NMF-based Source Separation Utilizing Prior Knowledge on Encoding Vector

Kisoo Kwon*, Jong Won Shin† Nam Soo Kim*
*School of Electrical Engineering and INMC, Seoul National University, Seoul, Korea
†School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Korea
E-mail: kskwon@hi.snu.ac.kr, jwshin@gist.ac.kr, nkim@snu.ac.kr

Introduction

- Nonnegative matrix factorization (NMF) has shown impressive performance in the single channel source separation.
- In the training phase of NMF, the **encoding matrix** $\boldsymbol{H}^{train}$ is usually **discarded** after training.
- However, it bears useful information on how often each basis was utilized.
- In [K. Willson, 2008], the distribution of the logarithm of the encoding vector is modeled as a multivariate **Gaussian** distribution.
- Our analysis on $\boldsymbol{H}^{train}$ revealed that each row of this matrix was also highly sparse.
- In this paper, we **propose** the penalty terms based on the **prior knowledge on** $\boldsymbol{H}$ in the separation phase for NMF-based source separation.

NMF-based enhancement

- The magnitude spectra, KLD, Multiplicative update rule
- **In training phase** : obtain $\boldsymbol{W}_s$ and $\boldsymbol{W}_n$ from training DB.
- **In enhancement phase**

$$V(t) \approx WH(t) \qquad V(t) = |Y(t)|,$$
$$W = [W_s, W_n], \;\; H(t) = [H_s{}^T(t), H_n{}^T(t)]^T$$

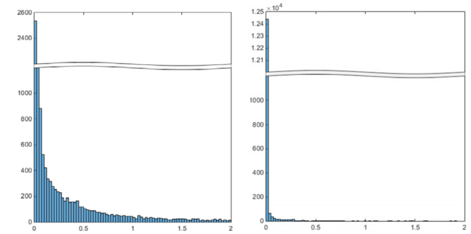
  - Update $H(t)$ with fixed $\boldsymbol{W}$ (# of max iter. =30)

$$H(t) \leftarrow H(t) \otimes \frac{\boldsymbol{W}^T(t)\frac{V(t)}{\boldsymbol{W}(t)H(t)}}{\boldsymbol{W}^T(t)\mathbf{1}}$$

  $\mathbf{1}$: *a square matrix of suitable size with all elements equal to one*

  - After obtaining $H(t)$
    $\hat{S}(t) = \boldsymbol{W}_s H_s(t), \quad \hat{N}(t) = \boldsymbol{W}_n H_n(t)$
  - Gain function: $G(t) = \frac{|\hat{S}(t)|^2}{|\hat{S}(t)|^2 + |\hat{N}(t)|^2}$
  - Enhanced signal at t-th frame
    $X(t) = G(t)Y(t)$

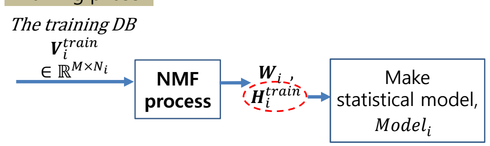
Utilizing prior knowledge of encoding vector

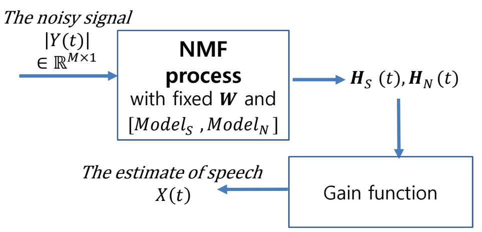


< The **histograms** of some rows of $\boldsymbol{H}_S^{train}$ corresponding to the most frequently and rarely used basis vectors.>

The shape of the histogram → **sparse distribution**
Sparse distributions → a **gamma** or an **exponential** distributions

$\boldsymbol{W}_i \in \mathbb{R}^{M \times r_i}$: the basis matrix of the source i
$\boldsymbol{H}_i^{train} \in \mathbb{R}^{r_i \times N_i}$: the encoding matrix of the source i
$\boldsymbol{V}_i^{train} \in \mathbb{R}^{M \times N_i}$: the training DB matrix of the source i
$\boldsymbol{H}_i\;(t) \in \mathbb{R}^{r_i \times 1}$: the encoding vector of the source i at t-th frame
$Model_i$: statistical model of $\boldsymbol{H}_i^{train}$





