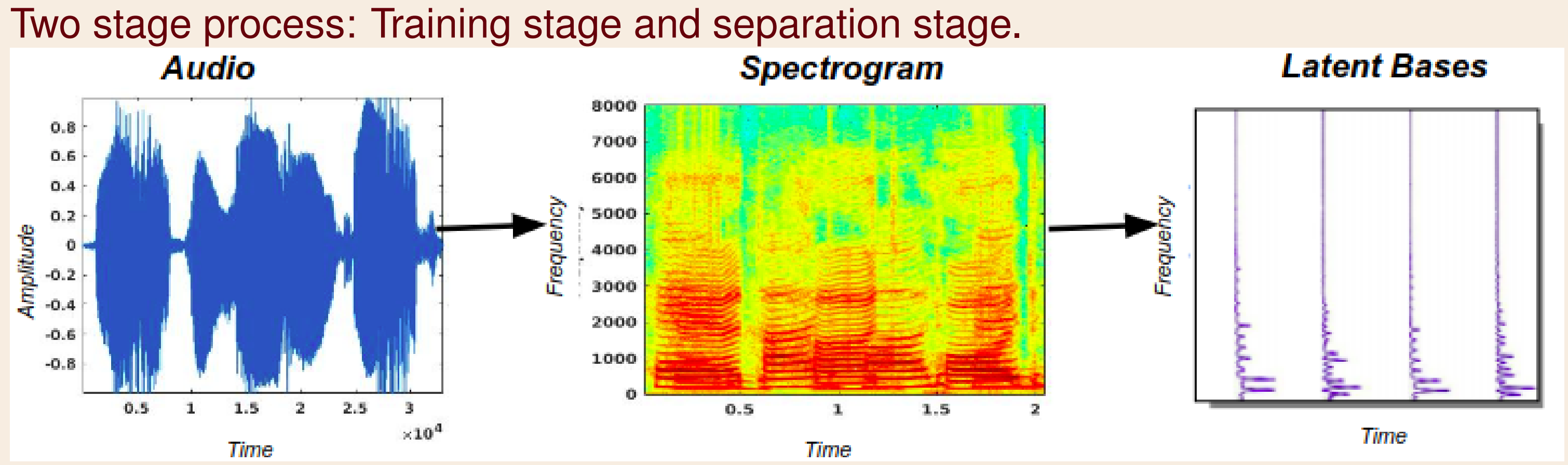
# A DYNAMIC LATENT VARIABLE MODEL FOR SOURCE SEPARATION

*Anurendra Kumar*[1], *Tanaya Guha*[1], *Prasanta Ghosh*[2]

[1]Indian Institute of Technology,Kanpur [2]Indian Institute of Science, Bangalore

## Introduction

- **Latent Variables**: Unobserved variables that explain the observed variables.
- **Supervised source separation**: Assumes small training data ( 15 sec) for each source.
  Popular methods: Latent Variable Model (LVM) and the Nonnegative Matrix Factorization (NMF).
  Two stage process: Training stage and separation stage.



- The latent bases for each sources are utilized to separate the sources.
- Latent variable models assumes mixture multinomial as likelihood and can be seen as probabilistic counterpart of non-negative matrix factorization.
- **Dynamic Modeling**: LVM and NMF assumes no temporal correlation in spectrogram. In past, exponential distribution as a dynamic prior [1]. Imperative to use Dirichlet as a prior since it is conjugate to multinomial. However, Dirichlet in its basic form yields negative updates

