# Rotational Reset Strategy for Online Semi-supervised NMF-based Speech Enhancement for Long Recordings

*Jun Zhou*[1,2], *Shuo Chen*[2], *and Zhiyao Duan*[2]

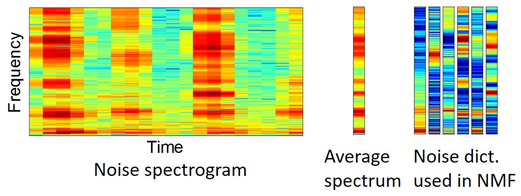
[1]Southwest University, Chongqing, China. • [2]University of Rochester, NY, USA.
*zhouj@swu.edu.cn* • *schen76@ur.rochester.edu* • *zhiyao.duan@rochester.edu*

## Introduction

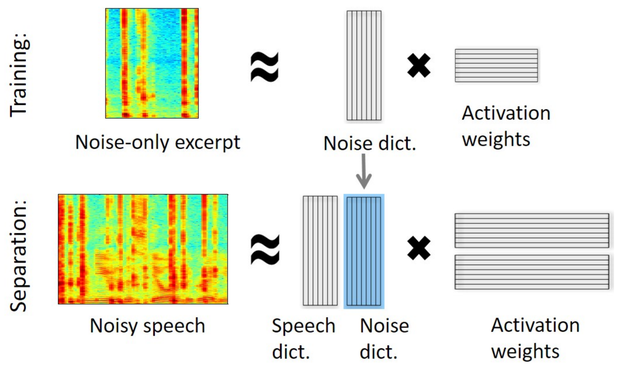
▶ Discover the **dictionary degradation problem** in online semi-supervised NMF-based speech enhancement approach.
▶ Proposes a **rotational reset strategy** for dictionary update to solve the degradation problem.
▶ Makes NMF-based approach truly outperform classical methods in **online semi-supervised** settings for **non-stationary noise**.
▶ First systematic experiment with long recordings (10-minute) for NMF-based speech enhancement.

## Online Semi-supervised NMF-based Speech Enhancement for Non-stationary Noise

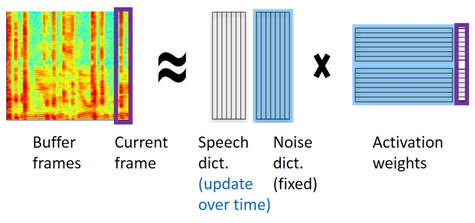
### ▶ NMF-based non-stationary noise modeling




### ▶ Semi-supervised algorithm [2]

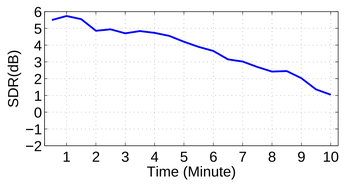



### ▶ Make it online [1]

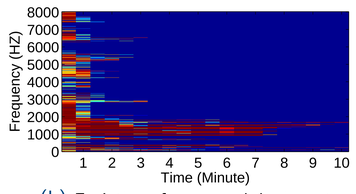



## The Speech Dictionary Degradation Problem

- The algorithm [1] works well when it was ran shorter than 1 minute.
- Severer degradation starts to happen after 2 minutes.
- The algorithm [1], and in fact all existing NMF-based methods, to our best knowledge, were evaluated using files shorter than 30 seconds.




(a) Enhancement performance over time.




(b) Evolution of one speech basis vector.

### ▶ Problem analysis

- Speech dictionary is updated using the multiplicative update rule.

$$P(f|z) \leftarrow \frac{1}{C_1}\Sigma_{s\in\mathcal{B}\cup\{t\}}V_{fs}P_s(z)\cdot P(f|z),\ \text{for } z\in\mathcal{S}, \qquad (1)$$

- Zero (close-to-zero) values will never be (be very slowly) updated.
- Once a value goes close to zero, it will be hard to come back.

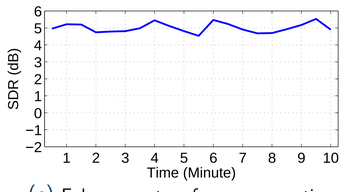
## Proposed Rotational Reset Strategy

- Rotationally reset $M$ elements of the speech dictionary in every $T$ seconds.
- Average reset rate is $M/T$.

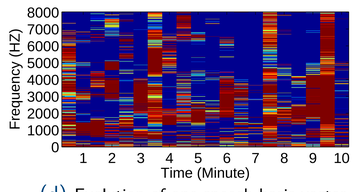
### ▶ Advantages

- Newly reset elements are recovering their abilities to adapt to new speech signal.
- Old elements keep the continuity of the dictionary to prevent sudden changes and performance drops in the enhanced speech signal.

### ▶ The degradation problem is now solved




(c) Enhancement performance over time.




(d) Evolution of one speech basis vector.

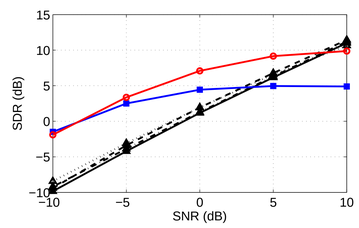
## Experiments

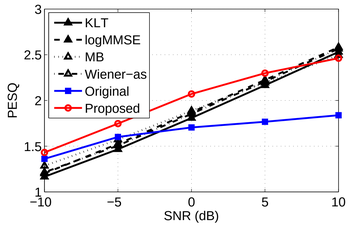
### ▶ Dataset

- 500 noisy speech files, 10 minute long each, sampled at 16 kHz.
- Each file is a concatenation of 10 speakers with alternating genders.
- 10 non-stationary noises: birds, casino, cicadas, computer keyboard, eating chips, frogs, jungle, machine guns, motorcycles, and ocean.
- 5 SNRs: -10, -5, 0, 5, 10 dB.

### ▶ Comparisons

- Four classical speech enhancement algorithms [3]
- Original algorithm in [1]




### ▶ Rotation parameter sensitivity analysis

| SDR (dB) | $M$=1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $T$=5s | 4.67 | 4.54 | 4.47 | 4.39 | 4.35 | 4.30 | 4.26 |
| 15s | **4.92** | **4.91** | 4.82 | 4.79 | 4.76 | 4.71 | 4.66 |
| 30s | 4.84 | **4.95** | **4.98** | **4.97** | **4.91** | **4.92** | 4.88 |
| 60s | 4.46 | **4.94** | **4.95** | **5.03** | **5.01** | **5.00** | **4.99** |
| 120s | 4.34 | 4.72 | 4.73 | 4.75 | 4.85 | **4.91** | **4.91** |
| 240s | 4.00 | 4.14 | 4.22 | 4.35 | 4.24 | 4.50 | 4.74 |

- Similar reset rates show similar enhancement performance.
- Performance is not sensitive to reset rate within the range of [1/30, 1/10] elements/second.
- Degradation appears if rate $< 1/40$ elements/second.
- Enhancement is not as good if rate $> 1/5$ elements/second.

**References**
[1] Z. Duan, G. J. Mysore, and P. Smaragdis, "Online PLCA for real-time semi-supervised source separation," in *Proc. Latent Variable Analysis and Signal Separation (LVA-ICA)*, 2012, pp. 34âĂŞ41.
[2] N. Mohammadiha, P. Smaragdis, and A. Leijon, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2140âĂŞ2151, 2013.
[3] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC press, 2013.