
On Data-Derived Temporal Processing in Speech Feature Extraction

Michael L Shire, Barry Y Chen

International Computer Science Institute

University of California at Berkeley

Berkeley, California USA

{shire,byc}@icsi.berkeley.edu



ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Introduction

- Temporal processing is commonly used in speech feature extraction to aid in performance and robustness.
- LDA applied to temporal trajectories gives discriminant data-driven filters.[Hermansky, van Vuuren 1997]
- Data-driven temporal filters change in the presence of reverberation.
- Test LDA filters using RASTA-PLP, Modulation-Filtered Spectrogram (MFSG), and PLP-cepstra.



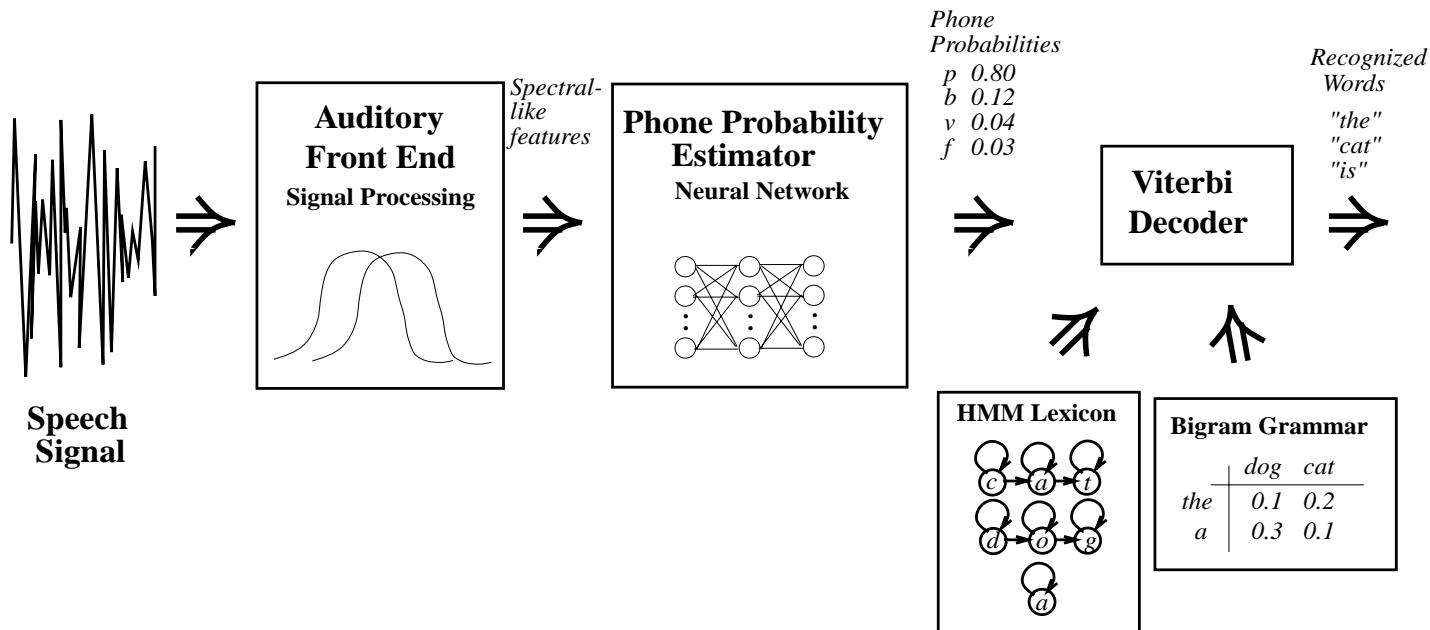
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Hybrid ANN-HMM System



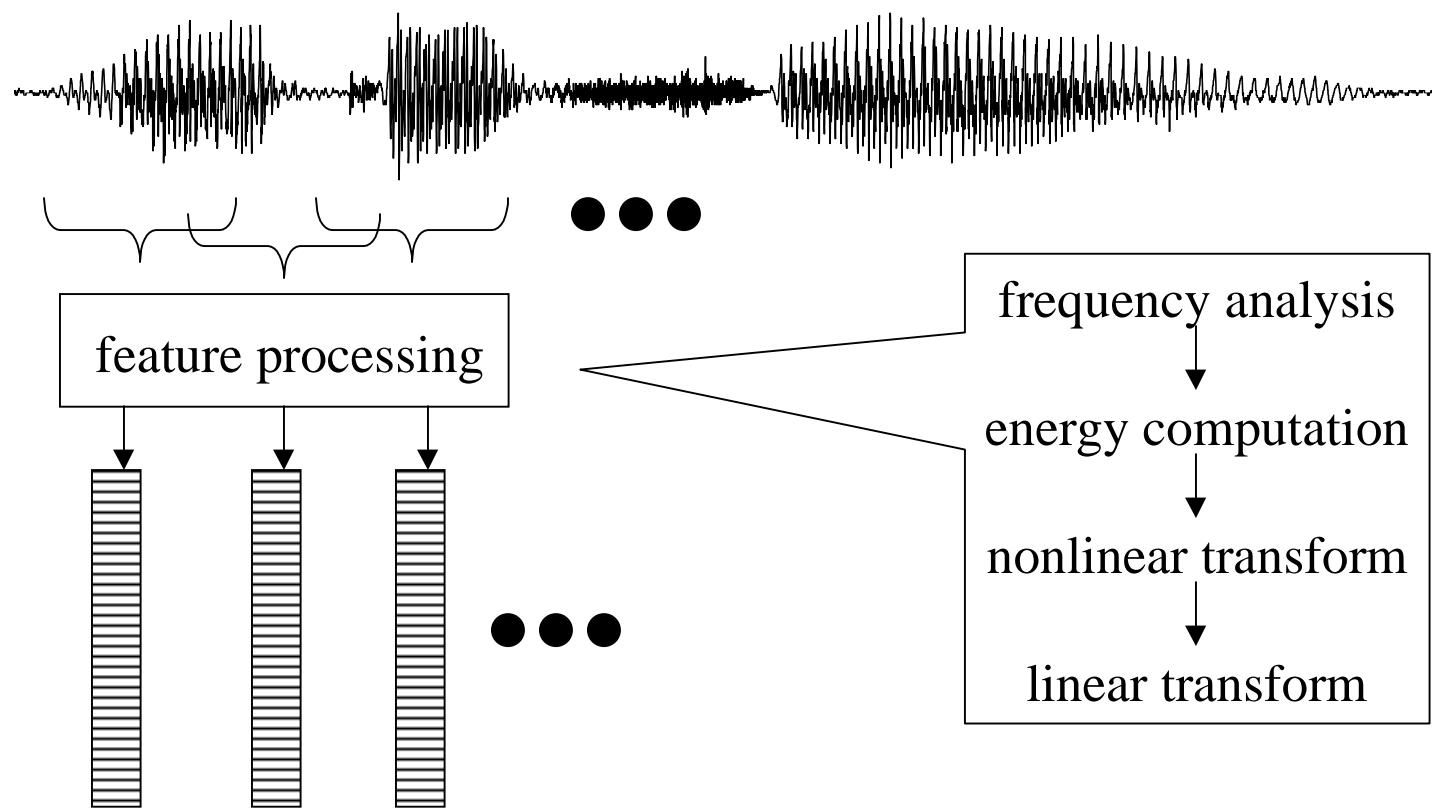
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Feature Extraction



