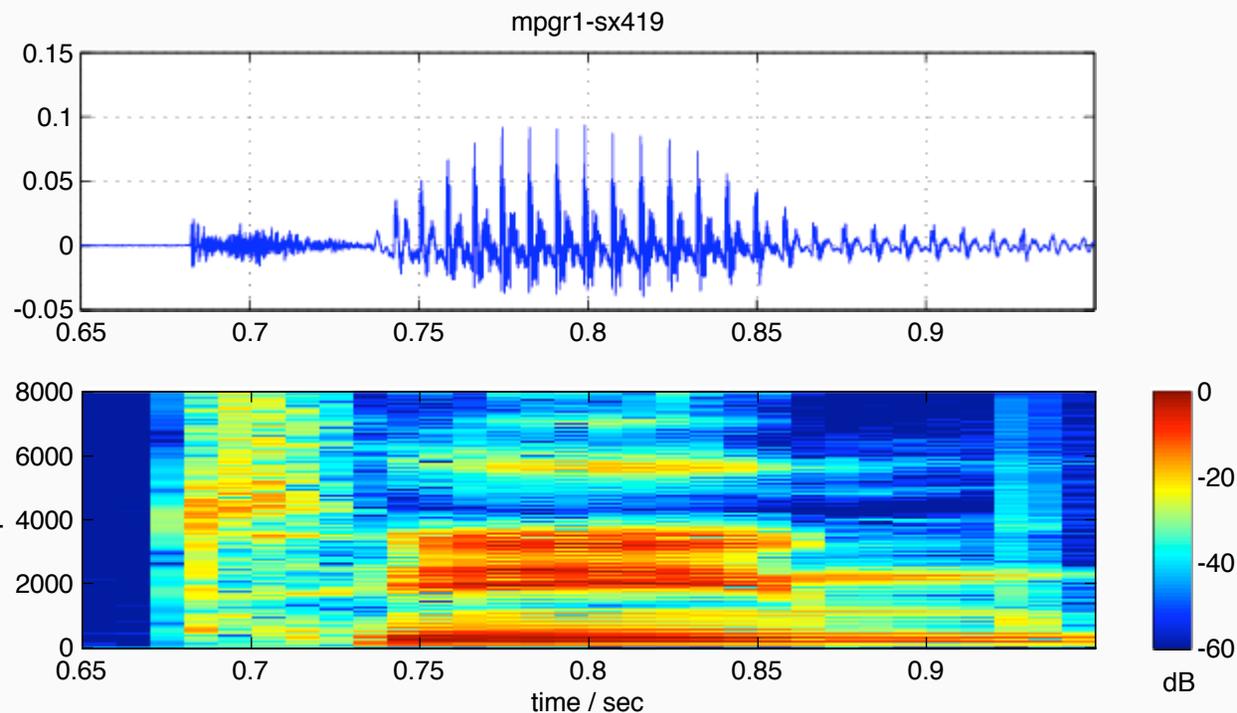


# *Dan's slides for EARS PI mtg*

- 4 slides on novel features based on linear predictor coefficients for the frequency (not time) domain
  - basic signal model accepted at ICASSP03
- A couple of slides on a very new idea to look for data-derived (ICA?) articulatory-style features

# *Temporal envelope features (Columbia)*

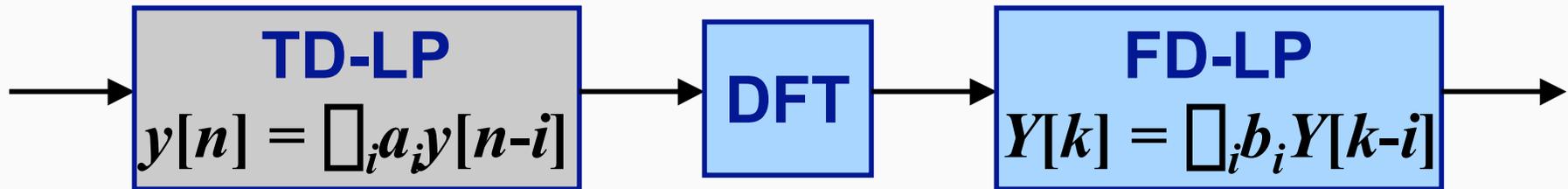
- Temporal fine structure is lost (deliberately) in STFT features:



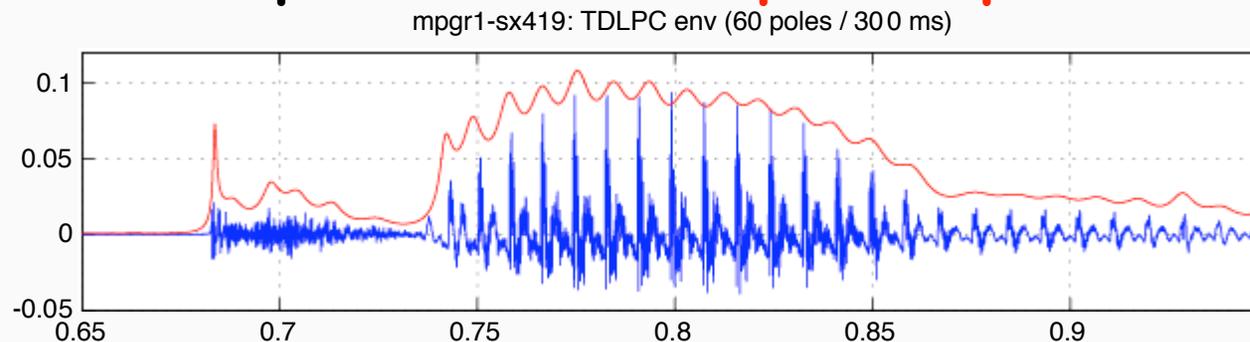
- Need a compact, parametric description...

# Frequency-Domain Linear Prediction (FDLP)

- Extend LPC with LP model of **spectrum**



- 'Poles' represent **temporal peaks**:



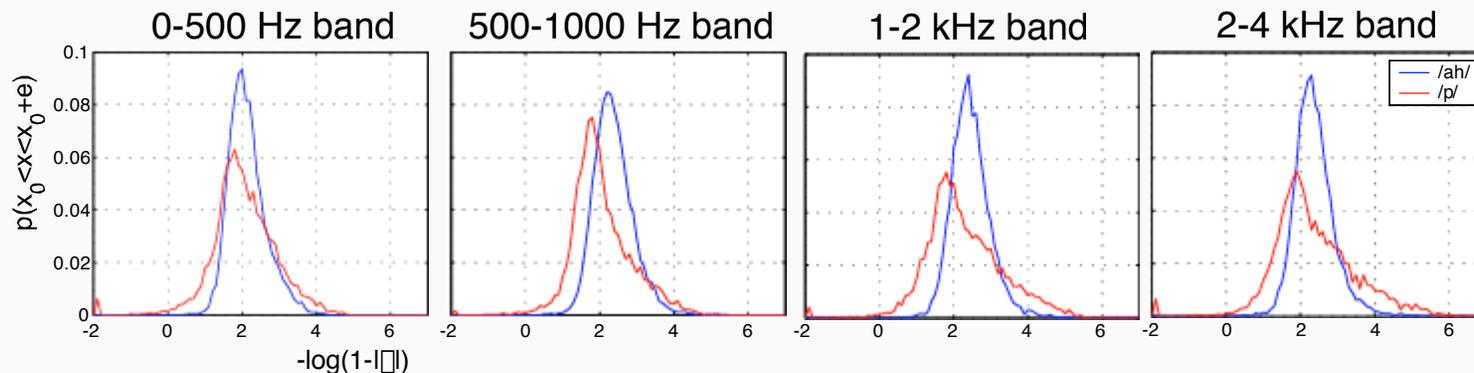
- Features  $\sim$  pole bandwidth, 'frequency'

# *FDLP features for speech*

- LP algorithm distributes fixed pole set within  $\sim 200$  ms time window
  - automatic selection of 'significant' times
- Pole **bandwidth**  $\square$  transient **sharpness**
  - $1 - \max(|\alpha_i|)$  in several bands as feature
  - help with classification of stop bursts etc.
- Pole **frequency**  $\square$  **timing** within window
  - $f_n - f_{n-1}$  as robust periodicity feature?

