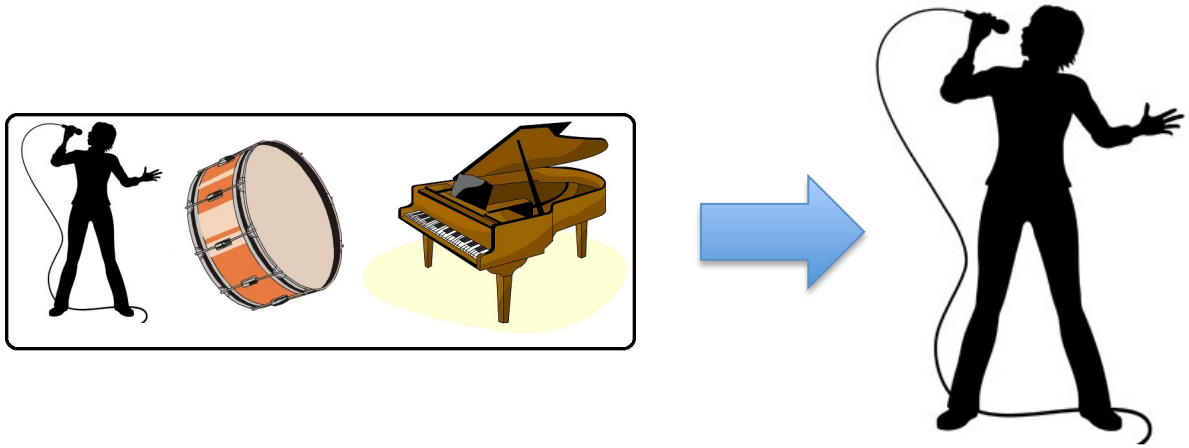
Non-negative Matrix Factorization and Local Discontinuity Measures for Singing Voice Separation



Presented by

Hatem Deif

Brunel University
Abu Dhabi University

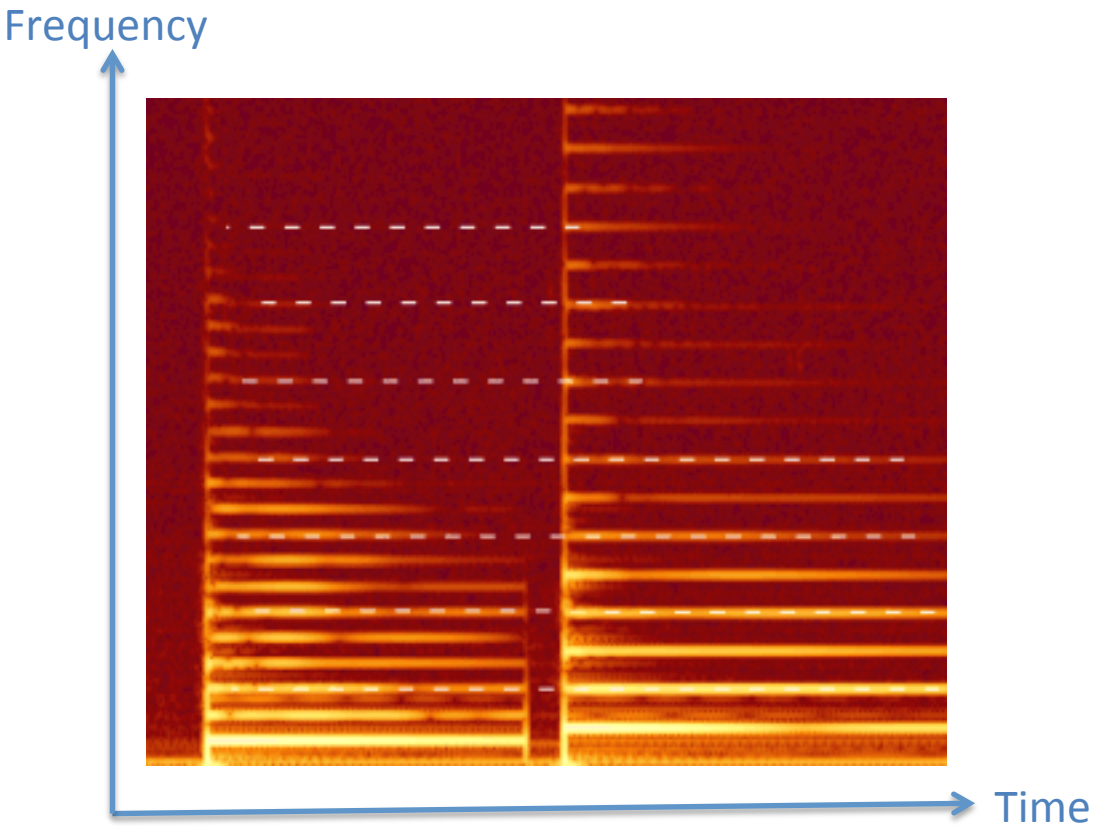
Outline

- Harmonic-Percussive-Voice Separation
- Using NMF and DM
- Proposed Algorithm
- Results

Harmonic Instruments

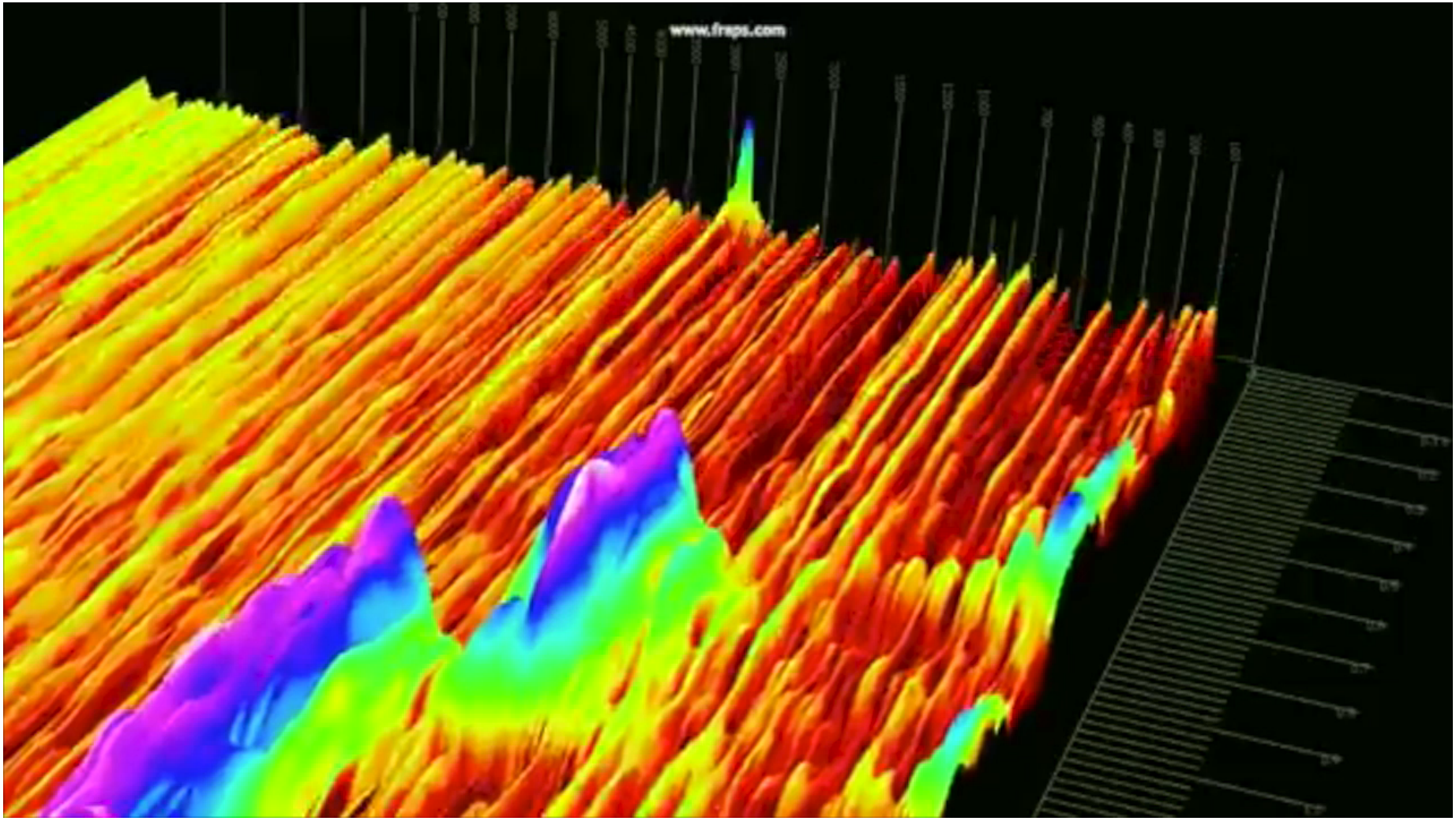
Percussion Instruments

Violin Spectrogram



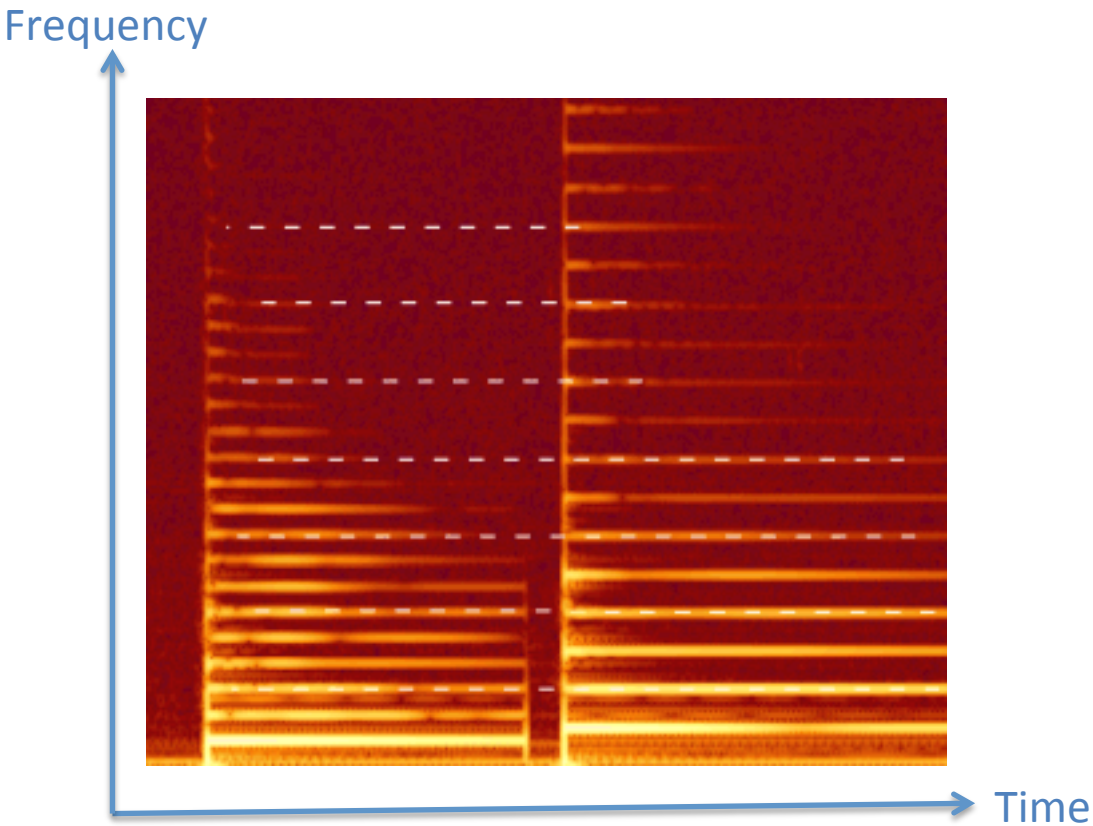
https://en.wikipedia.org/wiki/Musical_acoustics

Spectrogram Visualization



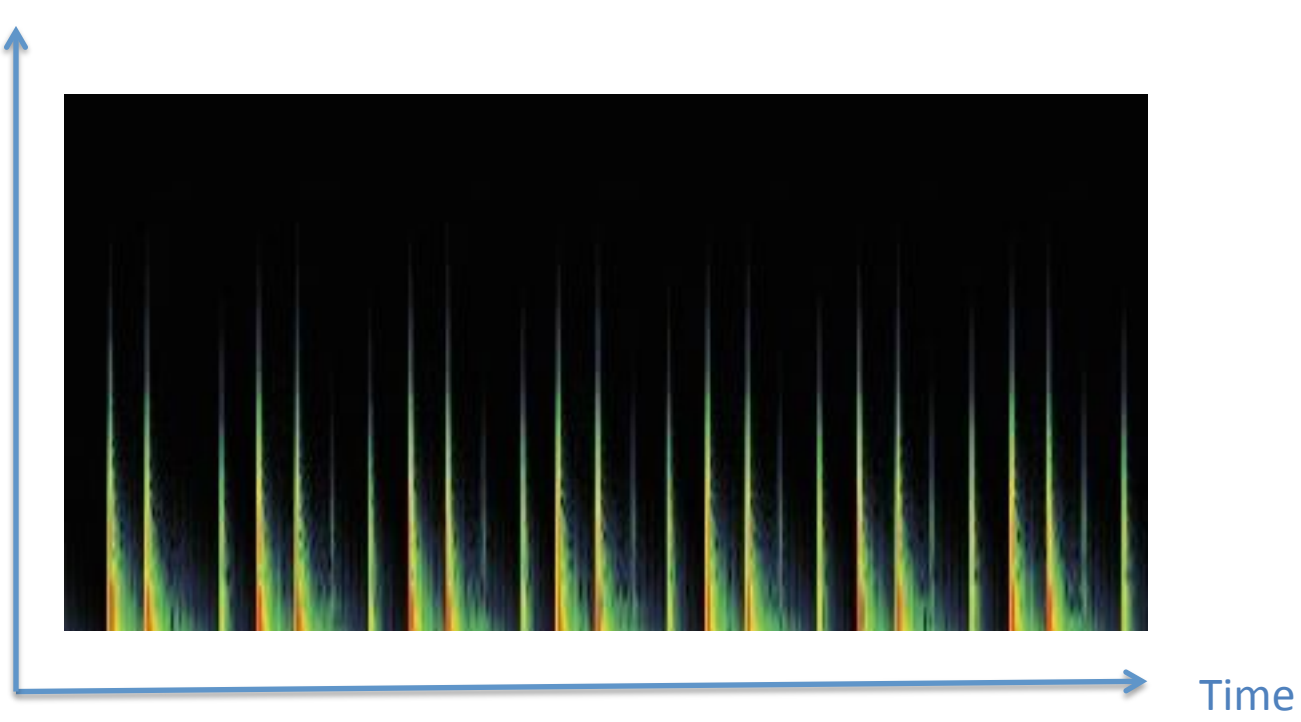
https://youtu.be/vvr9AMWEU-c Using iZotope Ozone 5 - 8192 FFT size, 93.75% overlap.