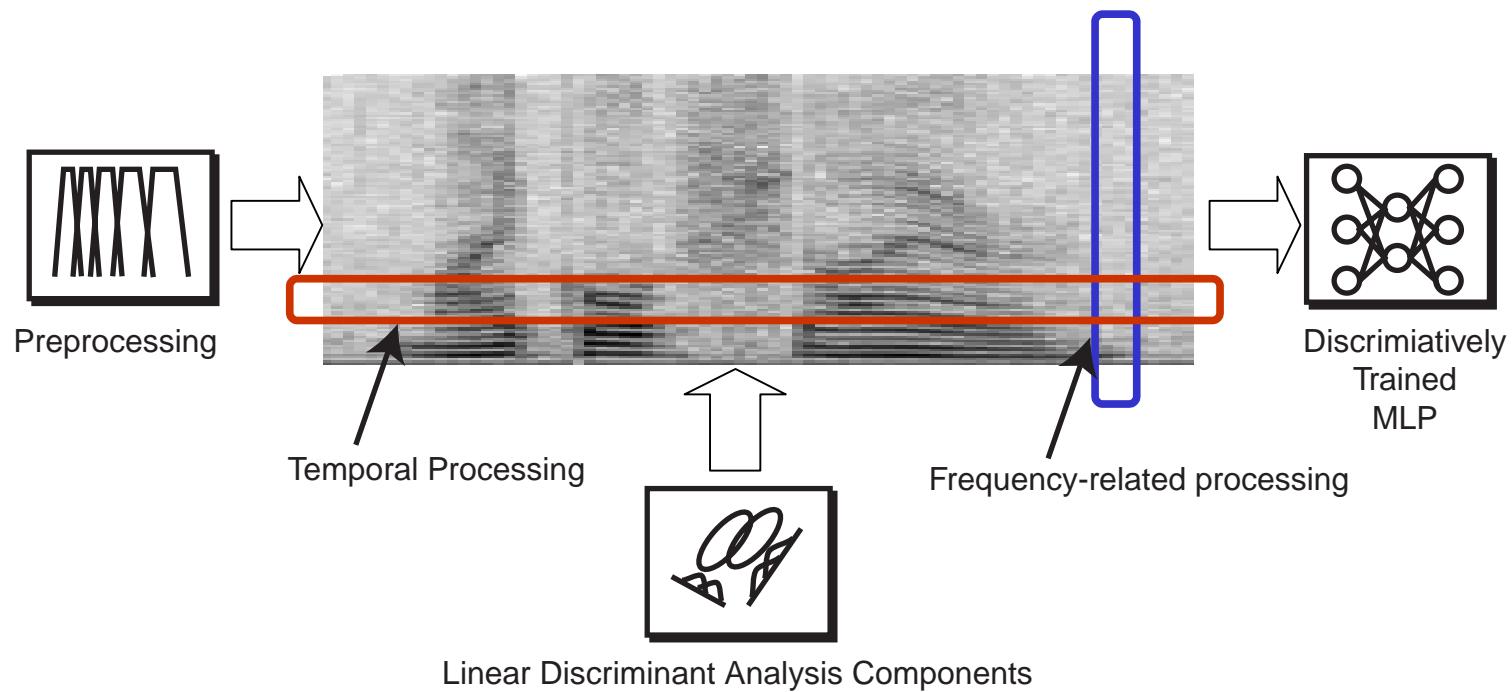
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Discriminant Training



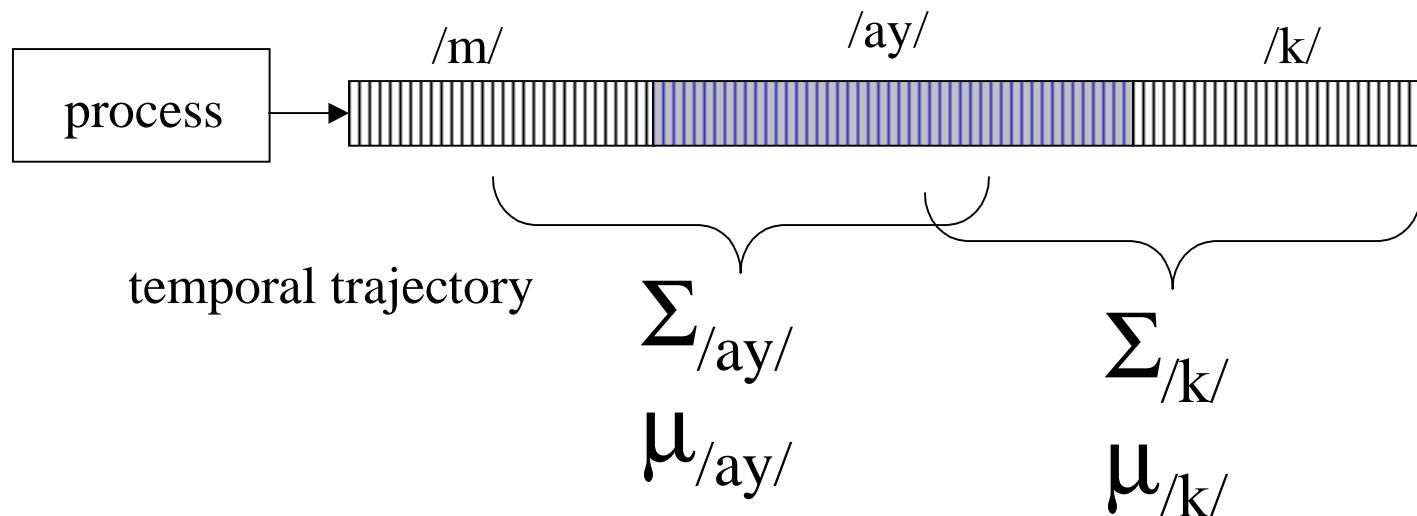
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



