# Improving melody extraction using Probabilistic Latent Component Analysis

Jinyu. Han[1] Ching-Wei. Chen[2]

[1]Interactive Audio Lab
Northwestern Univsersity, USA

[2]Media Technology Lab
Gracenote, Inc

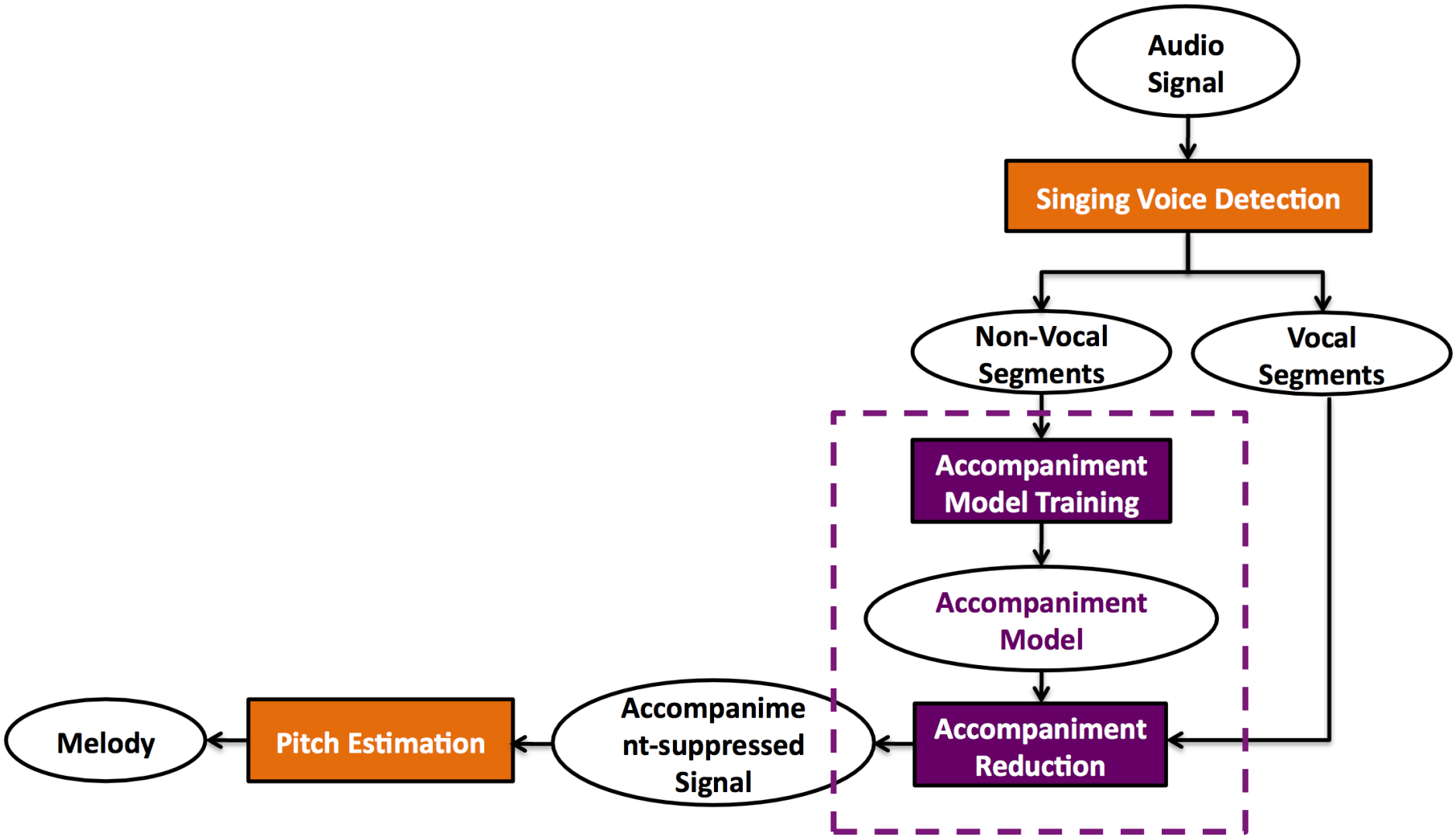
May 19, 2011

# Agenda

1. Introduction
2. Modeling the Spectrogram
   - Multinomial Model
   - Probabilistic Latent Component Analysis
3. System Description
4. Experiment Results
   - Illustration Example
   - System Comparison
5. Conclusion

