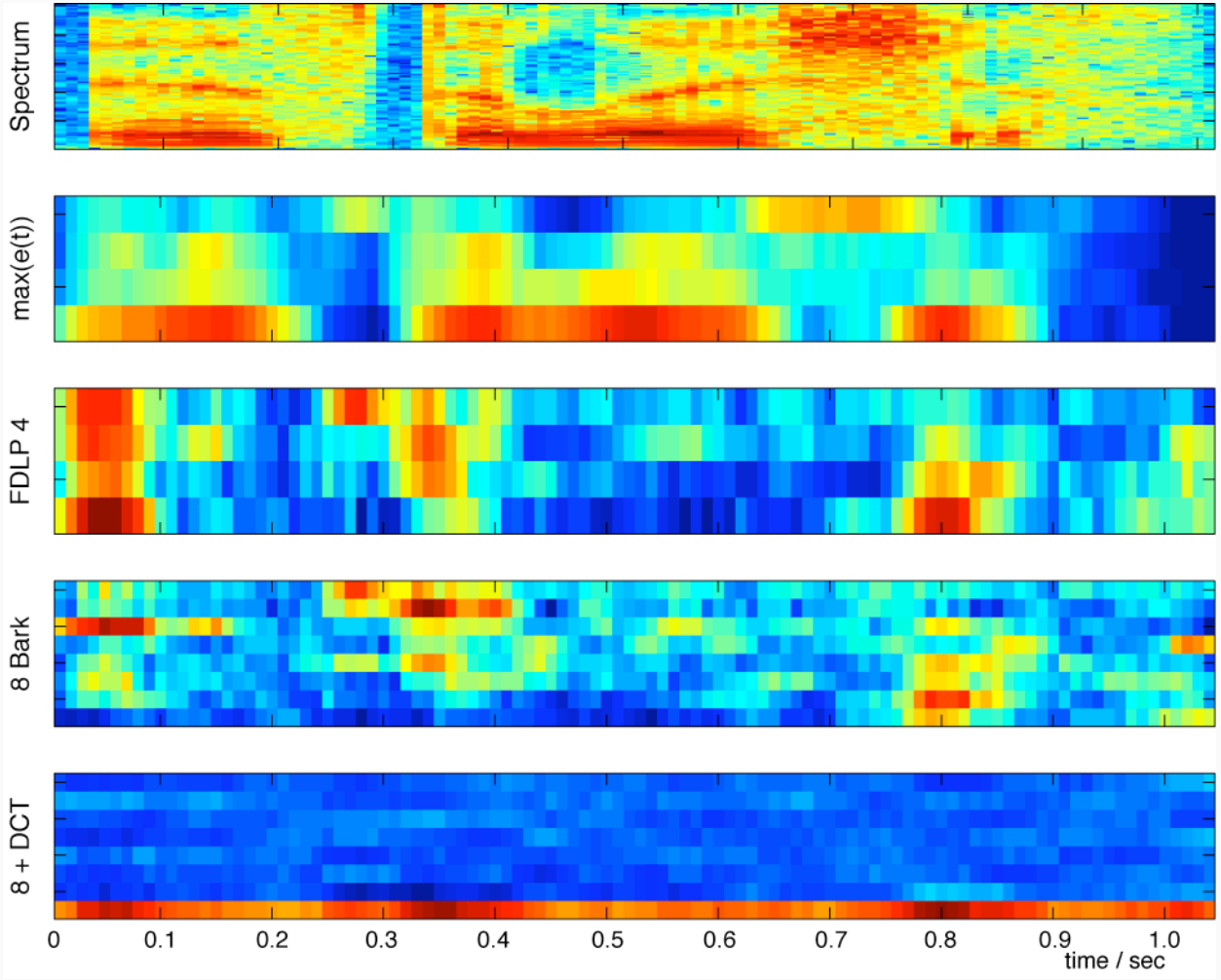


- Convert all-pole model to features
  - $\max(-\log(1 - |\square_i|)\cdot w_i)$ follows peak sharpness

for time resolution

FDLP Examples

- Source speech
- 4 band, $\max(e(t))$
- 4 band, $-\log(1-|\gamma_i|)$
- 8 band (Bark)
- 8 band + DCT