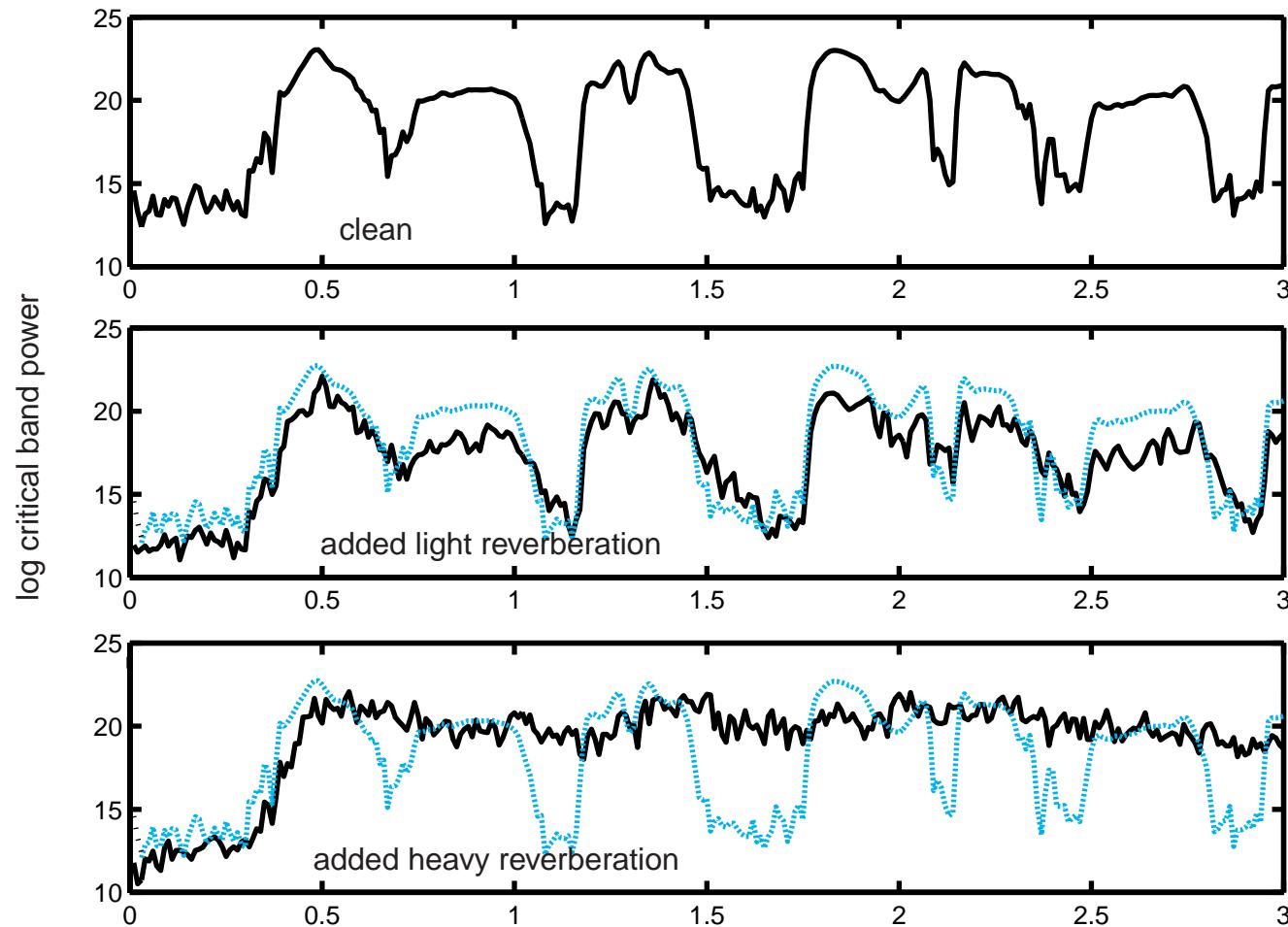
Temporal LDA Procedure



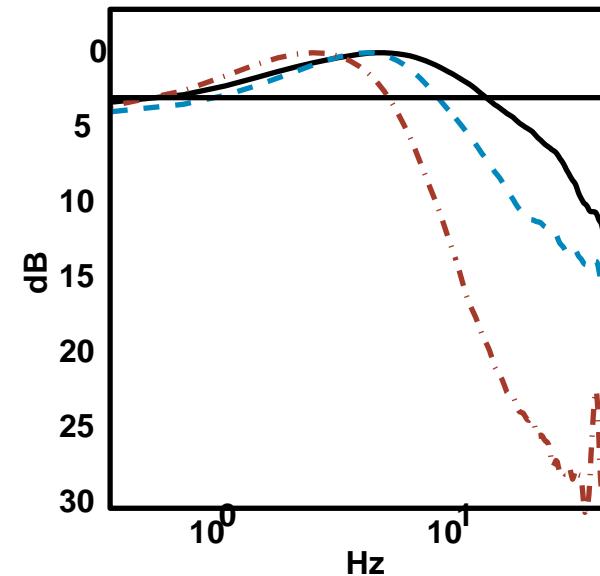
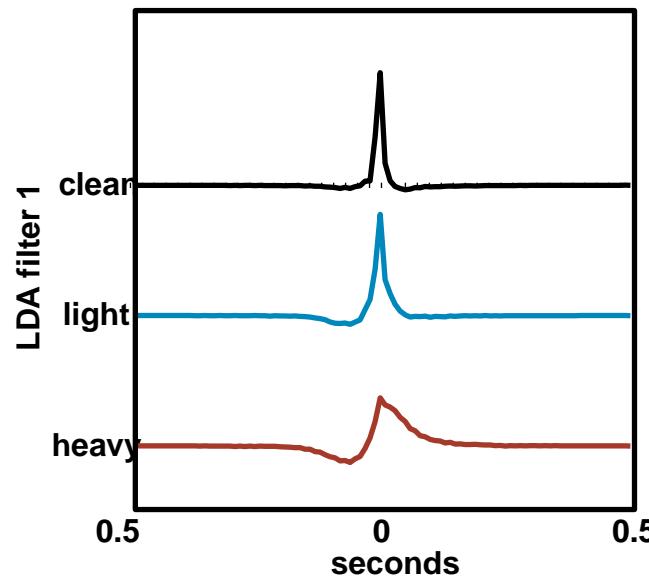
- Collect means and covariance matrices.
- Take LDA vectors as eigenvectors of ratio of covariance of class means μ_i to mean of class covariances Σ_i



Effect of reverberation



LDA-RASTA Filters



- Increasing severity of reverberation results in broader, smoother impulse responses and narrower modulation frequency pass-band.