## Proposed model

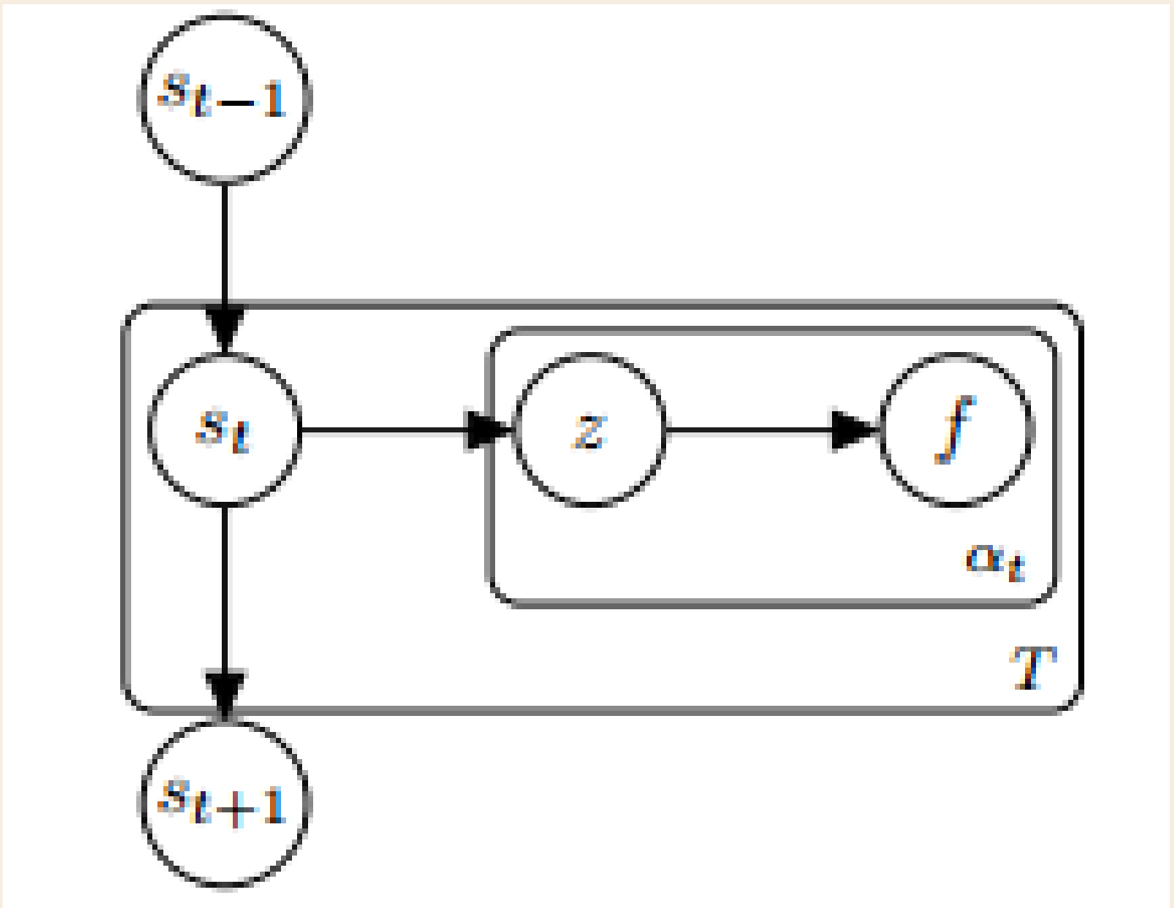


Figure : Plate notation for the proposed dynamic DLVM.

- $x(t) \Rightarrow$ signal, $\mathbf{X} \Rightarrow$ spectrogram, $\mathbf{N} \Rightarrow$scaled spectrogram

$$\mathbf{N} = \gamma|STFT(x(t))| = \gamma \mathbf{X} \tag{1}$$

- $\mathbf{N}$ as a surrogate of $\mathbf{X}$ for all analysis.
- Each count of frequency modeled as a mixture multinomial,

$$P_t(f) = \sum_{k=1}^{K} P_t(f, z_k) = \sum_{k=1}^{K} P_t(z_k) P(f|z_k) \tag{2}$$

- Let state $\mathbf{s}_t$

$$\mathbf{s}_t = [P_t(z_1), ..., P_t(z_K)]^{\mathsf{T}} = [s_t(1), s_t(2)..., s_t(K)]^{\mathsf{T}} \tag{3}$$

- We impose a Markovian dependence between states, which follows a dynamic Dirichlet distribution.

$$P(\mathbf{s}_t|\mathbf{s}_{t-1}, \mathbf{D}) = \mathrm{Dir}(\alpha_{t-1}\mathbf{D} \ \ \mathbf{s}_{t-1} + \mathbf{1}) \tag{4}$$

$$\text{where, } \ \alpha_t = \sum_f \mathbf{N}_{ft}, \ P(\mathbf{s}_1) = \mathrm{Dir}(\mathbf{1})$$

$\mathbf{D}_{kk} = d_k$, $1 \le k \le K$, where $d_k \in \mathbb{R}^+$ denotes the temporal dependence between two consecutive time instants for the $k$-th latent basis.