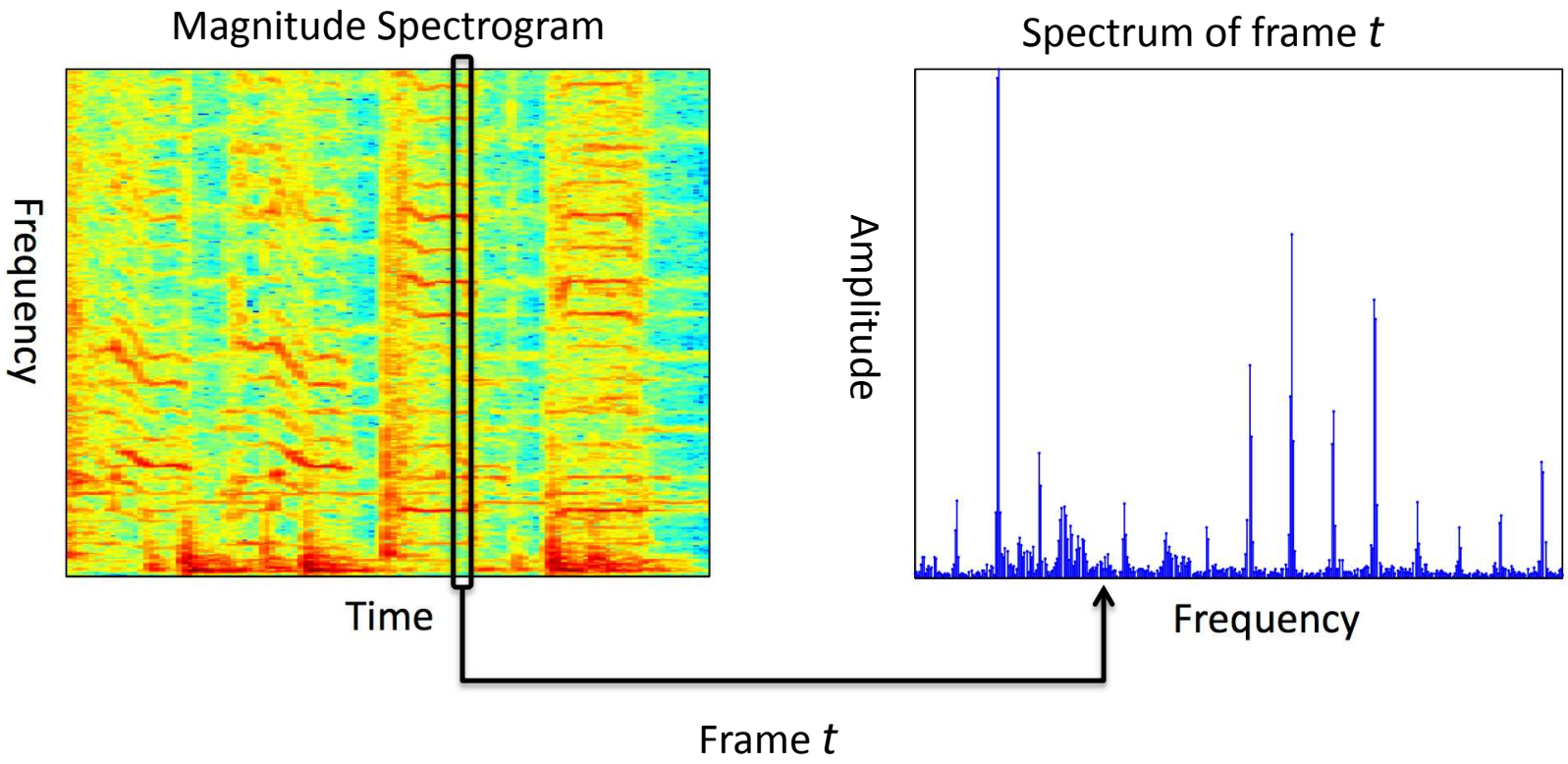
# Pick only the singing voice as the Melody