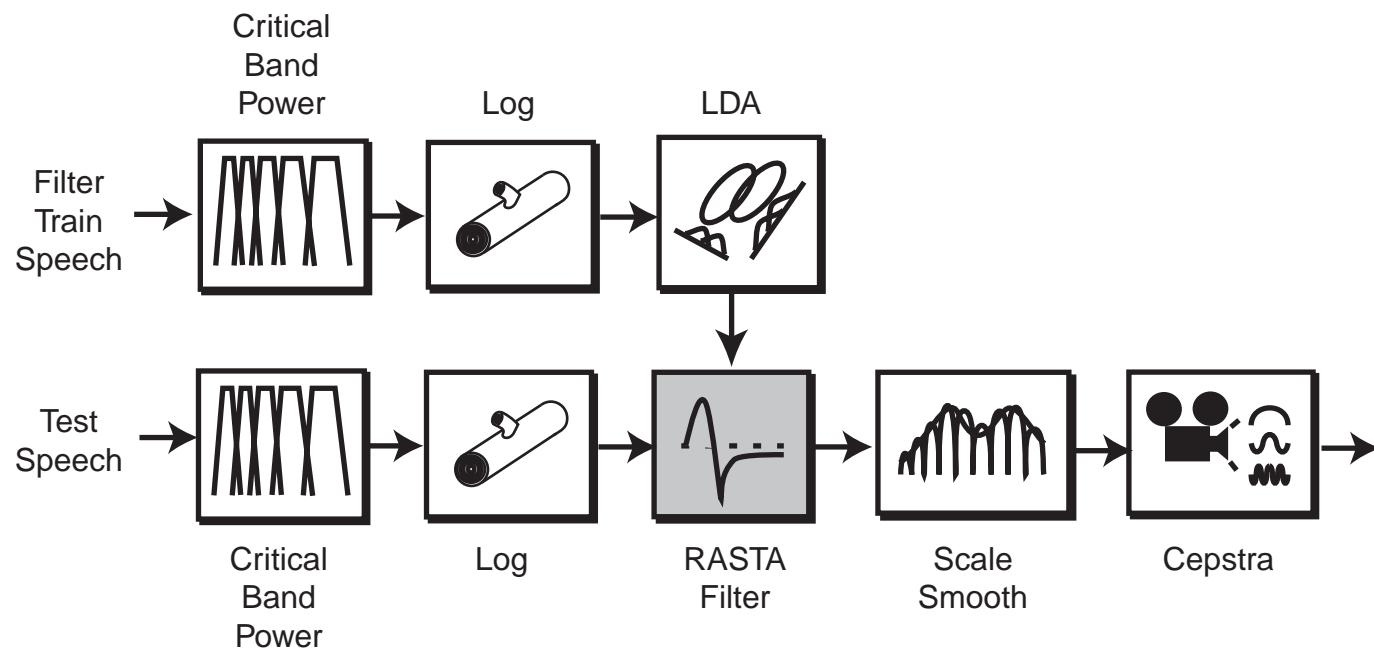
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



LDA-RASTA-PLP



- RASTA filter replaced with filters derived using LDA.
- Stories corpus used to train discriminant filters.
- Recognition tests used Numbers corpus.



ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



LDA-RASTA-PLP WER

WER(%) for different LDA filters on Numbers corpus
Single acoustic context to MLP

MLP		LDA filter		
train	test	clean	light	heavy
clean	clean	9.10	9.00	12.0-
light	light	18.90	16.40+	18.30
heavy	heavy	45.50	42.10+	38.90+

- 3 conditions: clean data, light reverberation, heavy reverberation.
- LDA filters derived in these conditions on independent corpus.
- MLP with only a single frame of acoustic features.
- Matched LDA-MLP training usually gave best results.



LDA-RASTA-PLP WER 2

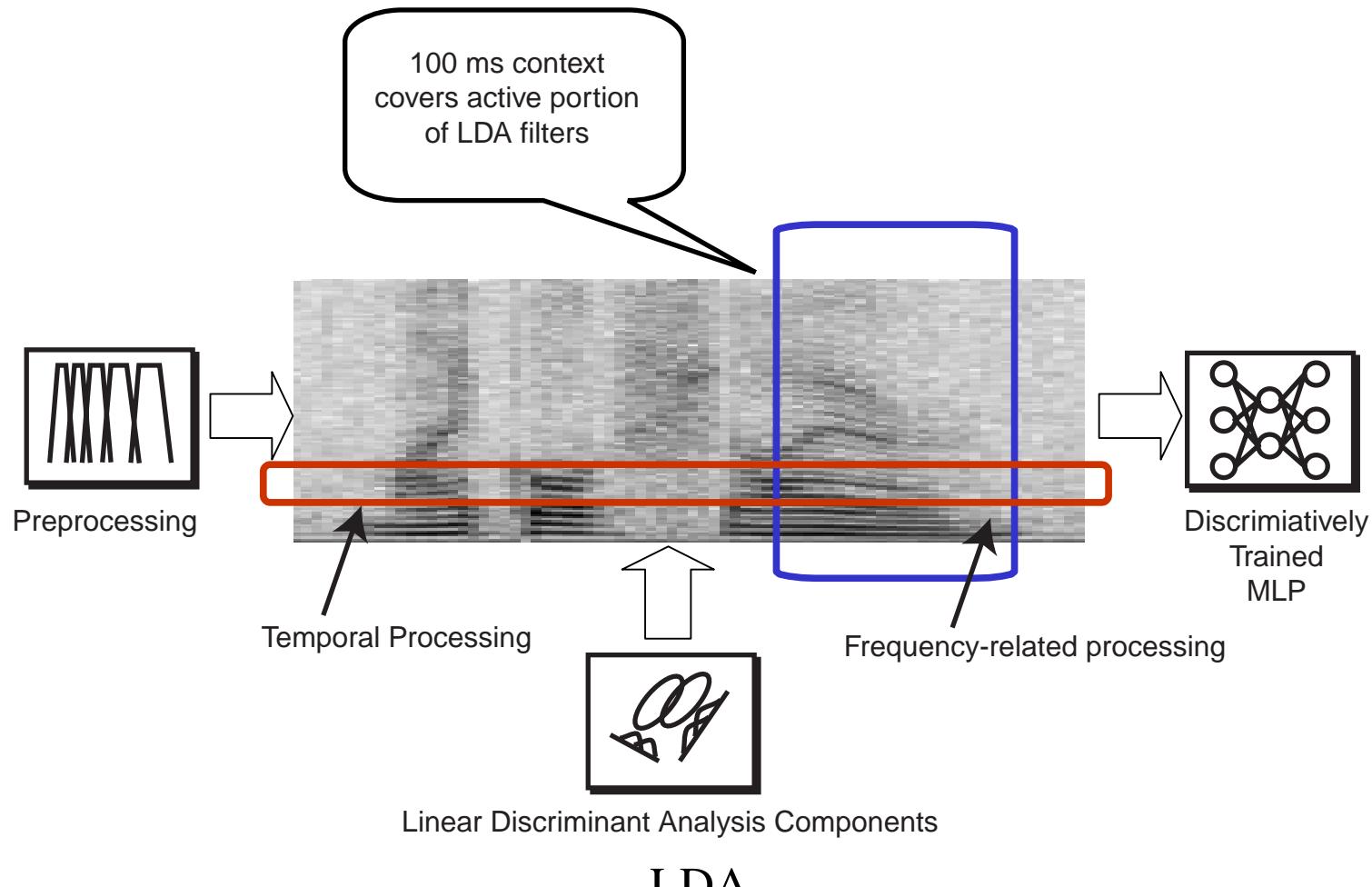
WER(%) for different LDA filters on Numbers corpus
MLP acoustic context of 9 frames (~100ms)

MLP		LDA filter		
train	test	clean	light	heavy
clean	clean	5.20	5.30	7.00-
light	light	11.10	10.70	12.40-
heavy	heavy	30.70	30.10	30.30

- Give MLP a wider acoustic context of 9 frames instead of 1.
- Difference between using different sets of filters reduces.
- Heavy filters are most restrictive and give worst results on mis-matched cases.
- MLP may be including its own temporal processing.