Violin Spectrogram



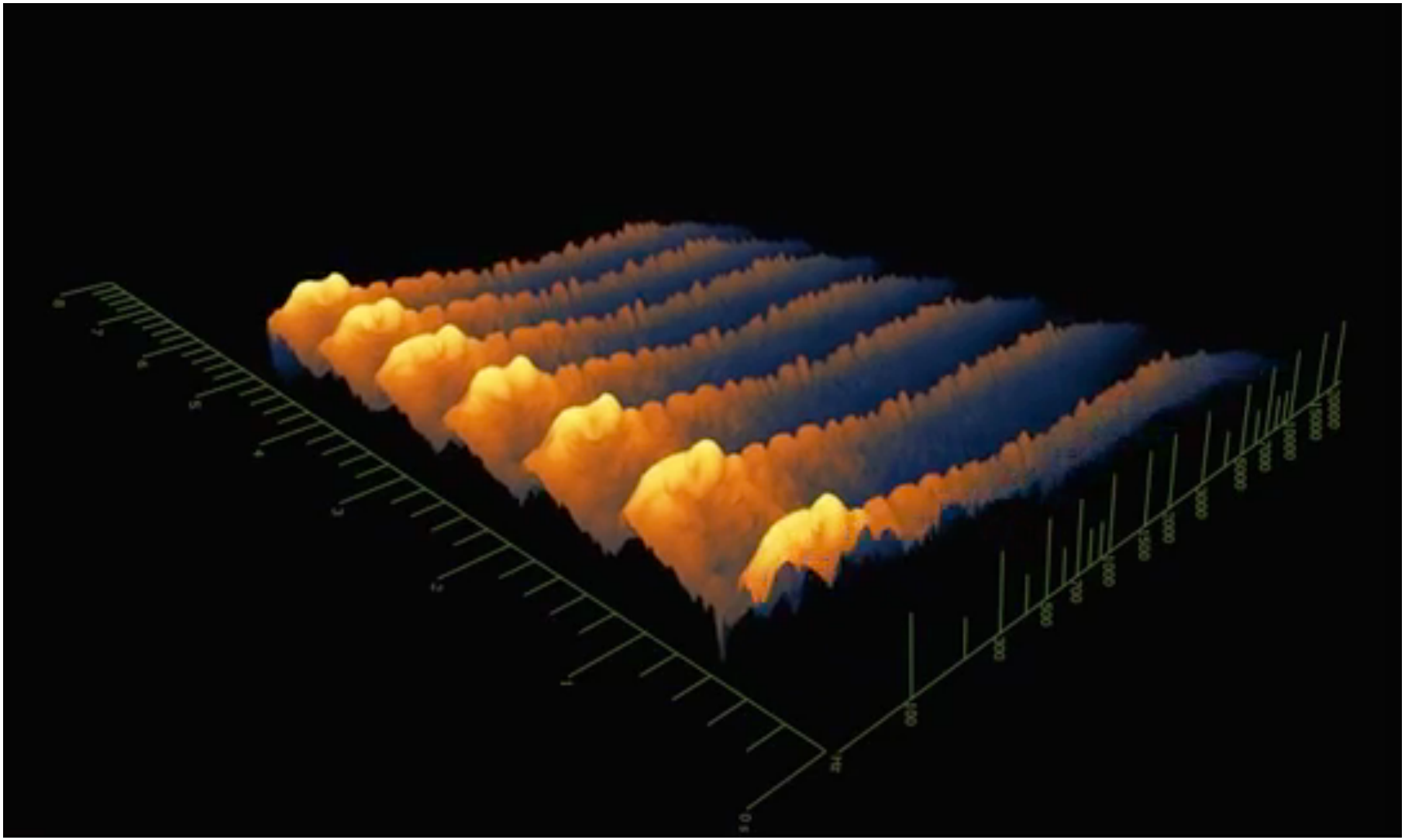
https://en.wikipedia.org/wiki/Musical_acoustics

