**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.

**Properties of Dynamic Dirichlet Distribution**

- Spectrogram at time \( i \) is modeled as count data over \( K \) bases. The dynamic Dirichlet prior allows us to have \( w_k \) extra pseudo-observations for each basis \( k \).
- Variance of each entry decreases as total number of observations at previous time instant increases.
- PLCA as a special case when no temporal dependence i.e. \( D=0 \).

**Dynamic DLVM as dynamic version of NMF**

- The EM algorithm can be viewed as a dynamic NMF algorithm.
- Dynamic DLVM as dynamic version of NMF.

**Algorithm 1 Dynamic DLVM as Dynamic NMF**

**Input**: \( X \)

**Output**: \( W, S, d \)

Randomly initialize \( W, S, d \) while Not converged do

- \( W_k = W_0 \sum_i X_{ft}(W_k)_s = W_0 S_k \)
- \( W_k \)

while Not converged do

- \( m_k = \alpha_k^{-1} S_k(1) \)
- \( S_d = S_0 \sum_i W_i (W_k)_s + m_k \)
- \( S_d \)

end $d$

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.

**Results & Discussion**

- **Table 1**: Comparison of different methods for noise separation.

<table>
<thead>
<tr>
<th>Evaluation Metric</th>
<th>Babble</th>
<th>Factory</th>
<th>White</th>
<th>Pink</th>
<th>Cockpit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average SNRI</td>
<td>5.63</td>
<td>2.60</td>
<td>5.07</td>
<td>2.04</td>
<td>2.78</td>
</tr>
<tr>
<td>Dynamic filtering</td>
<td>4.93</td>
<td>2.87</td>
<td>5.83</td>
<td>2.06</td>
<td>2.70</td>
</tr>
<tr>
<td>Dynamic smoothing</td>
<td>4.30</td>
<td>2.99</td>
<td>5.36</td>
<td>2.14</td>
<td>2.38</td>
</tr>
<tr>
<td>Dynamic DLVM</td>
<td>5.83</td>
<td>5.30</td>
<td>4.90</td>
<td>4.60</td>
<td>3.03</td>
</tr>
</tbody>
</table>

**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 \( K \).