Wider acoustic context to MLP



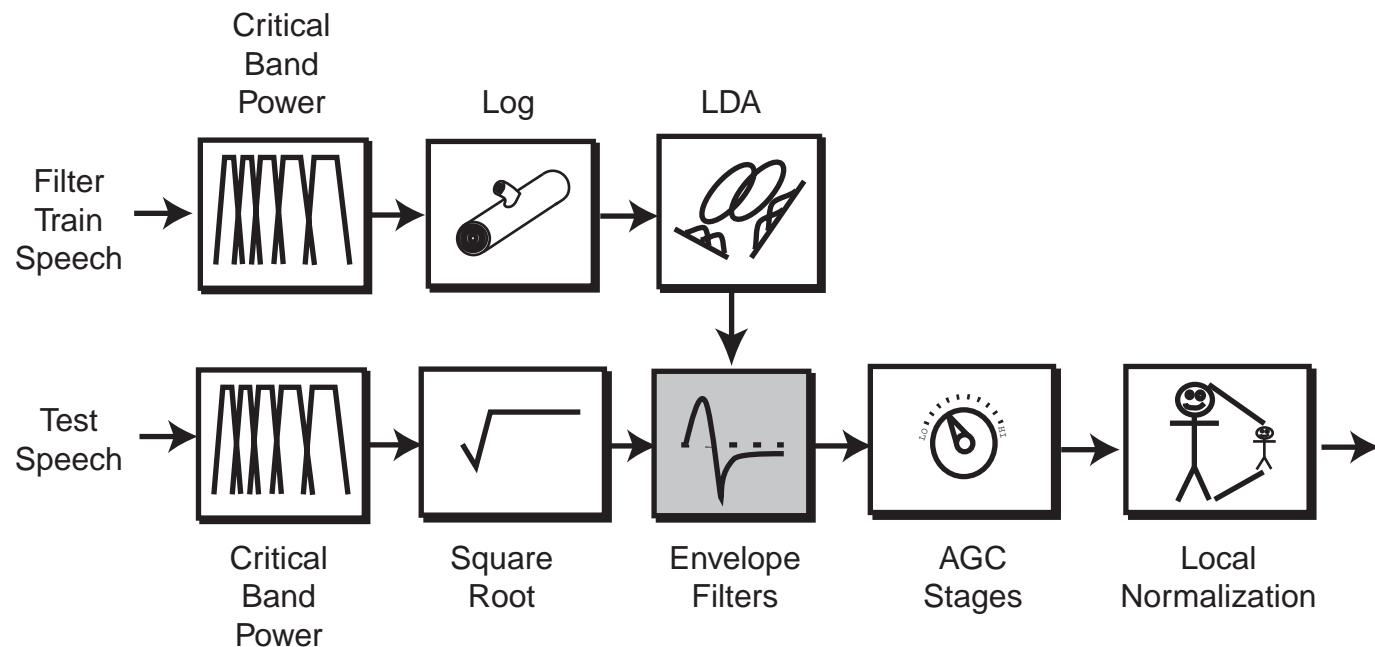
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



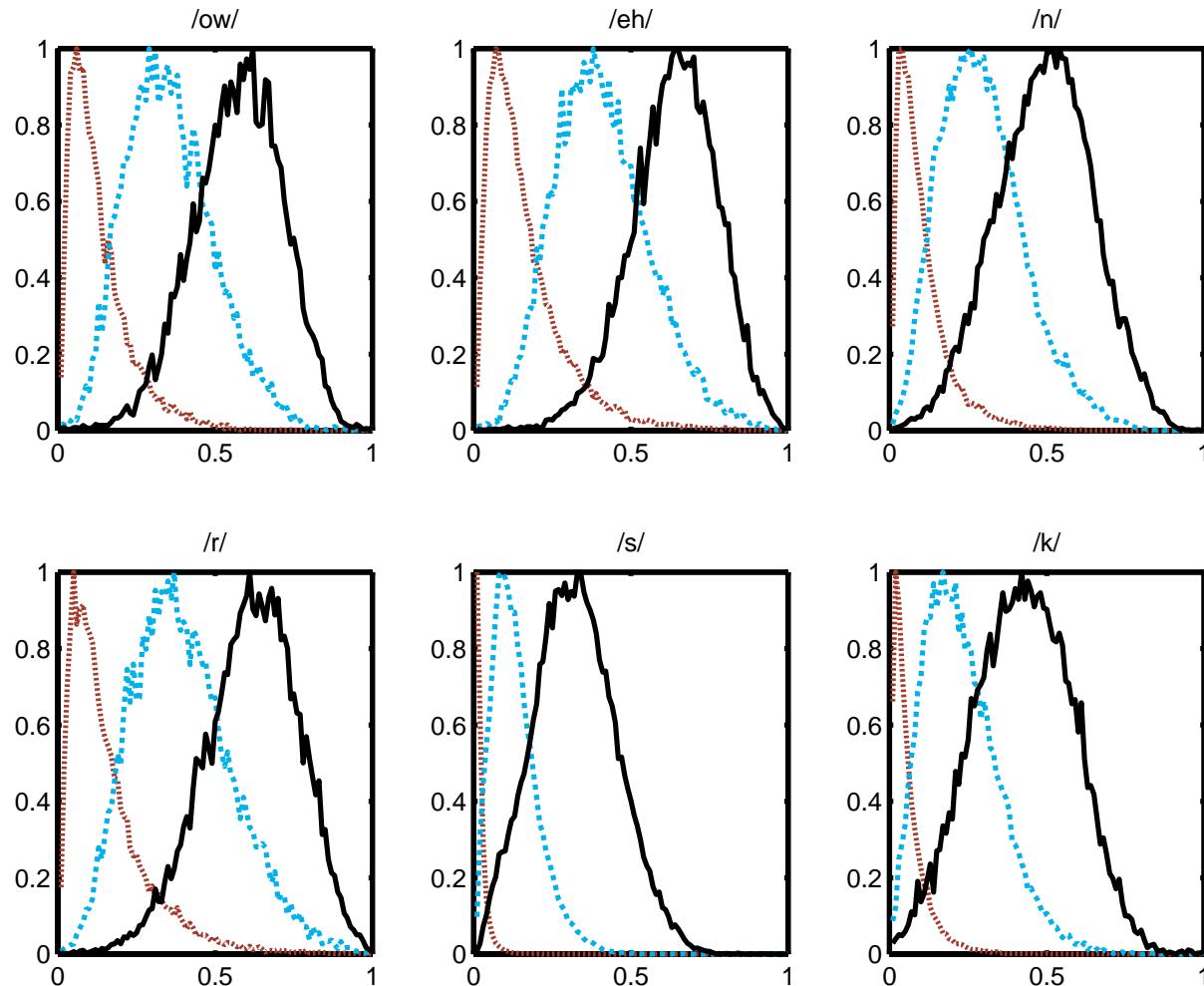
MFSG-LDA



- Original filters were 8HZ lowpass and 8-16Hz bandpass
- Replace with LDA-RASTA filters
- Notice oddity of designing from log instead of square root.



