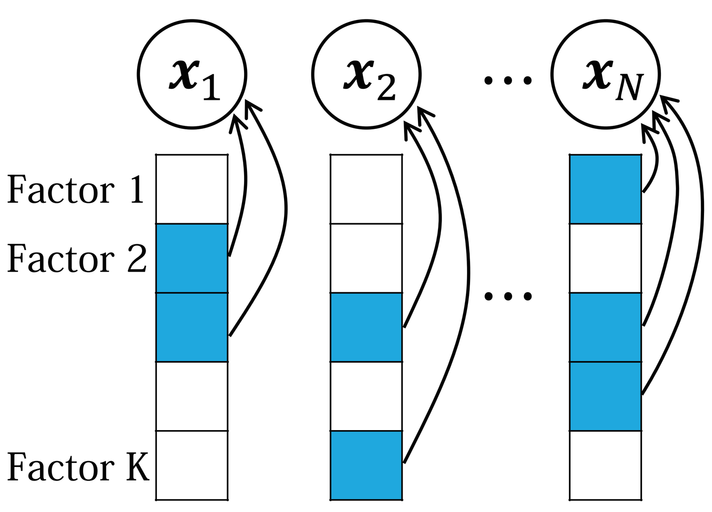
Infinite Probabilistic Latent Component Analysis for Audio Source Separation

Kazuyoshi Yoshii (Kyoto Univ. / RIKEN AIP)   Eita Nakamura   Katsutoshi Itoyama (Kyoto Univ.)   Goto Masataka (AIST)

Two Kinds of Latent Variable Models: Factor Models and Mixture Models

Factor models

Each sample is generated from multiple latent factors



Probabilistic models for decomposition

Example: PCA

Generate K basis vectors

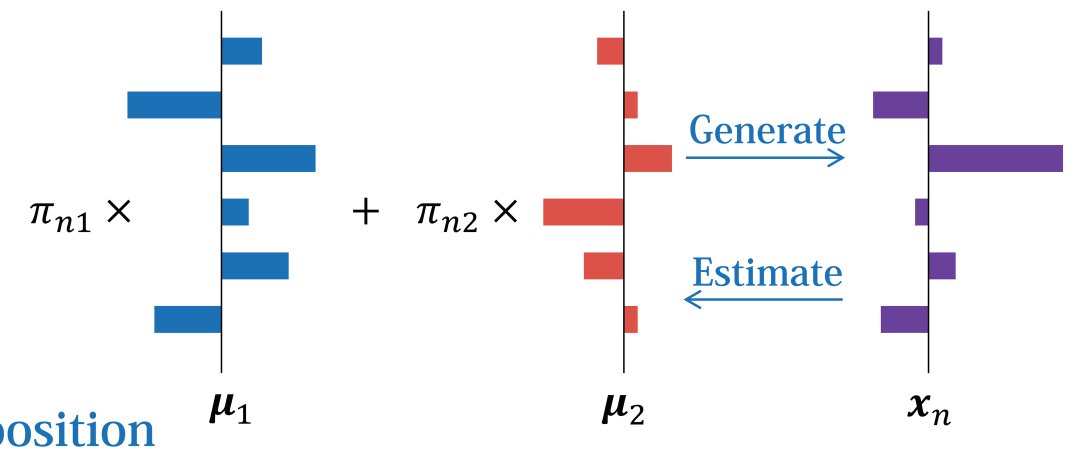
Parameters: $\boldsymbol{\theta} = \{\boldsymbol{\mu}_k\}_{k=1}^K$

