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 $H^{train}$ is usually discarded after training.
- However, it bears useful information on how often each basis was utilized.
- In [K. Wilson, 2008], the distribution of the logarithm of the encoding vector is modeled as a multivariate Gaussian distribution.
- Our analysis on $H^{train}$ 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 $H$ in the separation phase for NMF-based source separation.

NMF-based enhancement
- The magnitude spectra, KLD, Multiplicative update rule
- In training phase: obtain $W_t$ and $H_t$ from training DB.
- In enhancement phase:
  \[
  V(t) = WH(t) \quad V(t) = [v_1(t), \ldots, v_m(t)]^T
  \]
  \[
  W(t) = [W_1(t), \ldots, W_m(t)]^T
  \]
  \[
  H(t) \Rightarrow H(t) \otimes \frac{W(t)}{W(t)^T H(t)} I
  \]
  where $I$ is a square matrix of suitable size with all elements equal to one
- After obtaining $H(t)$
  \[
  \delta(t) = WH_t H(t)
  \]
  Gain function:
  \[
  G(t) = \begin{cases} 
  0 & \text{if } v < v_1 \\
  \frac{v - v_1}{v_m - v_1} & \text{if } v_1 < v < v_m \\
  1 & \text{if } v > v_m
  \end{cases}
  \]
  Enhanced signal at t-th frame
  \[
  \hat{x}(t) = G(t)v(t)
  \]

Utilizing prior knowledge of encoding vector
- The histograms of some rows of $H^{train}$ corresponding to the most frequently and rarely used basis vectors.
- The shape of the histogram is more gamma or an exponential distribution.
- $W_i \in \mathbb{R}^{m \times n_i}$: the basis matrix of the source $i$
- $H^{train}_i \in \mathbb{R}^{n_i \times n}$: the encoding matrix of the source $i$
- $H^{train}_i(t) \in \mathbb{R}^{n \times n_i}$: the encoding vector of the source $i$ at t-th frame
- $Model_1$: statistical model of $H^{train}$

Training phase
- The training DB:
  \[
  V(t) = W^{train} H^{train} \quad V(t) = [v_1(t), \ldots, v_m(t)]^T
  \]
  $W^{train}$ process
  Make statistical model, $Model_1$

Separation phase
- The noisy signal
  \[
  [y(t)] \in \mathbb{R}^m
  \]
  $Model_1$ process with fixed $W$ and $[Model_1, Model_2]$
  The estimate of speech
  \[
  \hat{x}(t) = G(t)y(t)
  \]
  Gain function

- 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 $H^{train}$
- The new objective function is given by
  \[
  f(H) = D(V|WH) - \gamma \sum_{i=1}^m (\hat{x}_i - 1) \log H_i - \frac{\alpha_i}{\beta_i}
  \]
  where $k$ and $\theta$ indicate shape and scale parameter, respectively.
- The MuR with KLD is now modified to
  \[
  H_1 \leftarrow H_1 \sum_{i=1}^m W_{ki}^2 w_{ki}^{\gamma_i} g^\gamma_i \left[ \frac{\alpha_i}{\beta_i} \right]
  \]

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: norm of $L$ with $\alpha = 1$
  - lognormal: the negative log-likelihood of log$H$ assuming that $H$ follows lognormal distributions where $\log a$ denotes element-wise logarithm.
    \[
    \text{gamma}: \text{the negative log-likelihood of } H \text{ in which the PDF of } H \text{ is modeled as an independent gamma distribution.}
    \]
  - exponential: the negative log-likelihood of $H$ in which the PDF of $H$ is modeled as an independent exponential distribution.

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.