FDLP Results

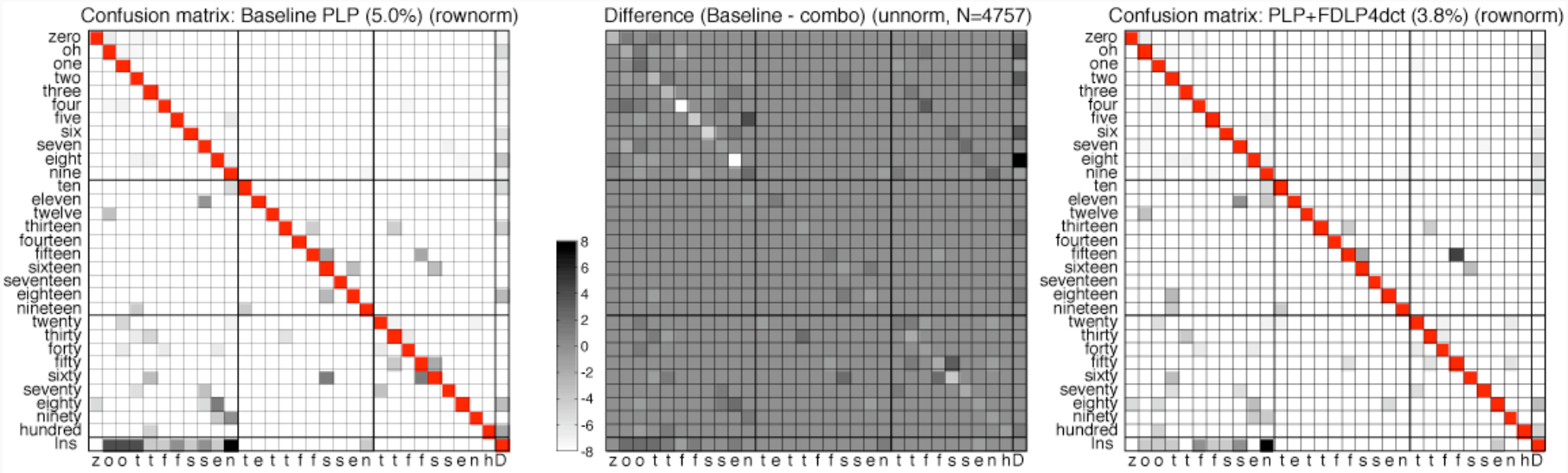
- Train on CTS, test on Numbers task
  - reduced 20k training set, speaker norm'd
- WER on N95-test:

plp12 (baseline)	5.0%
plp12 + FDLP4 log	4.1%
plp12 + FDLP4 log + dct	3.8%
plp12 + FDLP8 bark	4.1%
plp12 + FDLP8 + dct	4.4%

FDLP Results

- Confusion comparison:

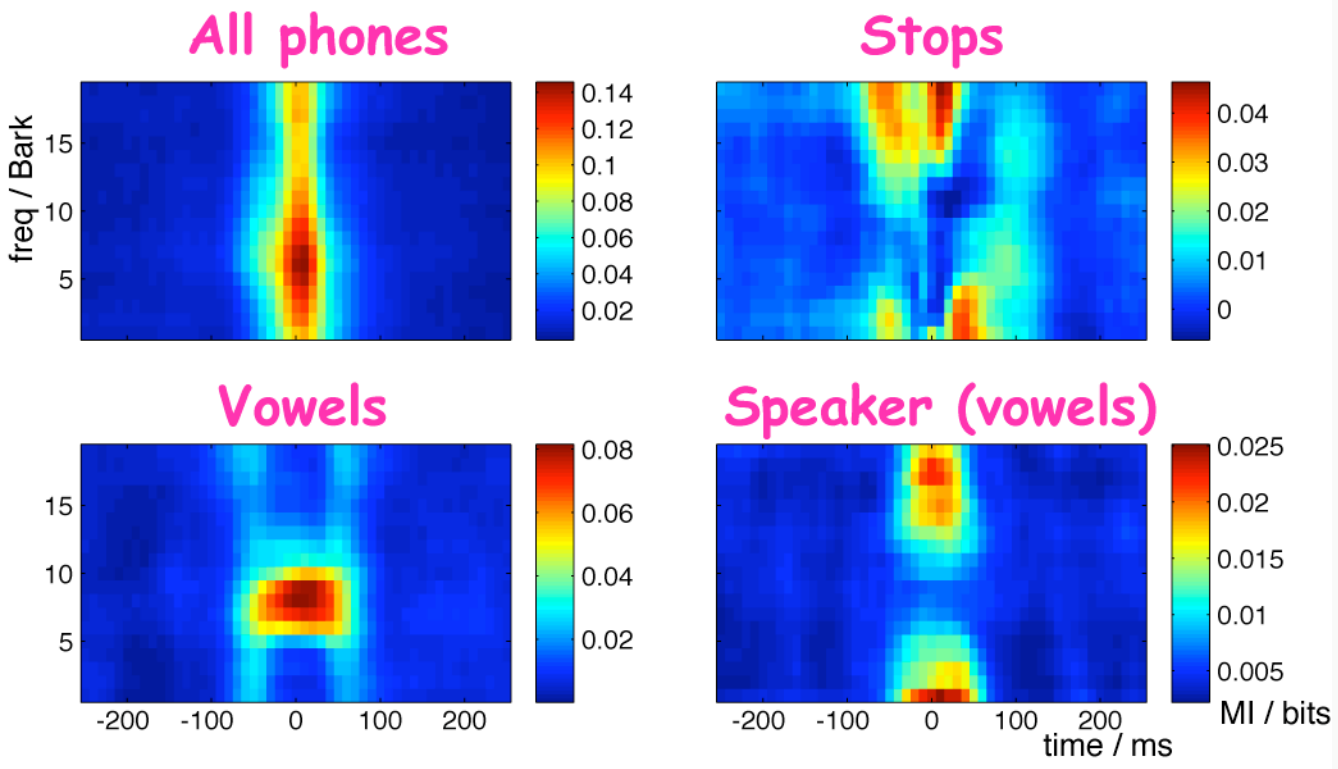
plp12 base | Difference | plp12+FDLP8dct



- Gains on Del "eight", "forty" "four"