Superimpose these basis vectors

Parameters: $\boldsymbol{\pi} = \{\boldsymbol{\pi}_n\}_{n=1}^N$

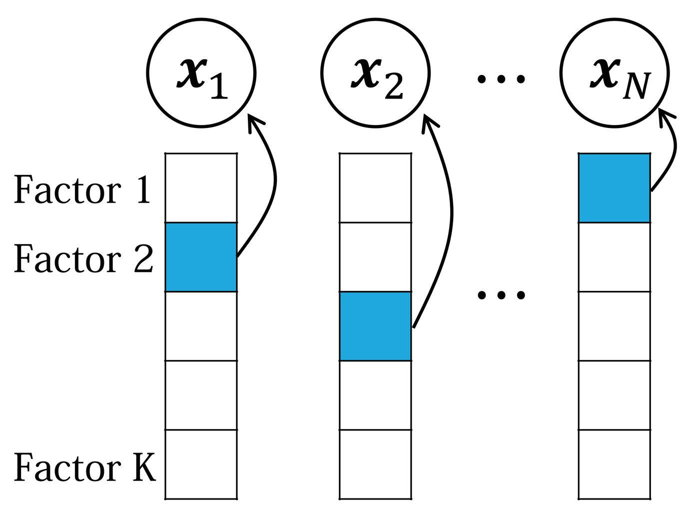
$$\boldsymbol{x}_n \sim N\left(\sum_{k=1}^K \pi_{nk}\boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right)$$

Inside sum operation: Sum of random variables



Mixture models

Each sample is generated from one of latent factors



Probabilistic models for clustering

Example: GMM

Generate K Gaussians

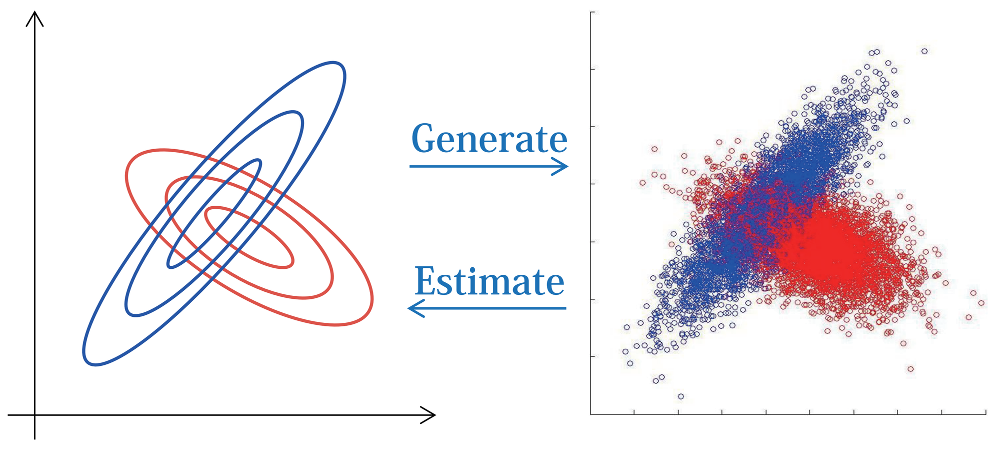
Parameters: $\boldsymbol{\theta} = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$

Superimpose these Gaussians

Parameters: $\boldsymbol{\pi} = \{\pi_k\}_{k=1}^K$

$$\boldsymbol{x}_n \sim \sum_{k=1}^K \pi_k N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Outside sum operation: Sum of prob. distributions



Nonnegative Matrix Factorization (NMF) and Probabilistic Latent Component Analysis (PLCA)

Factor models can be justified for source separation → NMF

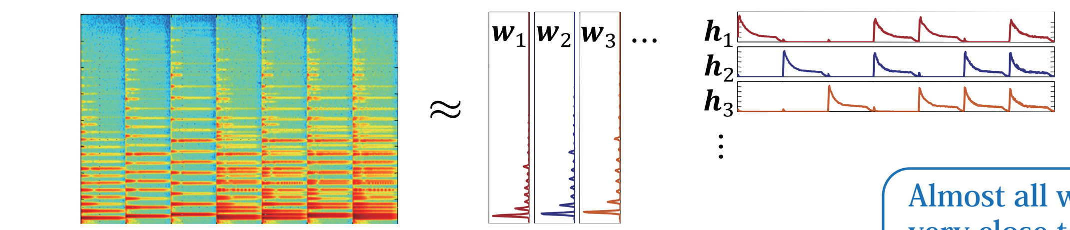
The observed mixture spectrum of each frame (sample) is given by the weighted sum of multiple source spectra

The mixture spectrum is decomposed into multiple source spectra

Gamma Process NMF (GaP-NMF) [Hoffman 2010]