# System Overview



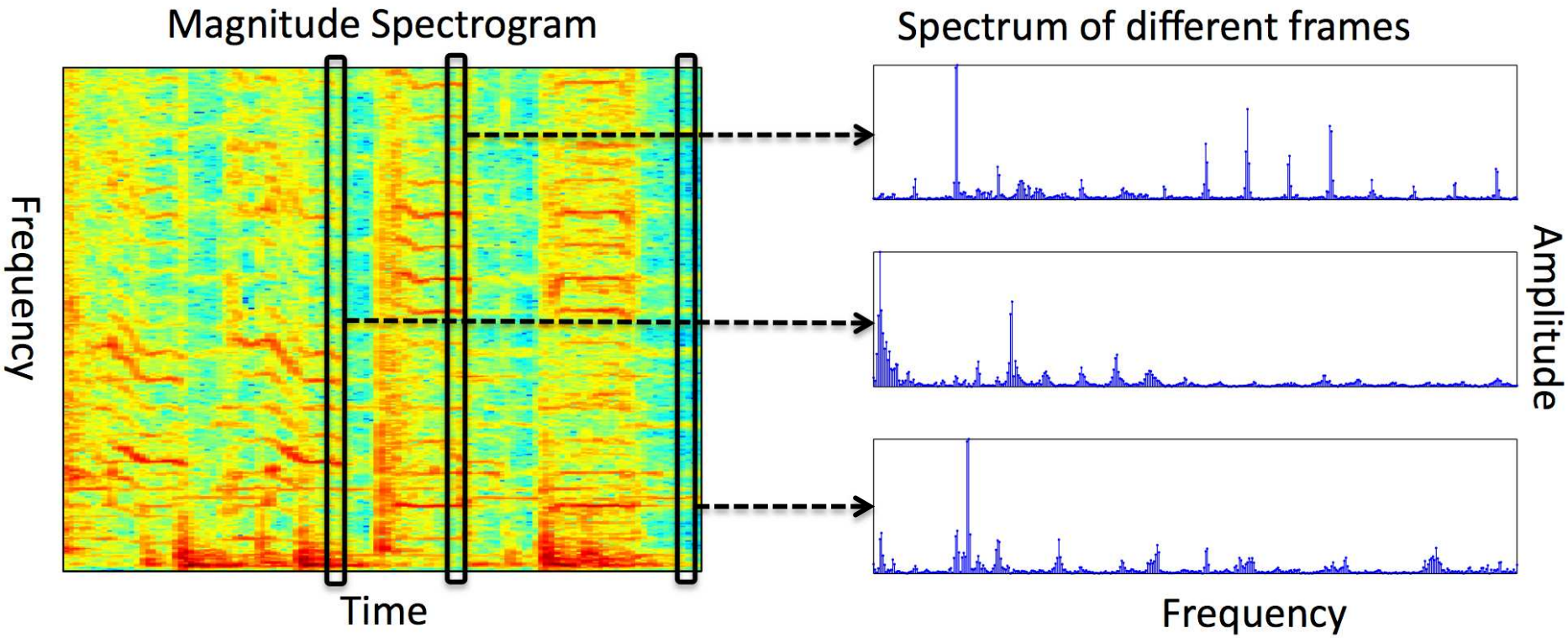
# Multinomial Distribution for Spectrogram

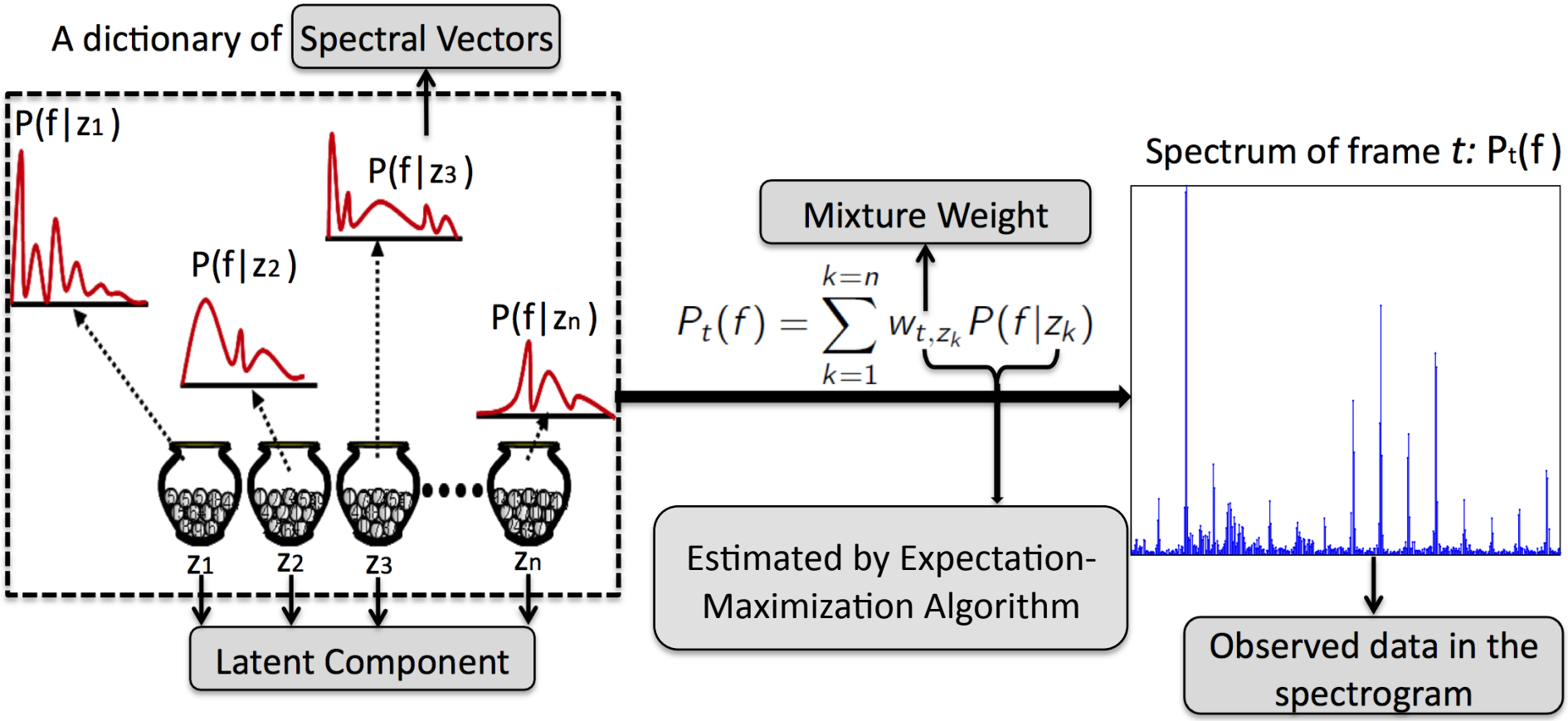
Figure: Probability distribution underlying the $t$-th spectrum