# ***FDLP preliminary results***

- Distribution of pole magnitudes for different phone classes (in 4 bands):



- NN Classifier Frame Accuracies:

plp12N	57.0%
plp12N+FDLP4	58.4%

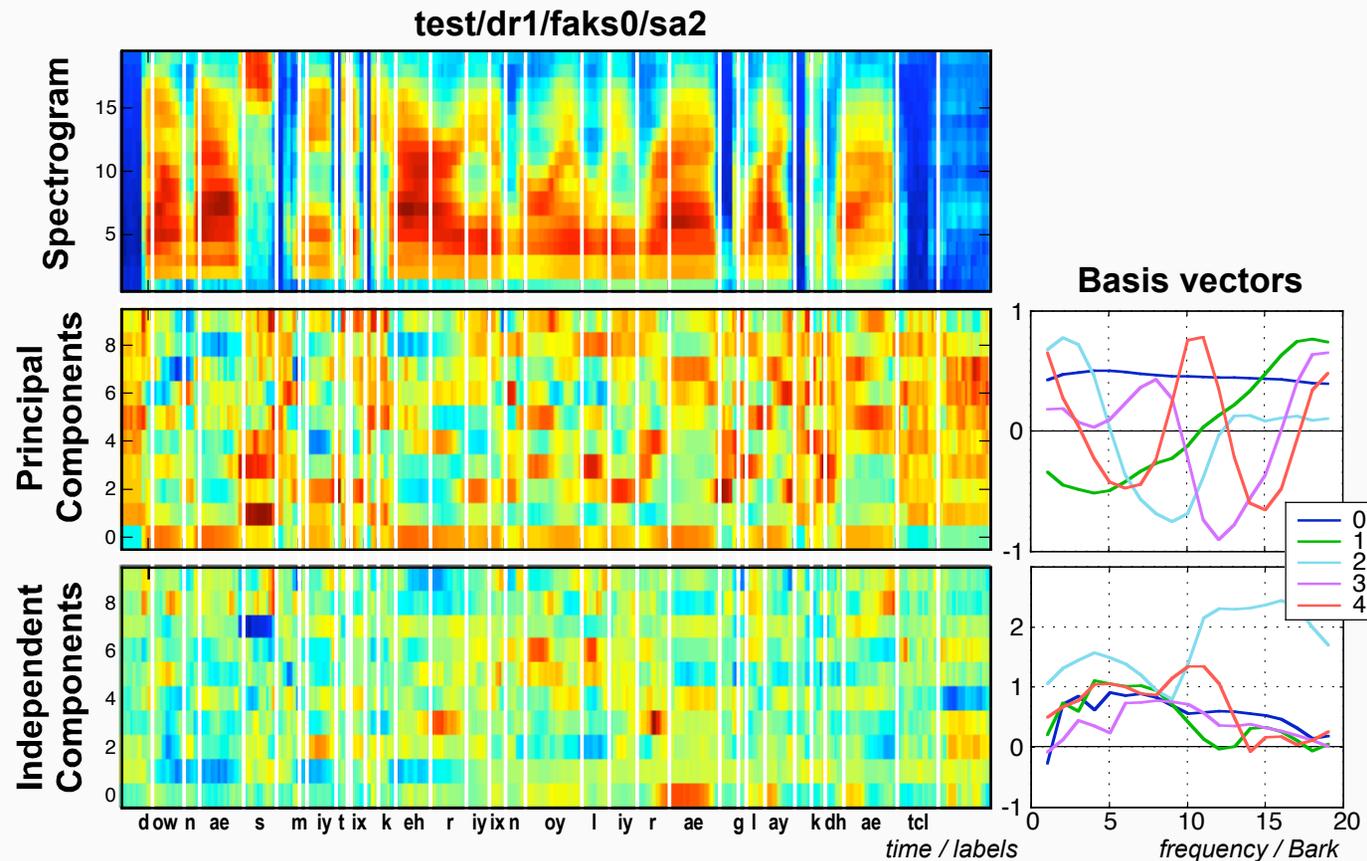
# *Data-derived phonetic features (Columbia)*

