Gammatone Sub-band Magnitude Domain Dereverberation for ASR

Kshitiz Kumar, Rita Singh,
Bhiksha Raj, and Richard M. Stern
Carnegie Mellon University

ICASSP 2011, 26th May 2011

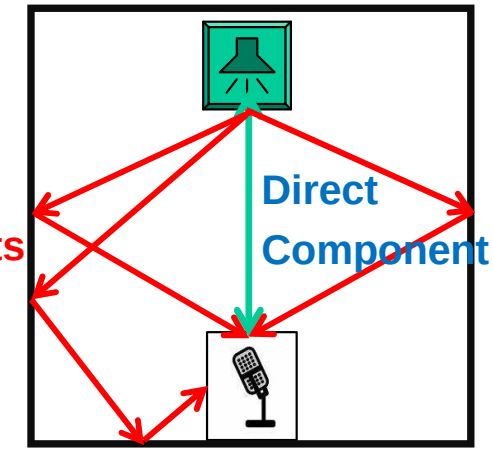
Electrical & Computer ENGINEERING

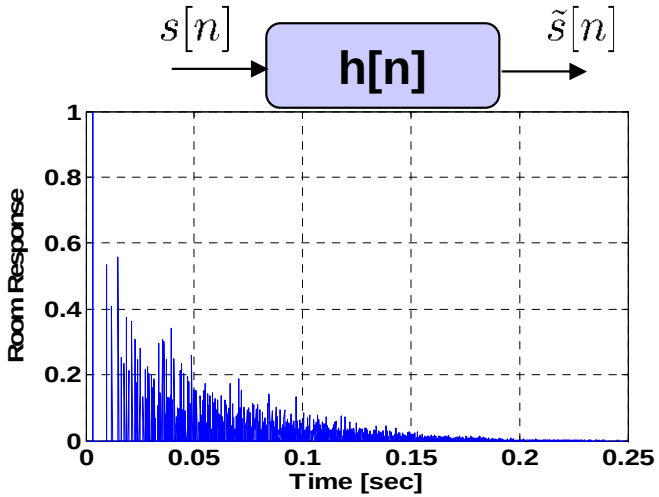
Robustness in Speech Technologies

- Speech is a natural medium of communication for humans
- Speech Technologies
  - Speech Recognition, Translation
  - Speaker Identification, Dictionary Systems
  - Voice Retrieval
- Work great in controlled/lab conditions
  - Performance degrades in real-world conditions including noise and reverberation

The Problem of Reverberation

- Reverberation is a superposition of delayed and attenuated signals
- It is modeled in terms of room impulse response (RIR)

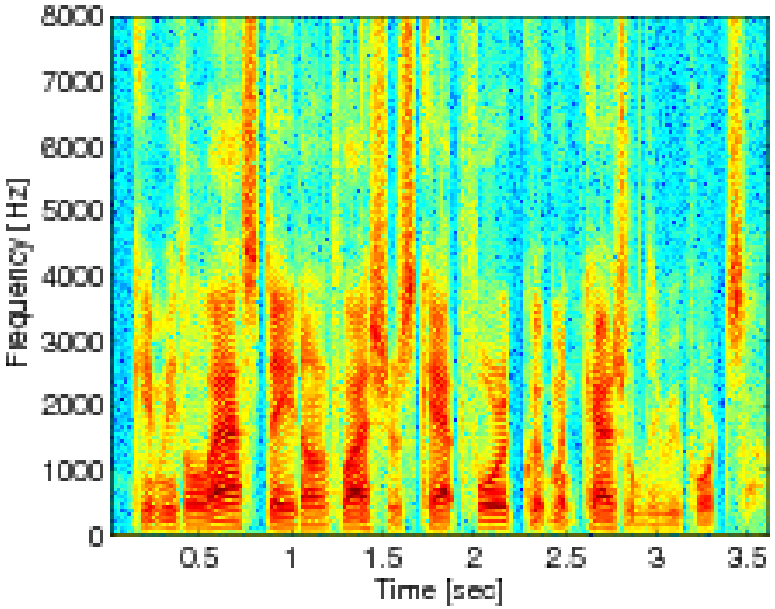




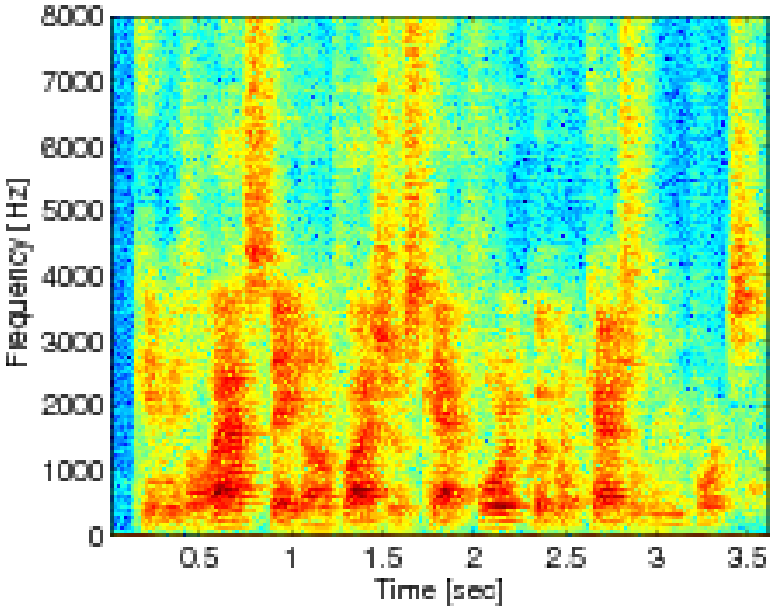
Simulated RIR 5x4x3 m

- Room reverberation time (RT) - time for signal energy to decay by 60-dB

Impact of Reverberation on Speech Spectra



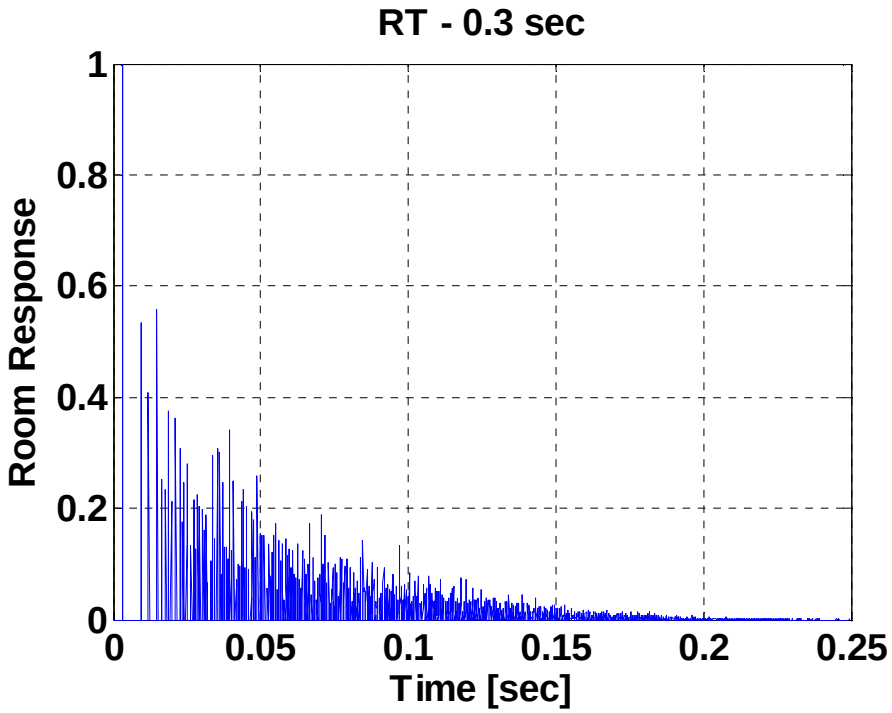
Clean Condition

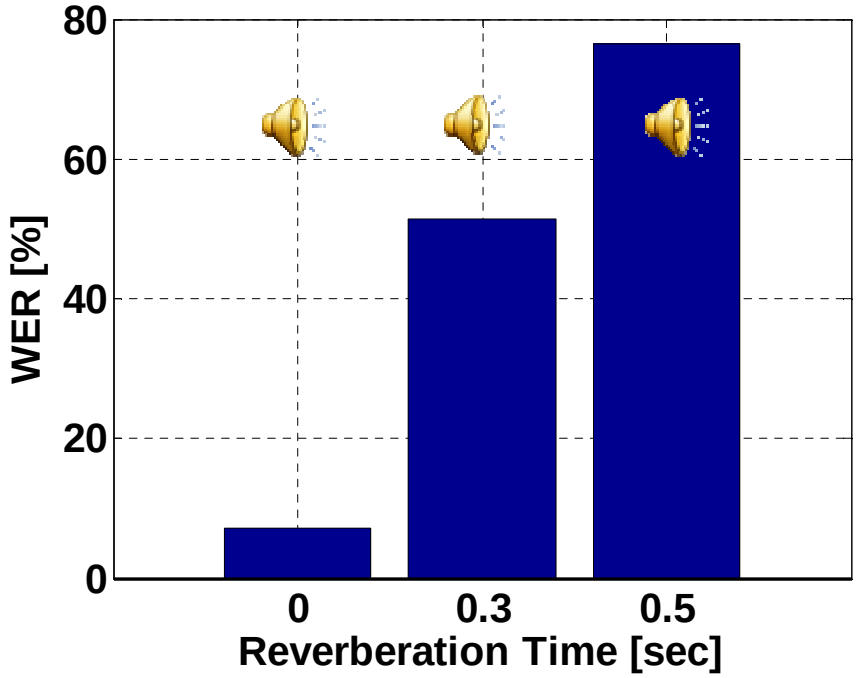


Reverberation Condition [RT of 300 ms]

Impact of Reverberation on ASR

- WER increases from 7% (Clean) to 52% (RT 0.3 sec)