Bass Drum Spectrogram



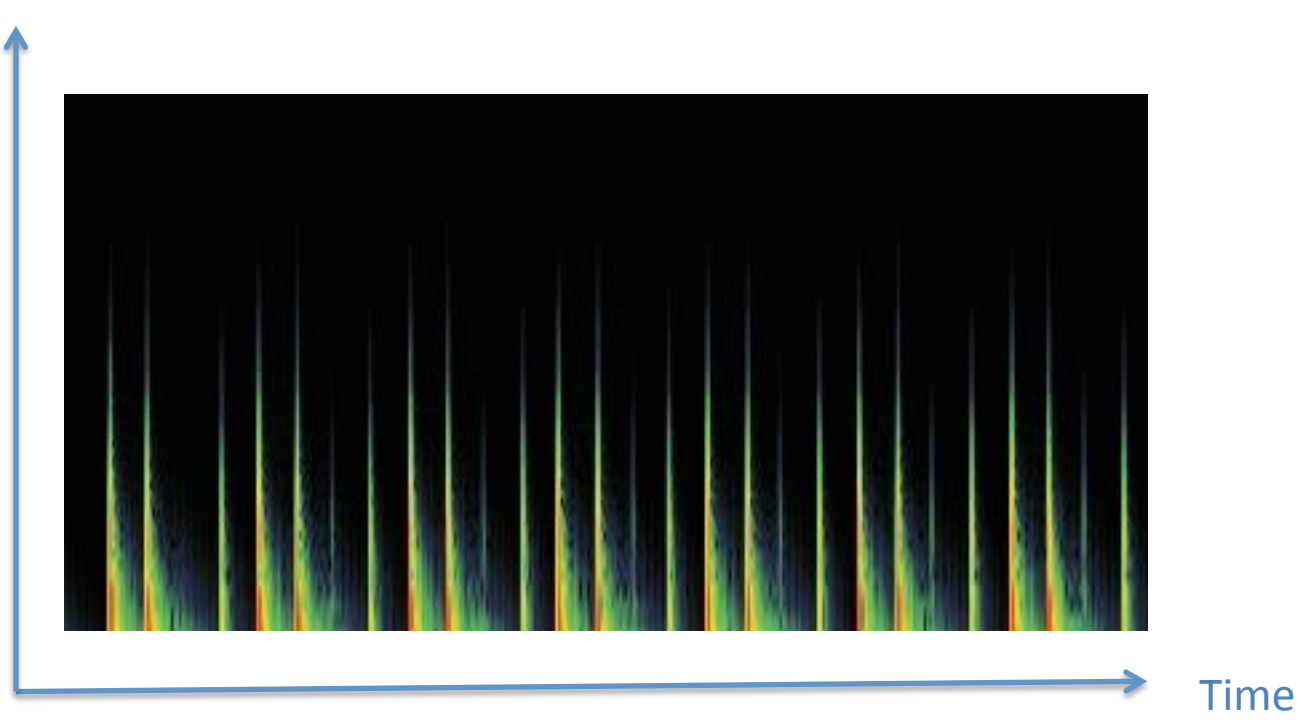
Freesound.org

Spectrogram Visualization



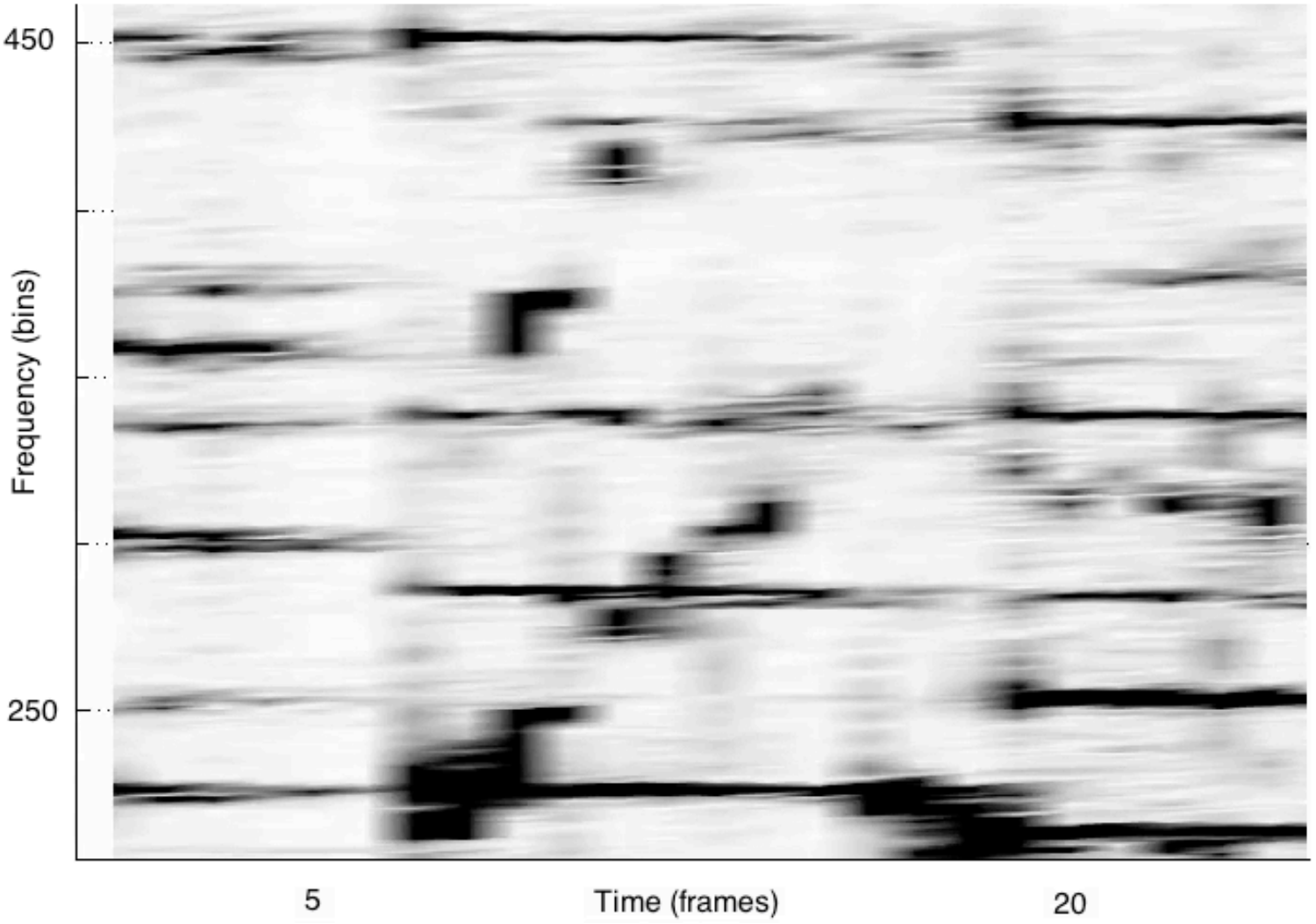
https://youtu.be/0etE3oK0HCY

Bass Drum Spectrogram



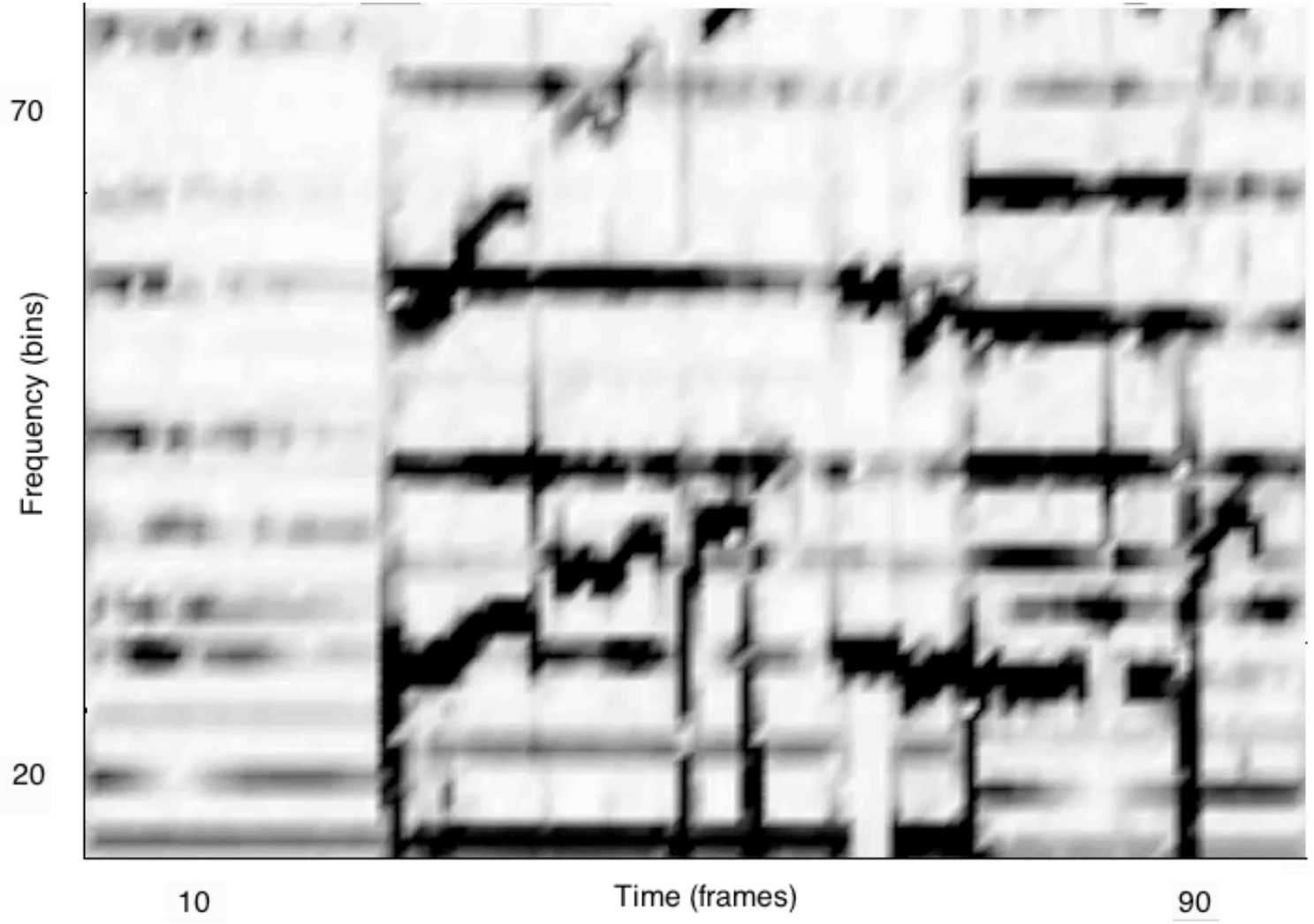
Freesound.org

Voice – High Frequency Resolution Spectrograms



STFT Window Length = 8 K

Voice – Low Frequency Resolution Spectrograms

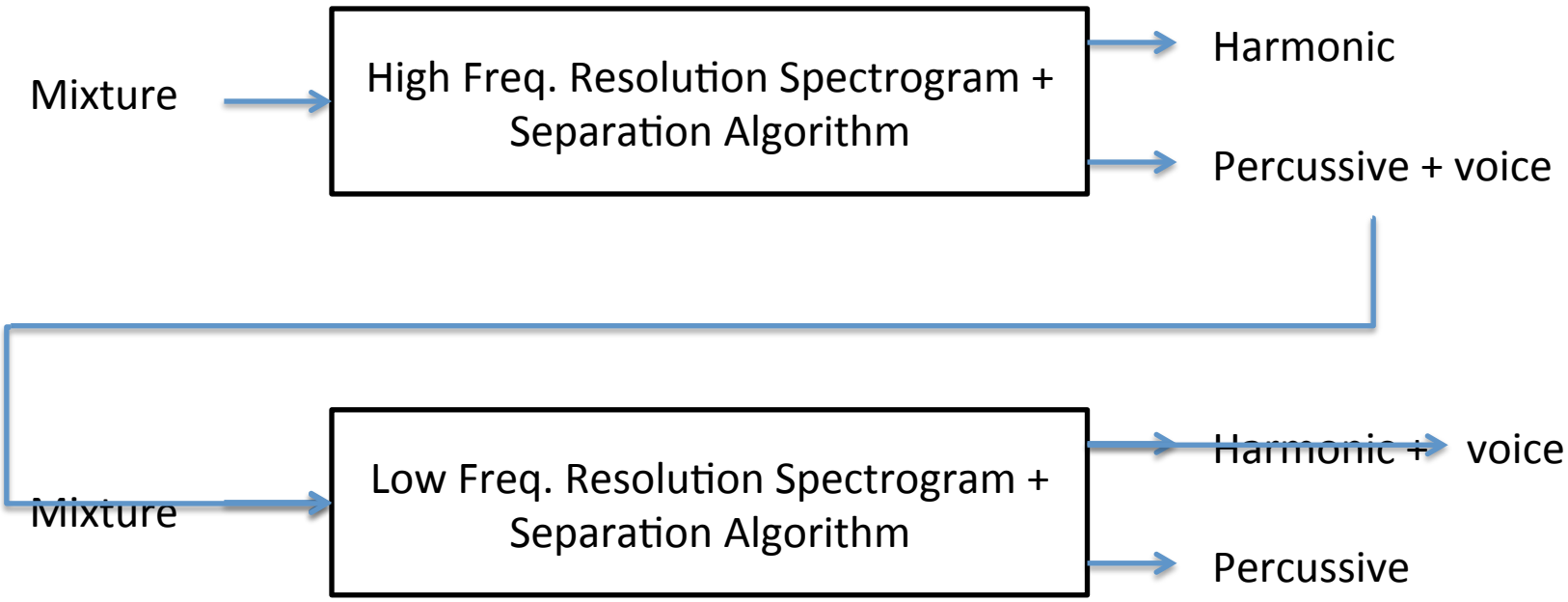


STFT Window Length = 2 K

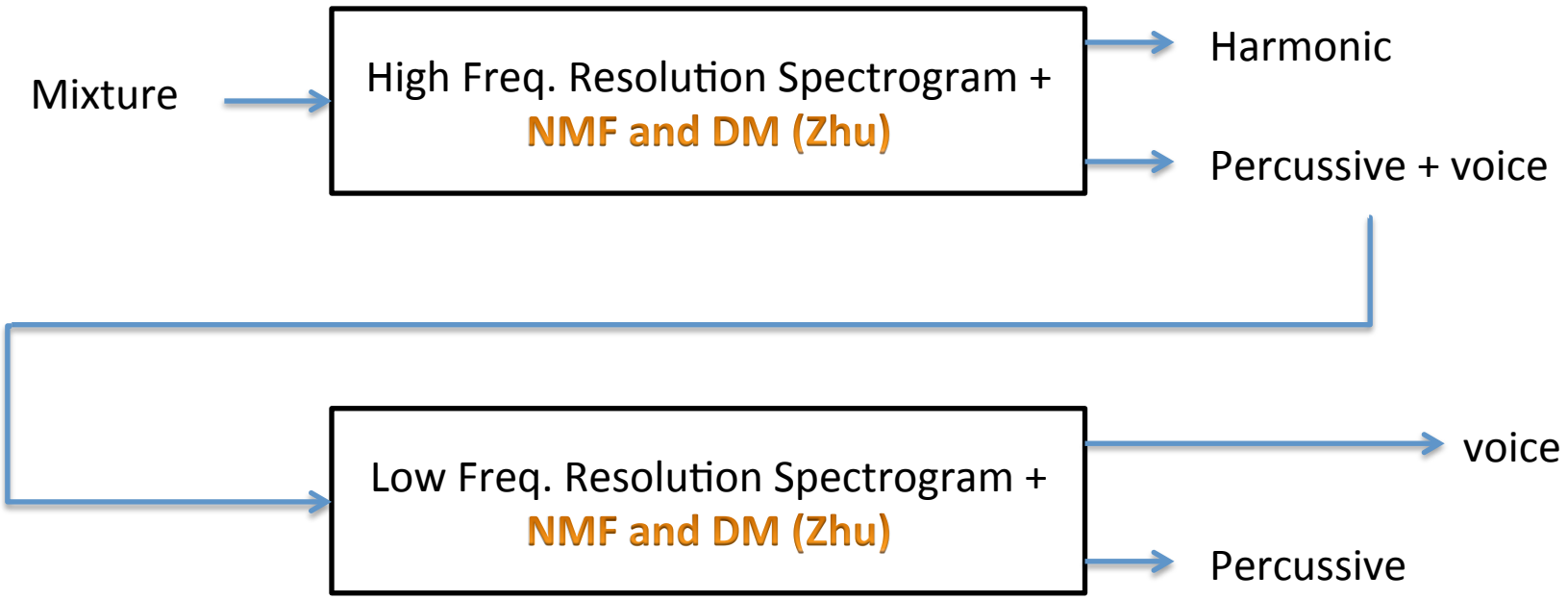
Harmonic - Percussive Separation

- HPSS Algorithm by Ono et al.
- Median Filtering by Fitzgerald
- NMF with DM by Zhu et al.

