Speech Dereverberation by Constrained and Regularized Multi-Channel Spectral Decomposition: Evaluated on REVERB Challenge

Audience

Microsoft Research 微软亚洲研究院

Meng Yu, Audience Inc. Frank K. Soong, Microsoft Research Asia

Overview

- Multi-channel speech dereverberation
- Cross-channel cancellation
- Spectral decomposition
- Convolution in spectral magnitude domain
- Generalized Kullback-Leibler divergence
- Constrained and regularized non-negative deconvolution
- Solved by a multiplicative algorithm
- Evaluations on "speech enhancement task"

Algorithm

Goal: Estimate the spectral magnitude of clean speech S

- Reverberated speech spectral magnitude X
- Actual observation Z by assuming noise
- $Z[n,k] \approx X[n,k] = S[n,k] * H[n,k]$,

Single-change solution:

By imposing the non-negativity and sparsity constraints (H. Kameoka et al, 2009 and K. Kumar, et al, 2011)

$$\{S,H\} = argmin_{S,H}.J$$

$$where\ J = \sum_i (Z[i,k] - \sum_m S[m,k]H[i-m,k])^2 + \lambda \sum_i S[i,k] \quad (1)$$

$$s.t.\ S[n,k] \geq 0, H[n,k] \geq 0, \sum_n H[n,k] = 1$$

Further motivation:

- Advantage of KL divergence for estimating small values in between spectral peaks compared to l_2 norm. $D(x|y) = x \log \frac{x}{y} - x + y$.

Algorithm *cont.*

- Cross-channel cancellation enforces the filters $H_i, i = 1,2,...,N$, to resolve the spatial difference between channels. The cross-channel cancellation error is to be minimized.

Multi-channel solution:

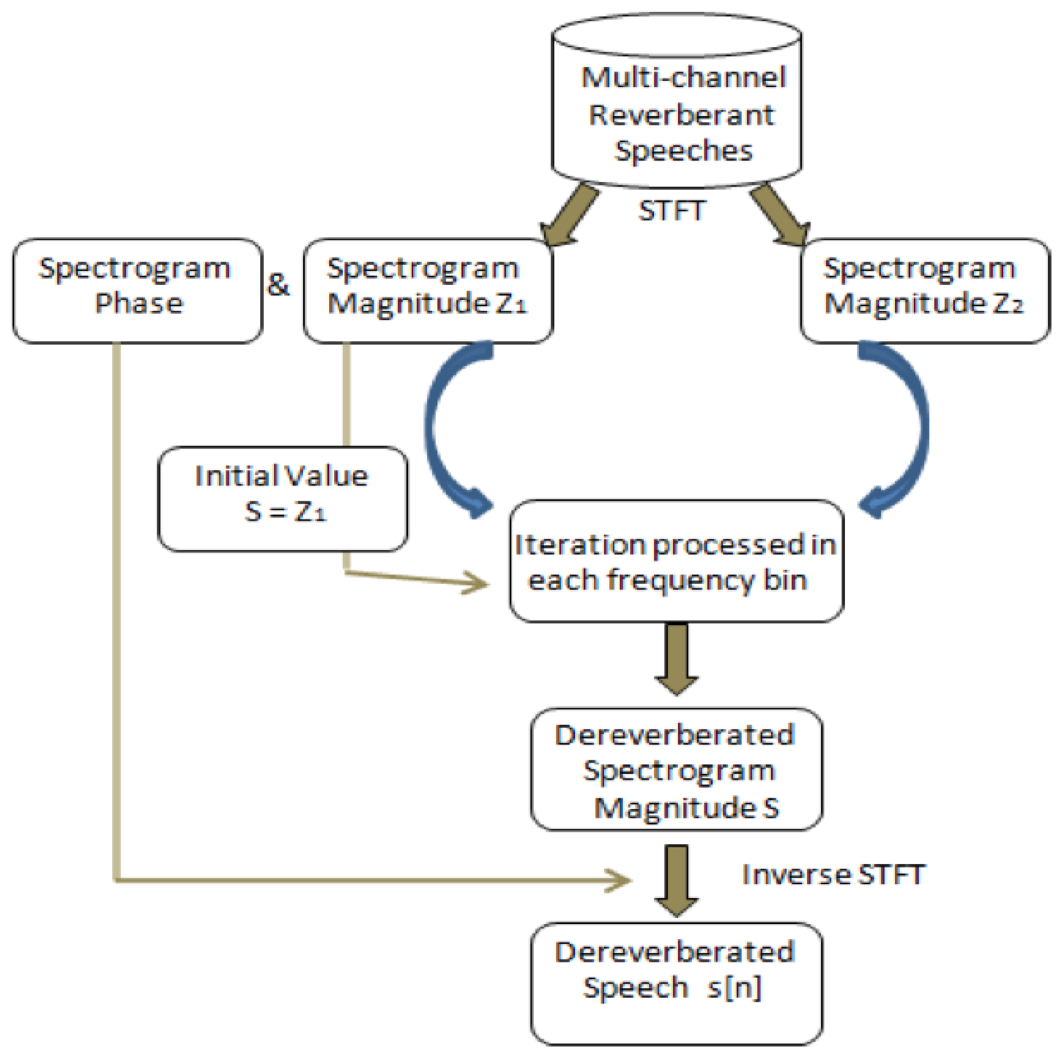
$$J = \sum_{j=1}^{2}\sum_i (Z_j[i,k] \log \frac{Z_j[i,k]}{\sum_m S[m,k]H_j[i-m,k]} - Z_j[i,k] + \sum_m S[m,k]H_j[i-m,k]) + \beta \sum_i (\sum_m X_1[m,k]H_2[i-m,k] - \sum_m X_2[m,k]H_1[i-m,k])^2 + \lambda \sum_i S[i,k] \quad (2)$$

$$s.t.$$

$$S[n,k] \geq 0, H_j[n,k] \geq 0, \sum_n H_j[n,k] = 1, j = 1,2$$

X_j is substituted by Z_j in the above model as the solution $\hat{S} * \hat{H}_j$ is expected to converge to Z_j. Easy to be extended to N-channel ($N > 2$).

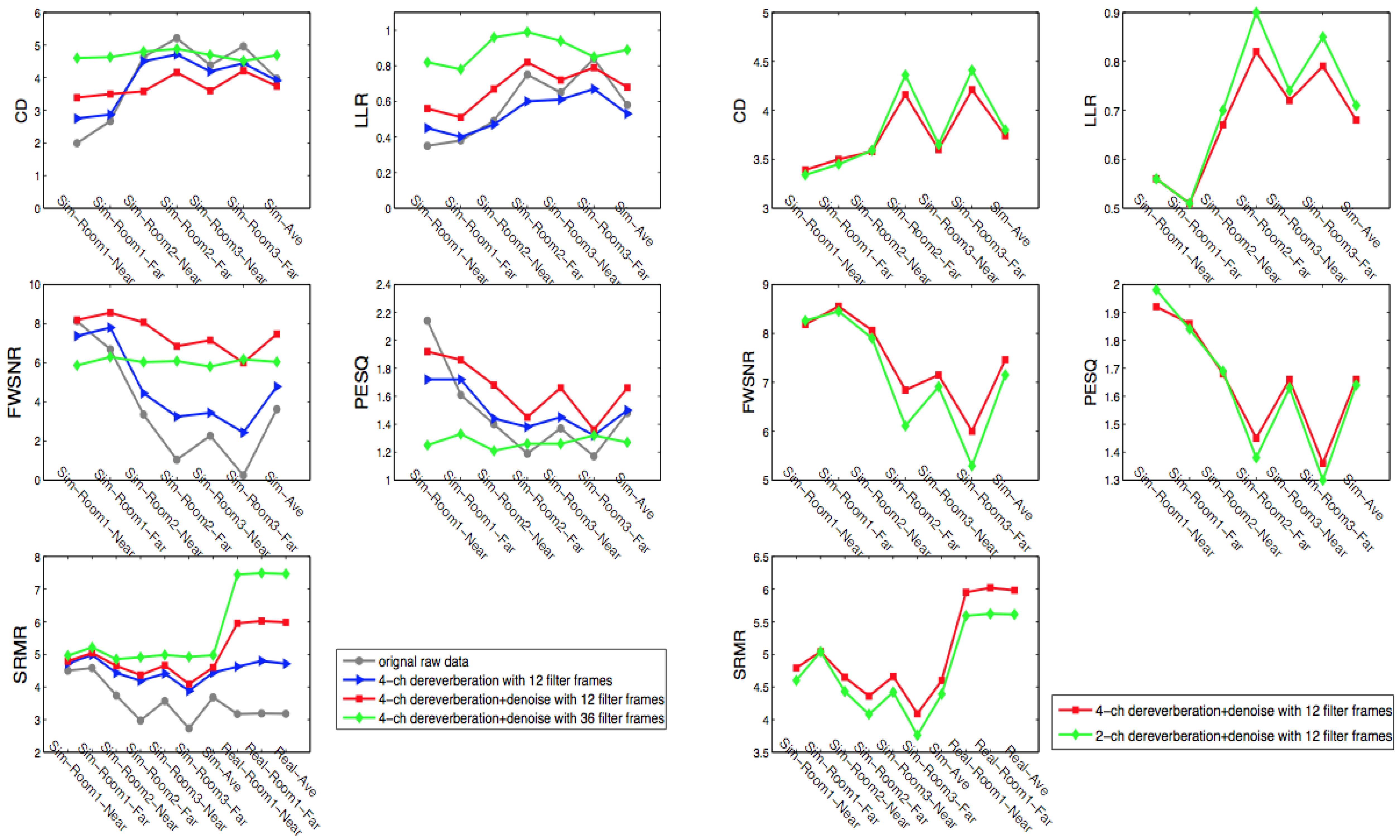
Diagram:



Experimental Evaluation

Channel: Select channel 1 and 2 for dual-channel processing, while 1, 2, 3, and 4 for 4-channel processing.

Post processing: A noise suppression post processing (optimally-modified log-spectral amplitude speech estimator (I. Cohen 2003) is applied to the dereverberated signal to suppress the background noise as an option.



Metrics Analysis

	CD	LLR	FWSNR	SRMR	PESQ
CD	-	0.15	0.66	0.58	0.69
LLR	0.15	-	-0.45	-0.49	-0.33
FWSNR	0.66	-0.45	-	**0.99**	**0.87**
SRMR	0.58	-0.49	**0.99**	-	**0.82**
PESQ	0.69	-0.33	**0.87**	**0.82**	-

- Consistent improvement of the proposed algorithm is proven by the metrics, such as SRMR, FWSNR and PESQ.
- The performance cross-correlation between those metrics is shown in the table, indicating SRMR, FWSNR and PESQ perform consistently.