Sparse learning of the nonnegative weights for infinitely many sources

Gamma process (GaP): A prob. dist. on the nonnegative weights

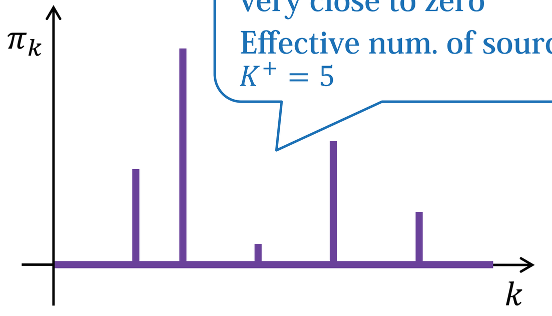


$$\boldsymbol{x}_n \sim \text{Poisson}\left(\sum_{k=1}^{\infty} \pi_k h_{kn}\boldsymbol{w}_k\right)$$

$$\boldsymbol{\pi} \sim \text{GaP}(\alpha)$$

When α is large K^+ tends to be large

Almost all weights are very close to zero. Effective num. of sources $K^+ = 5$

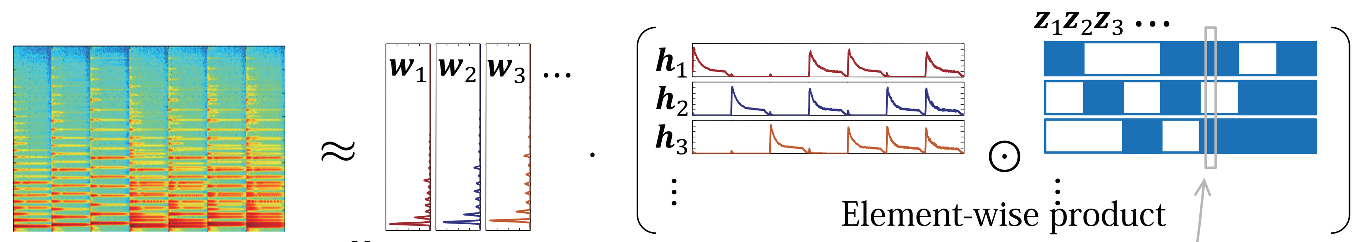


Beta Process NMF (BP-NMF) [Liang 2014]

Sparse learning of the binary activations for inifinitely many sources

Beta process (BP): A prob. dist. on the infinitely many coin-toss probs.

Bernoulli process (BeP): A prob. dist. on the binary activations

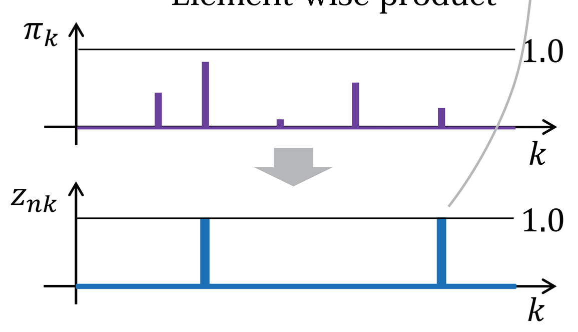


$$\boldsymbol{x}_n \sim \text{Poisson}\left(\sum_{k=1}^{\infty} z_{nk} h_{kn}\boldsymbol{w}_k\right)$$