- Let pseudo-observation for each basis $k$ as $\mathbf{m}_{tk} = \alpha_{t-1}\mathbf{d}_k\mathbf{s}_{t-1}(k)$

## Properties of Dynamic Dirichlet Distribution

- Spectrogram at time $t$ is modeled as count data over $K$ bases. The dynamic Dirichlet prior allows us to have $\mathbf{m}_{tk}$ extra pseudo-observations for each basis $k$ .
- Variance of each entry decreases as total number of observations at previous time instant increases.

$$Var(\mathbf{s}_t(k)|\mathbf{s}_{t-1}) \propto \frac{1}{(\sum_k \mathbf{m}_{tk} + K)^2 (\sum_k \mathbf{m}_{tk} + K + 1)}$$

- PLCA as a special case when no temporal dependence i.e. $\mathbf{D}$=0

## Dynamic DLVM as dynamic version of NMF

- The EM algorithm can be viewed as a dynamic NMF algorithm.
- $\mathbf{W}_{fk} = P(f|\mathbf{z}_k)$; $\mathbf{S}_{kt} = P_t(z_k)$
- $\mathbf{X}_{F\times T} = \mathbf{W}_{F\times K}\mathbf{S}_{K\times T}\mathbf{G}_{T\times T} = \mathbf{W}_{F\times K}\mathbf{H}_{K\times T}$;

**Algorithm 1** Dynamic DLVM as Dynamic NMF

```
Input: $\mathbf{X}$
Output: $\mathbf{W}, \mathbf{S}, \mathbf{d}$
Randomly initialize $\mathbf{W}, \mathbf{S}, \mathbf{d}$
while Not converged do
    $\mathbf{W}_{fk} = \mathbf{W}_{fk} \sum_t \frac{\mathbf{X}_{ft}}{(\mathbf{WS})_{ft}} \mathbf{S}_{kt}$
    $\mathbf{W}_{fk} = \mathbf{W}_{fk} / \sum_k \mathbf{W}_{fk}$
    while Not converged do
        $\mathbf{m}_{tk} = \alpha_{t-1}\mathbf{d}_k\mathbf{s}_{t-1}(k)$
        $\mathbf{S}_{kt} = \mathbf{S}_{kt} \sum_f \mathbf{W}_{fk} \frac{\mathbf{X}_{ft}}{(\mathbf{WS})_{ft}} + \mathbf{m}_{tk}$
        $\mathbf{S}_{kt} = \mathbf{S}_{kt} / \sum_t \mathbf{S}_{kt}$
        Update $\mathbf{d}$
    end
end
```

## Experimental Setup

- Speaker source separation and Speech noise separation.
- **Speaker source separation:** Around 25 seconds of speech (8 to 9 sentences) from 10 speakers (5 male, 5 female) from TIMIT. First 17 seconds for training. Tested on 45 synthetic mixtures by digitally adding the speech from two speakers.
- **Speech noise separation:** Five noise types: Babble, Factory, White, Pink and Cockpit.
- Evaluation Metric: Signal to noise ratio improvement (SNRI), Source to Distortion ratio (SDR), Source to interference ratio (SIR), Source to Artifact ratio (SAR). SDR, SIR and SAR are perceptual metrics.

## References

1. N. Mohammadiha, P. Smaragdis, G. Panahandeh, and S. Doclo, A state-space approach to dynamic nonnegative matrix factorization, IEEE Transactions on Signal Processing, vol. 63, no. 4, pp. 949959, 2015
2. P. Smaragdis, B. Raj, and M. Shashanka, A probabilistic latent variable model for acoustic modeling, NIPS, vol. 148, pp. 81, 2006.
3. N. Mohammadiha, P. Smaragdis, and A. Leijon, Prediction based filtering and smoothing to exploit temporal dependencies in NMF, in ICASSP. IEEE, 2013, pp. 873877.