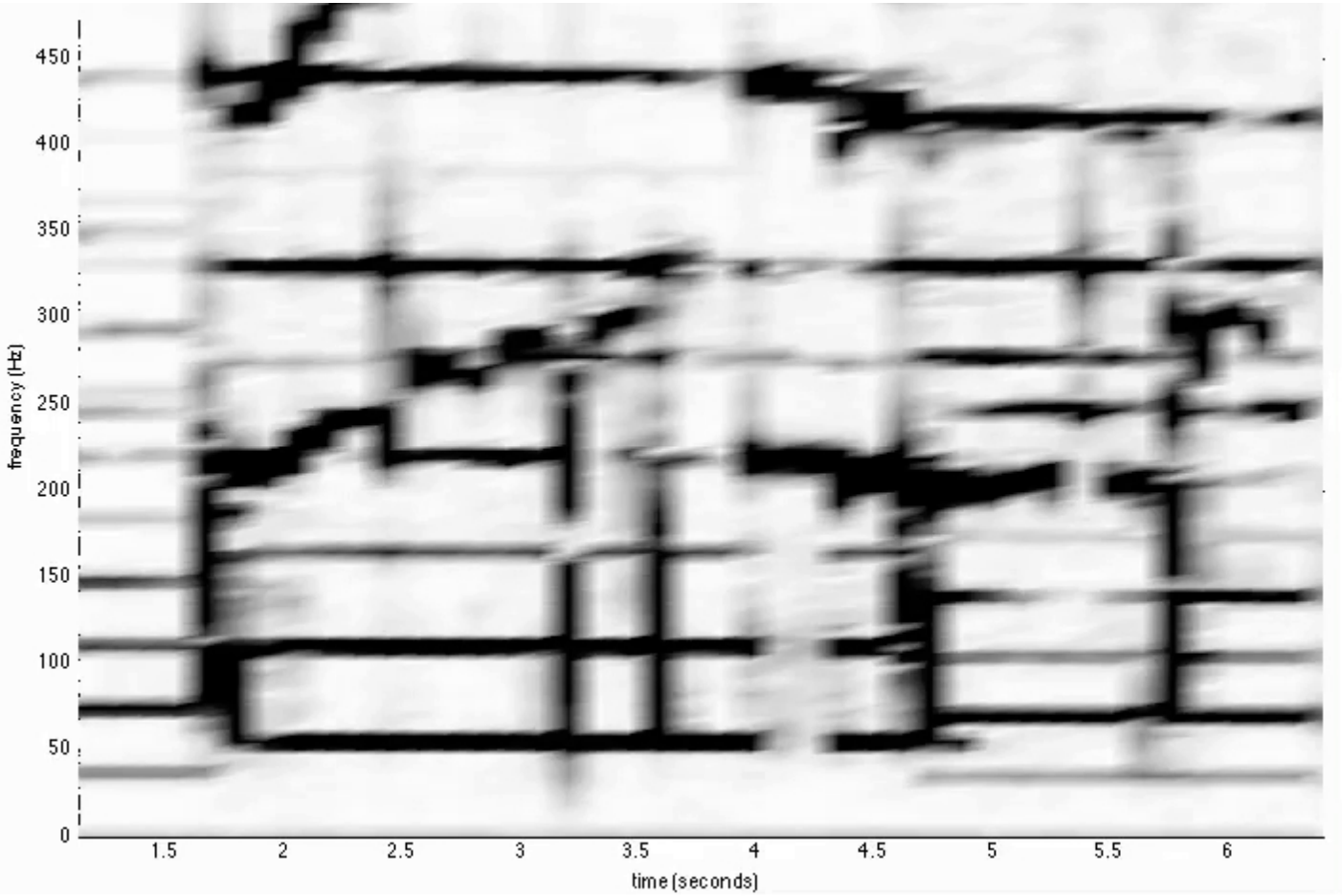
Two stages for separating voice



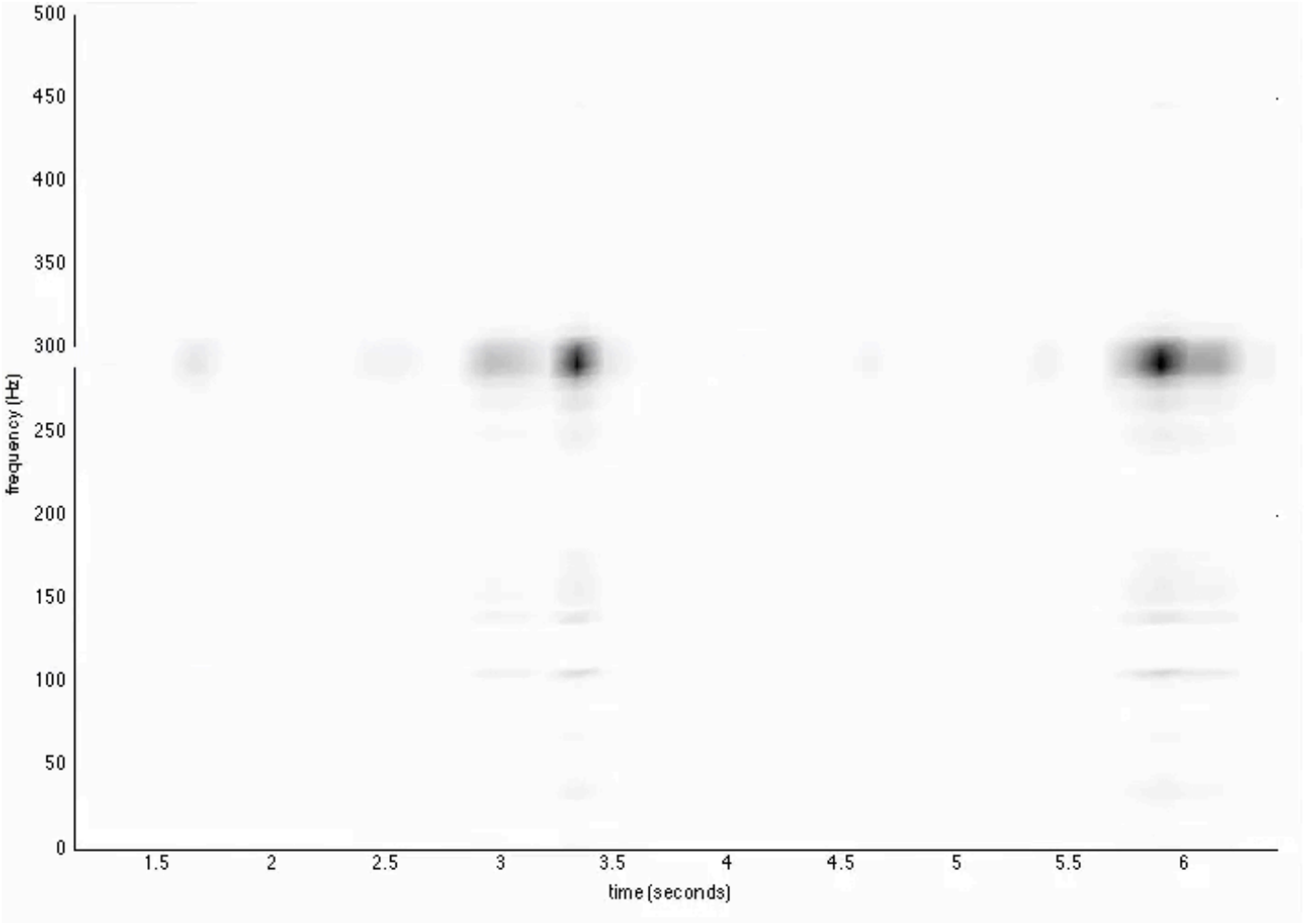
Two stages for separating voice



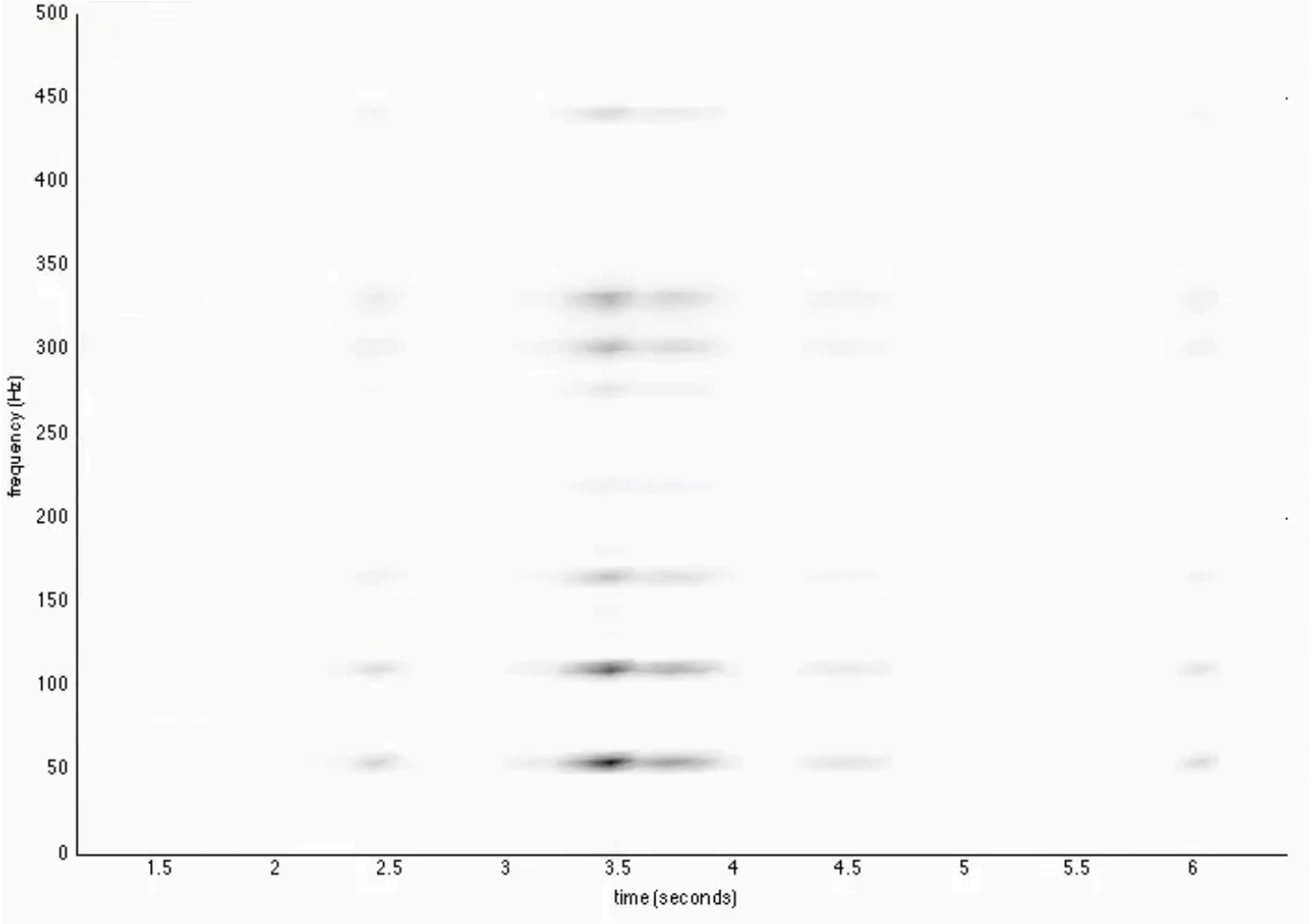
NMF of The Magnitude Spectrogram X



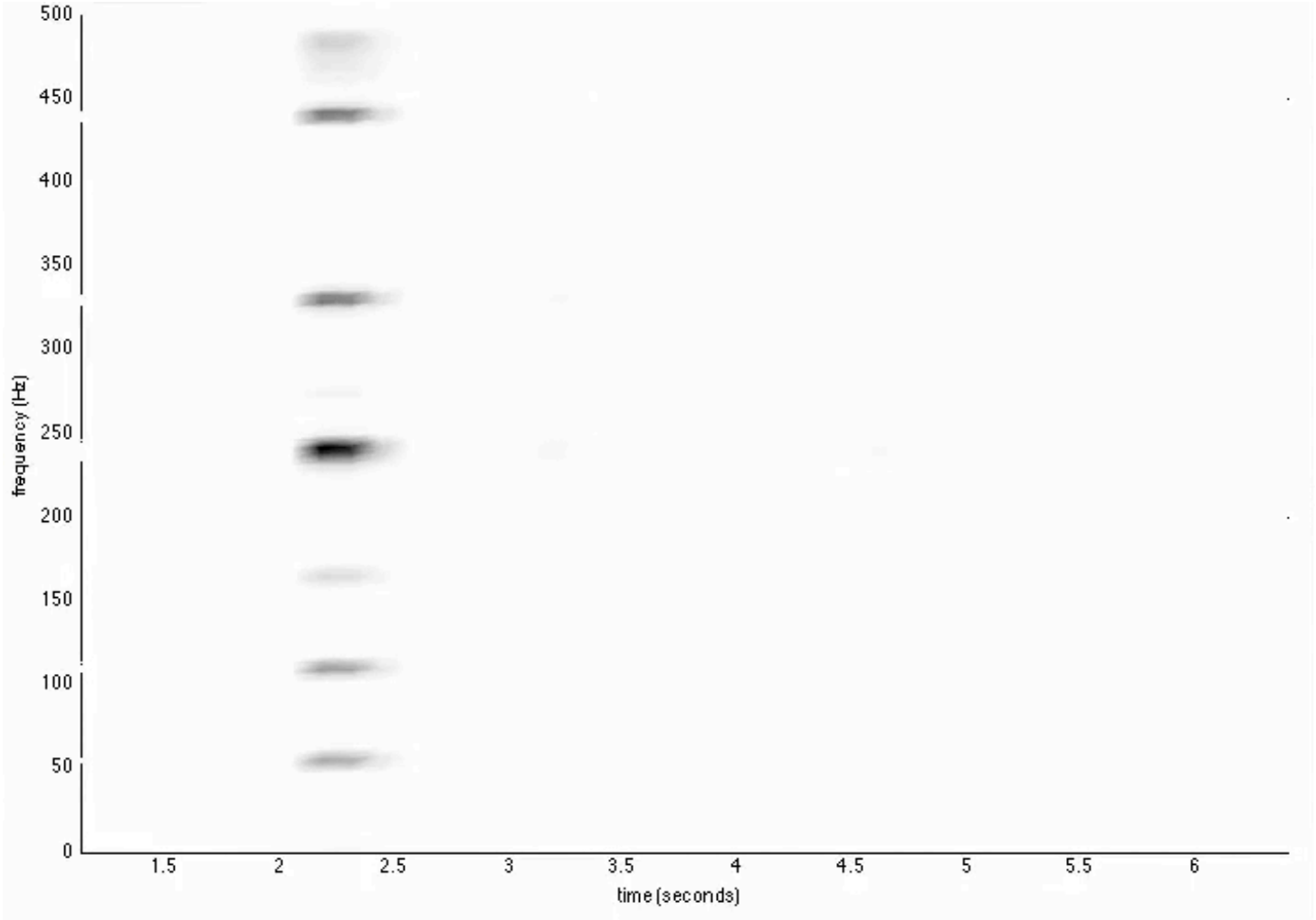
Components X^j j=1,2,3,4,5,6,…..,15



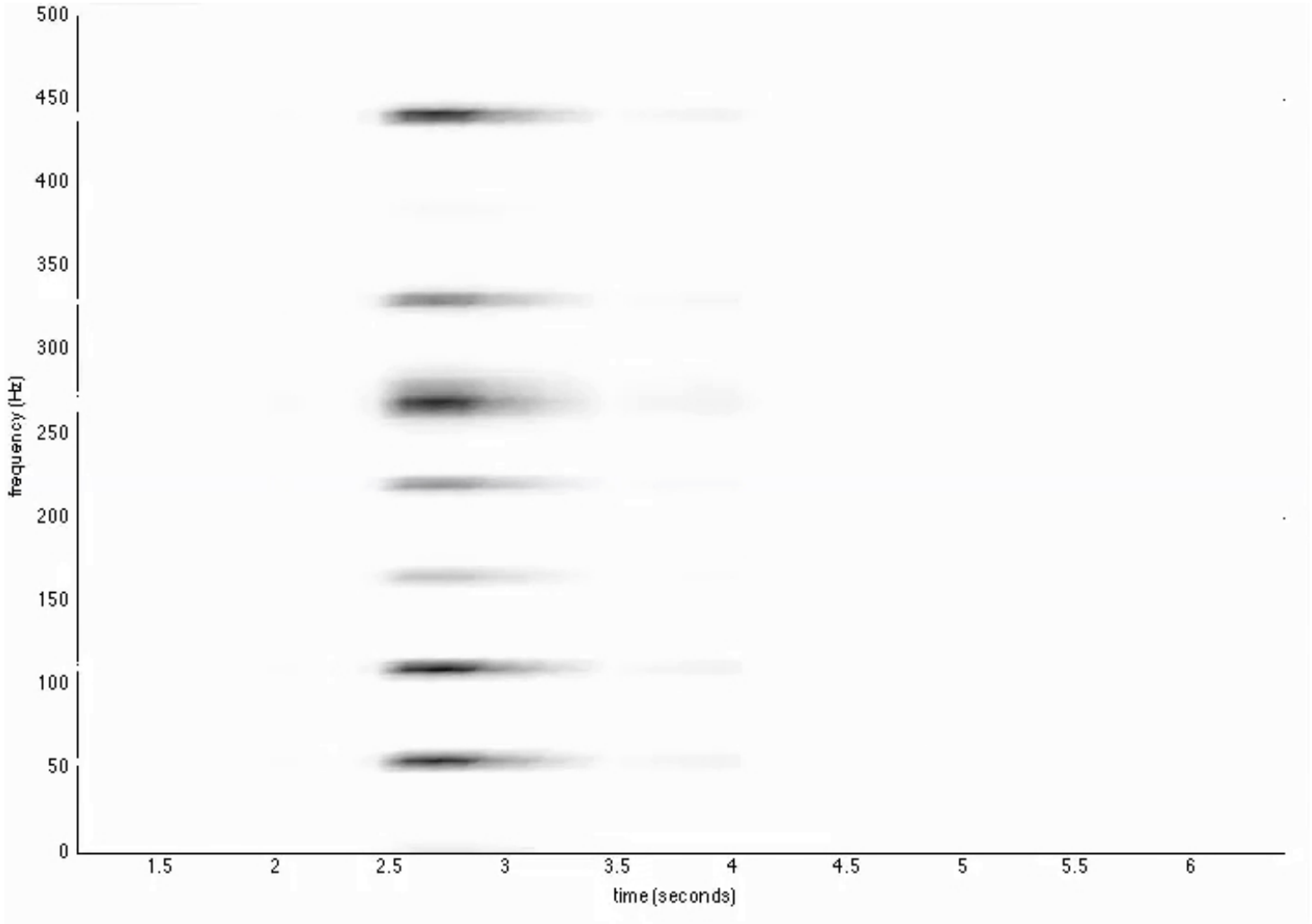
Components X^j j=1,2,3,4,5,6,…..,15



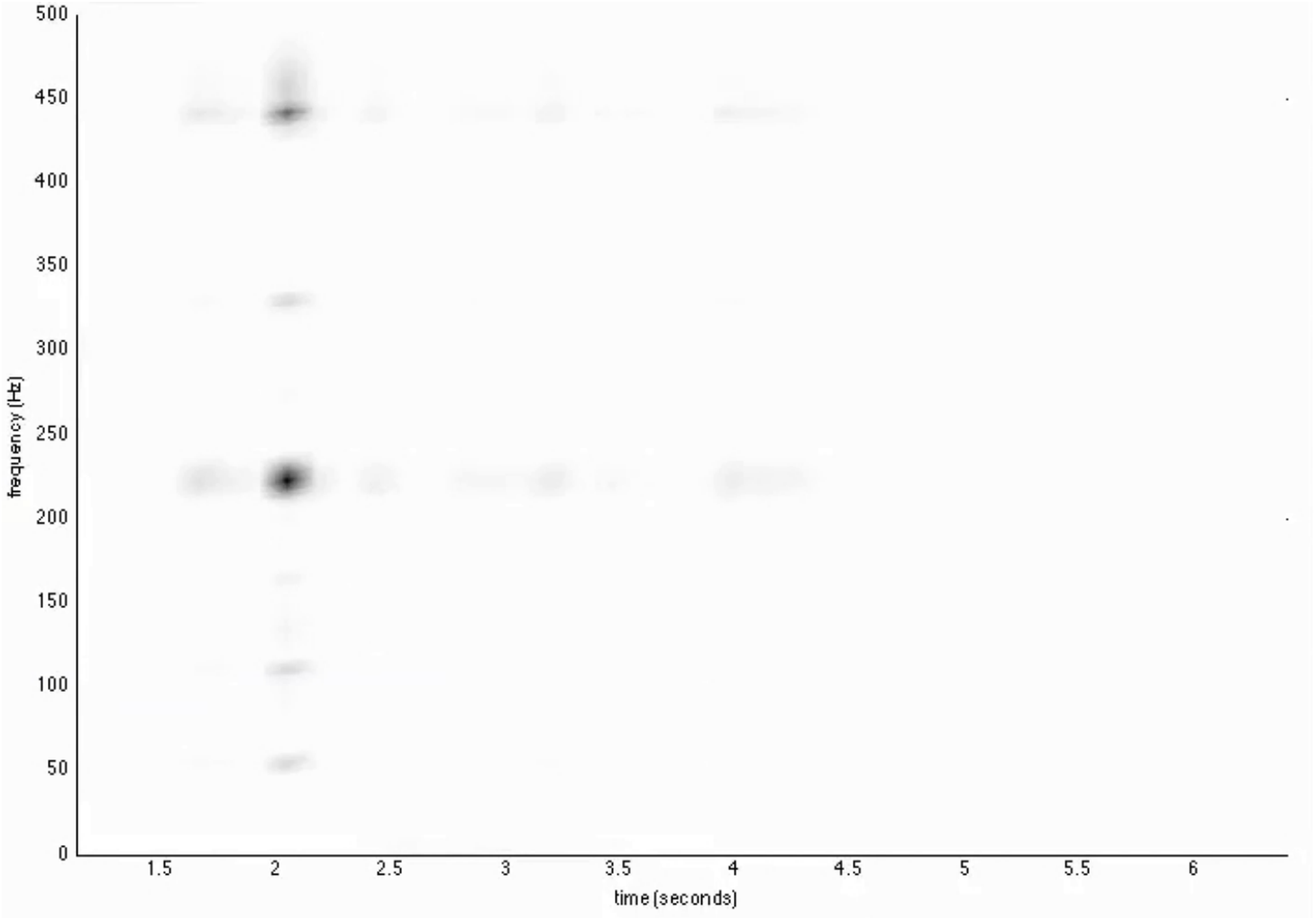
Components X^j j=1,2,3,4,5,6,.....,15