$$\boldsymbol{\pi} \sim \text{BP}(\alpha)$$

$$\boldsymbol{z}_n \sim \text{BeP}(\boldsymbol{\pi})$$

When α is large K^+ tends to be large

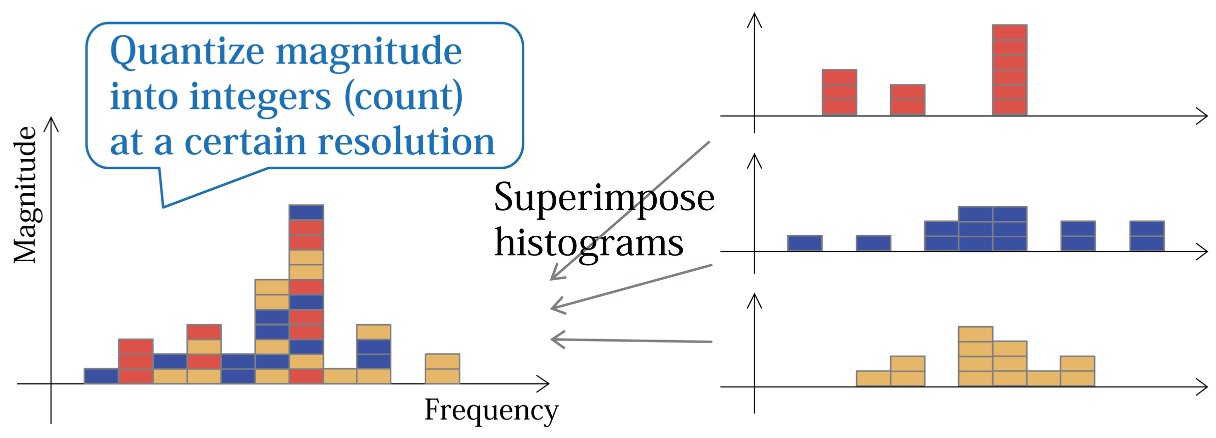


Mixture models have also been used → PLCA

Tricky interpretation
Mixture spectrum = Histgoram of sound quanta (samples)

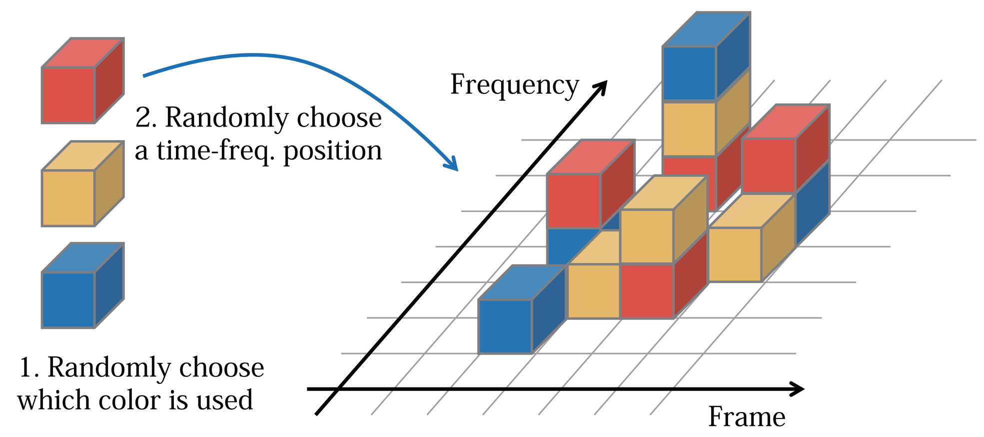
Each sample is categorized into one of sources

The spectrum as a whole can include multiple sources



The resolution of quantization shoud be appropriately specified in PLCA and KL-NMF (the resolution can be optimized [Hoffman 2012])