## Results & Discussion

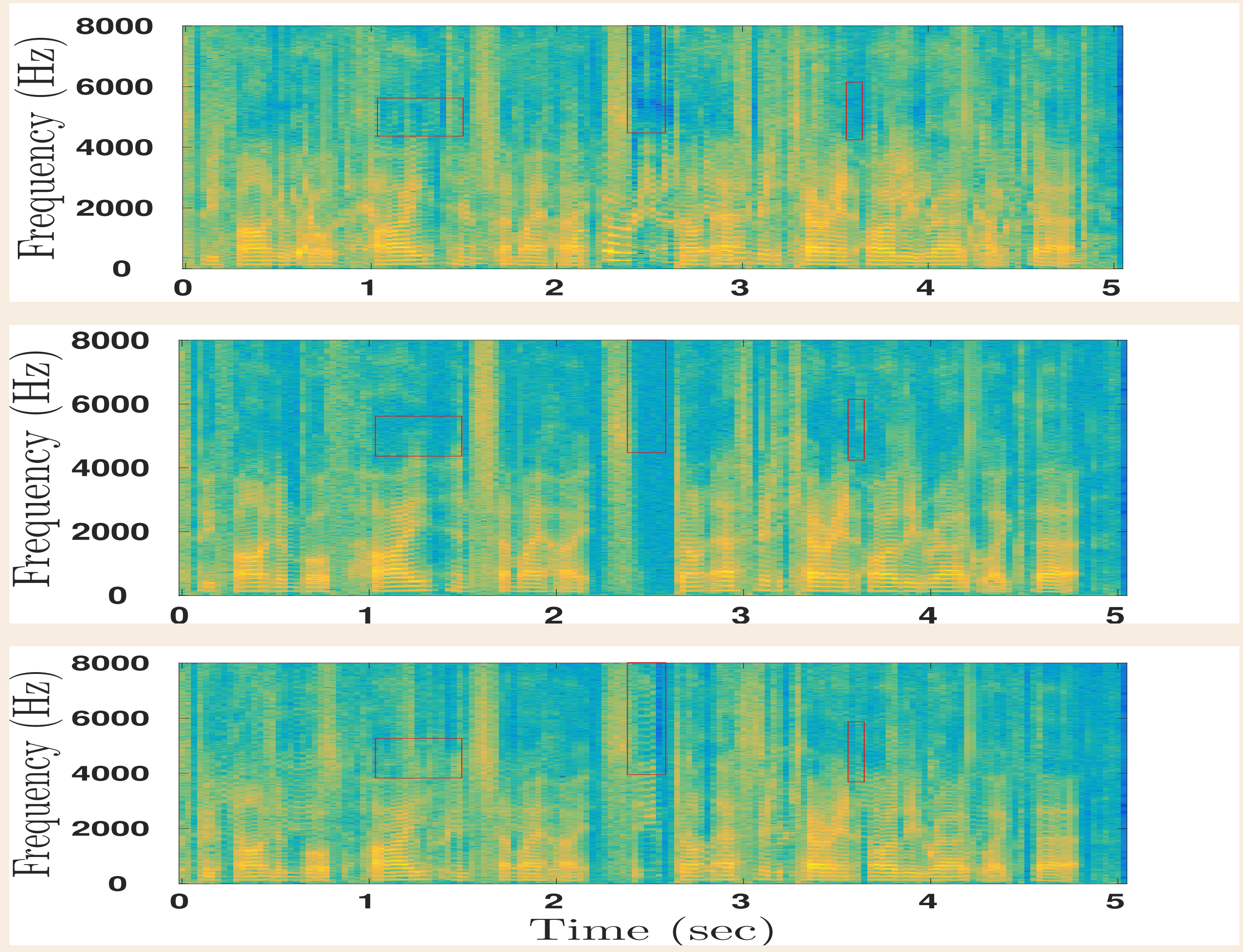


Figure : Original source, recovered source using PLCA, and recovered source using dynamic DLVM. Dynamic DLVM recovers a smoother spectrogram (areas of significant differences are highlighted).