Critical-band class distributions



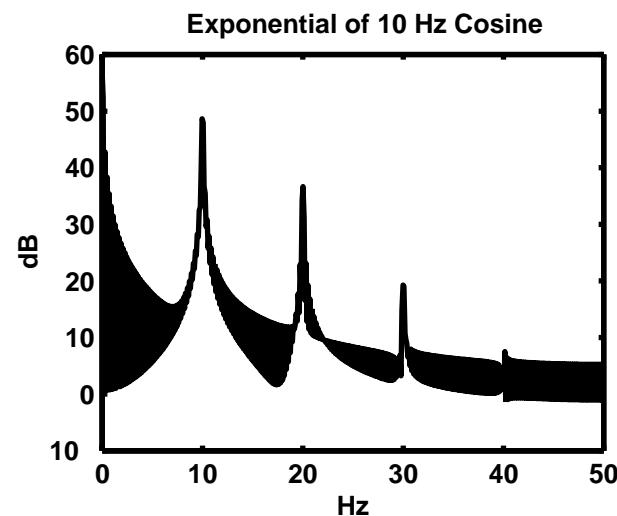
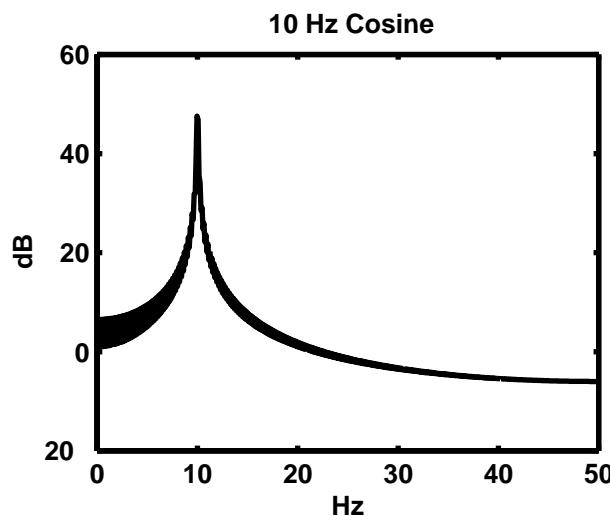
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



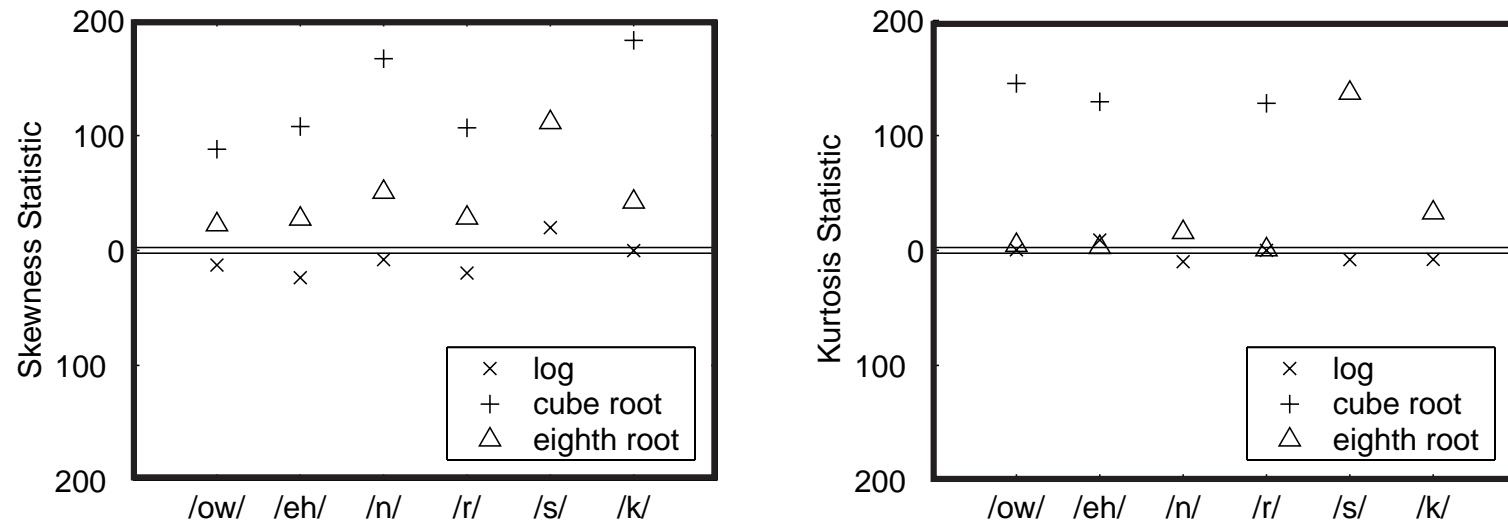
Domain



- Must use log-domain to better obey LDA assumptions.
- Effect of non-linearity is to add harmonics.
- Raises, noise floor but original fundamental remains.
- Also, shape invariance of log to powers: $\log(x^p)=p \log(x)$



Gaussianity test



$$S = \frac{1}{\sqrt{6T}\hat{\sigma}^3} \sum_{t=1}^T (x_t - \hat{\mu})^3$$

$$K = \frac{1}{\sqrt{24T}\hat{\sigma}^4} \sum_{t=1}^T (x_t - \hat{\mu})^4 - \sqrt{\frac{3T}{8}}$$

- Skewness and Kurtosis statistics should be zero for Gaussianity
- Log is closer to meeting criterion than cube and eighth root



MFSG-LDA

WER (%)