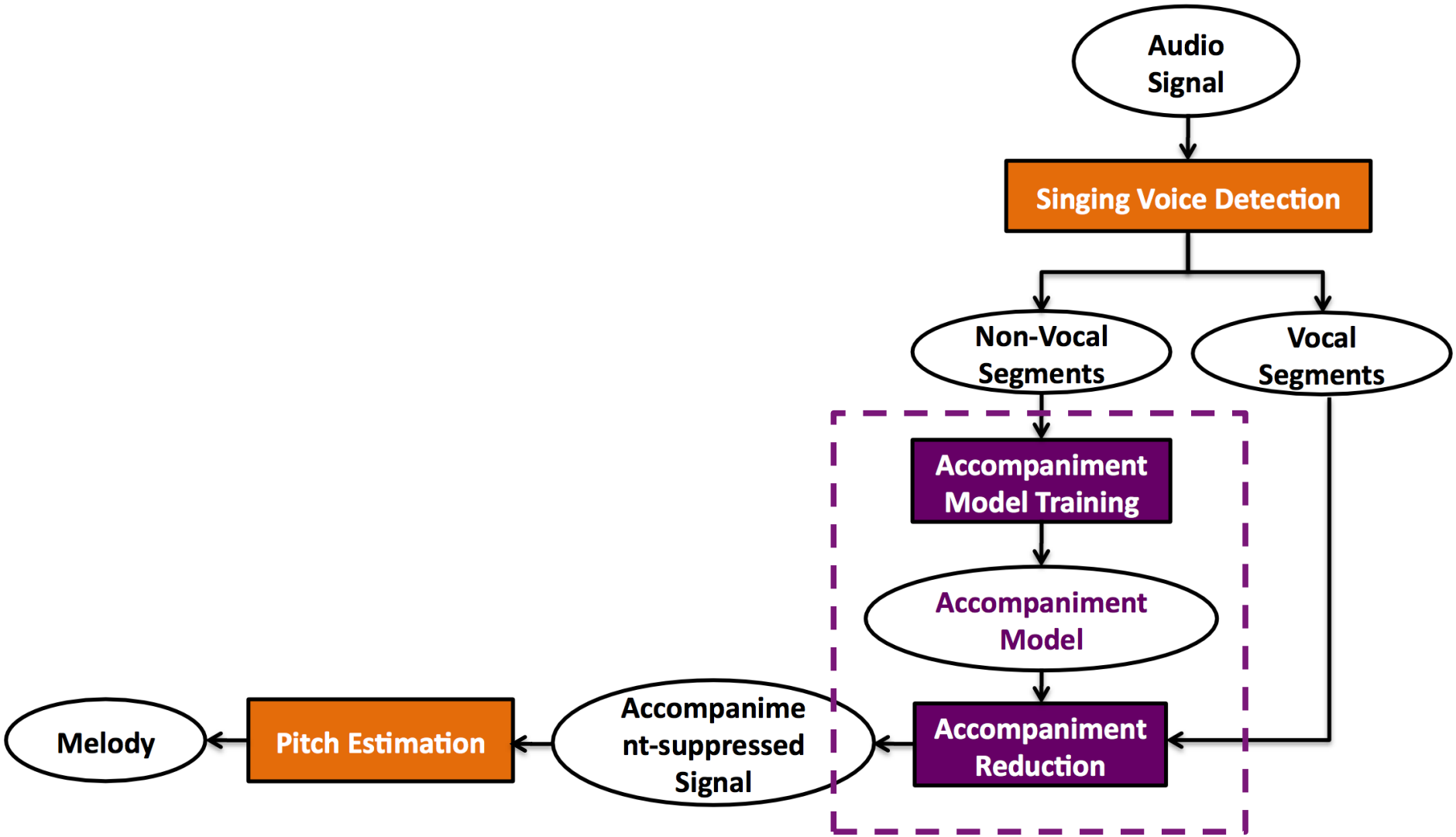

- Treat the spectrum in each time slice as a histogram
- Treat the histogram as a probability distribution