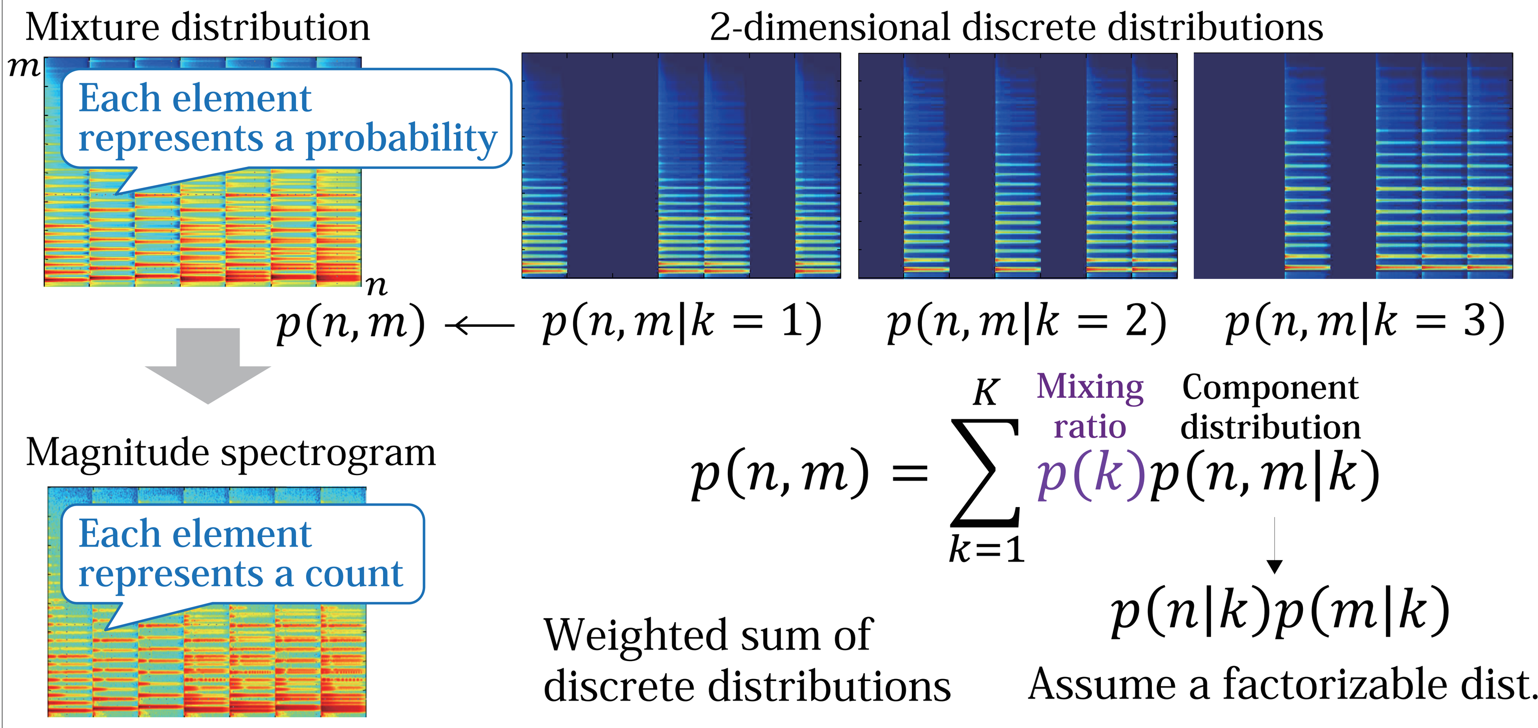
Each sound quantum is stochastically located in the time-frequency domain



Dirichlet Process PLCA (DP-PLCA) [Proposed Method]

Sparse learning of the normalized nonnegative weights for infinitely many sources

Dirichlet process (DP): A prob. dist. on the normalized nonnegative weights



$p(n,m) \longleftarrow p(n,m|k=1) \quad p(n,m|k=2) \quad p(n,m|k=3)$

$$p(n,m) = \sum_{k=1}^{K} p(k) p(n,m|k)$$

Mixing ratio — Component distribution

Weighted sum of discrete distributions

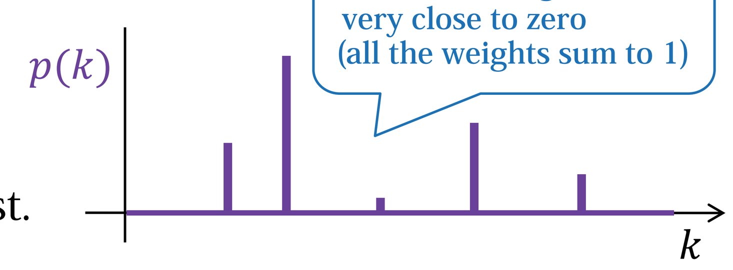
$p(n|k)p(m|k)$ — Assume a factorizable dist.

Dirichet distribution: $p(m|k) \sim \text{Dirichlet}(\gamma)$

Dirichlet distribution: $p(n|k) \sim \text{Dirichlet}(\beta)$

Dirichlet process: $p(k) \sim \text{DP}(\alpha)$

Almost all weights are very close to zero (all the weights sum to 1)



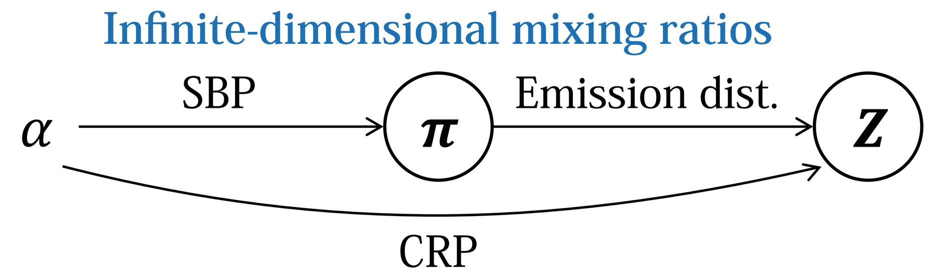
The hierarchical Dirichlet process (HDP) is not necessary unlike latent Dirichlet allocation (LDA)