Simulated RIR for a 5x4x3 m room

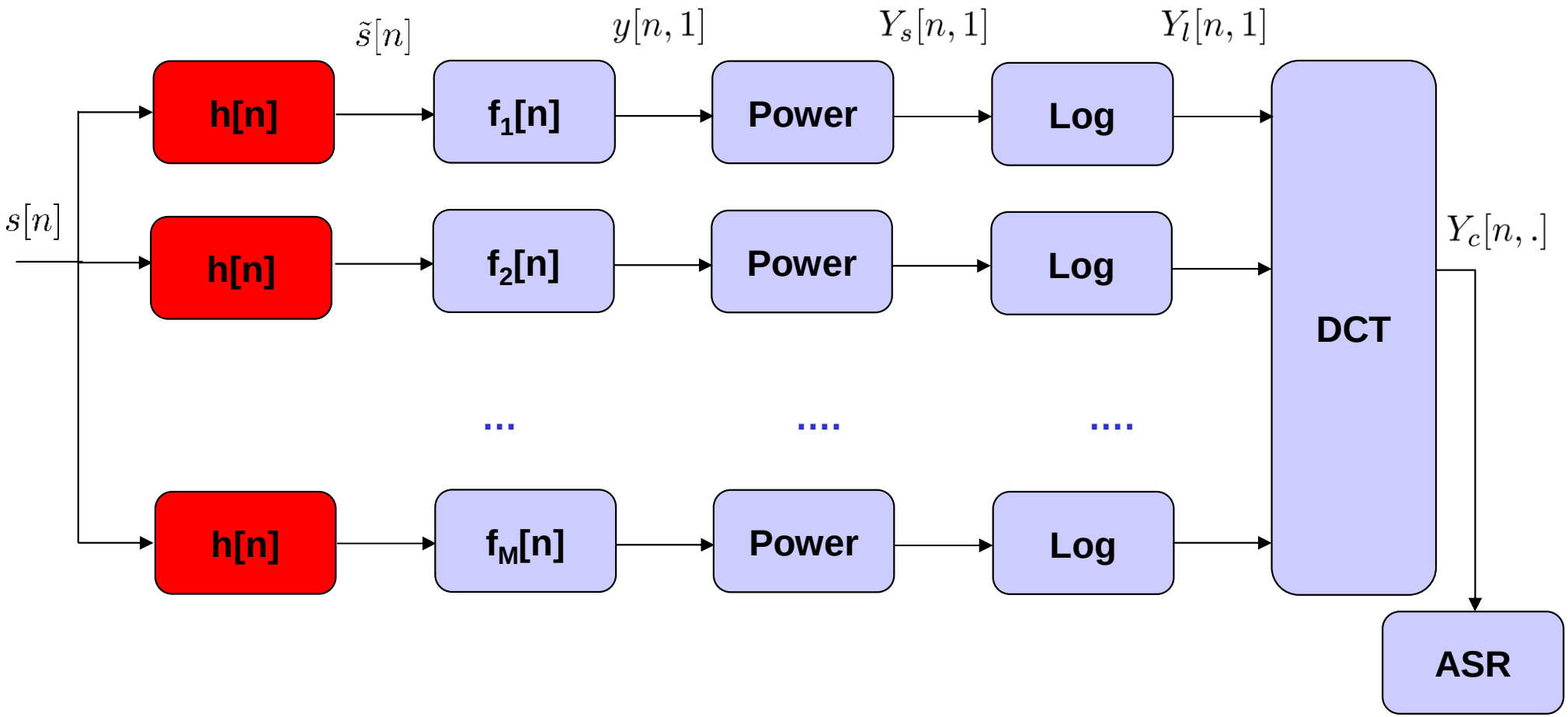
RM1 Database, Sphinx 3 Decoder

Train on Clean, test across environments

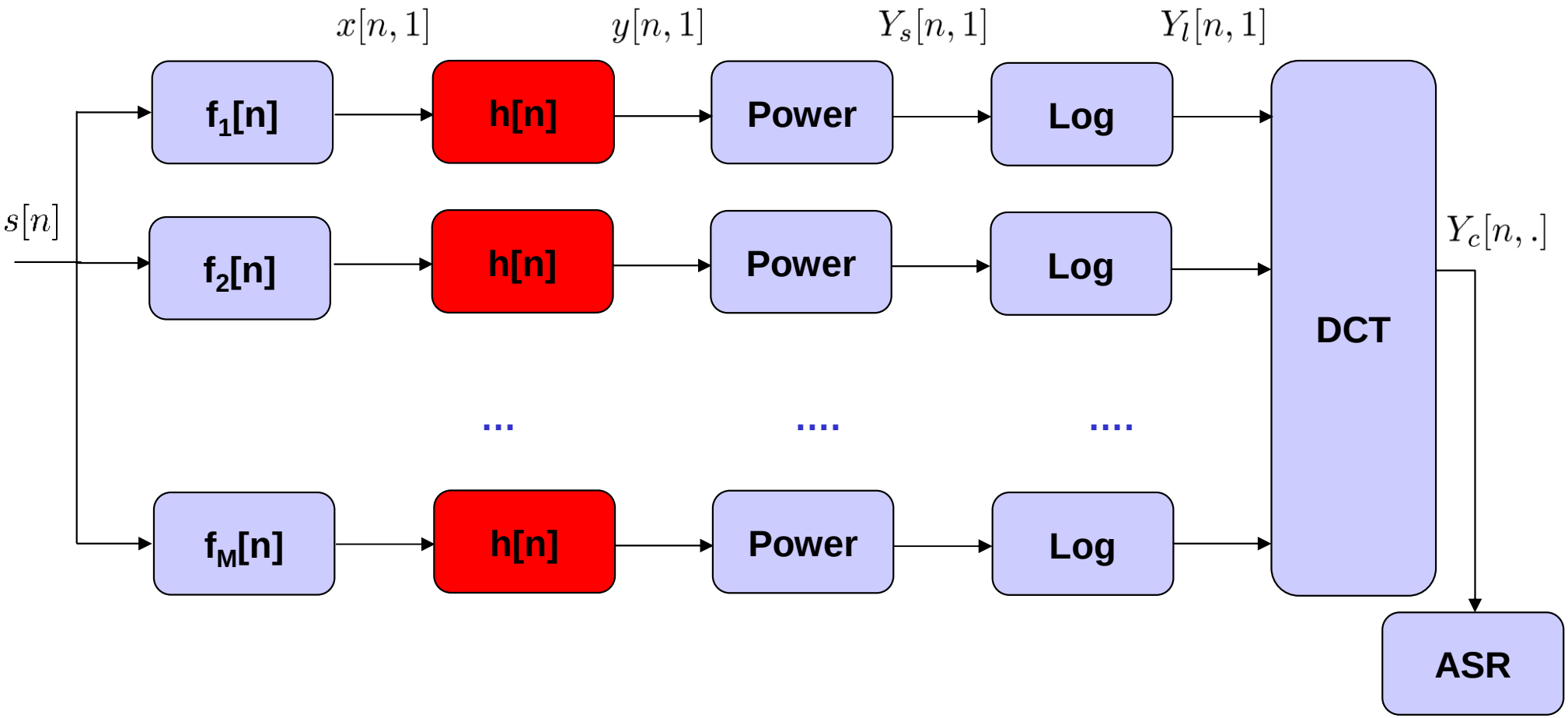
Electrical & Computer ENGINEERING

Studying Reverberation in Spectral Domain

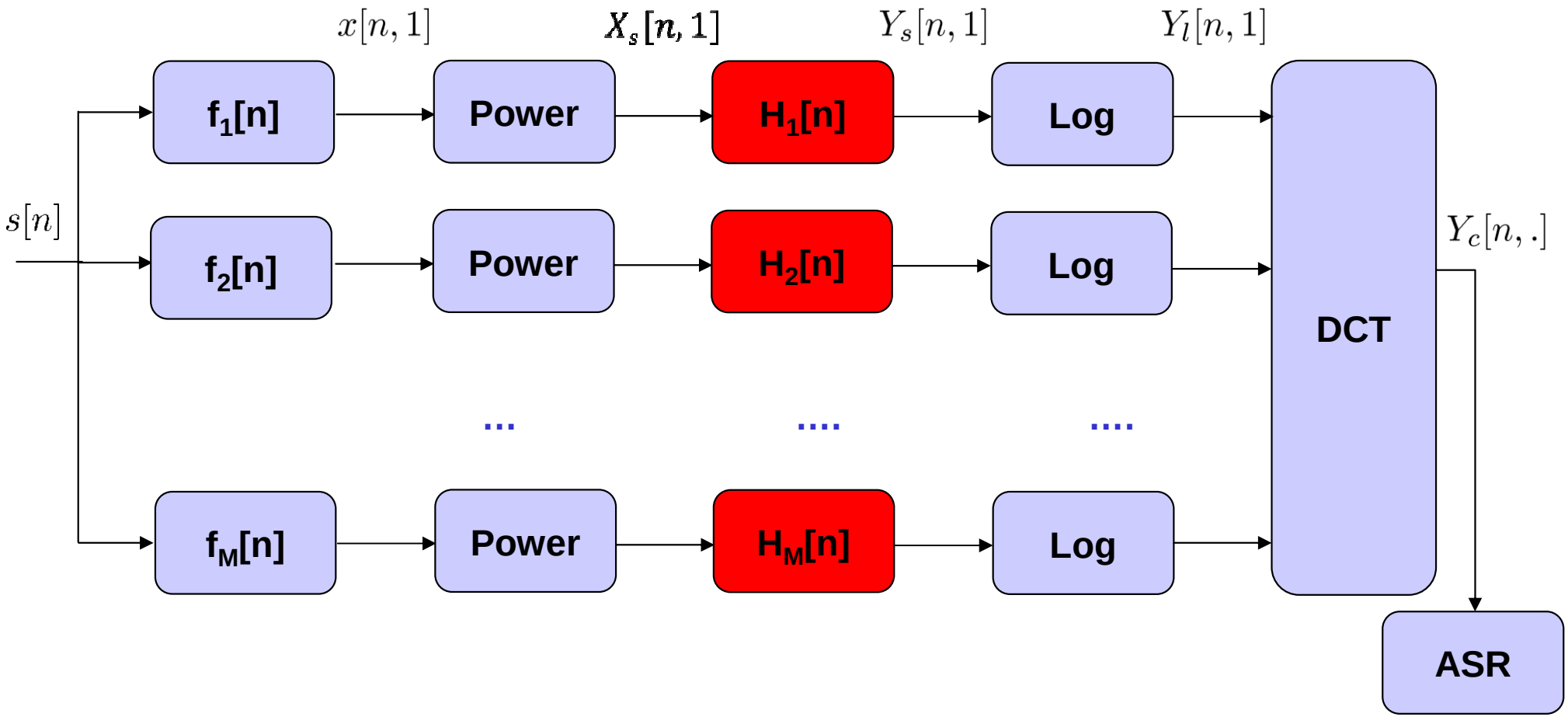
Speech Features: Short Duration Spectral Power