- Find a set of **independent** attributes to account for phonetic (lexical) distinctions
  - **phones** replaced by **feature streams**
- Will require new **pronunciation models**
  - asynchronous feature transitions (no phones)
  - mapping from phonetics (for unseen words)

Joint work with Eric Fosler-Lussier

# ICA for feature bases

- PCA finds **decorrelated** bases;  
ICA finds **independent** bases

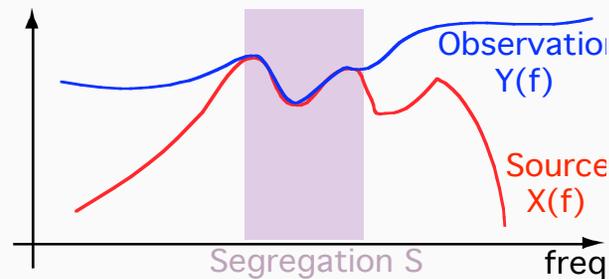


- Lexically-sufficient ICA basis set?

***Extra Slides***

# Speech Fragment Recognition (Columbia)

- Model match for missing features:



$$P(M,S|Y) = P(M) \int P(X|M) \cdot \frac{P(X|Y,S)}{P(X)} dX \cdot P(S|Y)$$

*joint prob. of model & seg.*      *likelihood 'boost'*      *segregation likelihood*

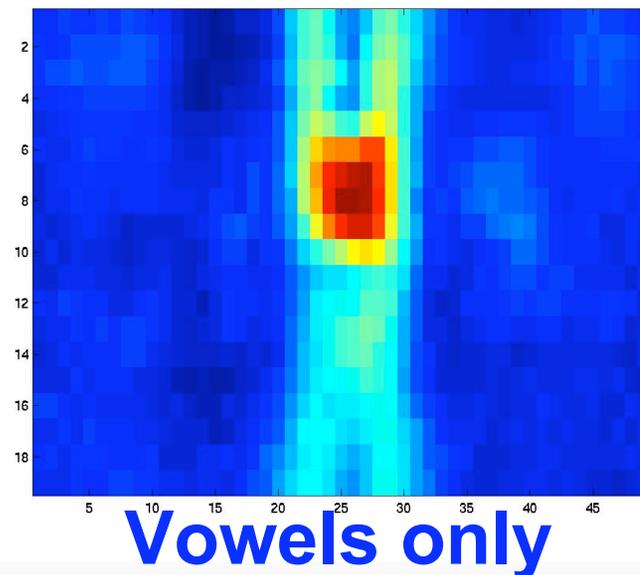
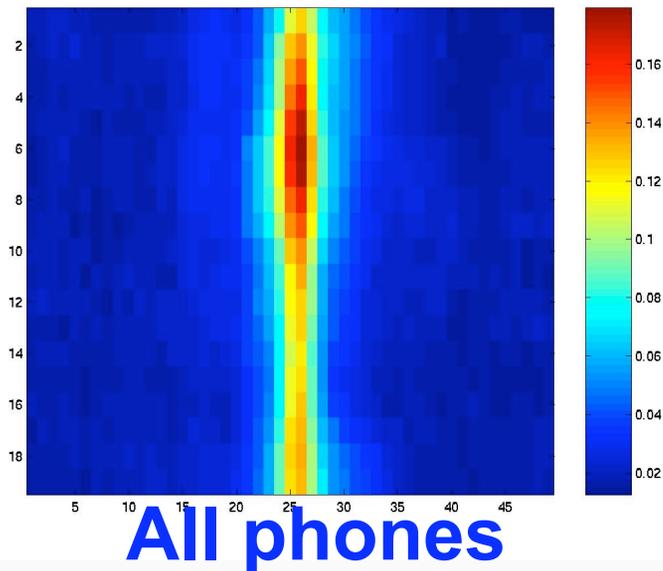
- .. for partial observations in noise
- .. or integrating partially-seen streams