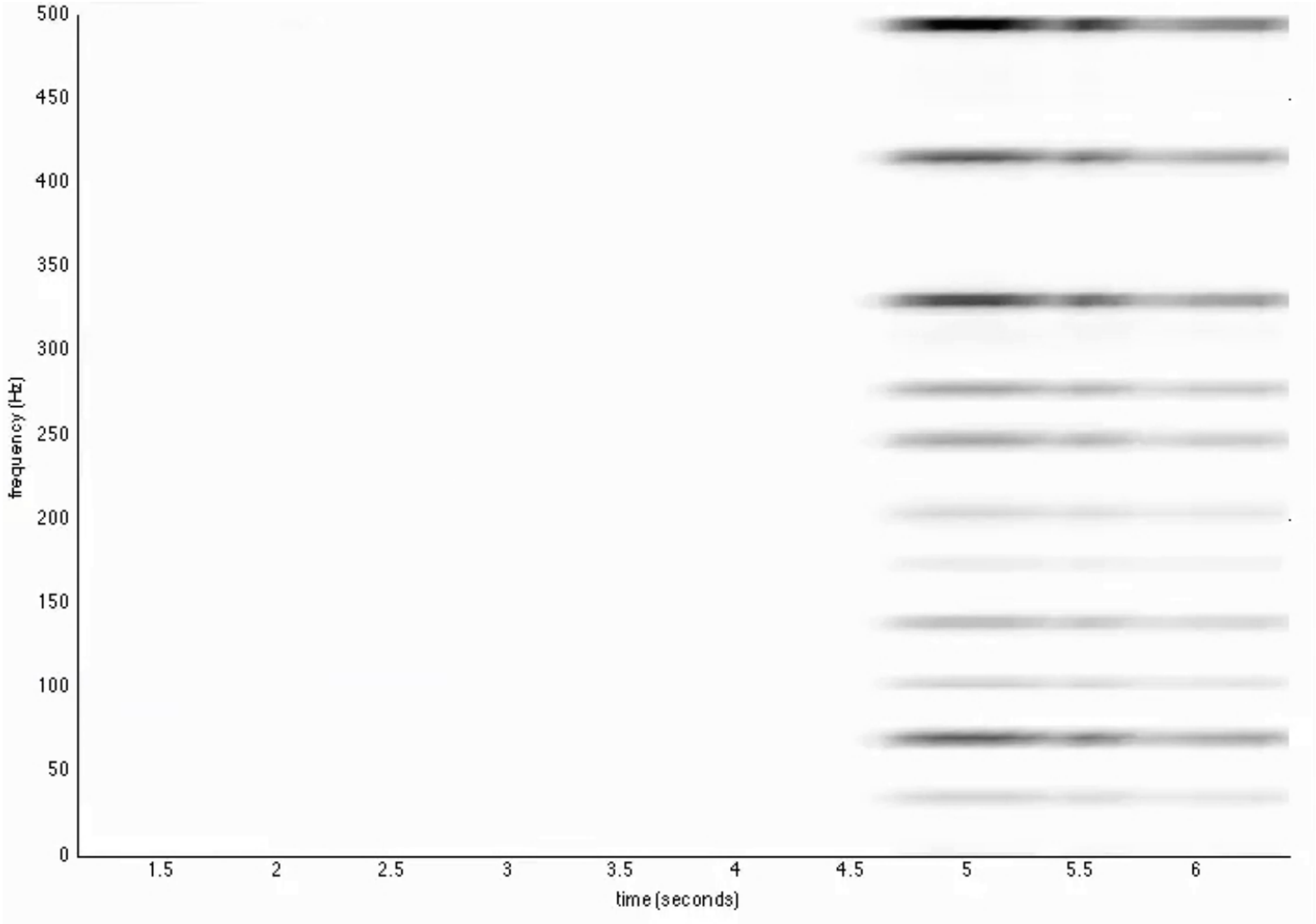
Components X^j j=1,2,3,4,5,6,……,15



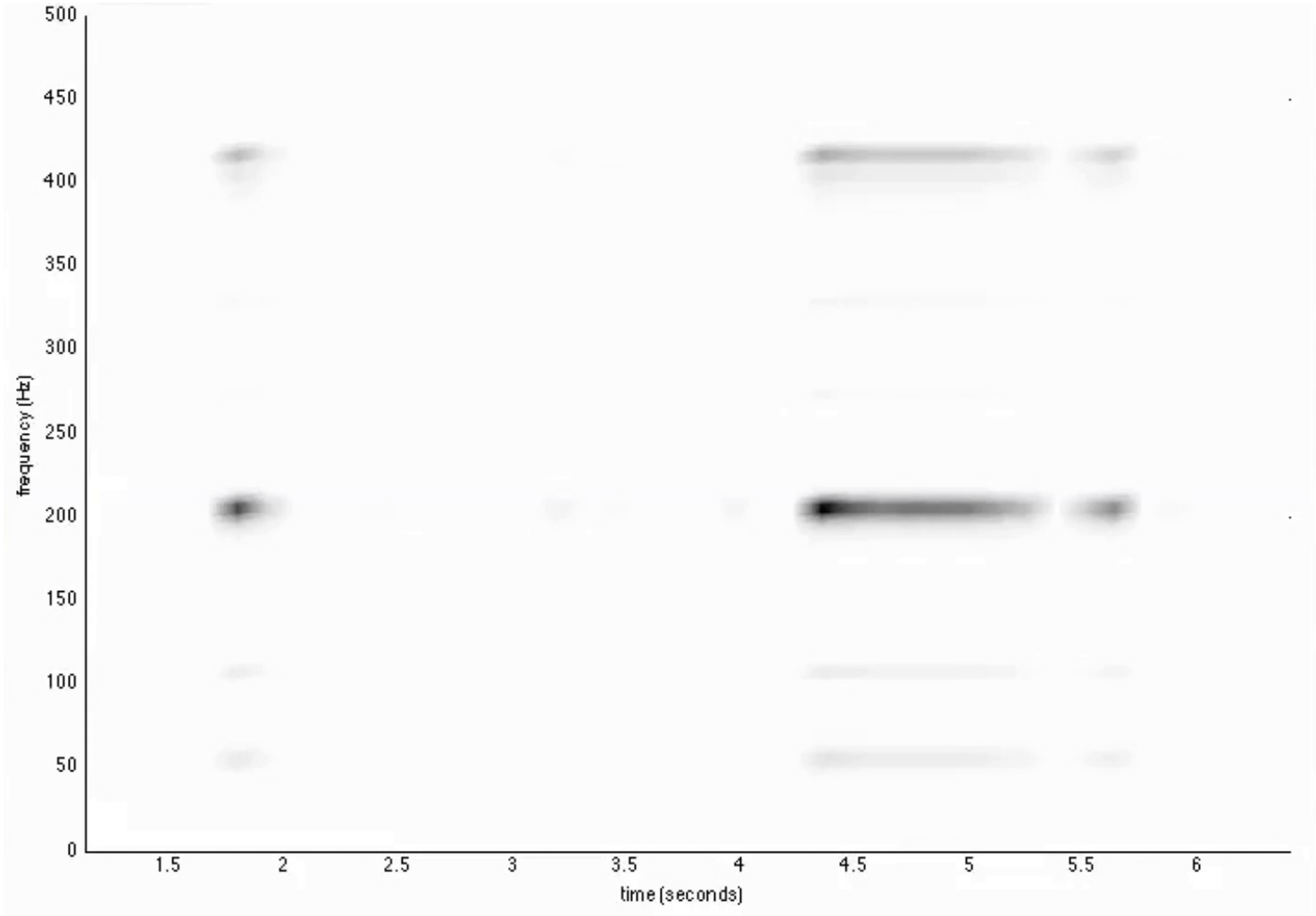
Components X^j j=1,2,3,4,5,6,…..,15



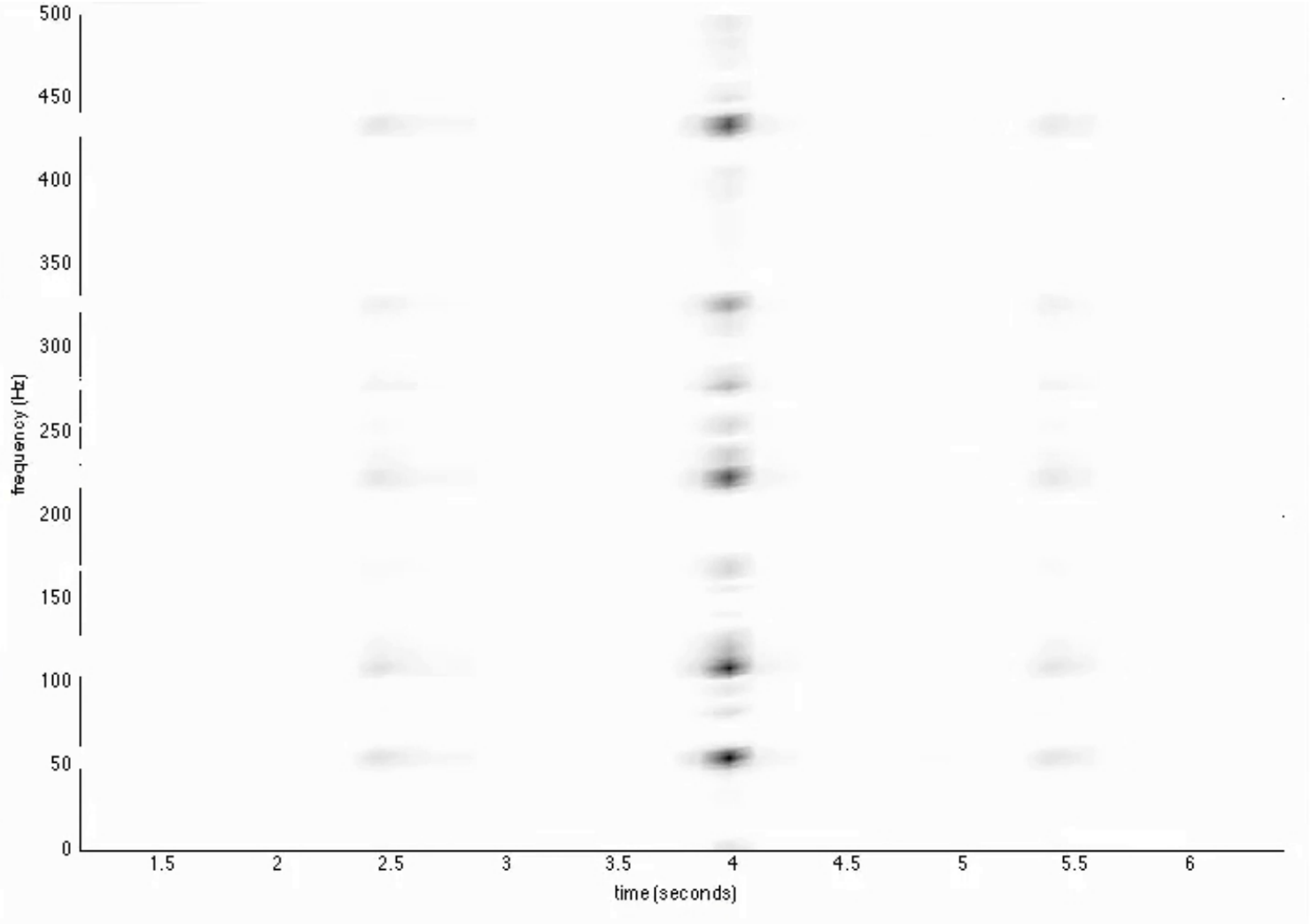
Components X^j j=1,2,3,4,5,6,.....,15



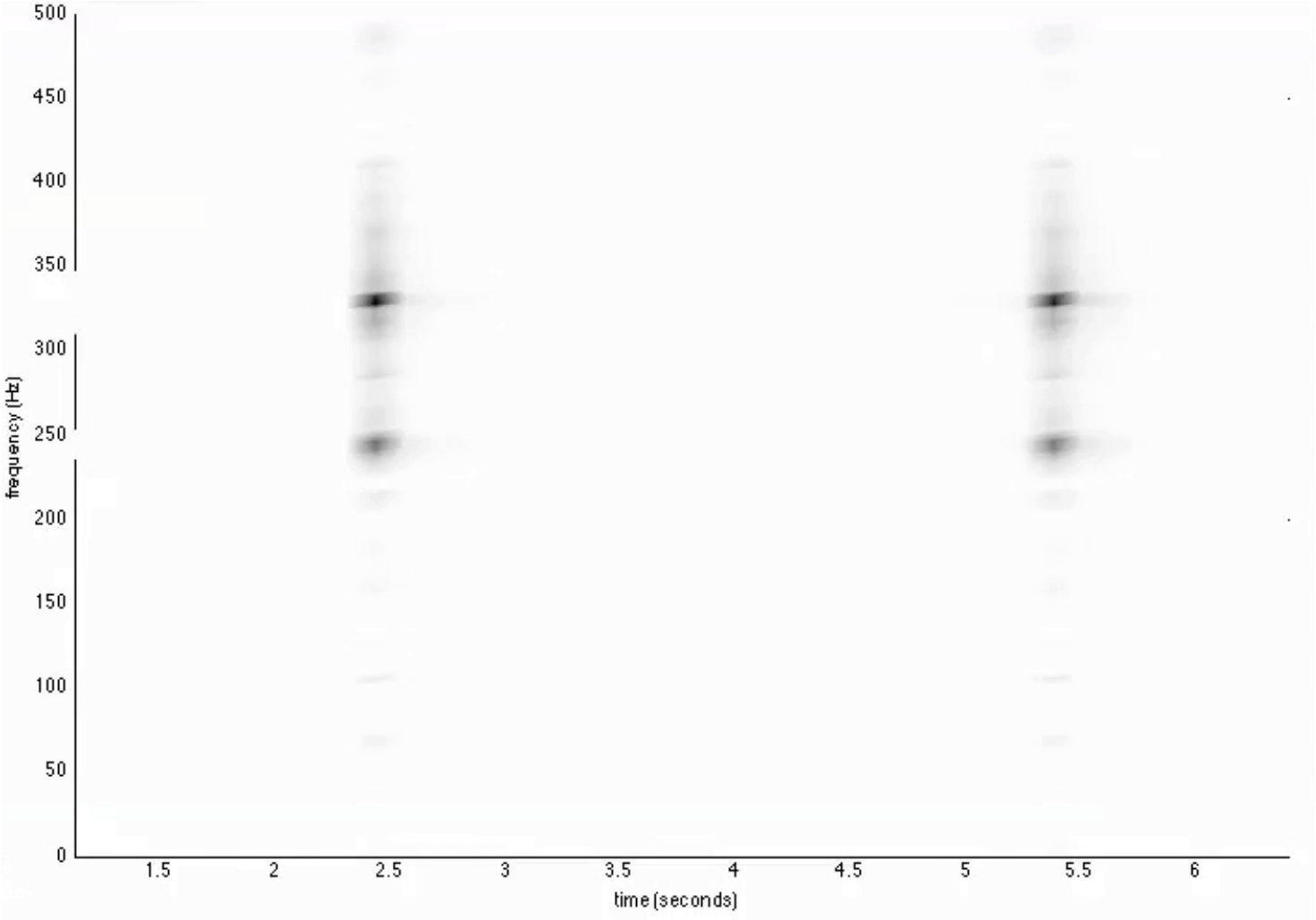
Components X^j j=1,2,3,4,5,6,.....,15



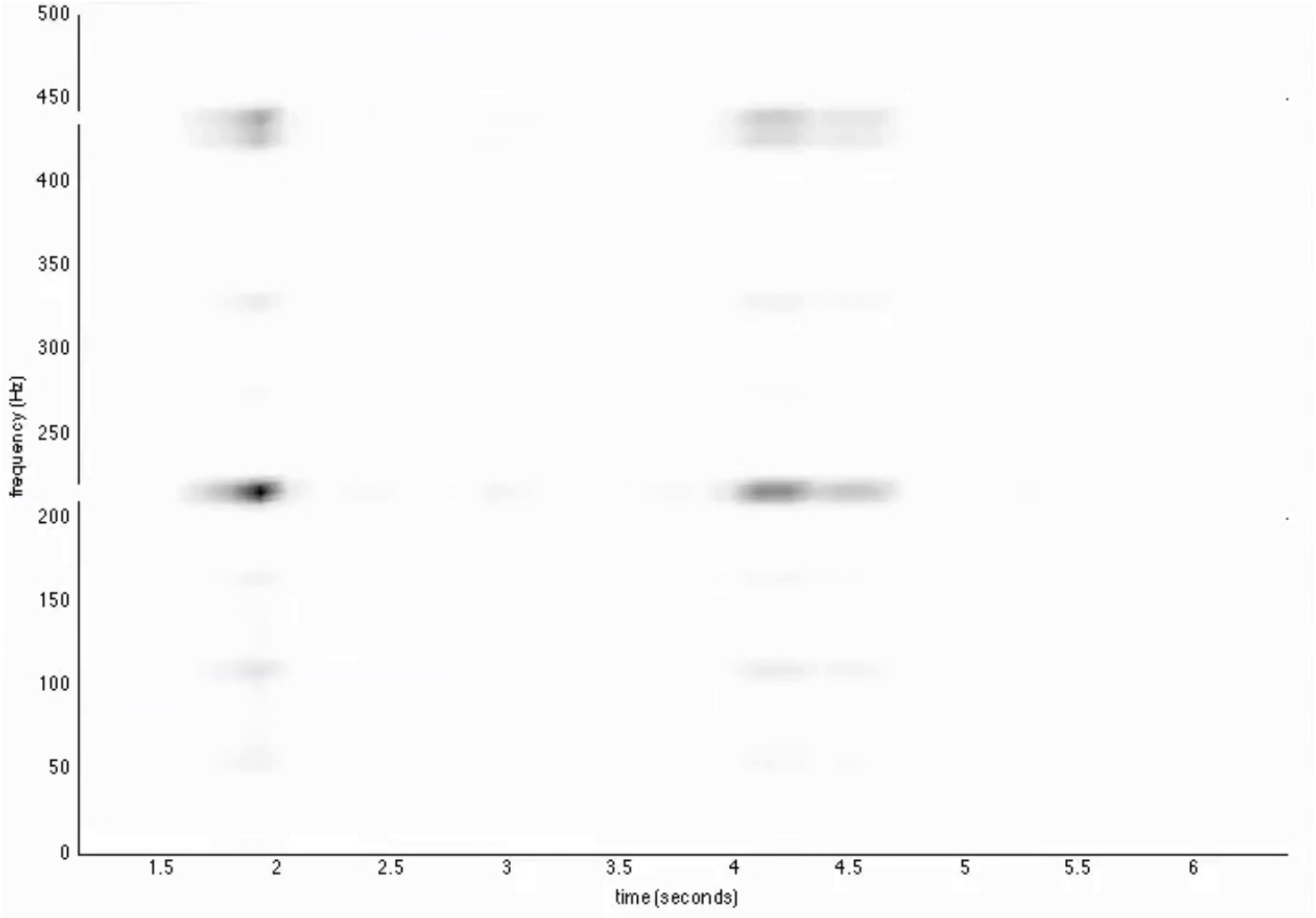
Components X^j j=1,2,3,4,5,6,.....,15



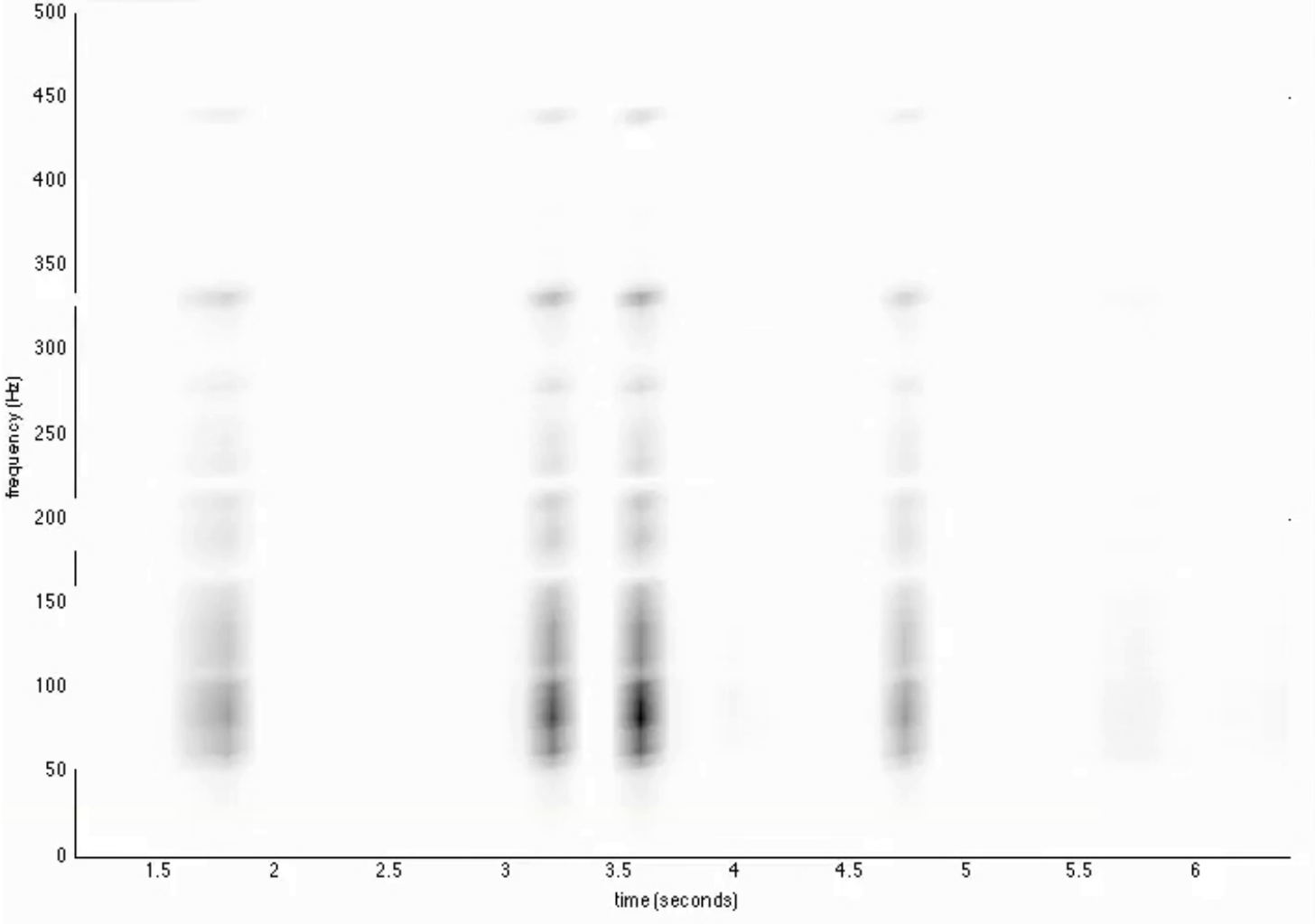
Components X^j j=1,2,3,4,5,6,.....,15



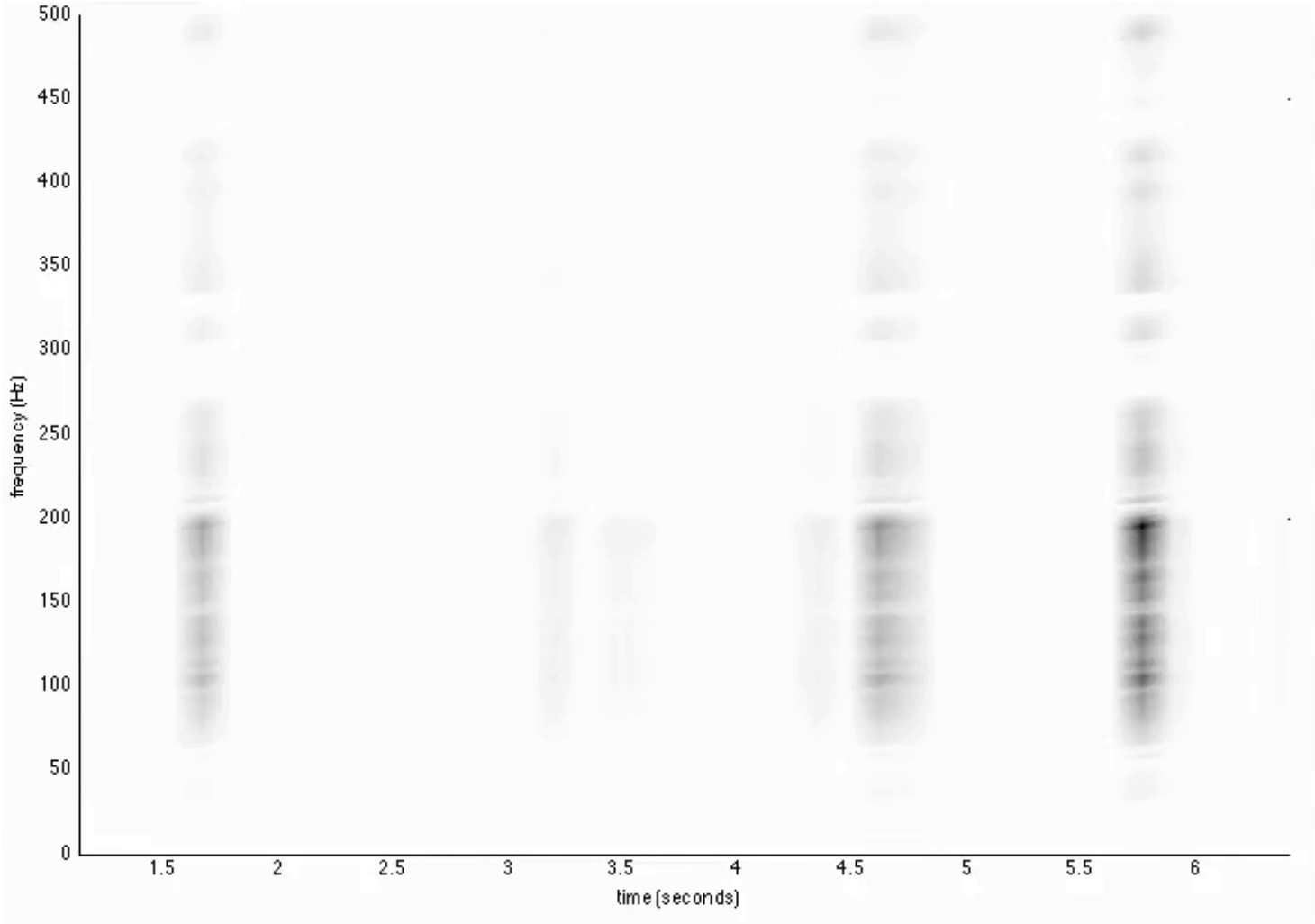