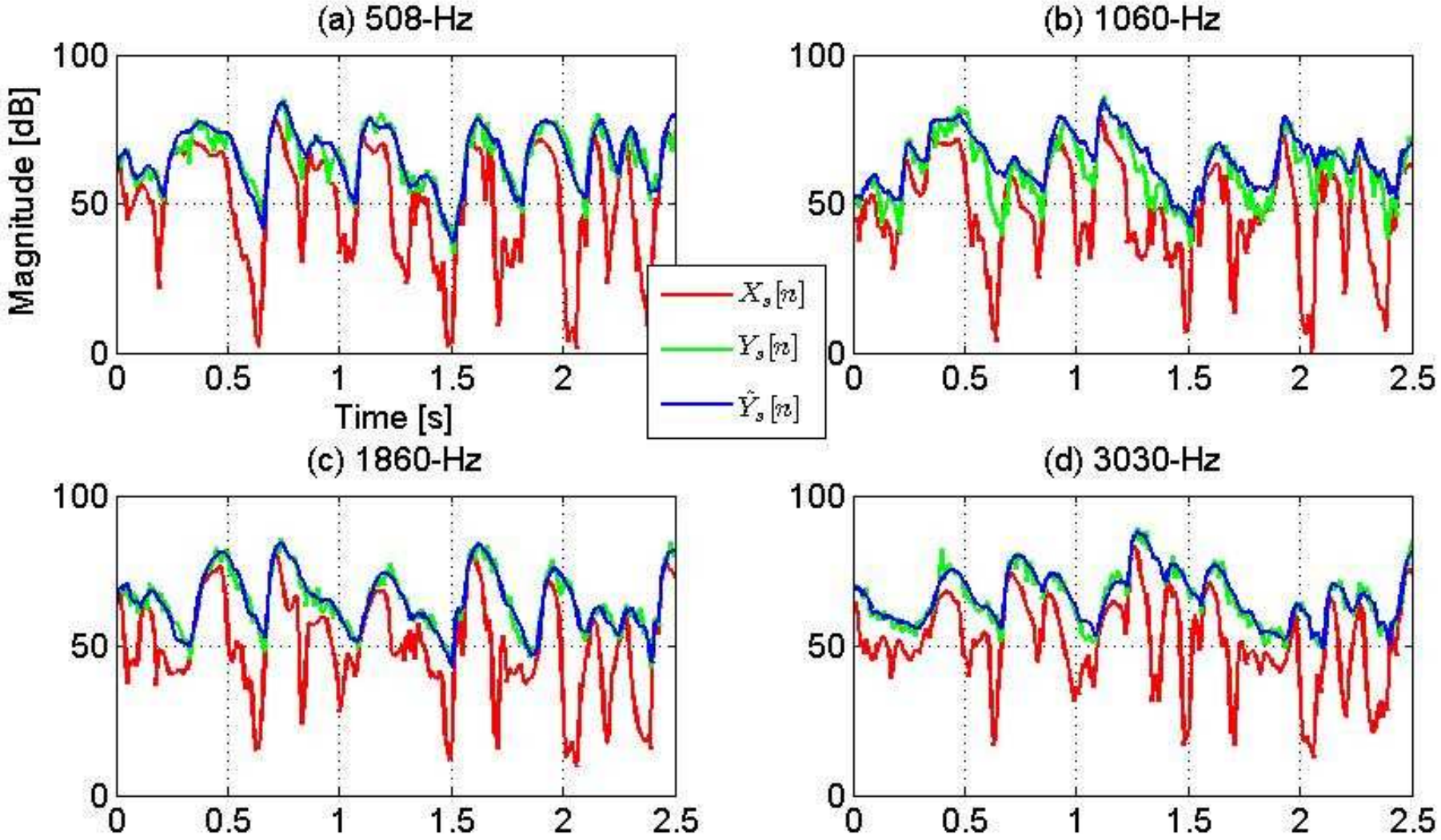
Speech Features: Short Duration Spectral Power



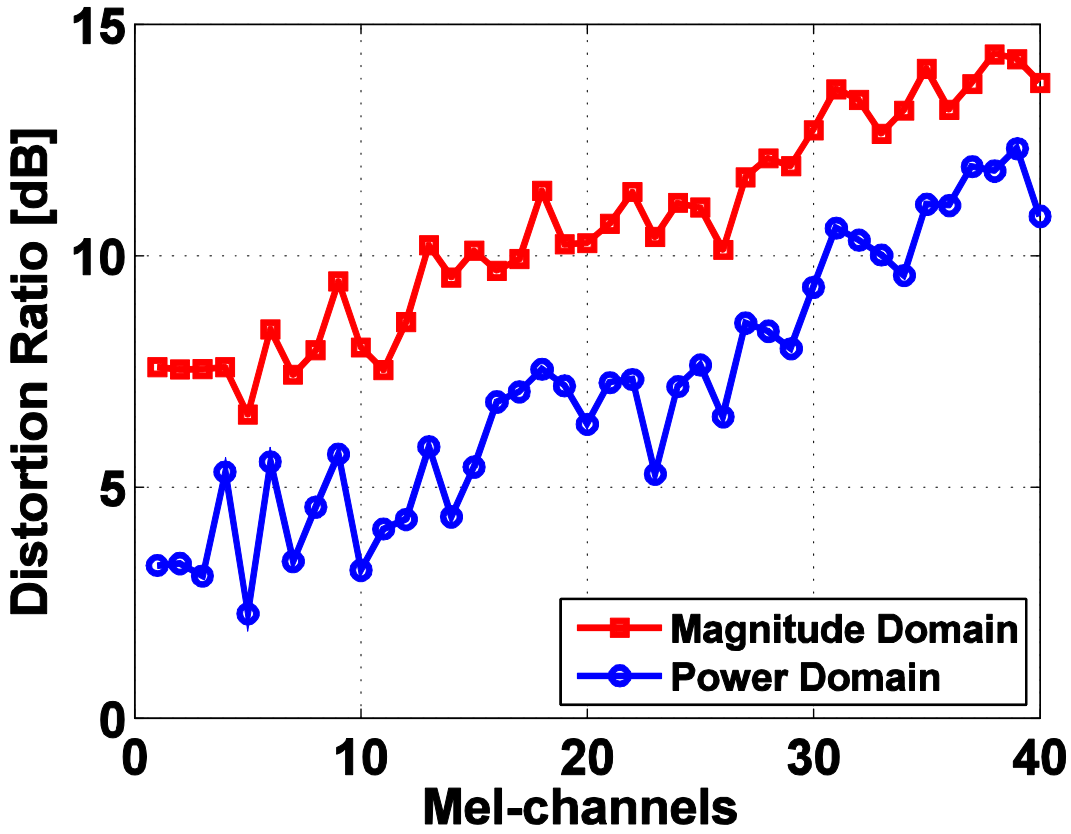
Speech Features: Short Duration Spectral Power



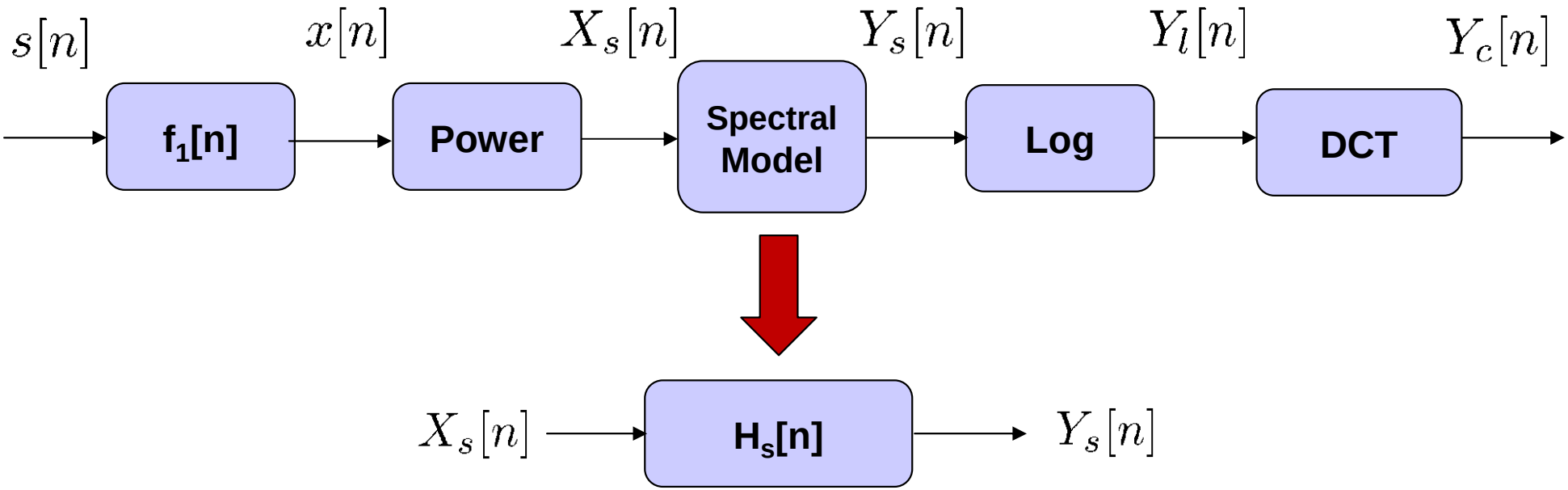
Proposed Short-Term Power Model



Power-Sequence-to-Distortion Ratio for the Model



GAMMATONE SUB-BAND NON-NEGATIVE MATRIX FACTORIZATION (NMF)



Spectral Convolution Model for NMF

- Convolution Model in Spectral domain

$$Y_s[n] = \sum_m H_s[m] X_s[n-m]$$

- Only $Y_s[n]$ is observed
- Multiple possible decompositions of $Y_s[n]$

- To confine solution space, incorporate constraints like non-negativity, sparsity

Dereverberation Problem Formulation

- Convolution Model

$$Y_s[n] = \sum_m H_s[m] X_s[n-m], \quad \text{Linear System Model}$$

$$Z_s[n] = Y_s[n] + N_s[n], \quad \text{Observation Model}$$

- Non-negativity constraint $X_s[n] \geq 0, \; H_s[n] \geq 0$
- (Optional) Sparseness $\sum_n X_s[n]^p$
- Objective Function

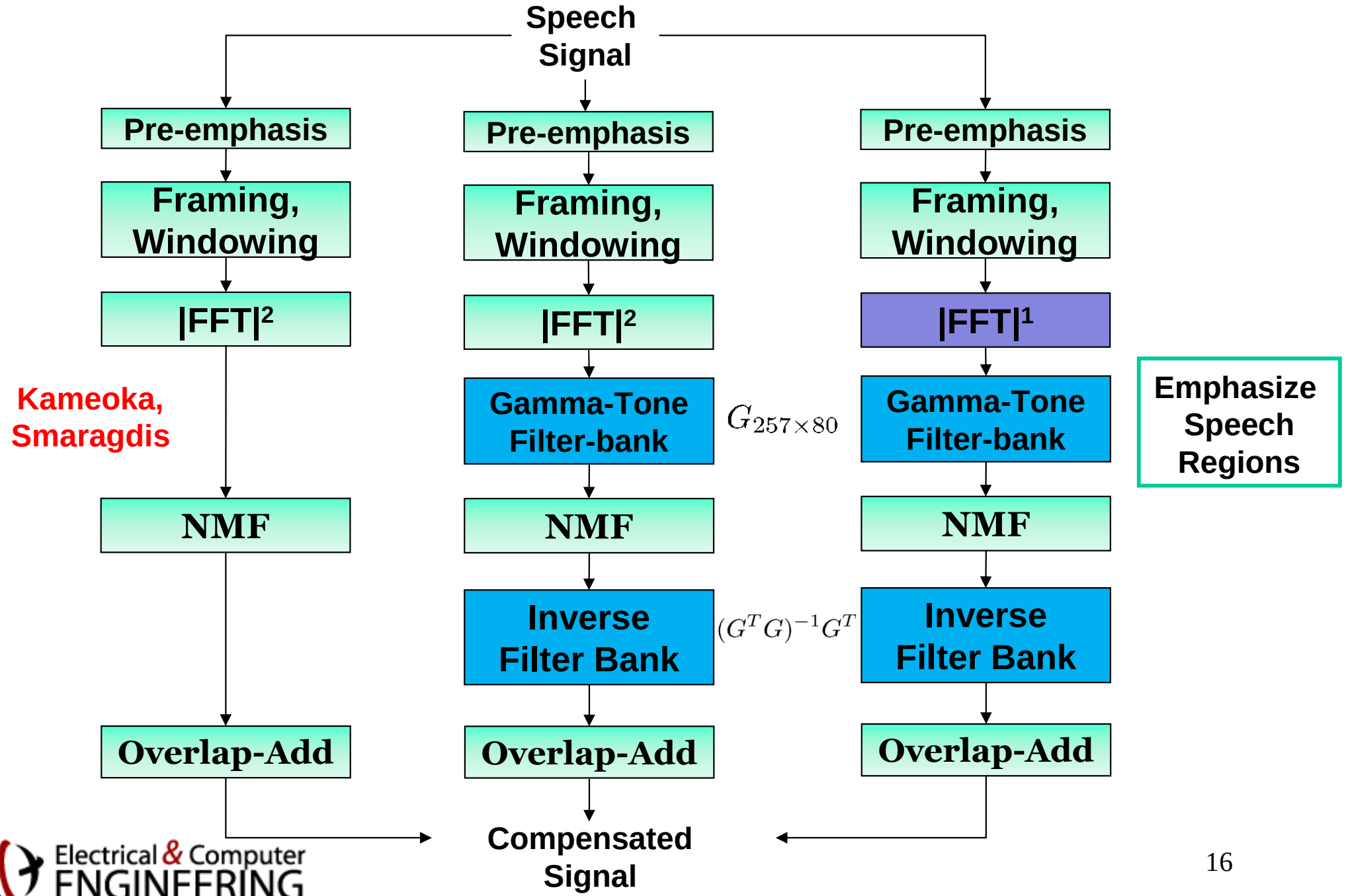
$$\text{minimize } E = \sum_i \left(Z_s[i] - \sum_m \left(X_s[m] H_s[i-m] \right) \right)^2 + \lambda \sum_i X_s[i]^p$$

$$\text{subject to } X_s[n] \geq 0 \; H_s[n] \geq 0 \; \sum_n H_s[n] = 1$$

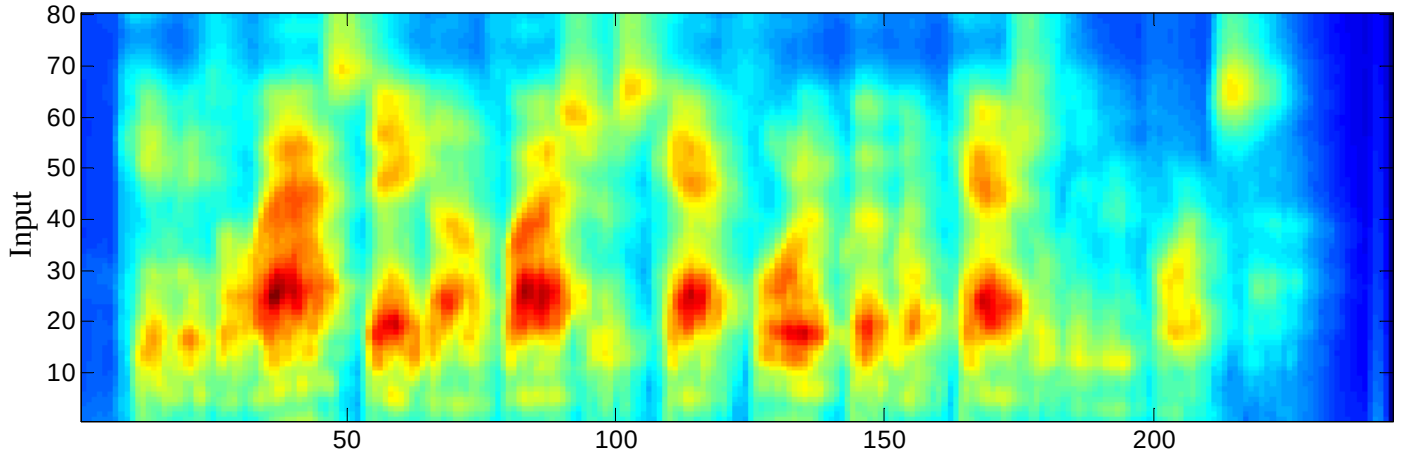
NMF Update Equations

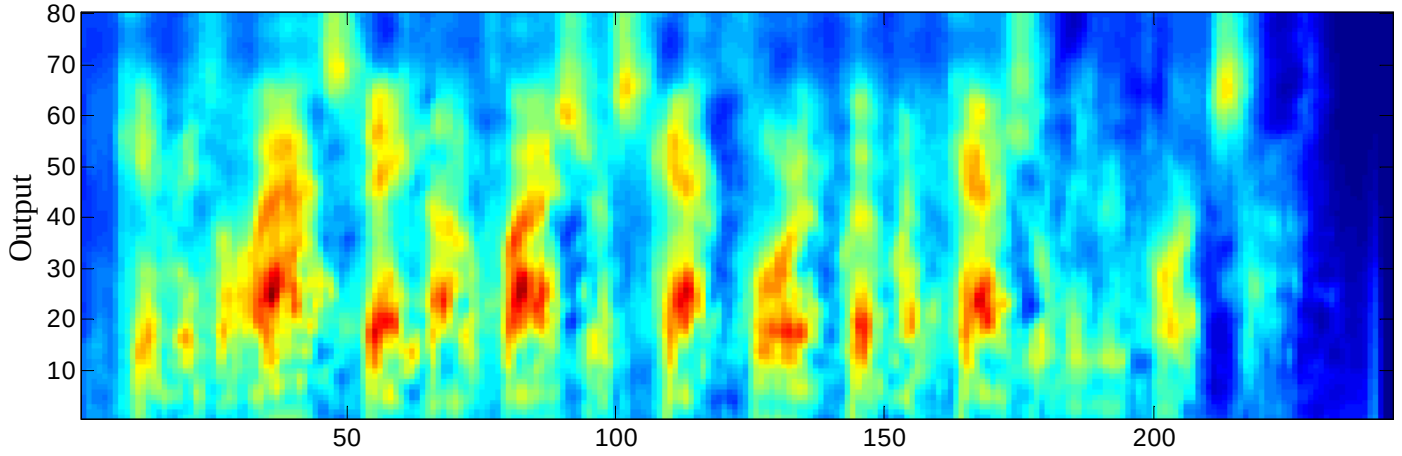
$$
\begin{aligned}
\frac{\partial E}{\partial X_s[n]} &= -2\sum_i \big(Z_s[i] - Y_s[i]\big)\, H_s[i-n] + \lambda \\
\bar{X}_s[n] &\leftarrow X_s[n] - \eta_s \frac{\partial E}{\partial X_s[n]} \\
&= 2\eta_s \sum_i Z_s[i] H_s[i-n] + X_s[n] - 2\eta_s \big(\sum_i Y_s[i] H_s[i-n] + \lambda/2\big) \\
\eta_s &= \frac{X_s[n]}{2\sum_i Y_s[i] H_s[i-n] + \lambda} \\
\bar{X}_s[n] &\leftarrow X_s[n] . \frac{\sum_i Z_s[i] H_s[i-n]}{\sum_i Y_s[i] H_s[i-n] + \lambda/2} \\
\bar{H}_s[n] &\leftarrow H_s[n] . \frac{\sum_i Z_s[i] X_s[i-n]}{\sum_i Y_s[i] X_s[i-n]}
\end{aligned}
$$

NMF Compensation: Gammatone Filters

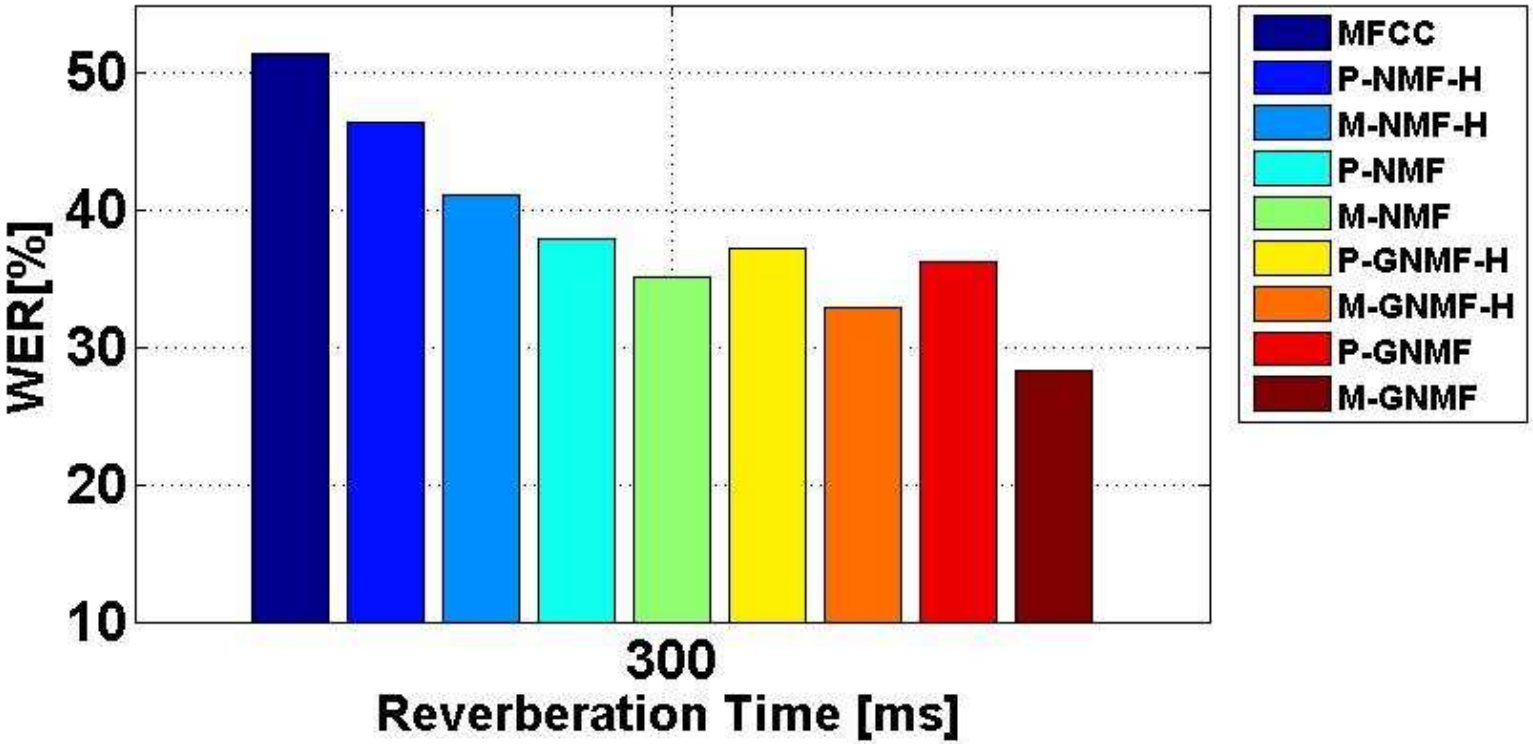


NMF Processing on Reverberant Speech (RT of 300 ms)





NMF Processing in Different Domains



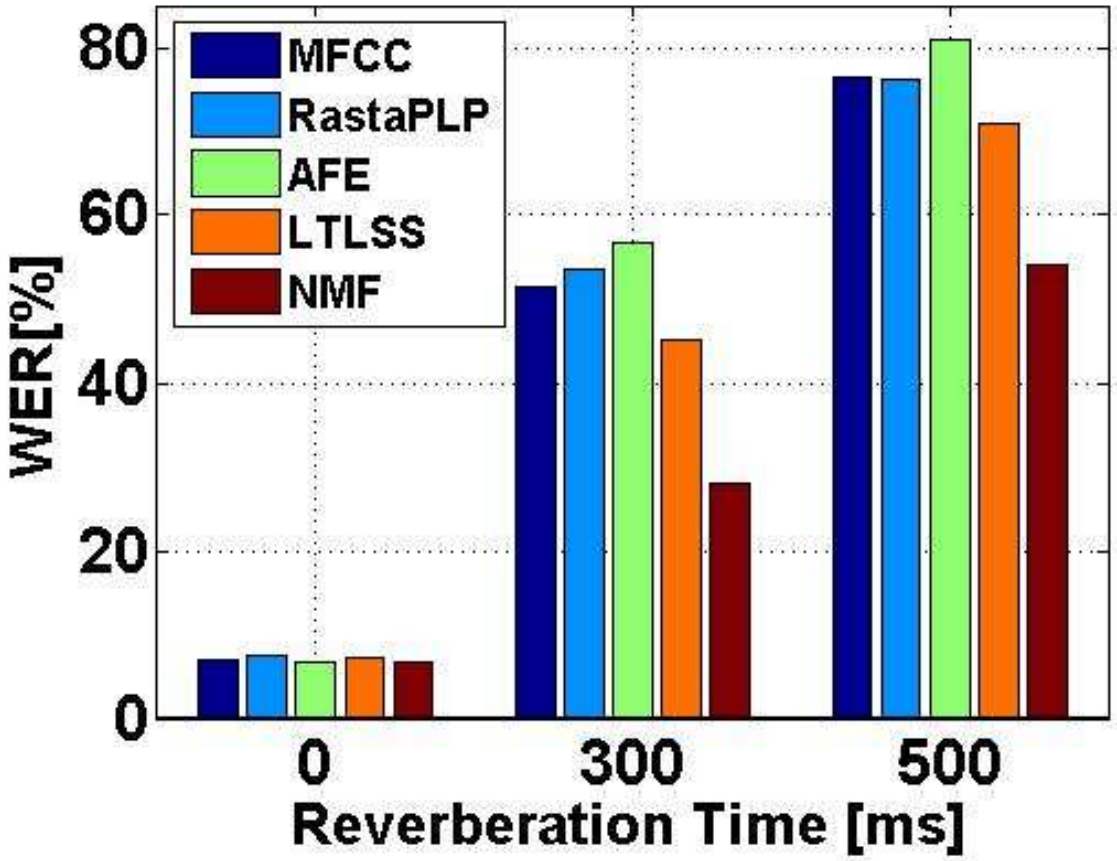
DARPA RM1 Database, Sphinx 3 decoder

GNMF (Gammatone domain NMF)

P-NMF (power-domain NMF), M-NMF (magnitude-domain NMF)

NMF-H (same H for all frequency bands)

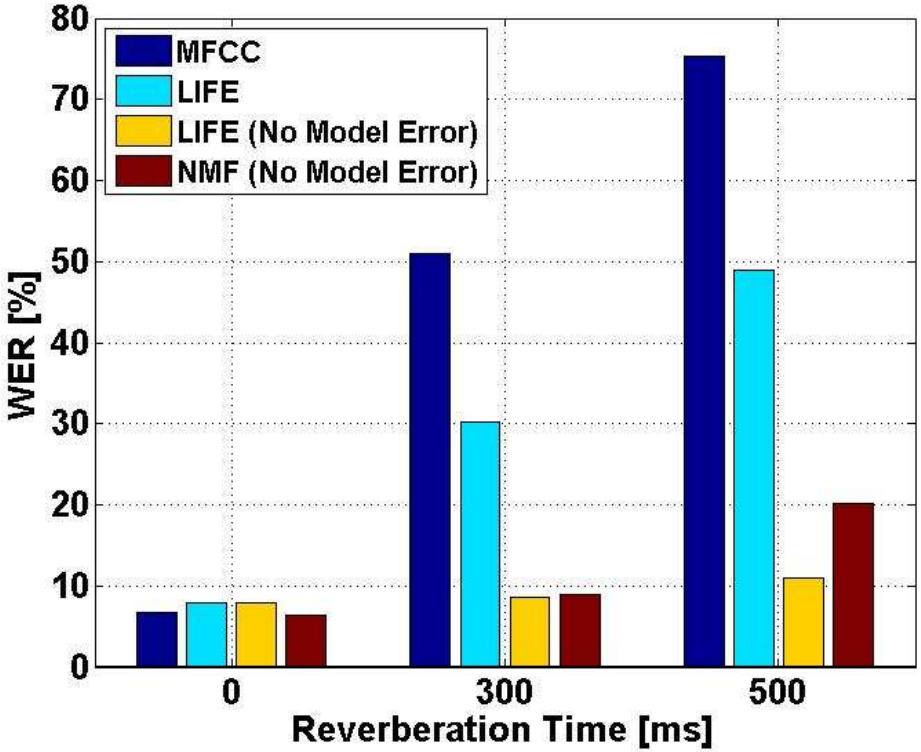
NMF WER Results



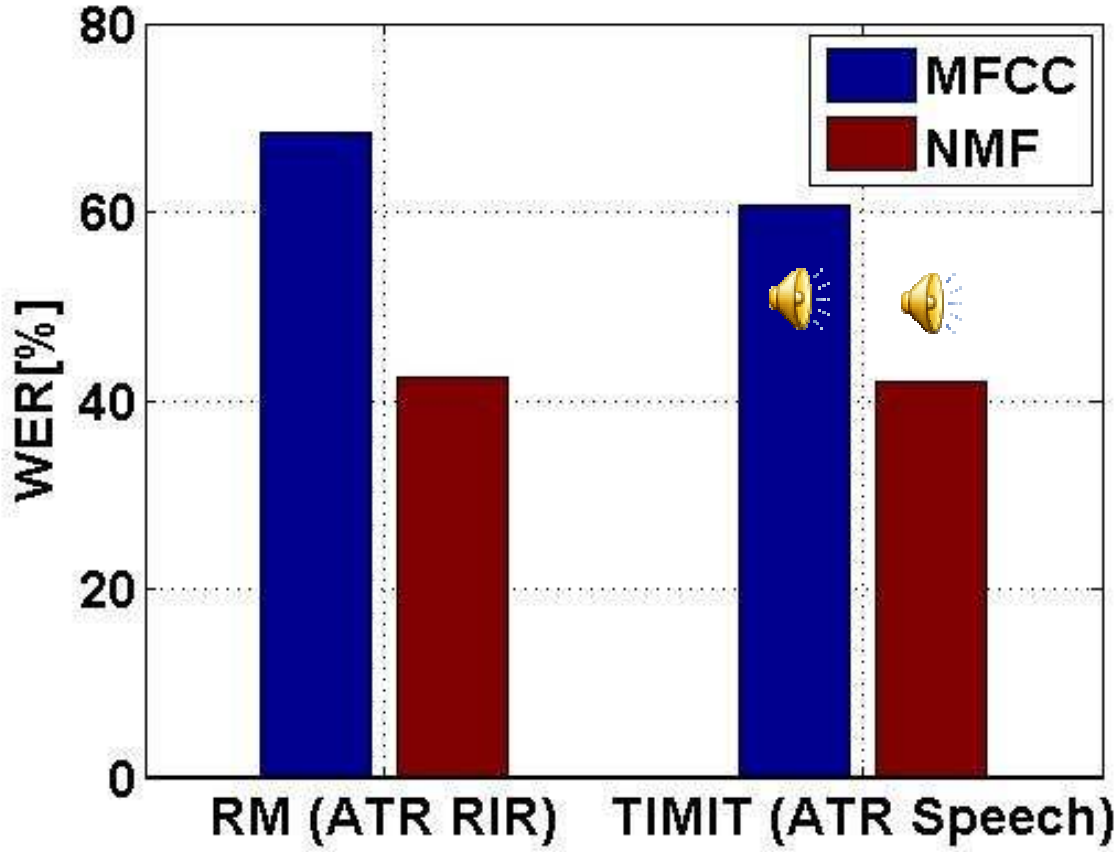
NMF: 45% relative reduction in WER at RT of 300 ms

An Experiment with zero Modeling Error

- Convolve cepstral sequences with a filter with an exponentially decaying filter-tap



NMF on Real Acoustic Recordings

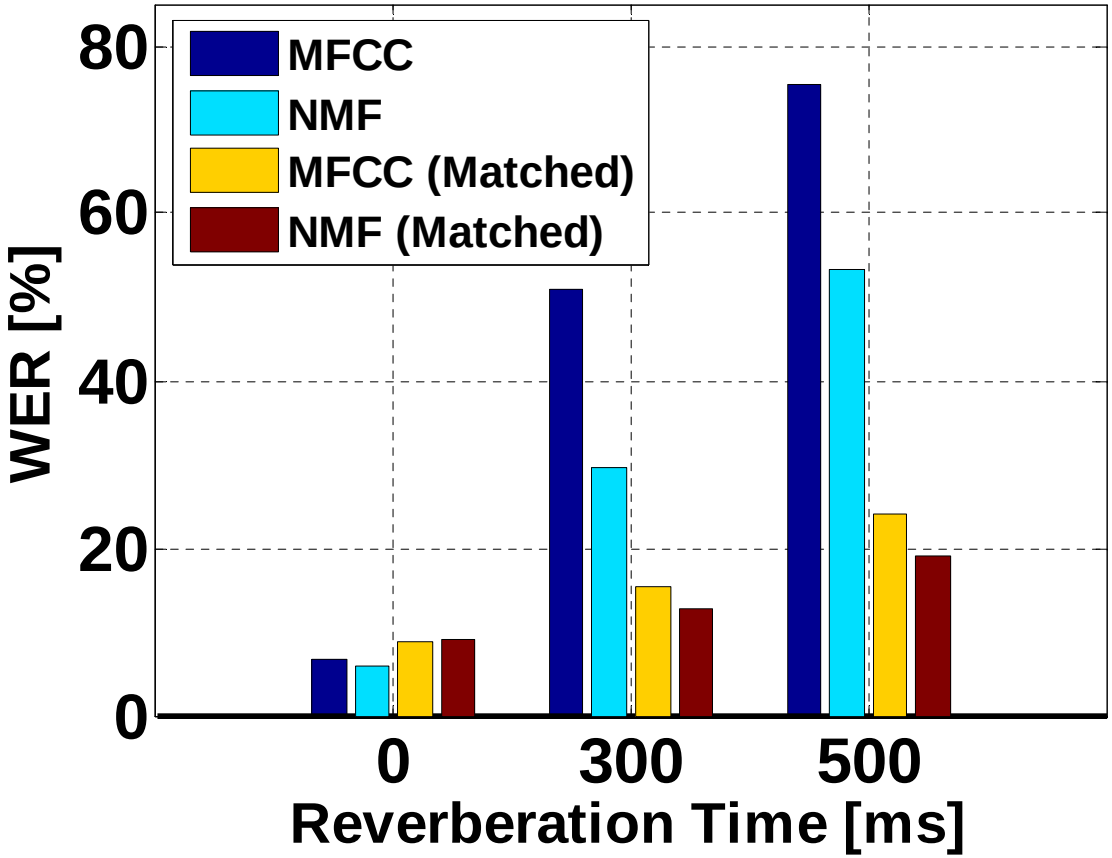


- ATR TIMIT Recordings
- Language Model from TIMIT Test Corpus

RT of 470 ms

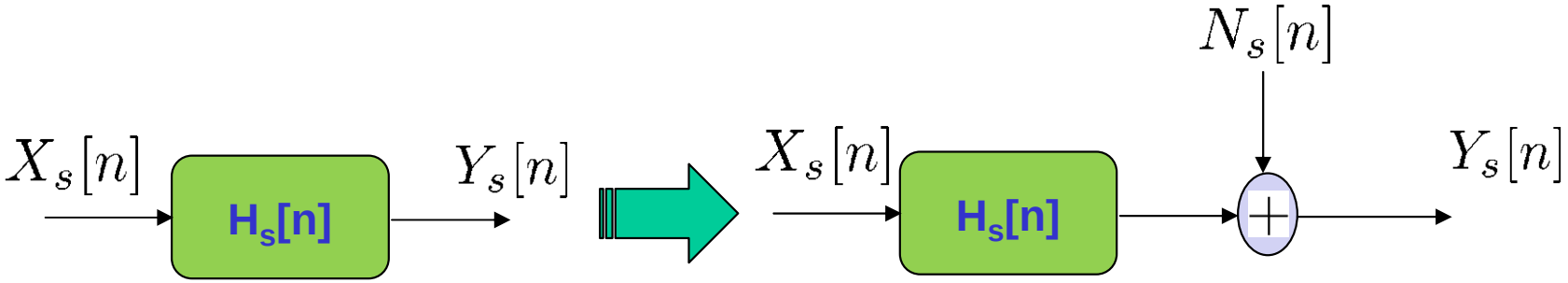
RT of 600 ms

RM Matched Condition Experiment



Joint Noise and Reverberation Model

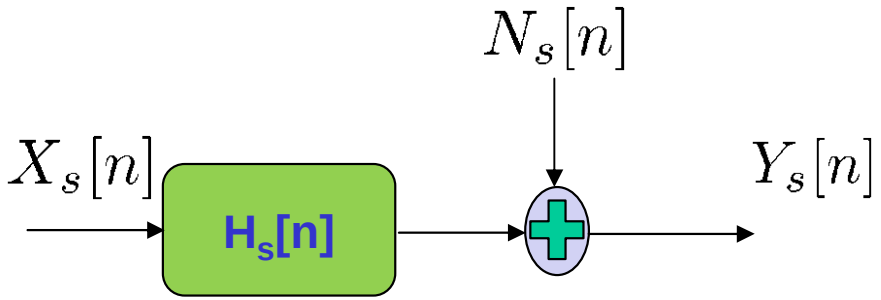
- Incorporate noise in reverberation Model



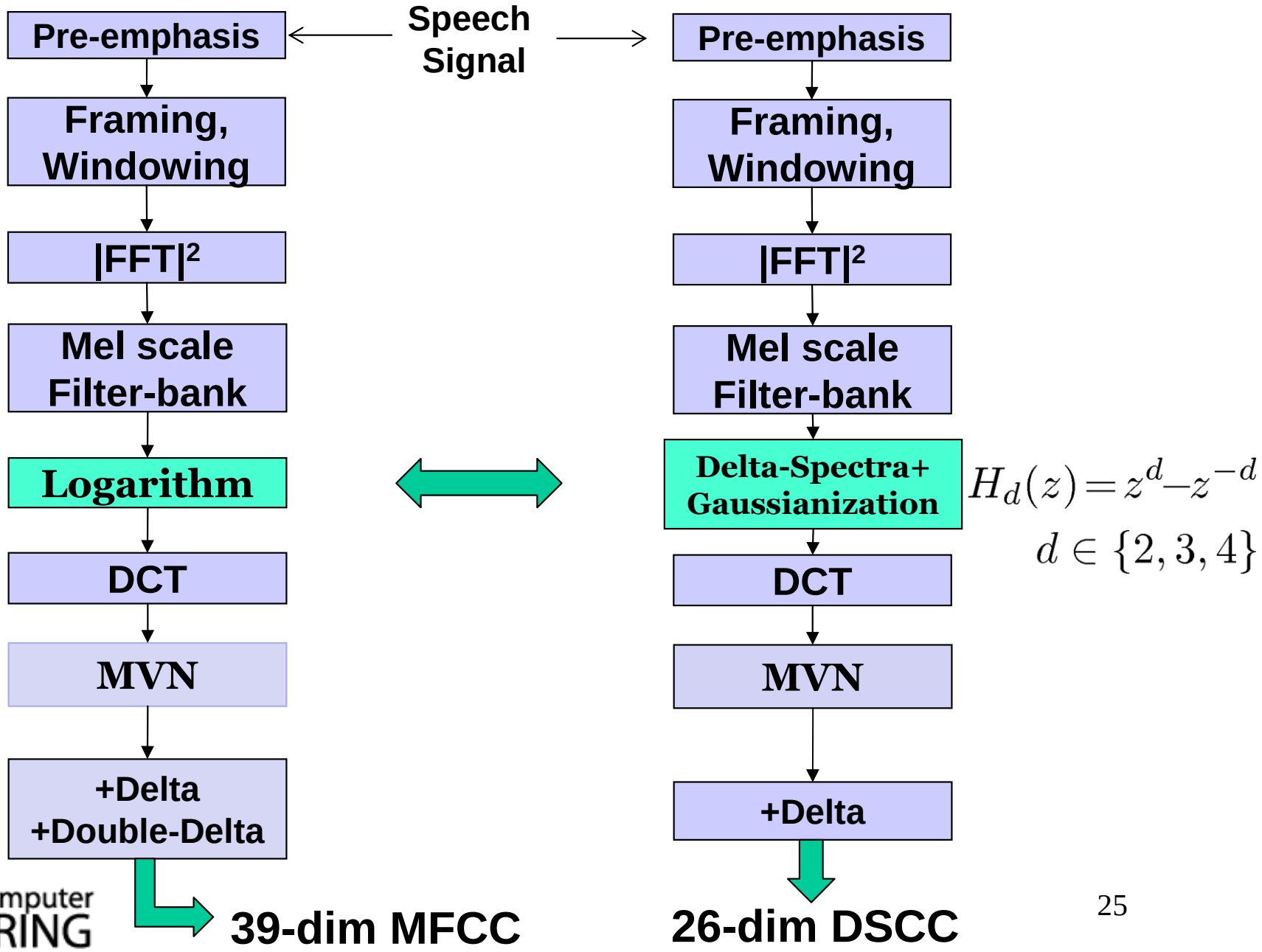
- Unified model for noise and reverberation in spectral domain

NMF and DSCC

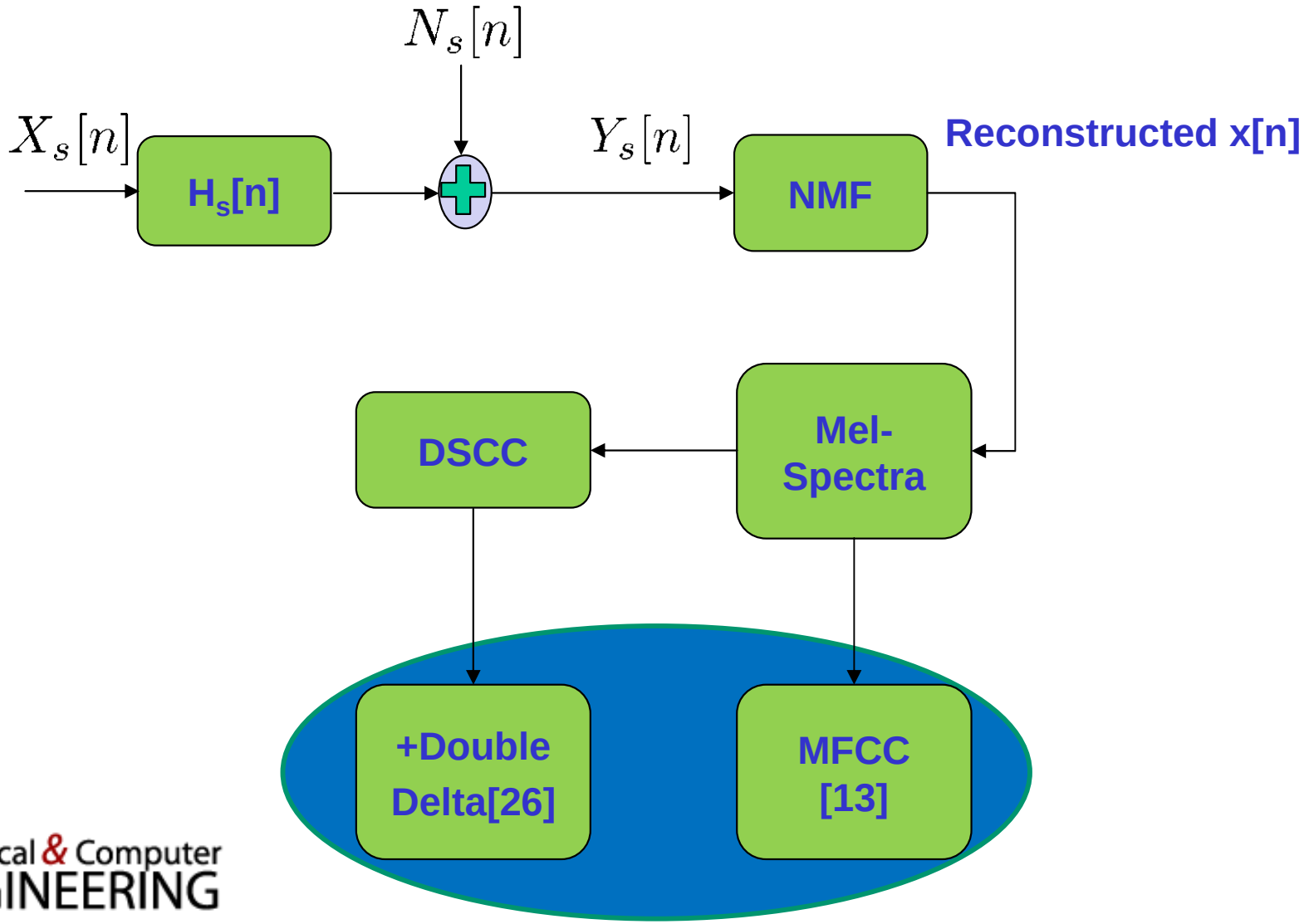
- NMF to compensate for $H_s[n]$
- DSCC to compensate for $N_s[n]$



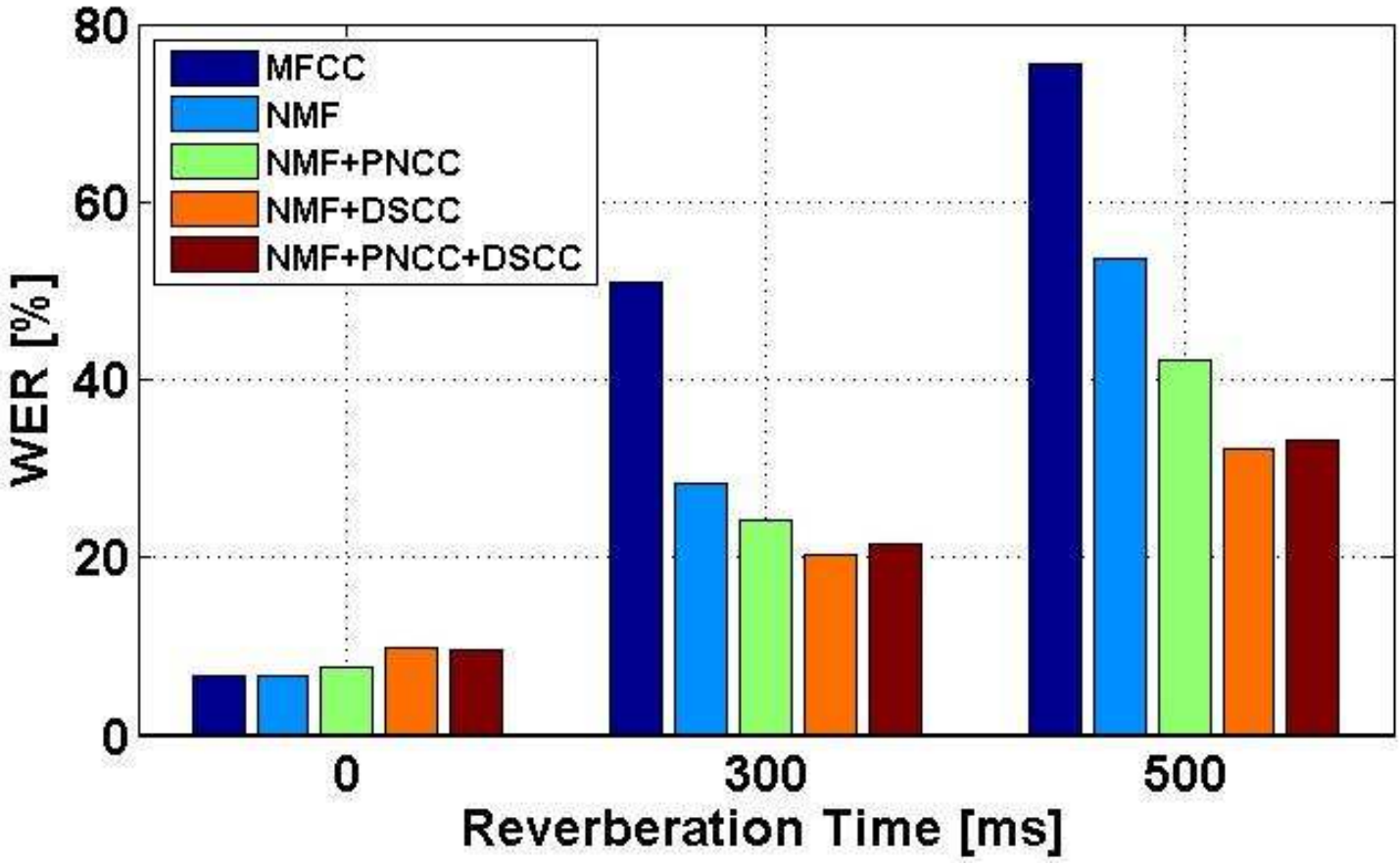
Delta Spectral Cepstral Coeff. (DSCC)



Joint NMF and DSCC

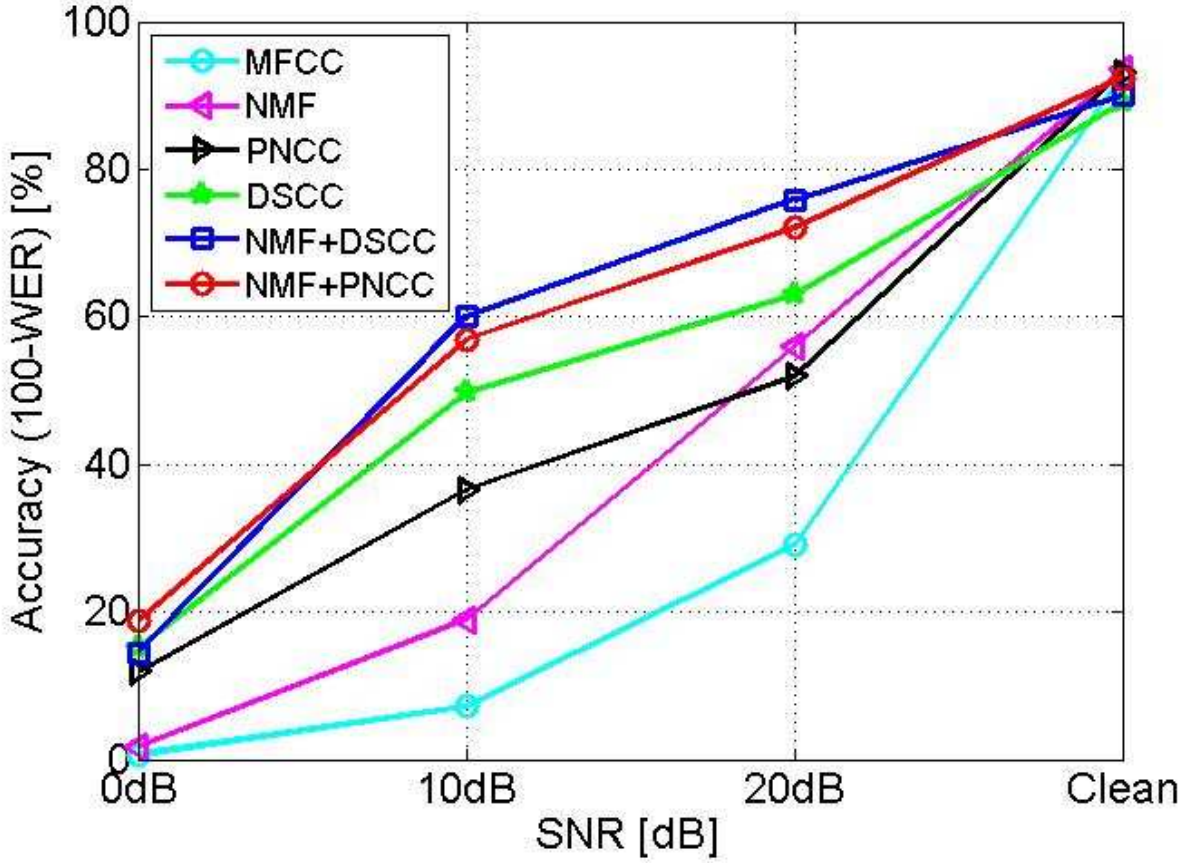


Reverb: NMF with DSCC



Reverb and Noise: NMF with DSCC

- RT of 300 ms



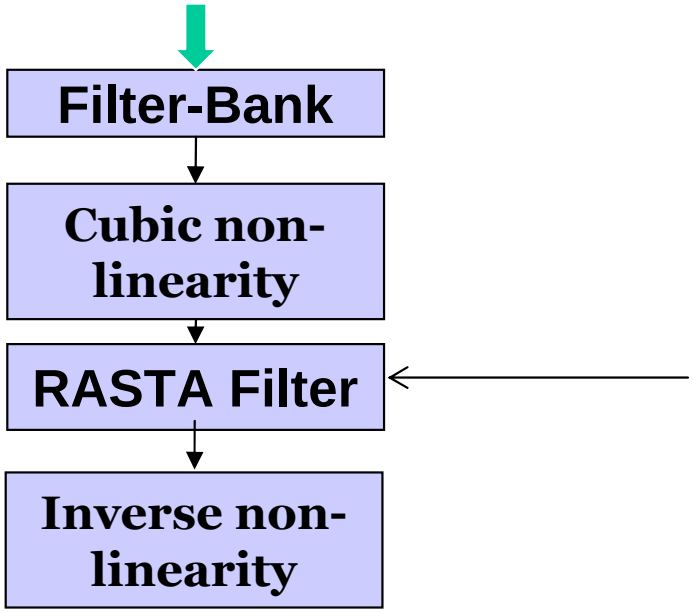
Summary

- We model reverberation in speech spectra
- NMF provides a framework for factorization of speech spectra
- Gammatone sub-band NMF provides substantial improvement over Fourier NMF
- Studied NMF on speech magnitude and power domains, we will next explore other non-linear domains.
- Studied a joint noise and reverberation problem. Integrated DSCC in the NMF framework for the joint problem

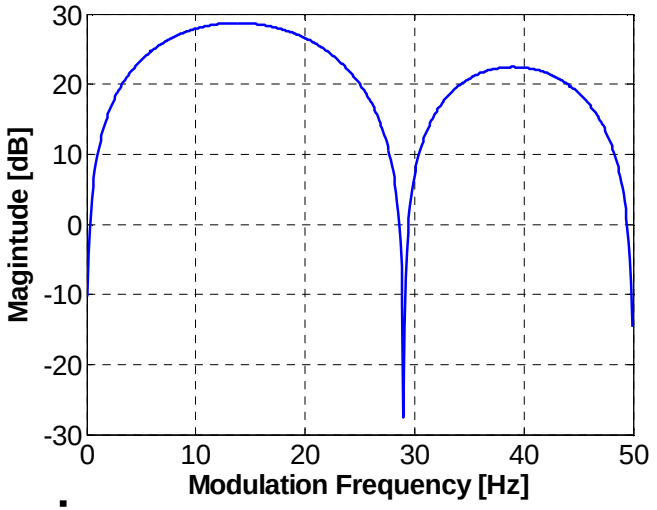
Thanks!!

Compensation in Modulation Frequency Domain: RASTA

Hermansky, Morgan, IEEE SAP 1994

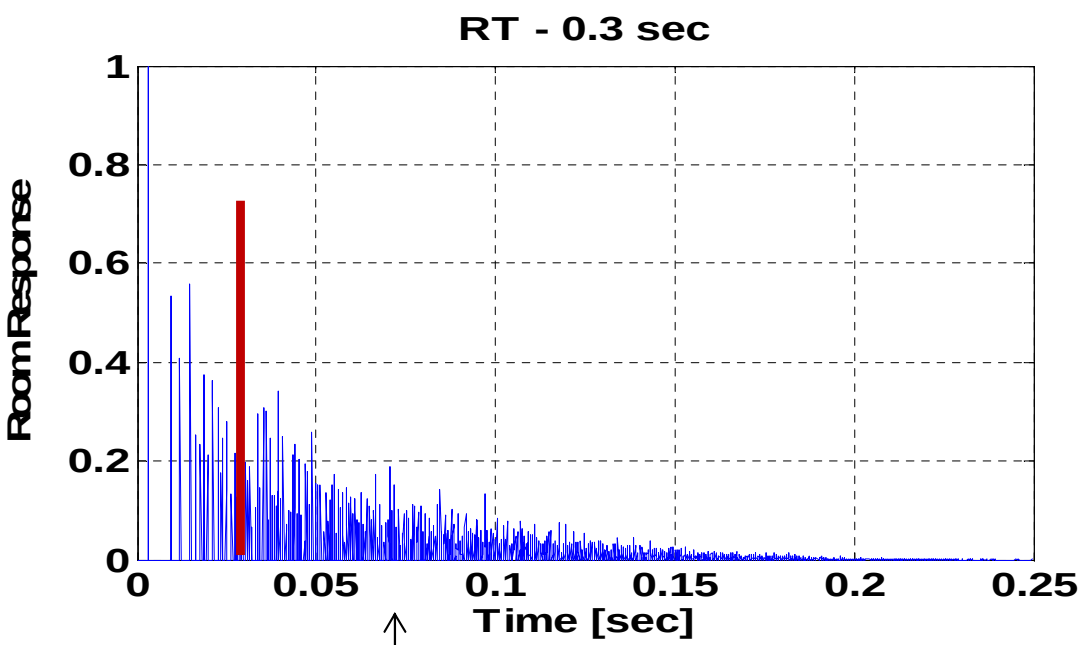


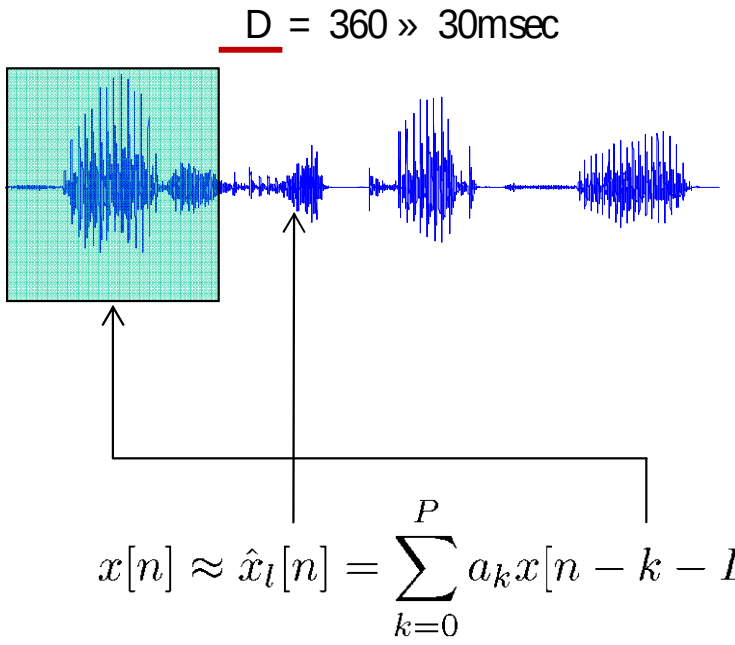
$$0.1z^4 \cdot \frac{2+z^{-1}-z^{-3}-2z^{-4}}{1-0.98z^{-1}}$$



- Filtering in modulation domain
- Speech lies in modulation frequency of 5-20 Hz
- Removes the DC component

Compensation in signal Domain: Suppression of Late Reverberation by MSLP





Spectral Subtraction

Sparsity Factor in NMF