# *Missing speech information*

- Noise is **not** our primary concern; **casual pronunciation** is a big issue
  - not missing **Spectral** information, but missing **Phonetic** information
- Can we model this as:
  - 'missing' (i.e. non-articulated)
  - 'features' (i.e. phonetic-style features) ... ?
- Need to *locate* information...  $P(S|Y)$

# *Class-dependent information*

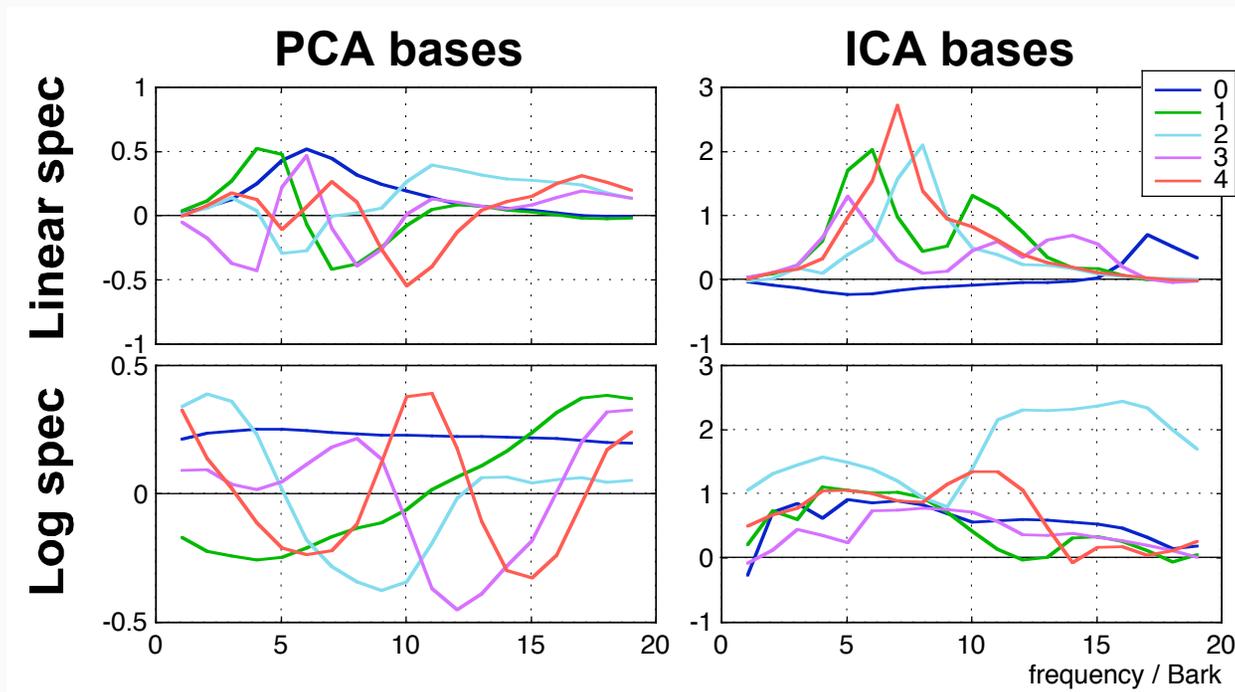
- Locate information per subword unit
- Mutual Information on time-frequency plane over different **phone classes**