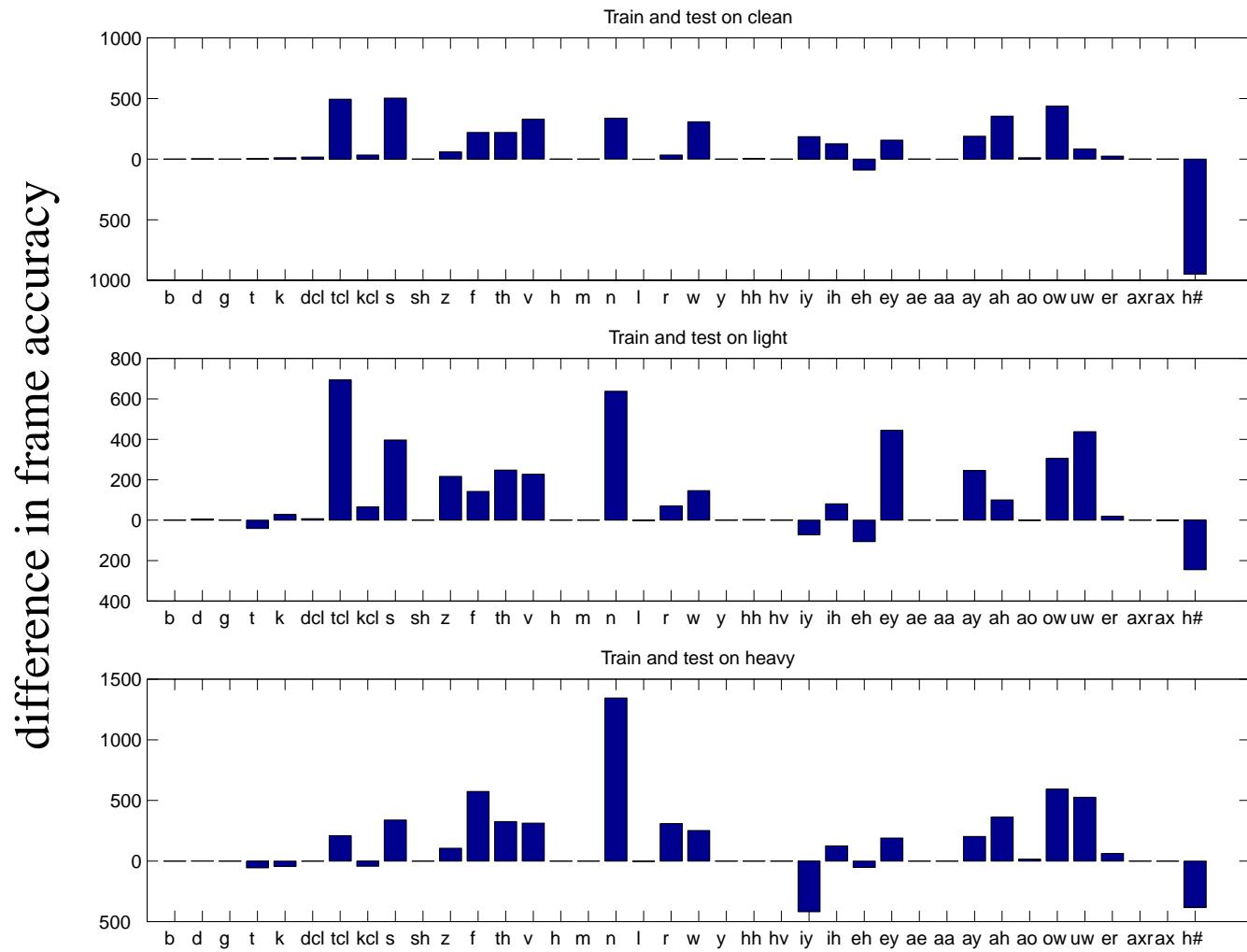
Train	Test	MFSG	MFSG-LDA
clean	clean	6.50	7.00
light	light	12.10	12.40
heavy	heavy	31.60	32.80

800HU MLP
9 frames context
Frame accuracy (%)

Train	Test	MFSG	MFSG-LDA
clean	clean	76.96	78.40+
light	light	70.95	72.82+
heavy	heavy	55.66	57.89+



MFSG vs MFSG-LDA



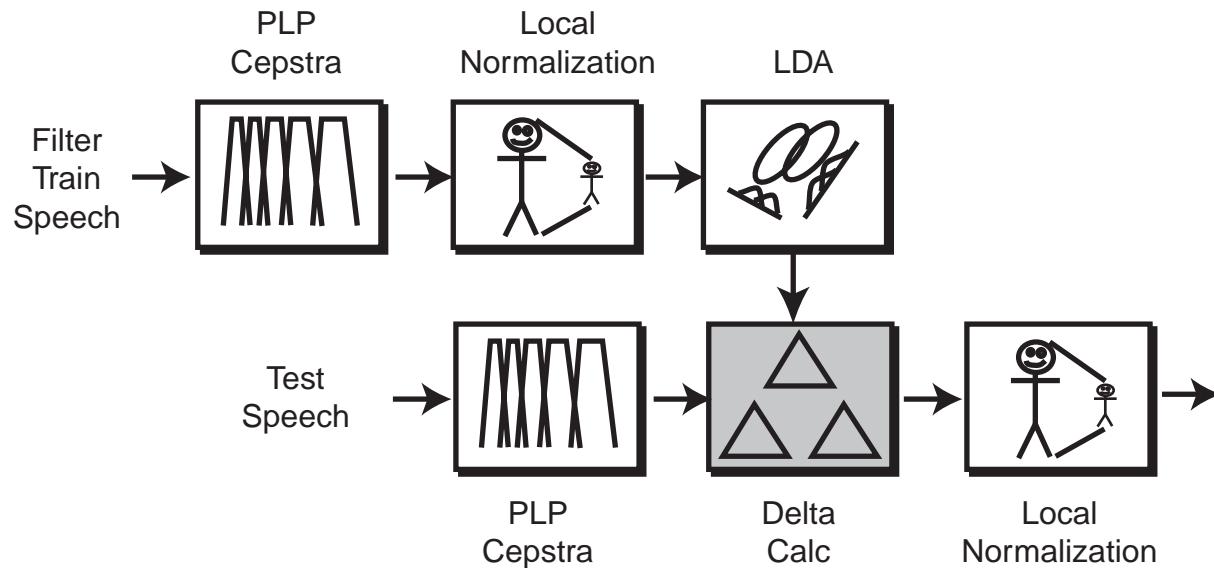
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