- The correlation coefficients among different components of the encoding vector were found not so significant. → apply **the independent exponential** or **gamma** distributions

- Employ the **gamma** distribution for $\boldsymbol{H}_i^{train}$
- The new objective function is given by

$$f(H) = D(V|WH) - \gamma_g \sum_{i=1}^{r} \left[(k_i - 1)\log H_i - \frac{H_i}{\theta_i}\right],$$

where k and θ indicate a shape and scale parameter, respectively.
- The MuR with KLD is now modified to

$$H_i \leftarrow H_i \frac{\sum_{k=1}^{M} \frac{W_{k,i} V_k}{\sum_{f=1}^{r} W_{k,f} H_f}}{\sum_{k=1}^{M} W_{k,i} + \gamma_g \left(\frac{1-k_i}{H_i} + \frac{1}{\theta_i}\right)}$$

- Employ the **exponential** distribution for $\boldsymbol{H}_i^{train}$
- The new objective function is given by

$$f(H) = D(V|WH) - \gamma_e \sum_{i=1}^{r} (\eta_i H_i),$$

where η indicates the rate parameter.
- The MuR with KLD is now modified to

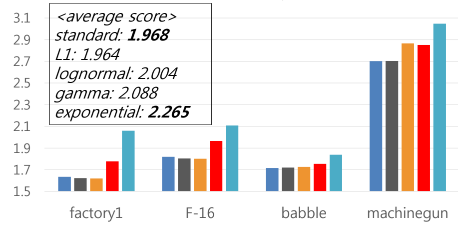
$$H_i \leftarrow H_i \frac{\sum_{k=1}^{M} \frac{W_{k,i} V_k}{\sum_{f=1}^{r} W_{k,f} H_f}}{\sum_{k=1}^{M} W_{k,i} + \gamma_g \eta_i}$$

Experiment

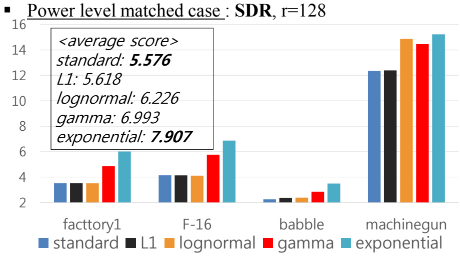
- Speech DB: TIMIT / noise DB : NOISEX-92
- 16kHz / 75% overlap / 512 FFT-size / r=128
- Measurement: **PESQ and SDR**

- The penalty terms used in the experiments were
  - ***standard***: no constraint to the separation phase
  - ***L1***: L_1 norm of L with $\eta_i = 1$
  - ***lognormal***: the negative log-likelihood of $logH$ assuming that H follows lognormal distributions where $logA$ denotes element-wise logarithm.
  - ***gamma***: the negative log-likelihood of H in which the PDF of H is modeled as an independent gamma distribution.
  - ***exponential***: the negative log-likelihood of H in which the PDF of H is modeled as an independent exponential distribution.

- Power level matched case : **PESQ**, r=128



- Power level matched case : **SDR**, r=128



- Power level **mismatched** case

distribution	Power level of the test data	PESQ	SDR
standard	original	1.9681	5.5763
lognormal	original	2.0043	6.2264
	+10dB	1.9904	6.1562
	-10dB	2.0052	6.0510
gamma	original	2.0878	6.9930
	+10dB	2.1717	7.4682
	-10dB	1.9320	4.0748
exponential	original	2.2648	7.9071
	+10dB	2.2415	7.9153
	-10dB	2.2188	7.8422

Conclusions

- We utilize the statistical information on the **encoding vector** obtained during the training.
- We propose an additional penalty term in the test phase.
  : based on a **sparse distribution** such as an **exponential** or a **gamma** distribution.
- Experiment results show that the proposed methods can **enhance the source separation performance**.