- $\pm 250\text{ms}$  / 19 bark, TIMIT phone ctrs

# ICA for feature bases

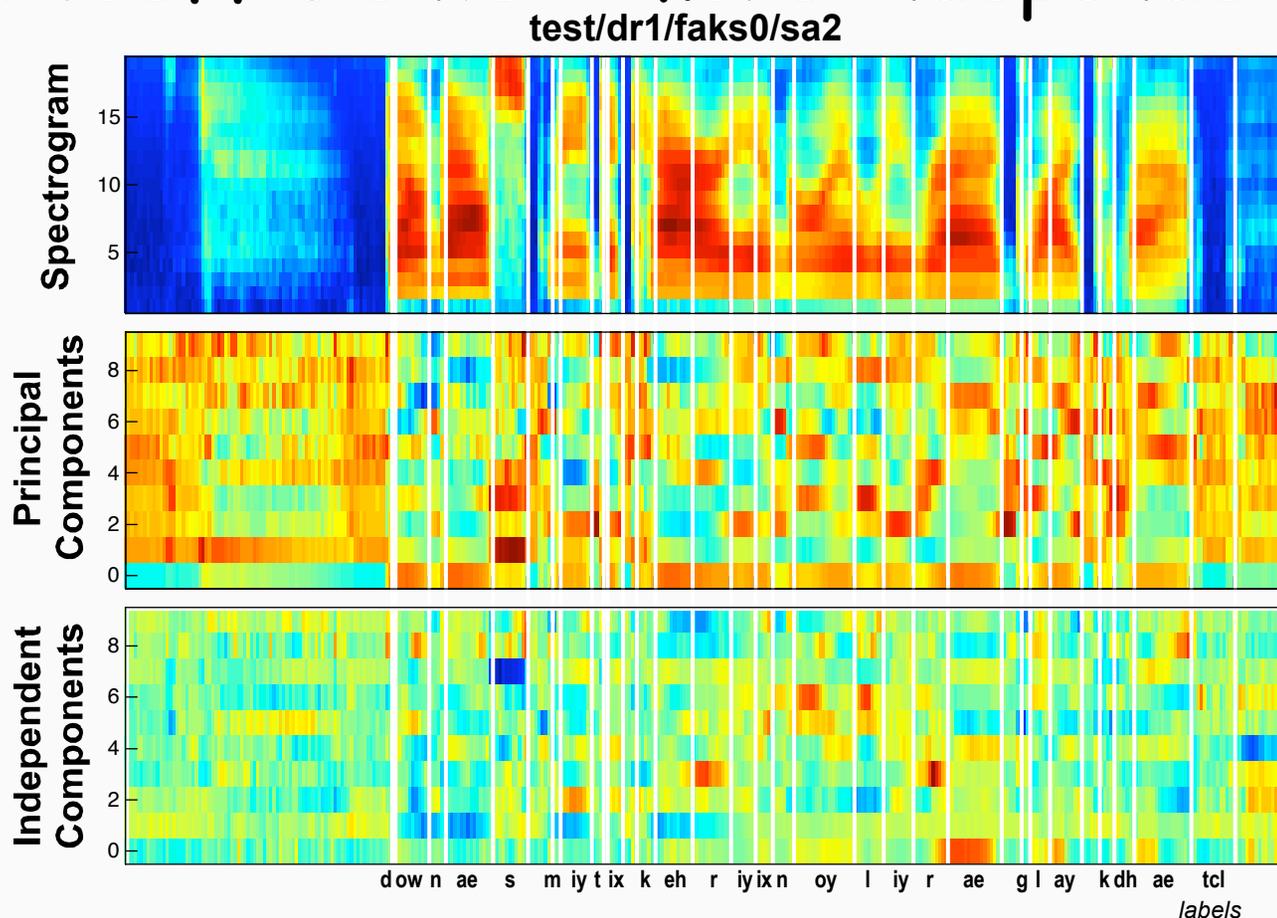
- PCA finds **decorrelated** bases;  
ICA finds **independent** bases



- Find lexically-sufficient ICA basis set?

# ICA for feature bases

- ICA coefficients ~ more independent:



- Looking for *orthogonal* subword features