# Multinomial Distribution for Spectrogram

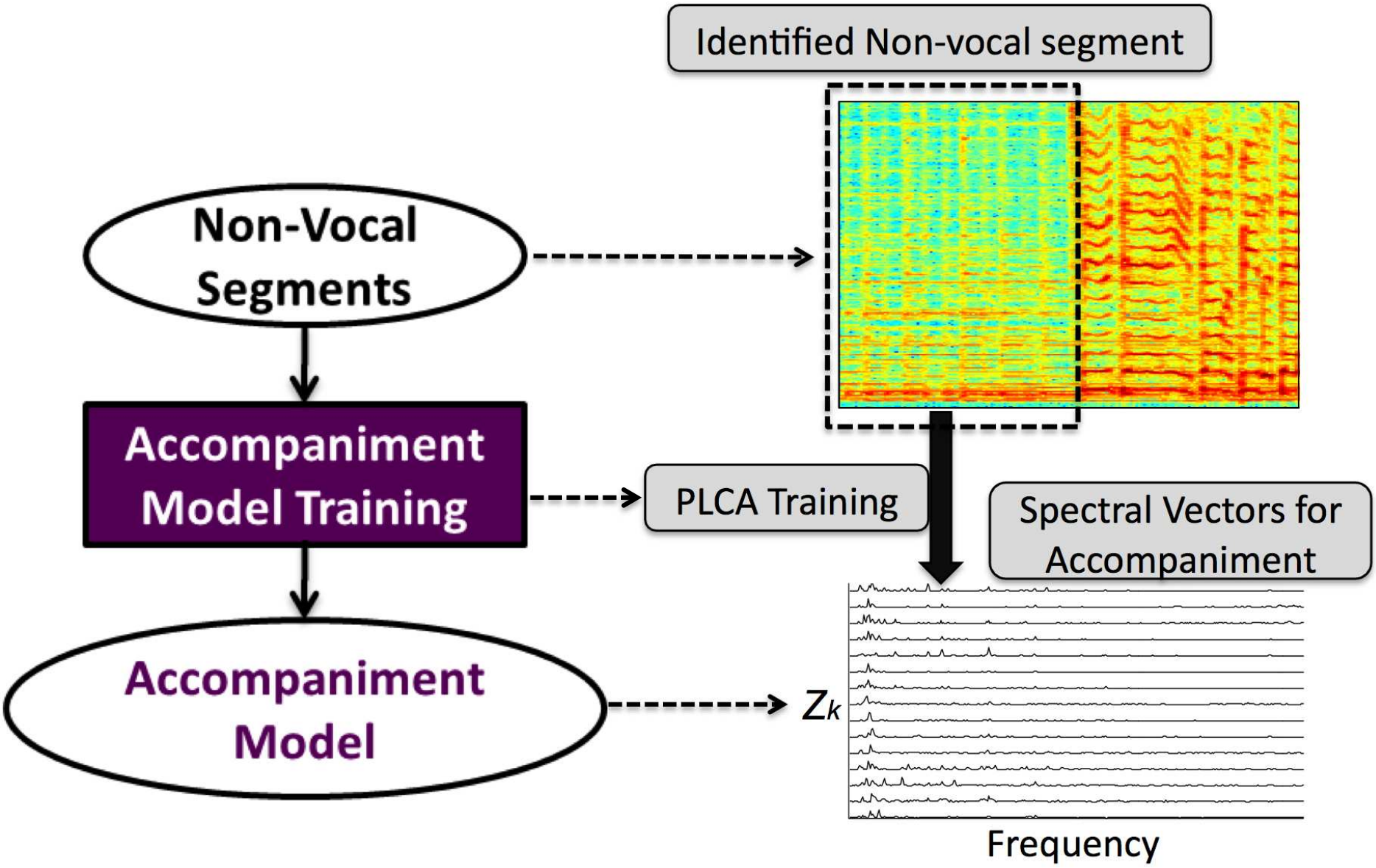



A dictionary of Spectral Vectors
P(f|z1)
P(f|z2)
P(f|z3)
P(f|zn)
z1
z2
z3
zn
Latent Component
Mixture Weight
$P_t(f) = \sum_{k=1}^{k=n} w_{t,z_k} P(f|z_k)$
Estimated by Expectation-Maximization Algorithm
Spectrum of frame t: Pt(f)
Observed data in the spectrogram