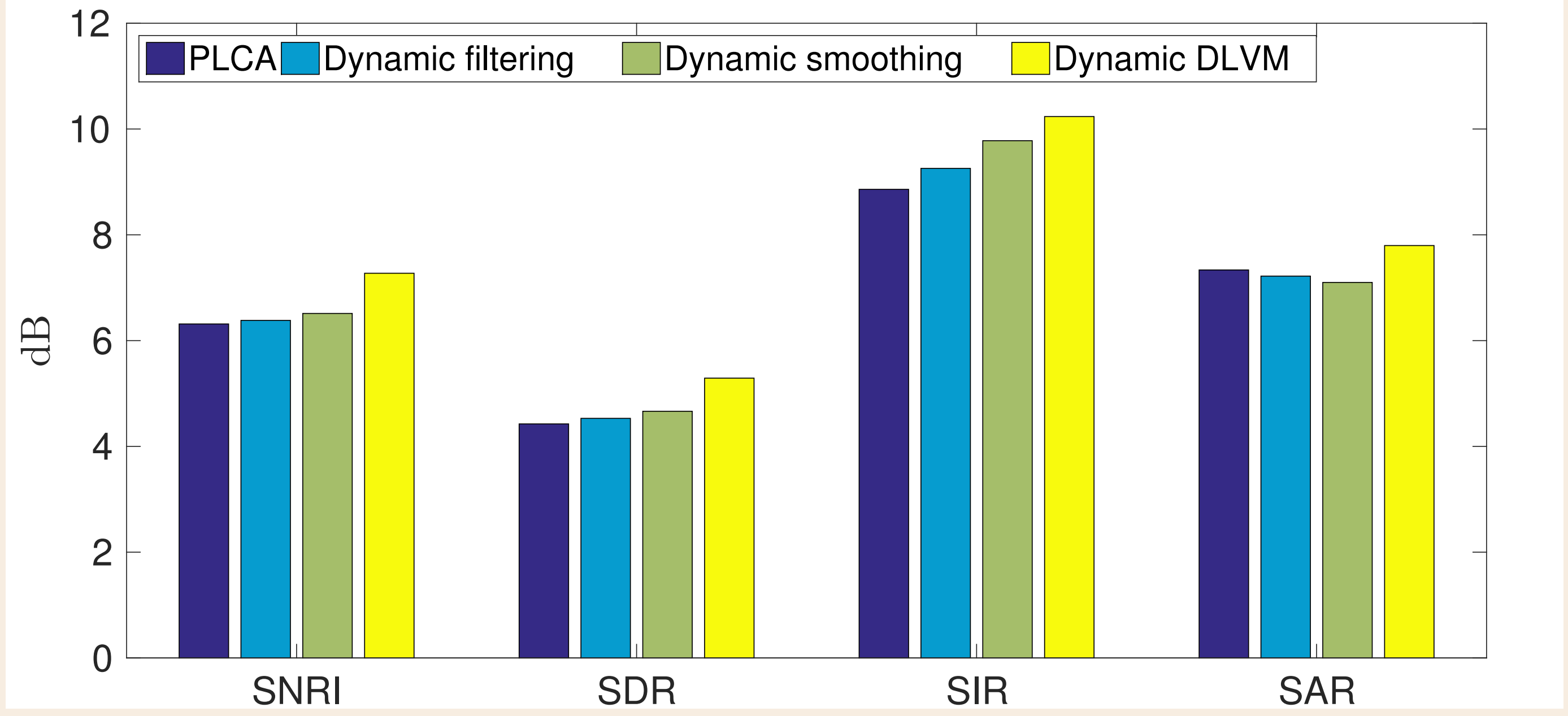


Figure : Results on speaker separation: Dynamic DLVM compared with three existing techniques in terms of four evaluation metrics. Our model outperforms PLCA by $0.96$ dB in SNRI, $0.87$ dB in SDR, $1.38$ dB in SIR, and $0.46$ dB in SAR.

Table : Comparison of different methods for noise separation

| | Average SNRI | | | | |
|---|---|---|---|---|---|
| | Babble | Factory | White | Pink | Cockpit |
| PLCA [2] | 5.63 | 2.60 | 5.07 | 2.04 | 2.78 |
| Dynamic filtering [3] | 4.93 | 2.87 | **5.83** | 2.06 | 2.70 |
| Dynamic smoothing [3] | 4.30 | 2.99 | 5.36 | 2.14 | 2.38 |
| **Dynamic DLVM** | **5.83** | **5.30** | 3.90 | **4.60** | **3.03** |
| | Average SAR | | | | |
| PLCA [2] | 6.69 | 8.14 | 8.30 | 7.82 | 7.84 |
| Dynamic filtering [3] | 6.44 | 7.73 | 5.25 | 5.97 | 4.36 |
| Dynamic smoothing [3] | 5.65 | 7.73 | 3.98 | 7.44 | 3.21 |
| **Dynamic DLVM** | **7.22** | **8.75** | **9.92** | **8.66** | **9.13** |

## Conclusion

1. Proposed a dynamic Dirichlet distribution particularly suitable for dynamic non-negative data.
2. Dynamic DLVM can be interpreted as dynamic NMF.
3. Our model does not require any free parameter apart from K.

https://anurendra.github.io/Homepage/

anu.ankesh@gmail.com, tanaya@iitk.ac.in,prasantg@iisc.ac.in