Components X^j j=1,2,3,4,5,6,.....,15



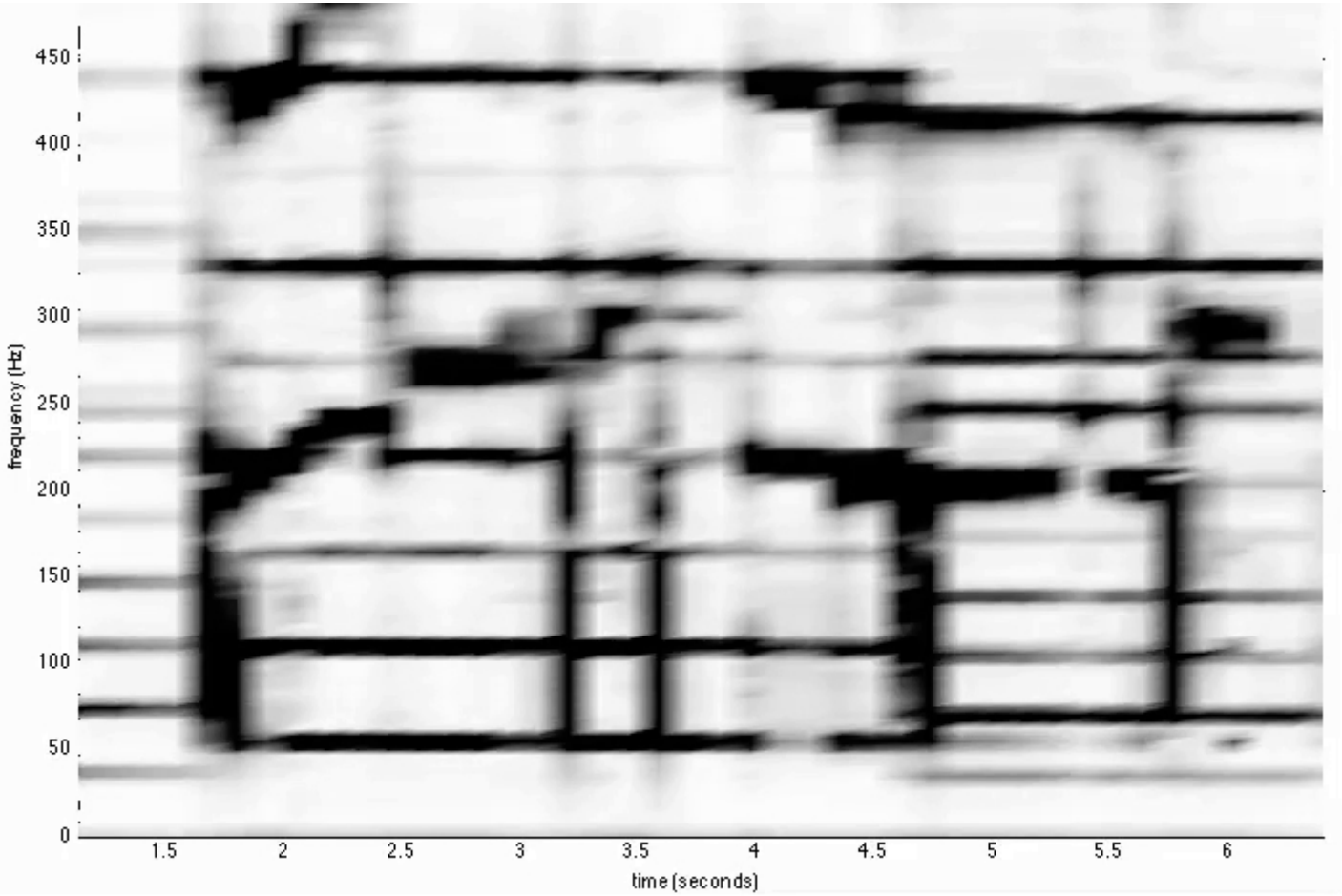
Components X^j j=1,2,3,4,5,6,.....,15



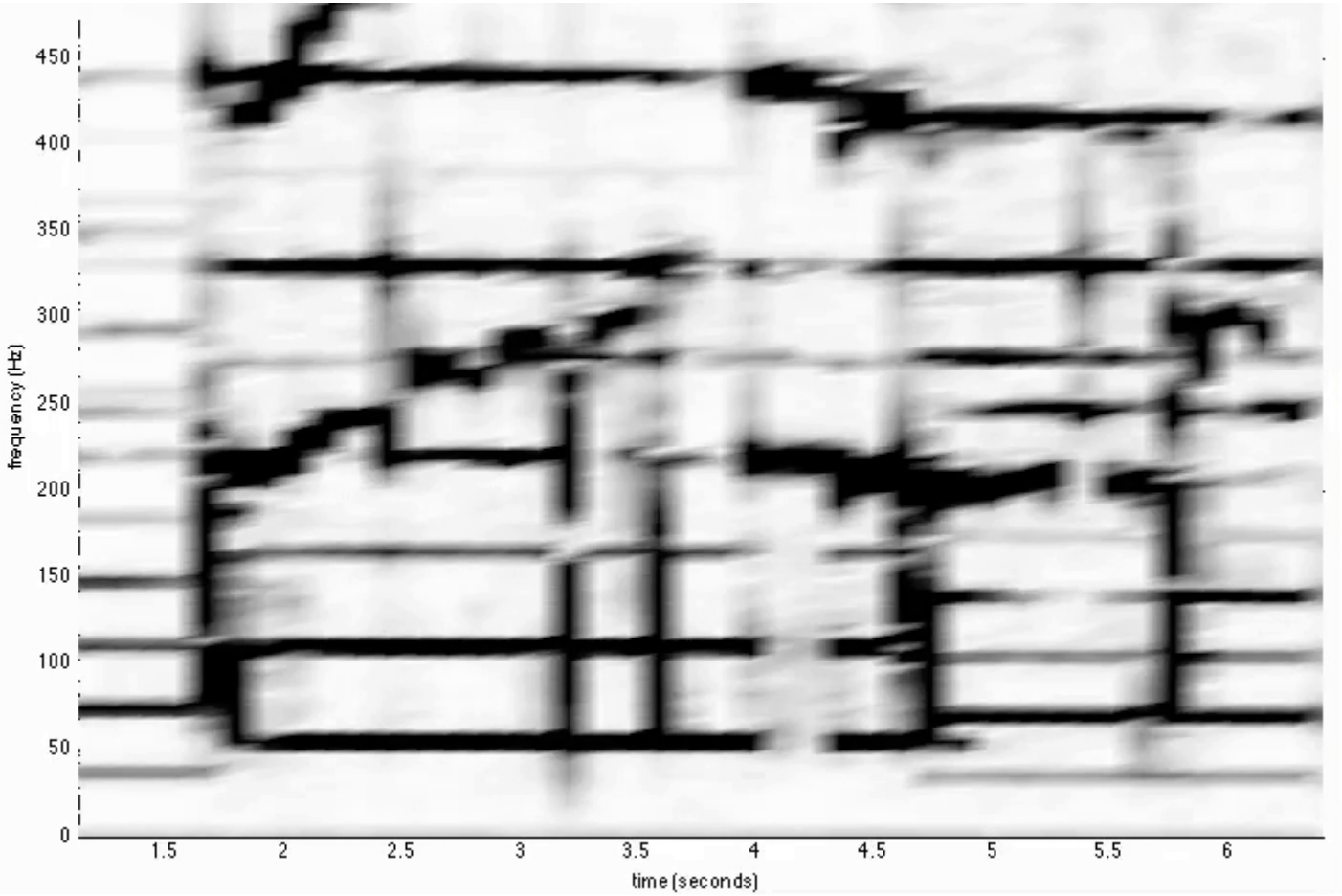
Components X^j j=1,2,3,4,5,6,……,15



$$\sum_j X^j$$


450
400
350
300
250
200
150
100
50
0
frequency (Hz)
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
time (seconds)

$$X$$


450
400
350
300
250
200
150
100
50
0
frequency (Hz)
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
time (seconds)