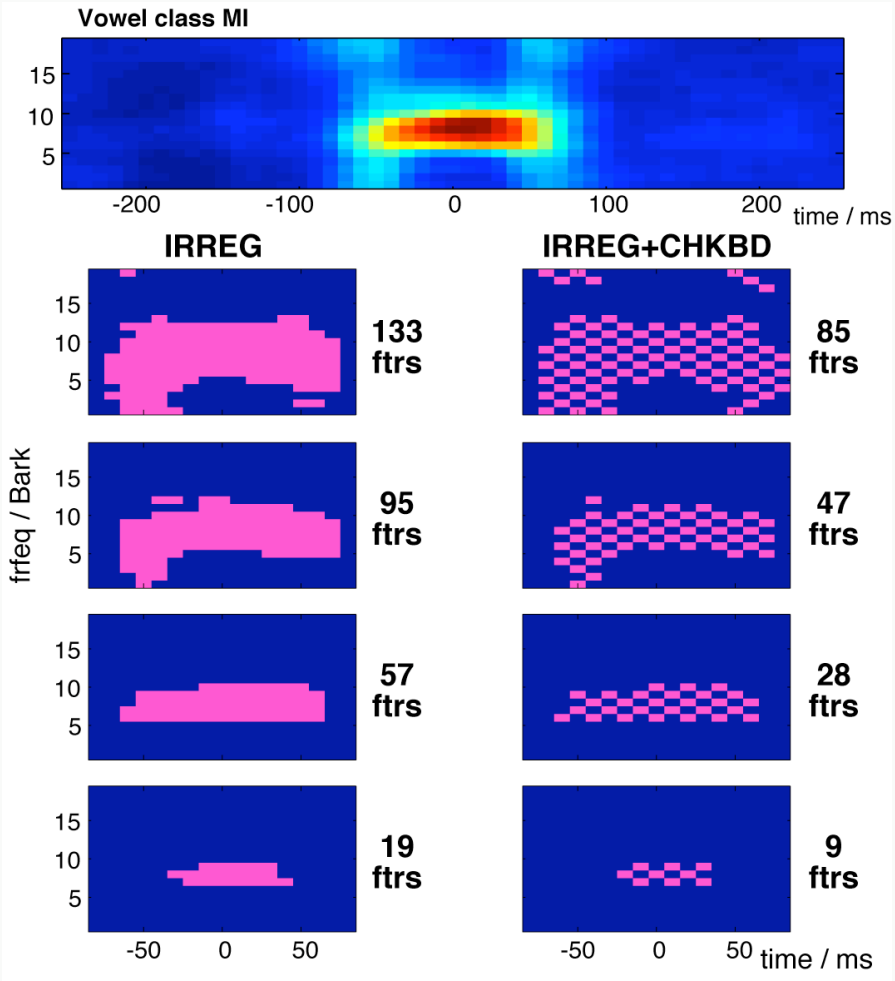
T-F Mutual Info (TFMI) for class feature selection

- MI(T-F cells; phone label) □ info. spread:
  - timit, bark log spec, ±250 ms, phone ctrs