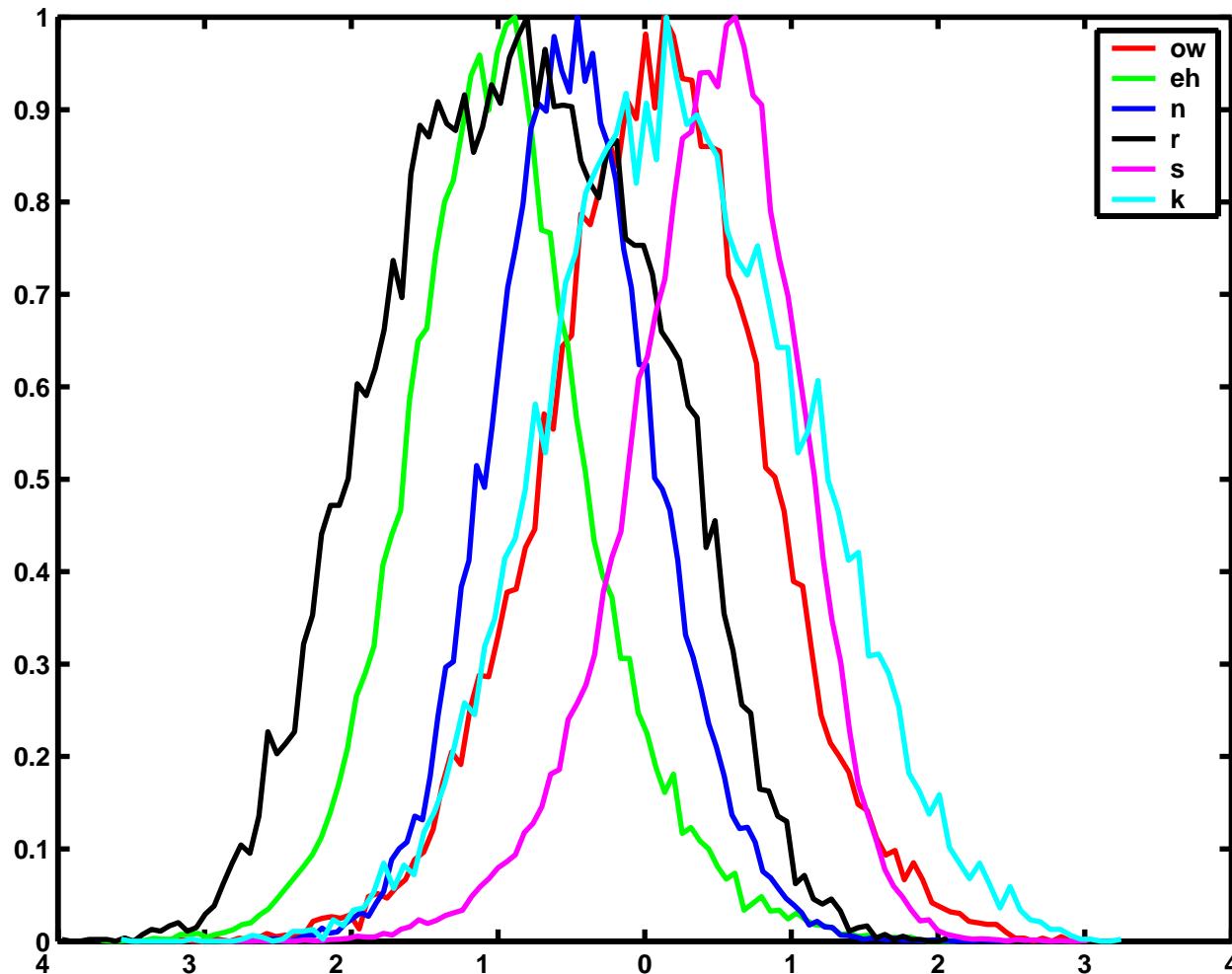
PLP-Cepstra-LDA (delta replacement)



- Features, delta, and double delta replaced by LDA filtered features
- Cepstra has implicit logarithm followed by linear transform.



PLP-Cepstra class distributions



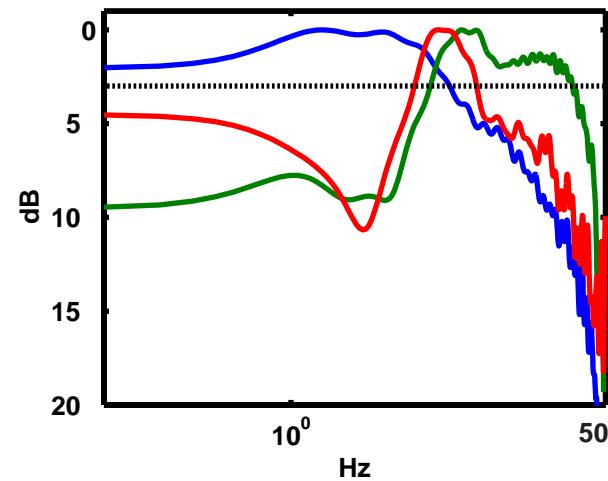
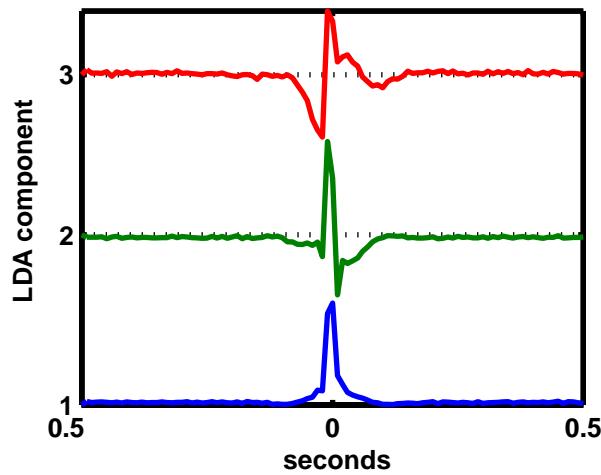
ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Sample PLP LDA filters



ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



PLP-LDA

WER (%)

Train	Test	PLP+Deltas	PLP-LDA
clean	clean	7.80	8.80 -
light	light	16.20	17.60 -
heavy	heavy	46.10	40.40 +

400HU MLP
single frame acoustic input
Frame accuracy (%)

Train	Test	PLP+Deltas	PLP-LDA
clean	clean	70.75	74.65 +
light	light	62.05	67.74 +
heavy	heavy	45.59	53.48 +



Summary

- RASTA-PLP
 - General improvement in WER using specific LDA filters.
 - Filters appear to work best in matched training and testing conditions.
 - Advantage mitigated by allowing contextual frames to MLP.
- MFSG, PLP-Cepstra
 - Consistent improvement in frame-level phone classification accuracy.
 - Inconsistent effect on WER.



ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Discussion

- LDA filters designed to discriminate between phone classes.
- Yields consistent improvement in frame accuracy.
- WER can be worse, mismatch with Viterbi assumption?
- Relation of frame and word accuracy difficult to analyze.



ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen



Conclusion

- Is environment specific temporal LDA useful in speech feature extraction in ASR?
 - Automatic means of selecting reasonable temporal filters.
 - Can help when adding acoustic context is prohibitive or expensive to the probability estimator.
 - Basic speech modulations below about 16Hz is important, further specific partitioning with filters can yield diminishing returns.



ICSI - UCB

ICSLP 2000 - Beijing China

M.L.Shire,B.Y.Chen