# System Overview

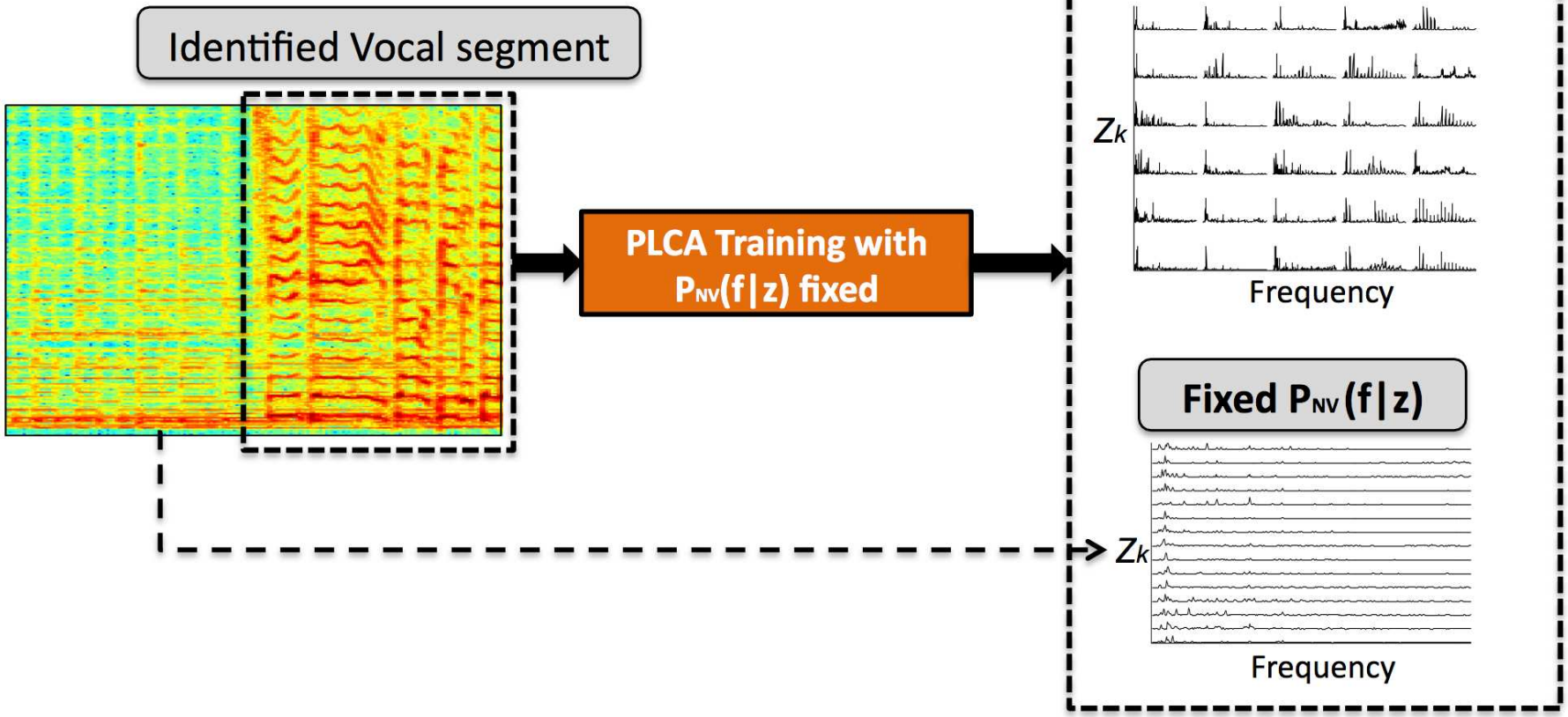


# Train $P_{nv}(f|z)$ from the non-vocal segment



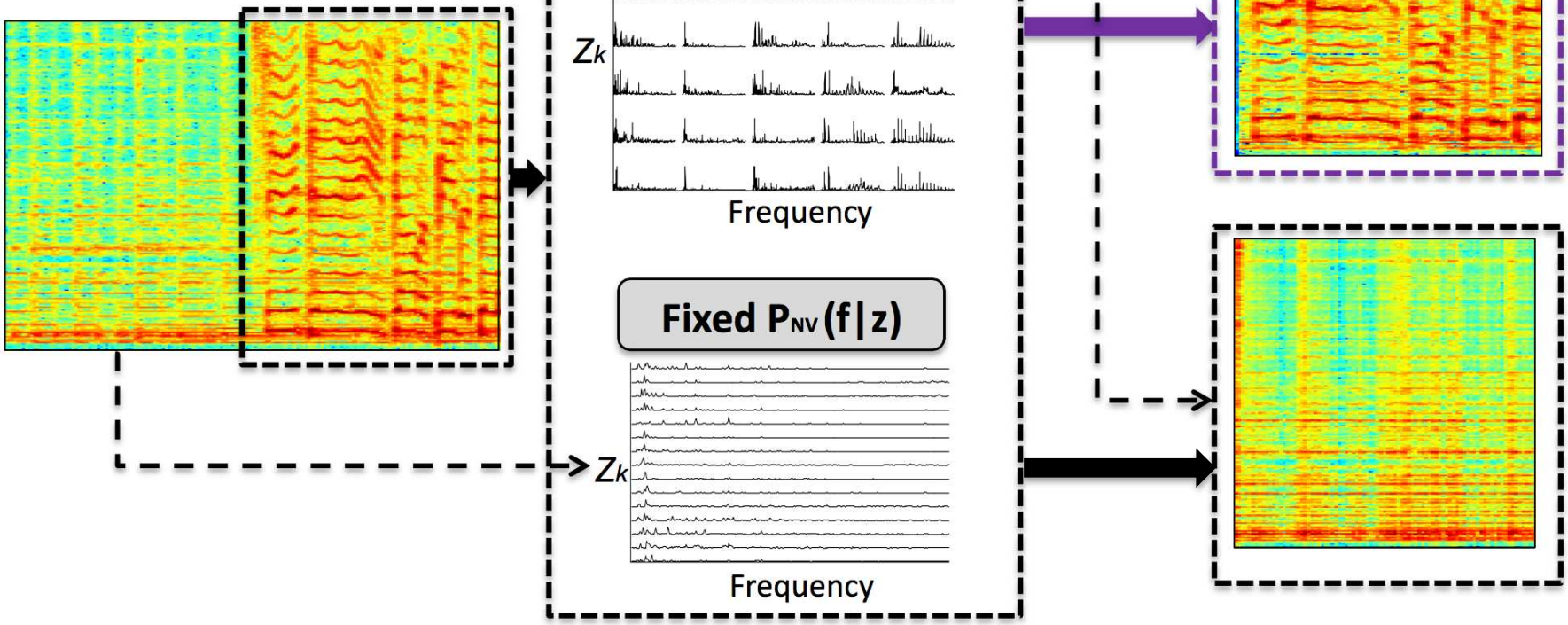
# Extract singing voice in the mixture

$$P_t(f) = \sum_{z_{nv}} w_{t,z_{nv}} P_{nv}(f|z) + \sum_{z_v} w_{t,z_v} P_v(f|z)$$