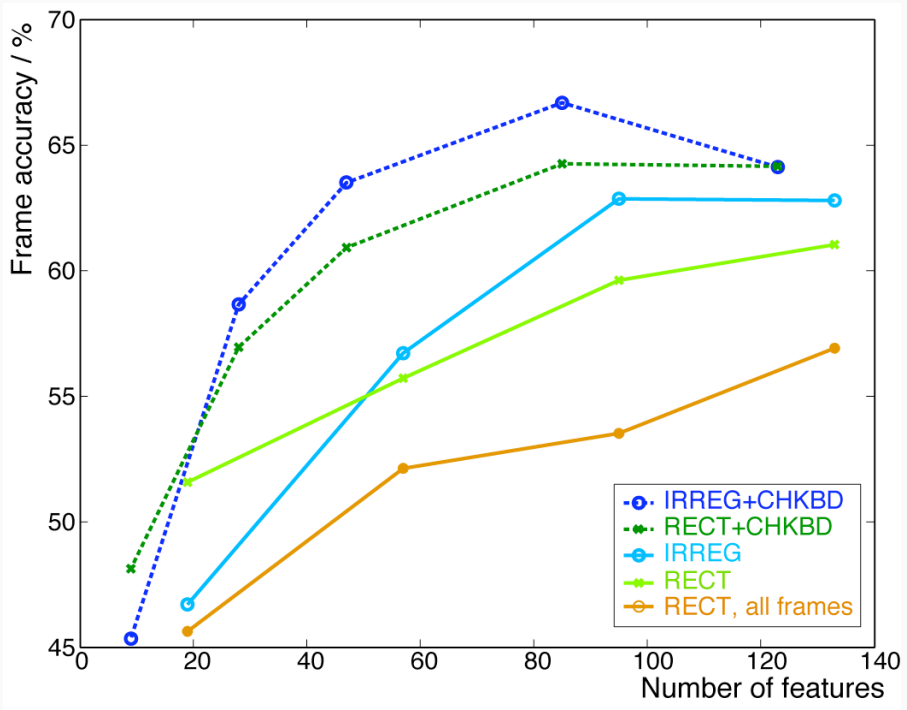


TFMI for Vowel Classifiers

- Choose TF cells with top MI
  - train NN on phone centers, c/w rectangular

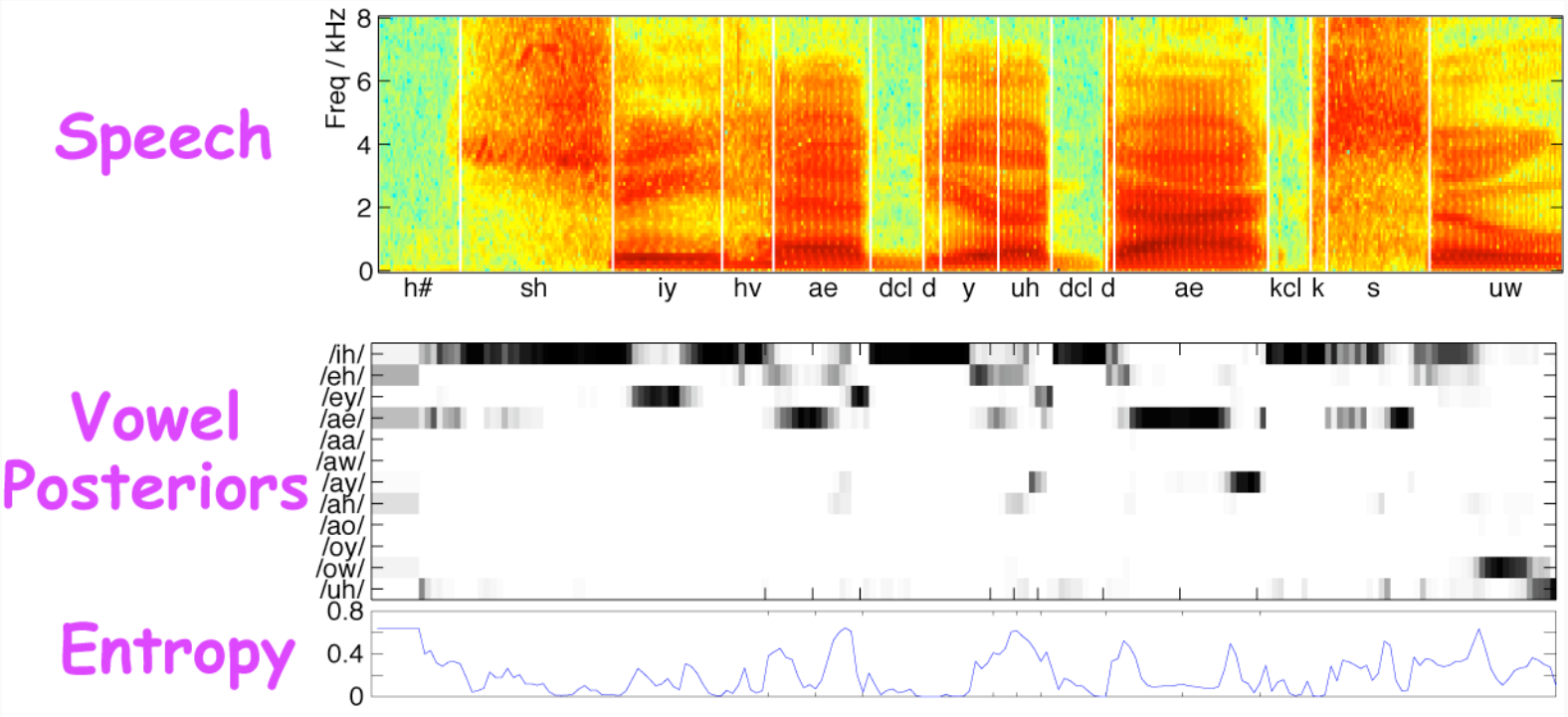


Classification
(given vowel center)



TFMI Classifier Recognition

- Vowel-center classifier on all frames:



- entropy to detect 'good' frames?
- separate p(vowel ctr) detector - GMM?

ICA for Feature Bases

- ICA on spectra finds independent bases
- Search for best & sufficient by MI to phone classes: