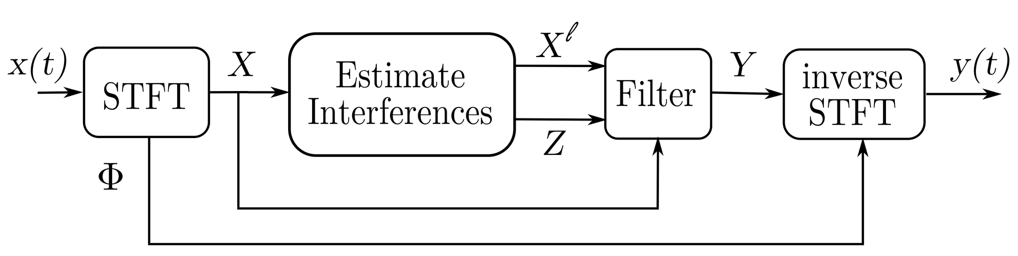
Reverberation Suppression based on sparse linear prediction in noisy environments

Nicolás LOPEZ[1,2], Gaël RICHARD[2], Yves GRENIER[2], Ivan BOURMEYSTER[1]

[1]Arkamys - Paris, France / [2]Institut Mines-Télécom ; Télécom ParisTech ; CNRS LTCI - Paris, France
nlopez@arkamys.com

Proposed Approach

A **single channel** late reverberation and noise suppression method is presented:



- Late reverberation estimated using frequency domain linear prediction with **sparse** constraints
- Background noise estimated when speech is absent
- **Blind processing** is assumed
- **Real time** method

Late Reverberation Estimation

- Observation model:

$$X_{k,n} = X_{k,n}^{early} + X_{k,n}^{late}$$

- Late reverberation model:

$$\hat{X}_{k,n}^{late} = \sum_{i=0}^{L-1} \alpha_{k,i} X_{k,n-i-\delta} = D_{k,n}\, \boldsymbol{\alpha_k}$$

L: model order, δ: delay

- Estimation with the LASSO :

$$\underset{\alpha_k}{\text{minimize}}\ ||X_{k,n} - D_{k,n}\, \boldsymbol{\alpha_k}||^2 \text{ s.t. } ||\boldsymbol{\alpha_k}||_1 \le \lambda$$

◇ **Sparse** prediction vector:

$$\boldsymbol{\alpha_k} = [\alpha_{k,0} \dots \alpha_{k,L-1}]^T$$

◇ **Signal-based** dictionary :

$$D_{k,n} = [\ X_{k,n-\delta}\ \ \dots\ \ X_{k,n-\delta-L+1}\ ] \quad (1)$$

- Solution with Least Angle Regression (LARS) algorithm
- Late reverberation *psd*:

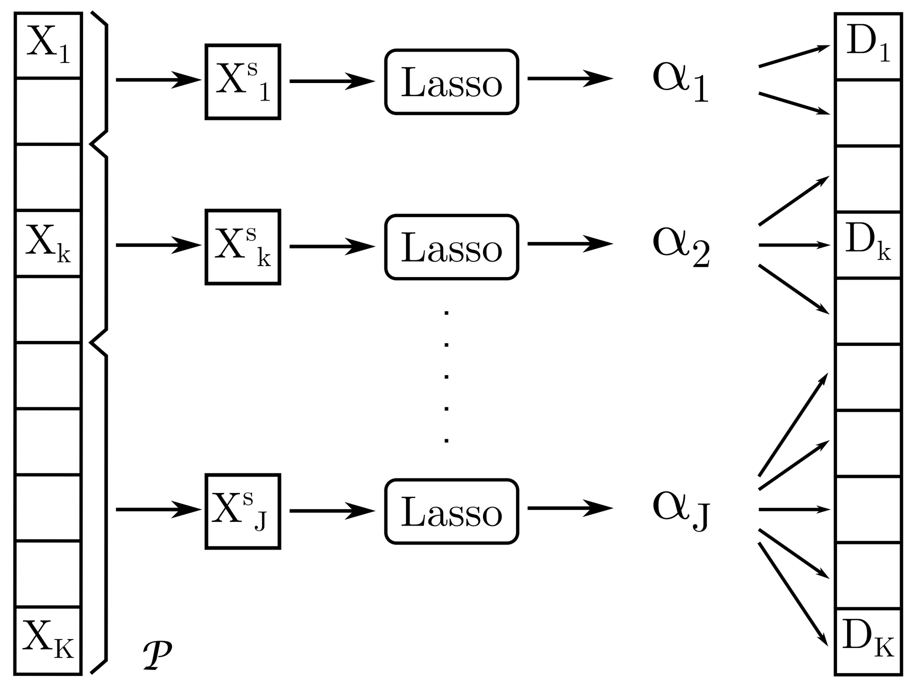
$$R_{k,n}^{late} = \beta_\ell\, R_{k,n-1}^{late} + (1-\beta_\ell)\, |\hat{X}_{k,n}^{late}|^2$$

Background Noise Estimation

- Use Voice Activity Detection with hard threshold
- Update noise *psd* if speech is absent : $Z_{k,n} = \beta_Z\, Z_{k,n-1} + (1-\beta_Z)|X_{k,n}|^2$
- If reverberation is high, it is likely to be estimated as noise ⇒ If $Z_{k,n} \approx R_{k,n}^{late}$: suppress reverberation only
- Filtering using the LSA estimator for multiple interferences

Reducing Complexity: Subband Gathering

- **Subsample spectrogram X from K to $J \ll K$ channels**
  - ◇ Define a J-segment partition $\mathcal{P}$ of $[1, K]$
  - ◇ Take average in each segment and obtain subsampled spectrogram X^s
  - ◇ Solve LASSO in each subsampled channel
- Estimate late reverberation using dictionary (1) and the J subsampled predictors, mapped to K channels



Reducing Complexity: Block-wise processing

- For each subsampled frequency j, **estimate one single predictor for N adjacent frames**:

◇ Define an observation vector:

$$V_{j,n} = [\ X_{j,n}^s\ \ \dots\ \ X_{j,n-N+1}^s\ ]^T$$

◇ Define a block-based dictionary:

$$D_{j,n}^V = [\ V_{j,n-\delta}\ \ \dots\ \ V_{j,n-\delta-L+1}\ ] \in \mathbb{R}^{N\times L}$$

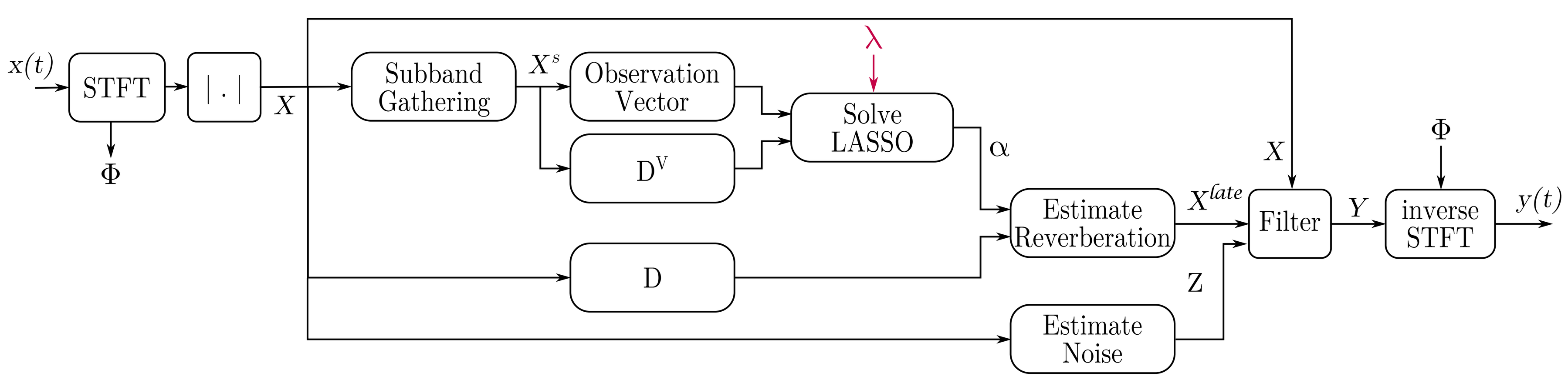
◇ Find j^{th} predictor and map to k channels:

$$\underset{\alpha_j}{\text{minimize}}\ ||V_{j,n} - D_{j,n}^V\, \boldsymbol{\alpha_j}||^2 \text{ s.t. } ||\boldsymbol{\alpha_j}||_1 \le \lambda$$

◇ Estimate late reverberation using dictionary (1) :

$$V_{k,n}^{late} = [\ X_{k,n}^{late}\ \ \dots\ \ X_{k,n-N+1}^{late}\ ]^T$$

Evaluation



Speech Enhancement Task: RTF : 9.41% on *SimData* and 9.29% on *RealData*

CD	Room 1		Room 2		Room 3		Ave.
	Near	Far	Near	Far	Near	Far	
Baseline	1.99	2.67	4.63	5.21	4.38	4.96	3.97
DRVNR	2.67	3.03	4.32	4.87	4.14	4.63	**3.94**
DRV	3.88	4.21	4.65	5.22	4.61	5.07	4.61
NR	4.45	4.82	4.41	5.35	4.86	5.64	4.92

SRMR	SimData							RealData		
	Room 1		Room 2		Room 3		Ave.	Room 1		Ave.
	Near	Far	Near	Far	Near	Far		Near	Far	
Baseline	4.50	4.58	3.74	2.97	3.57	2.73	3.68	3.17	3.19	3.18
DRVNR	6.96	8.19	6.59	7.21	6.23	6.28	6.91	7.40	7.68	7.54
DRV	5.91	6.39	5.83	5.94	5.70	5.77	5.92	9.05	8.83	**8.94**
NR	4.70	5.09	4.28	4.01	4.28	3.76	4.35	4.62	4.76	4.69

LLR	Room 1		Room 2		Room 3		Ave.
	Near	Far	Near	Far	Near	Far	
Baseline	0.35	0.38	0.49	0.75	0.65	0.84	0.58
DRVNR	0.42	0.45	0.51	0.72	0.67	0.81	**0.60**
DRV	0.79	0.83	0.81	1.02	0.94	1.06	0.91
NR	0.78	0.86	1.01	1.23	1.09	1.28	1.04

FWSNR	Room 1		Room 2		Room 3		Ave.
	Near	Far	Near	Far	Near	Far	
Baseline	8.12	6.68	3.35	1.04	2.27	0.24	3.62
DRVNR	6.47	6.29	4.05	2.91	3.51	2.42	4.27
DRV	4.95	4.63	5.16	3.90	4.62	3.54	**4.47**
NR	6.18	5.50	5.69	1.06	3.56	0.93	3.82

ASR Task: focus on DRVNR approach

WER	SimData							RealData		
	Room 1		Room 2		Room 3		Ave.	Room 1		Ave.
	Near	Far	Near	Far	Near	Far		Near	Far	
Baseline	12.93	17.72	24.03	72.54	30.46	79.72	39.53	83.16	84.48	83.81
AM_{clean}	17.54	22.42	24.04	45.60	30.78	56.92	32.87	74.58	71.71	73.14
AM_{mmc}	19.13	21.42	21.00	29.89	24.45	35.24	**25.35**	52.06	51.08	**51.57**

- AM_{clean}: Acoustic Model trained on clean data
- AM_{mmc} : Acoustic Model trained on Multi-Condition data processed with DRVNR approach.

- Good performance in far field and in big rooms
- Over subtraction in small rooms: consequence of blind processing