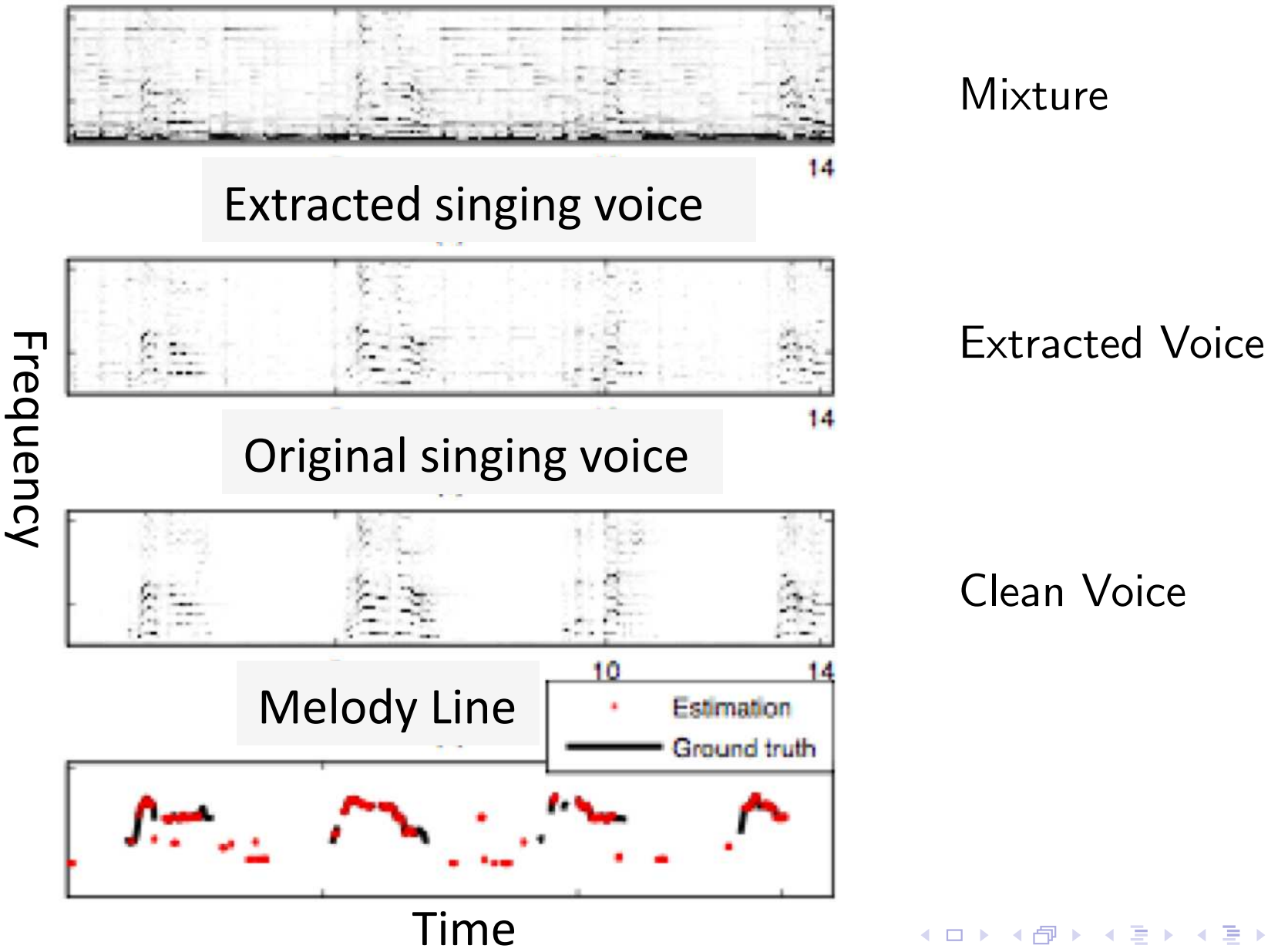


# Extract singing voice in the mixture

$$P_t(f) = \sum_{z_{nv}} w_{t,z_{nv}} P_{nv}(f|z) + \sum_{z_v} w_{t,z_v} P_v(f|z)$$



## A 14-s clip of “Simple Man” by Lynyrd Skynyrd



# Compare out system to DHP[1] and LW[2]

| | Precision | Recall | F-measure | Accuracy |
|---|---|---|---|---|
| DHP | **0.52** | 0.48 | 0.50 | 0.48 |
| LW | 0.09 | 0.086 | 0.09 | 0.19 |
| Proposed | 0.43 | **0.80** | **0.55** | **0.61** |

Parts of MIREX 2005 dataset: 9 recordings, totalling about 270 seconds of autio.

*Z. Duan, J. Han, and B. Pardo,* "Harmonically informed pitch tracking",in Proc. ISMIR, 2009.

*Y. Li and D. Wang,*
"Separation of singing voice from music accompaniment for monaural recordings", IEEE Trans. Audio, Speech, and Language

# Conclusion

- The Probabilistic Latent Variable Model is introduced to model the accompaniment and lead vocal adaptively
- Experimental results show that the melody of the singing voice in mixture aduio is successfully extracted to some extent.
- Future directions include improving the vocal/nonvocal segementation module and the pitch estimation algorithm.

# Acknowledgement


- The first author performed this work with Ching-Wei Chen while at the Gracenote Media Technology Lab. We thank Markus Cremer, Bob Coover, Phillip Popp, Trista Chen, and Peter Dunker for enlightening discussions.

- The authors would like to thank the reviewers for their comments that help improve the paper.

- We also want to thank Bryan Pardo, David Little, Zhiyao Duan, Zafar Rafii, and Mark Cartwright for their suggestions that improve the presentation.