Bayesian Inference for DP-PLCA Based on Collapsed Gibbs Sampling

Finite truncation is required	Gamma process (GaP)	Beta process (BP)	Dirichlet process (DP)
Weak-limit approximation	Infinitely many gamma dist.	Infinitely many beta dist.	Infinitely many Dirichlet dist.
Stick-breaking process	Complicated	Well-known	Well-known
Restaurant representation	Complicated	Indian Buffet process (IBP)	Chinese restaurant rocess(CRP)

Finite truncation is NOT required

Infinite-dimensional mixing ratios



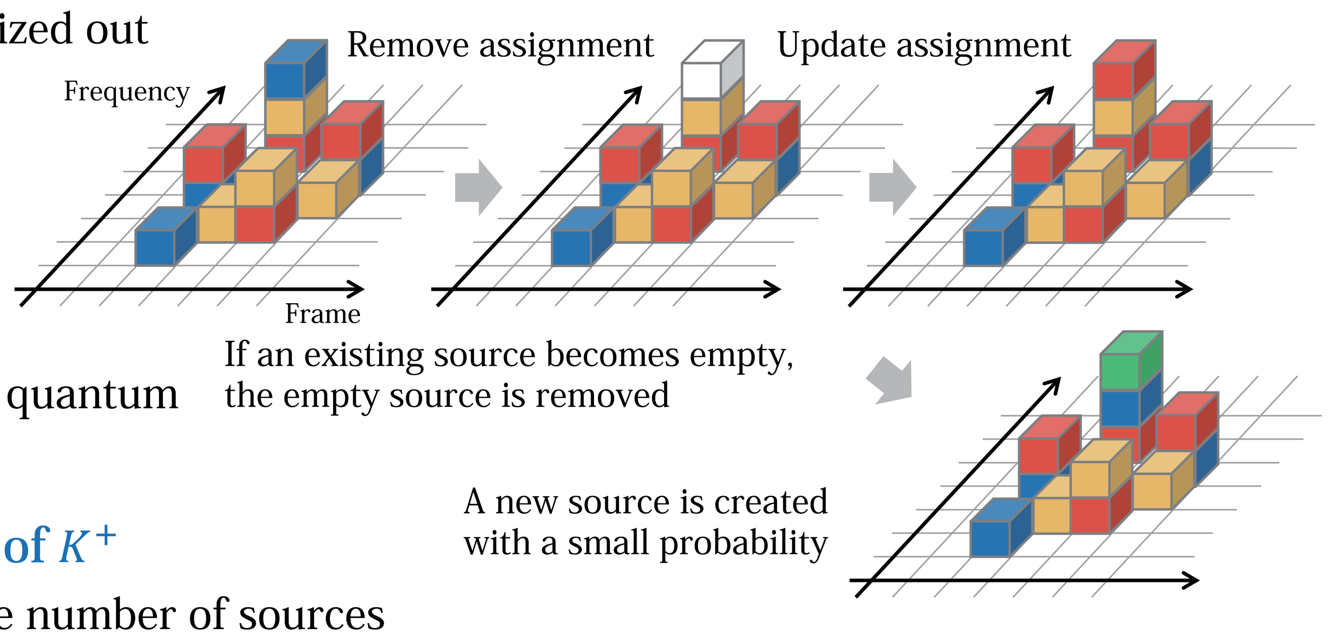
The latent variables (source assignment of individual sound quanta) can be estimated while estimating the effective number of mixtures (sources) K^+ without finite truncation

All the parameters can be marginalized out

Using the CRP representation, $p(k)$ is marginalized out

Using the conjugacy of the priors, $p(n|k)p(m|k)$ is marginalized out

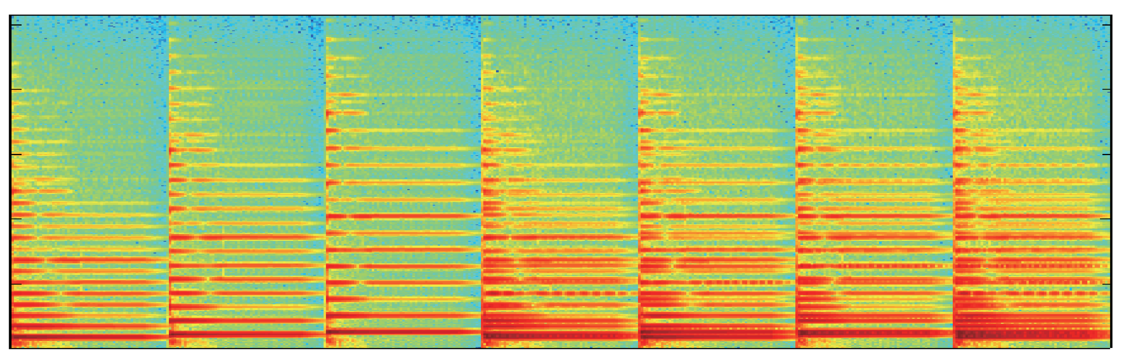
Only the assignment of each sound quantum is updated iteratively



Application of DP-PLCA to simple mixture signal

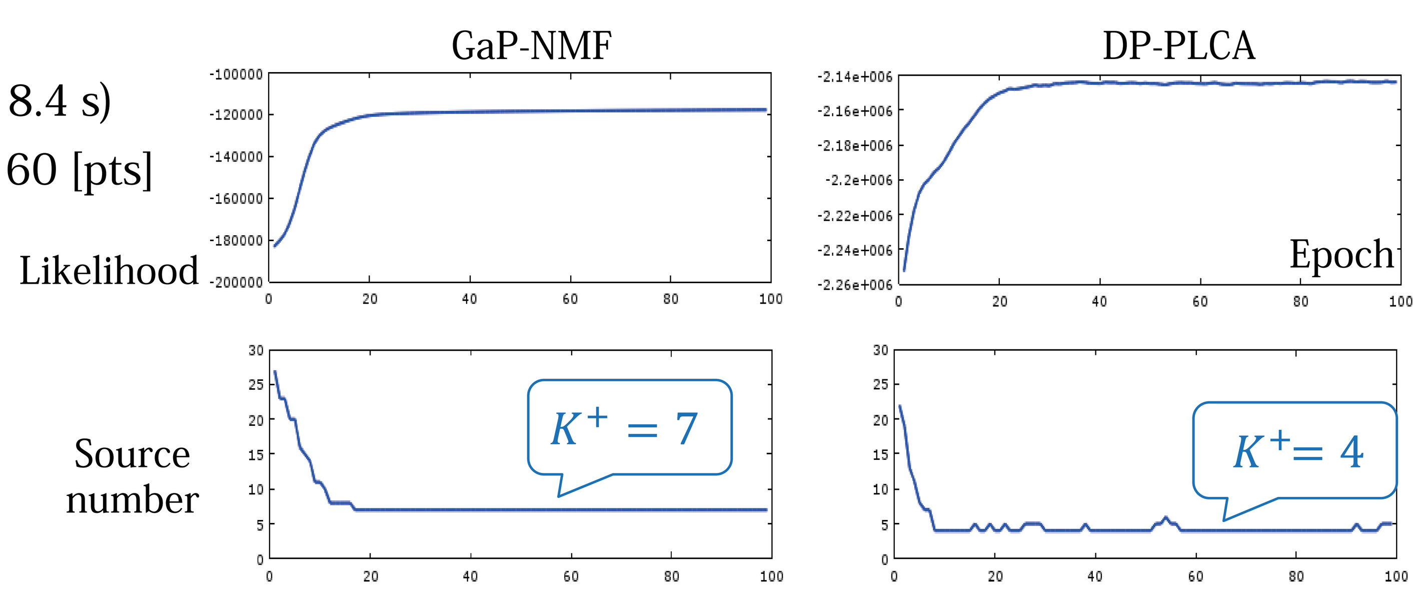
A mixture signal consisting of three sources (C, E, G) (1.2 s x 7 = 8.4 s)

16 [kHz]/16 [bits] mono, window size: 512 [pts], window shift: 160 [pts]



DP-PLCA tends to work better in estimation of K^+

GaP-NMF based on VB tends to overestimate the number of sources



Future Work

Optimize the resolution of quantization as in Poisson-uniform NMF [Hoffman 2012]

Investigate the connection to KL-NNF based on the Poisson distribution

Use more complicated priors that represent harmonic or percussive structures

Investigate physical meanings of sound quanta