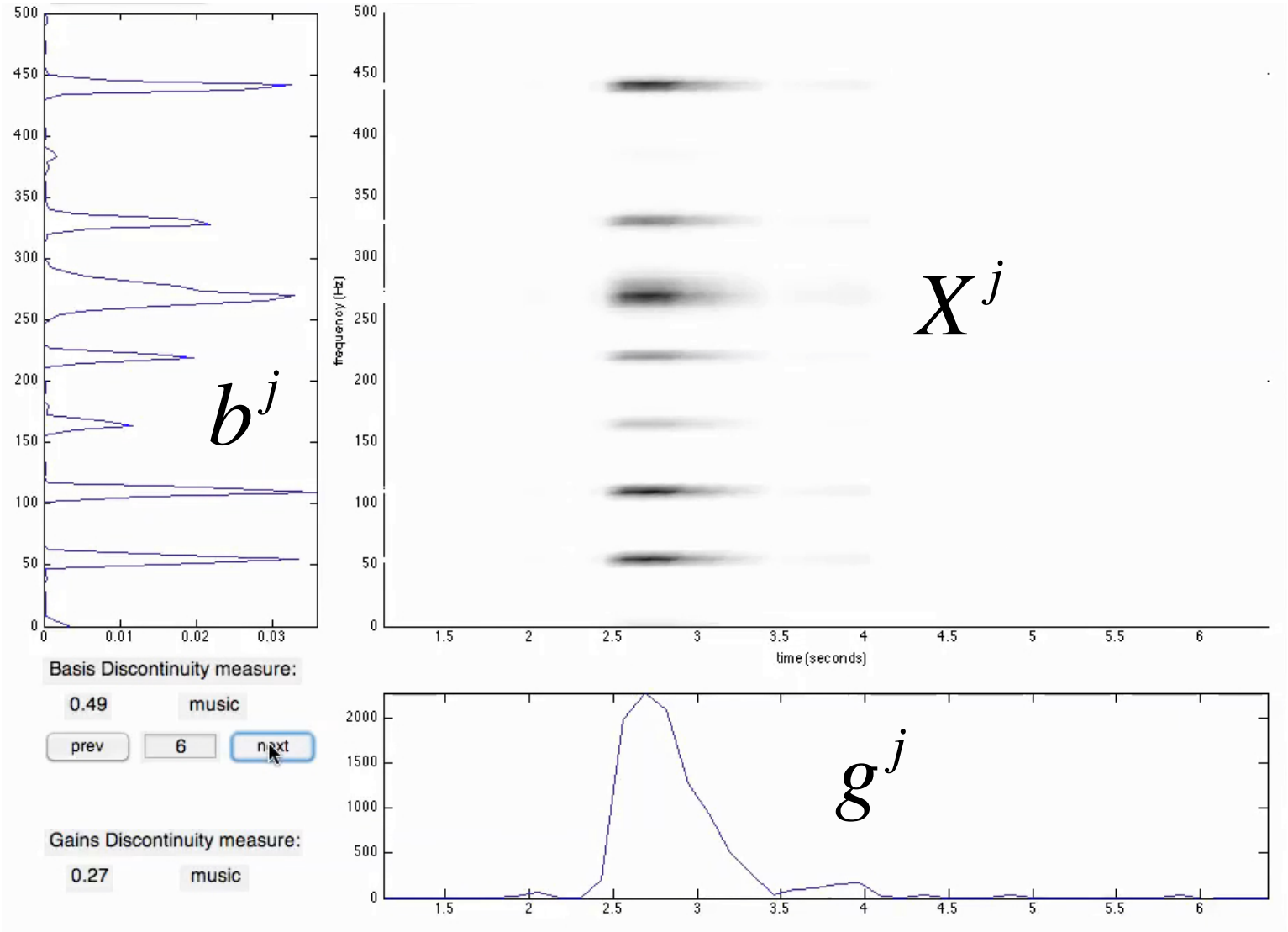
NMF of The Magnitude Spectrogram X

$$X \approx \sum_j X^j$$

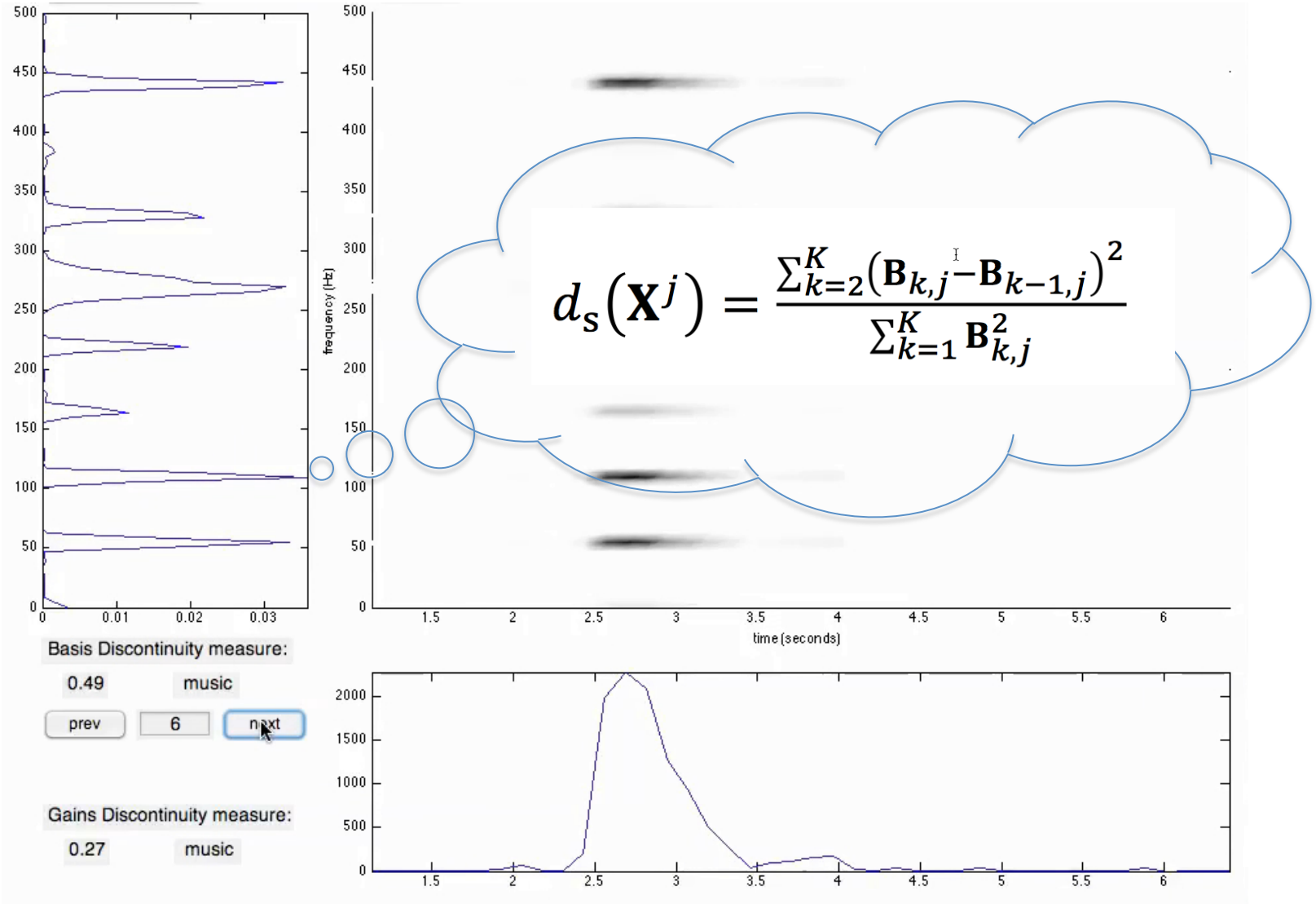
$$X \approx BG$$

$$X^j = b^j g^j$$

The Component $X^j = b^j g^j$

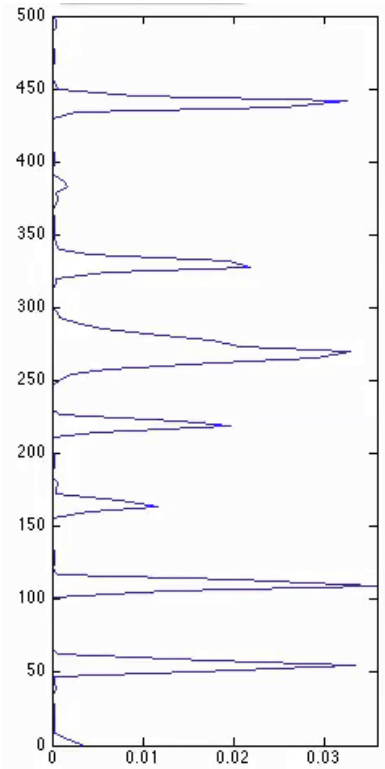


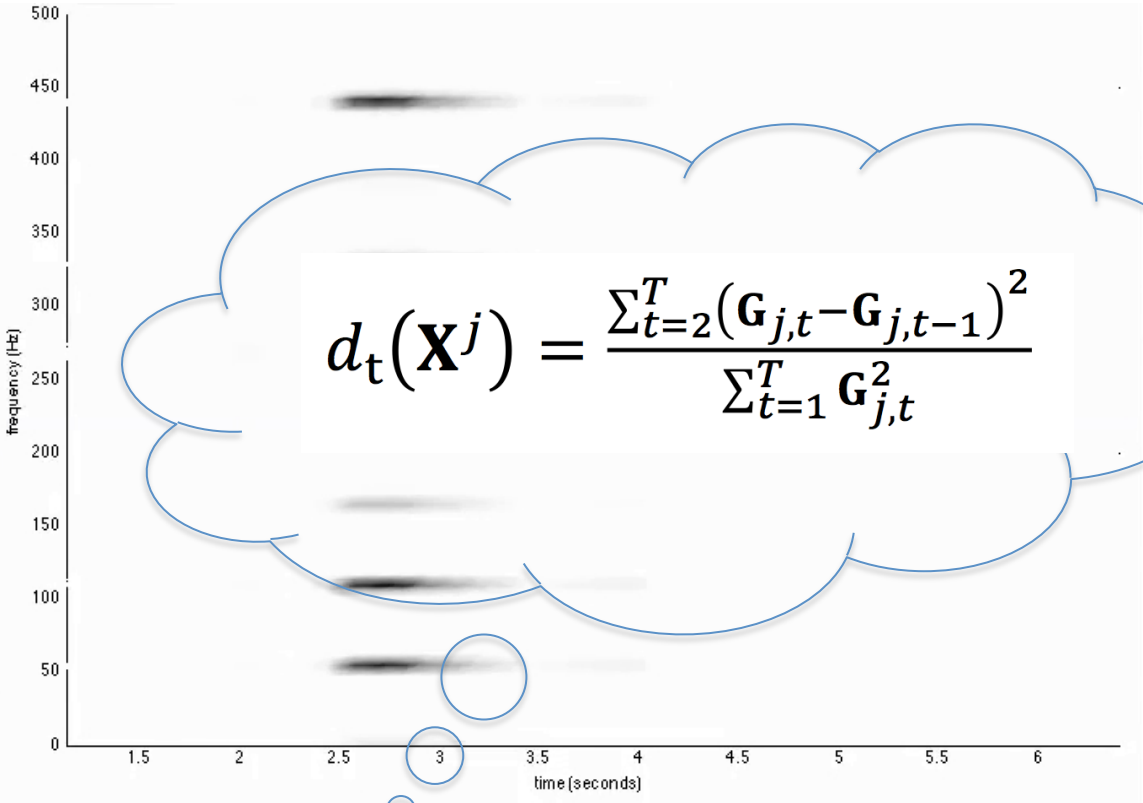
Classifying the Component to H/P using DM (Zhu)



$$d_s(\mathbf{X}^j) = \frac{\sum_{k=2}^{K}(\mathbf{B}_{k,j} - \mathbf{B}_{k-1,j})^2}{\sum_{k=1}^{K}\mathbf{B}_{k,j}^2}$$

Classifying the Component to H/P using DM (Zhu)





$$d_t(\mathbf{X}^j) = \frac{\sum_{t=2}^{T}(\mathbf{G}_{j,t}-\mathbf{G}_{j,t-1})^2}{\sum_{t=1}^{T}\mathbf{G}_{j,t}^2}$$

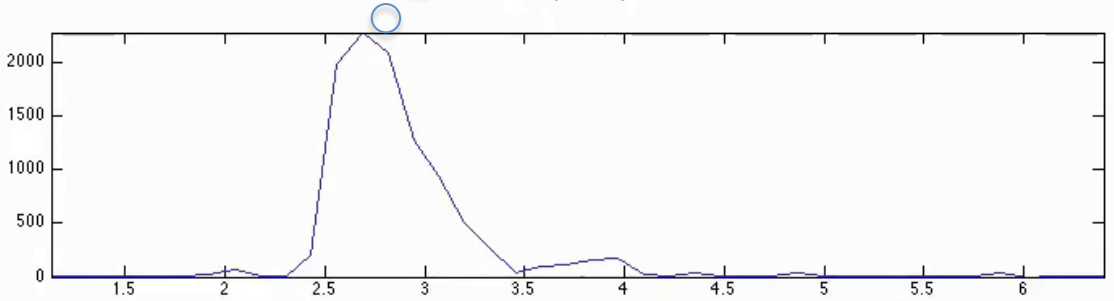
Basis Discontinuity measure:

0.49 music

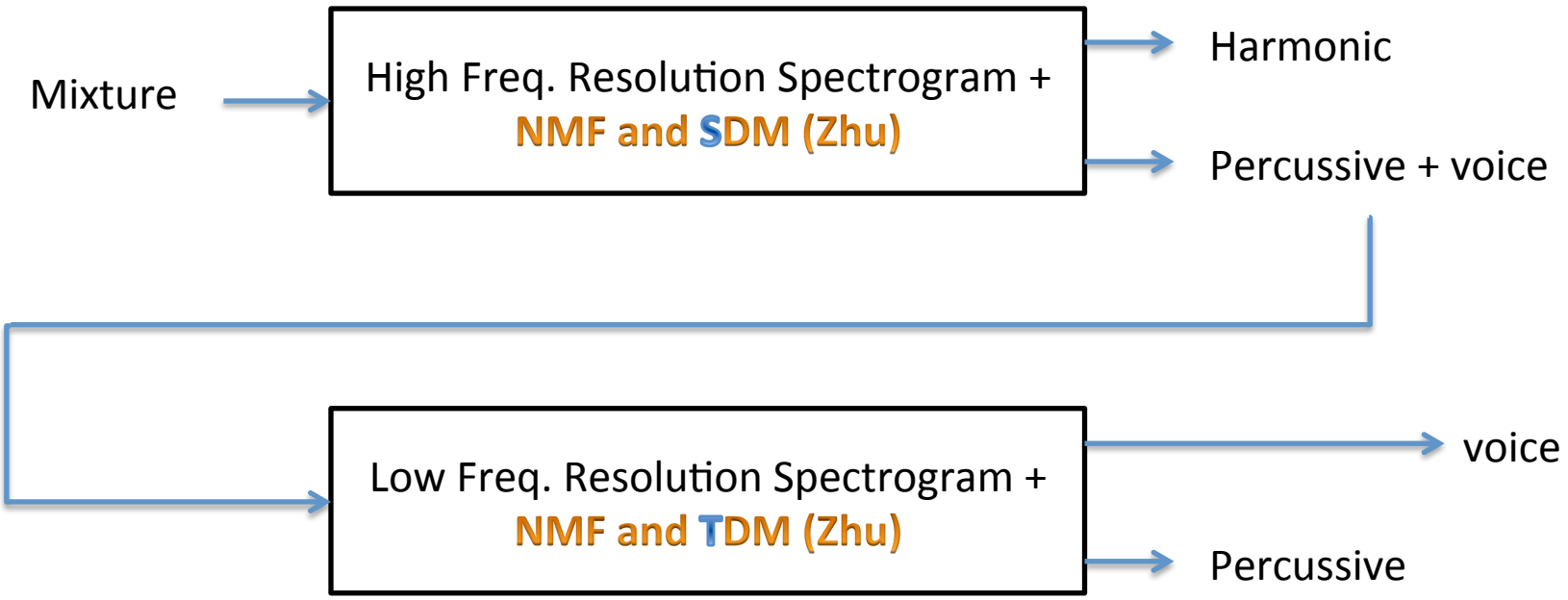
prev 6 next

Gains Discontinuity measure:

0.27 music



Separating Singing Voice



The Proposed Algorithm

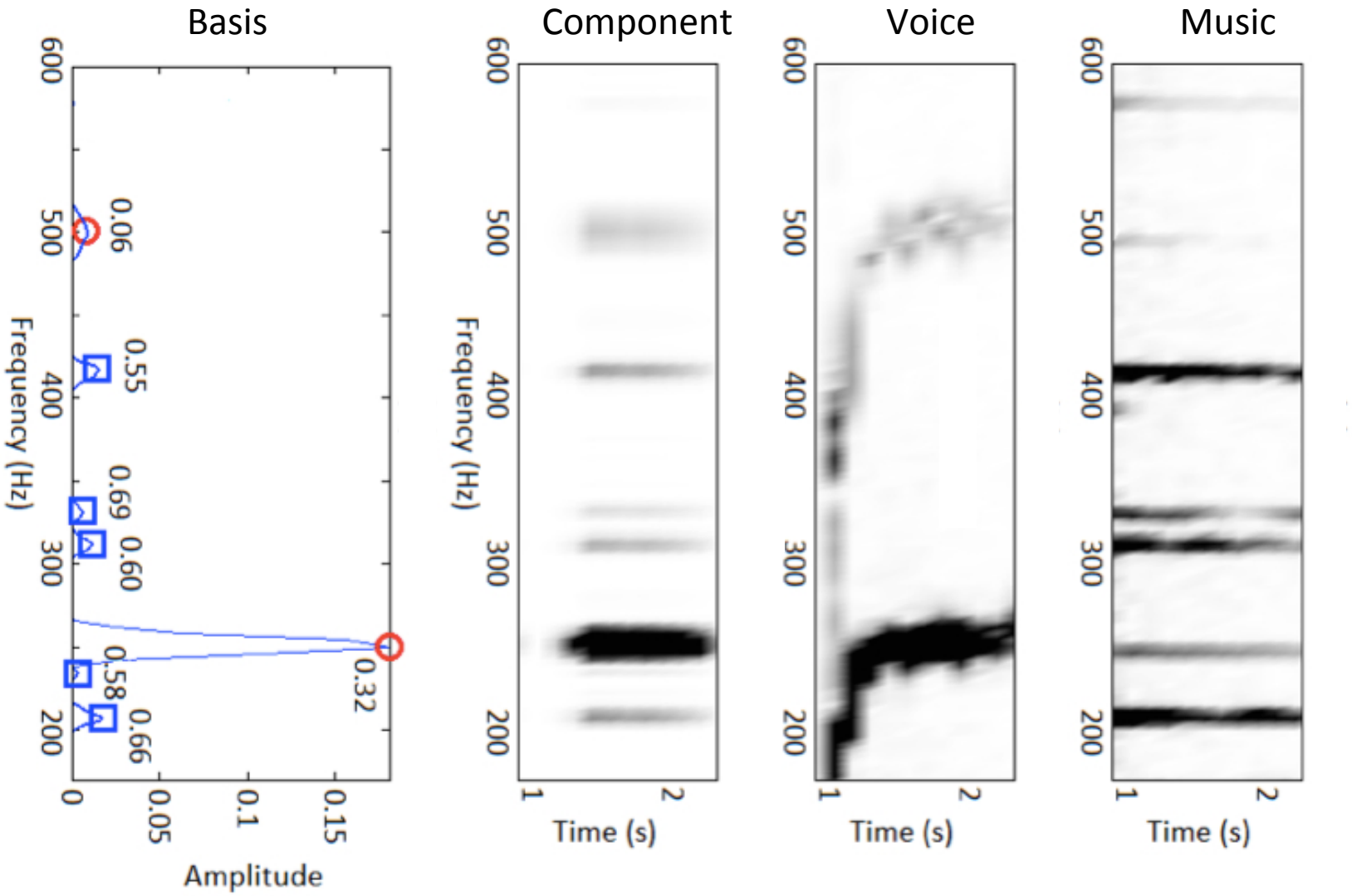
The Baseline Algorithm

1- Uses GDM to classify each component as a whole as either H or P

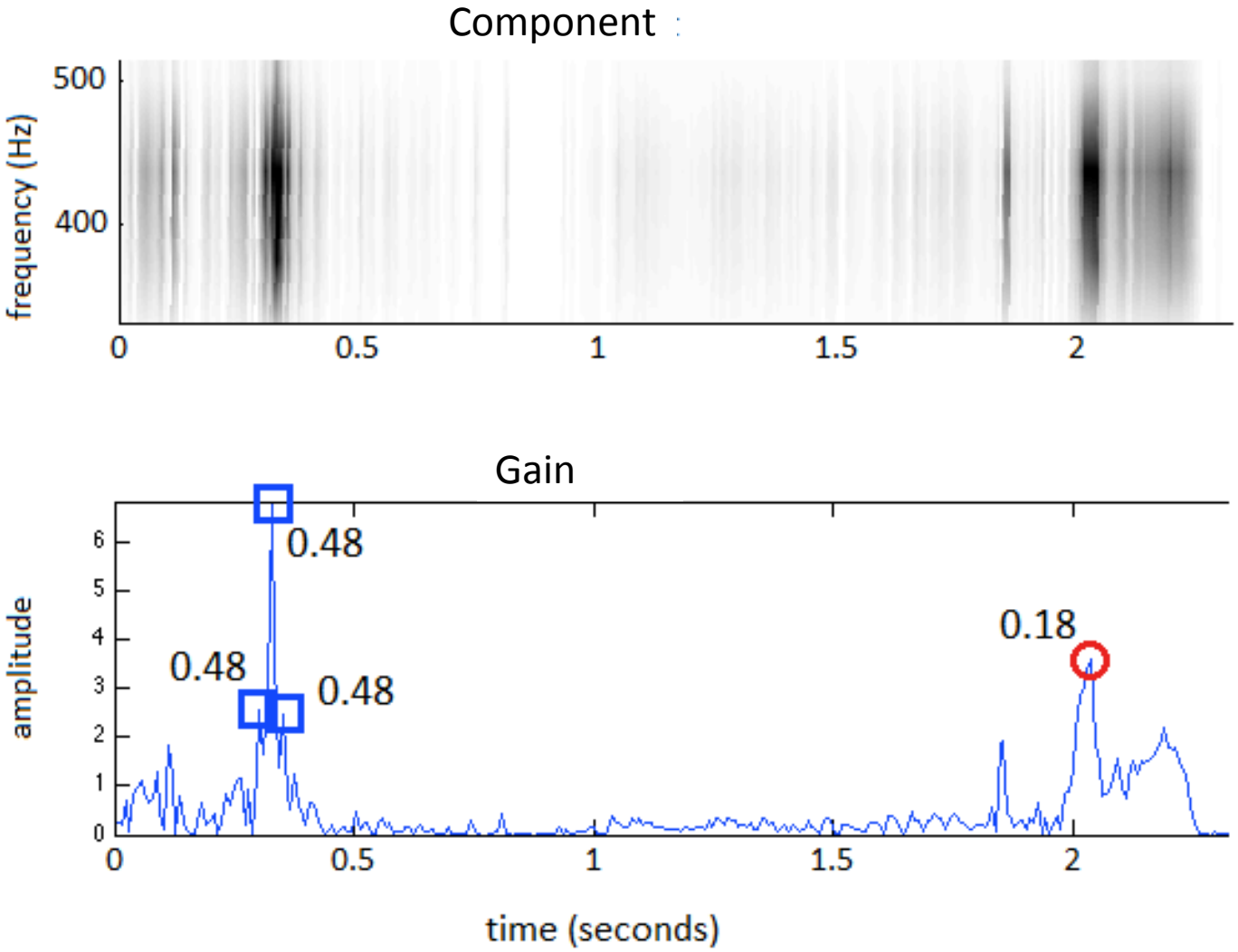
The Proposed Algorithm

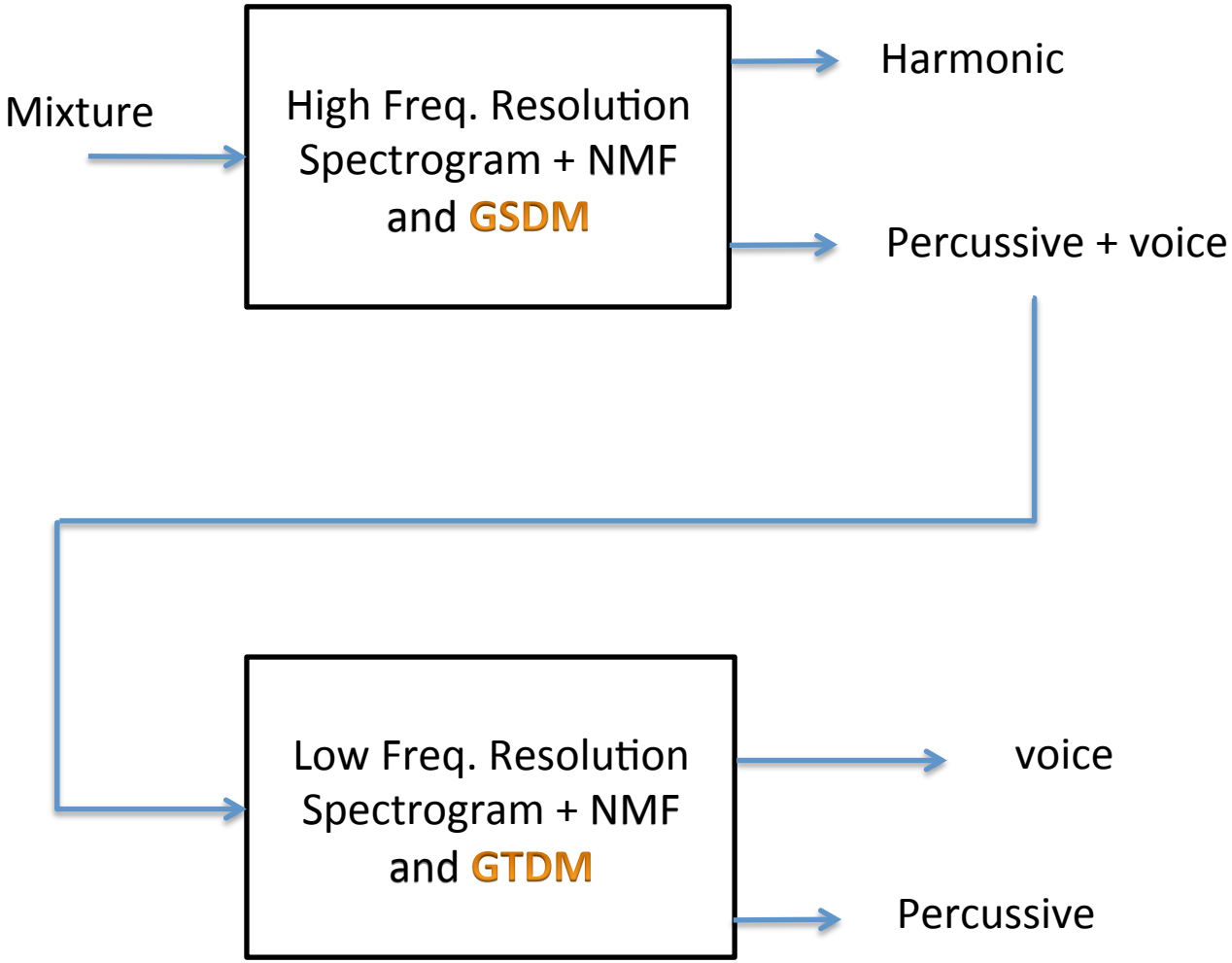
1- Uses LDM to decompose each component into two components one H and the other P

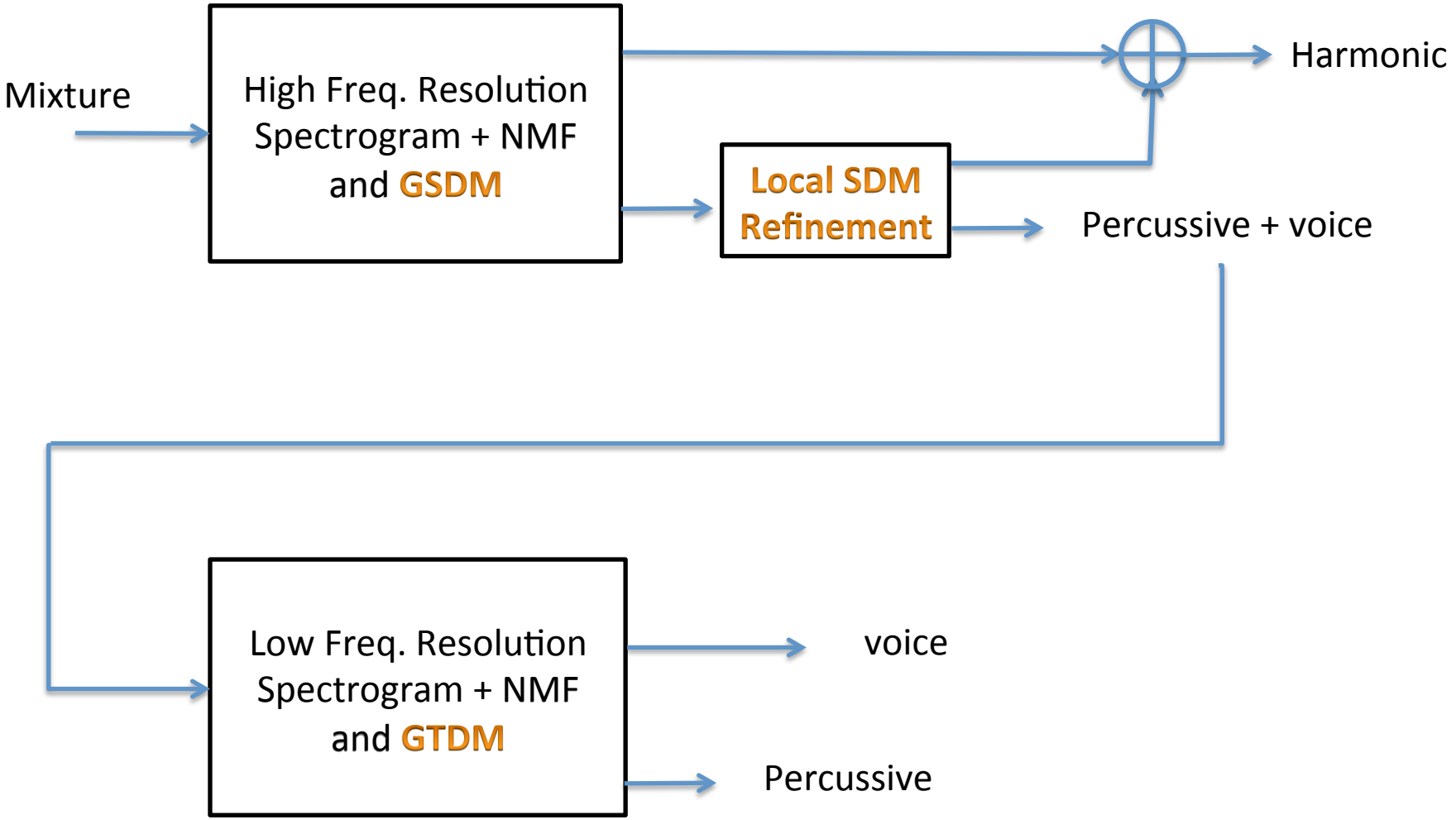
Local Spectral Discontinuity Measures

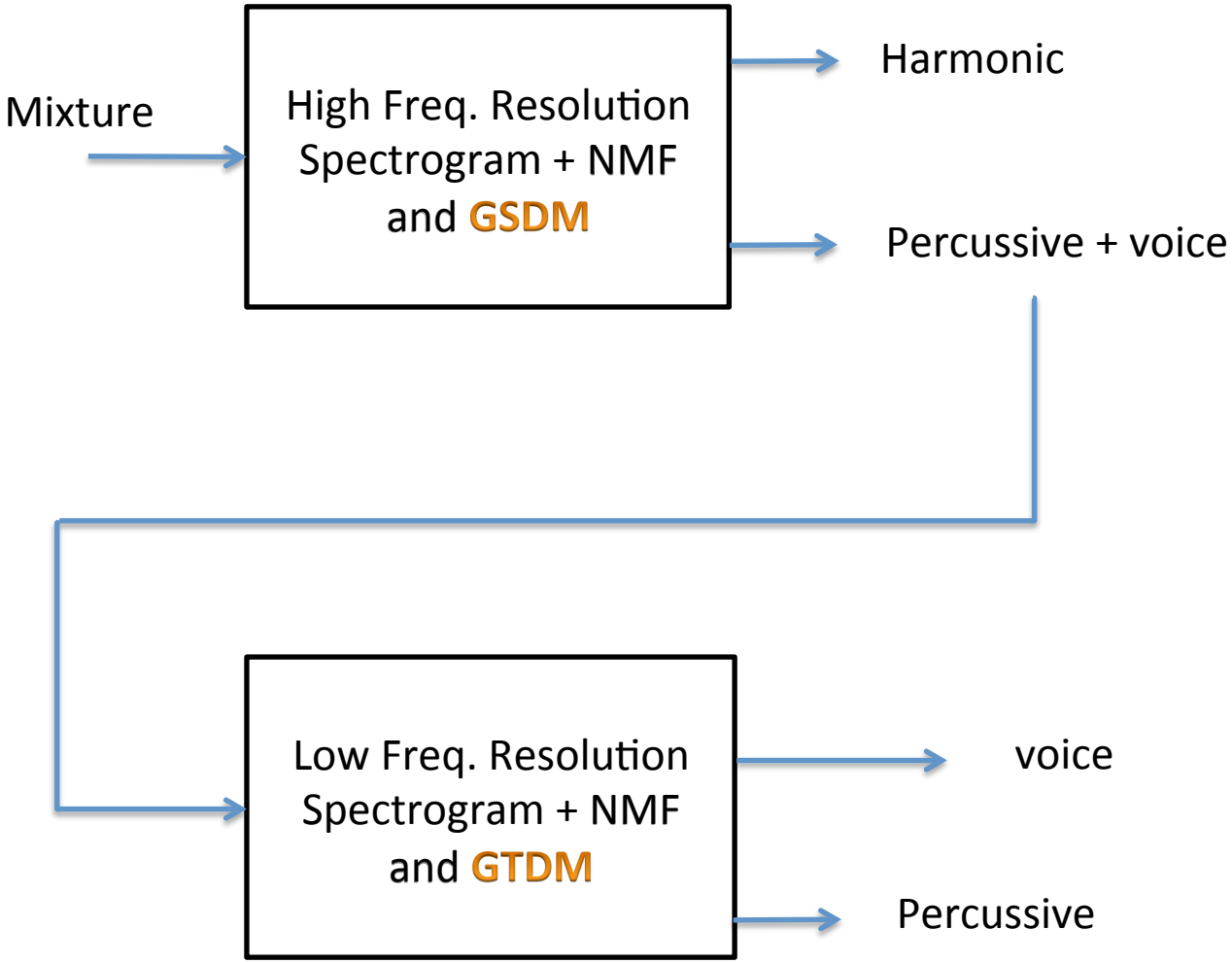


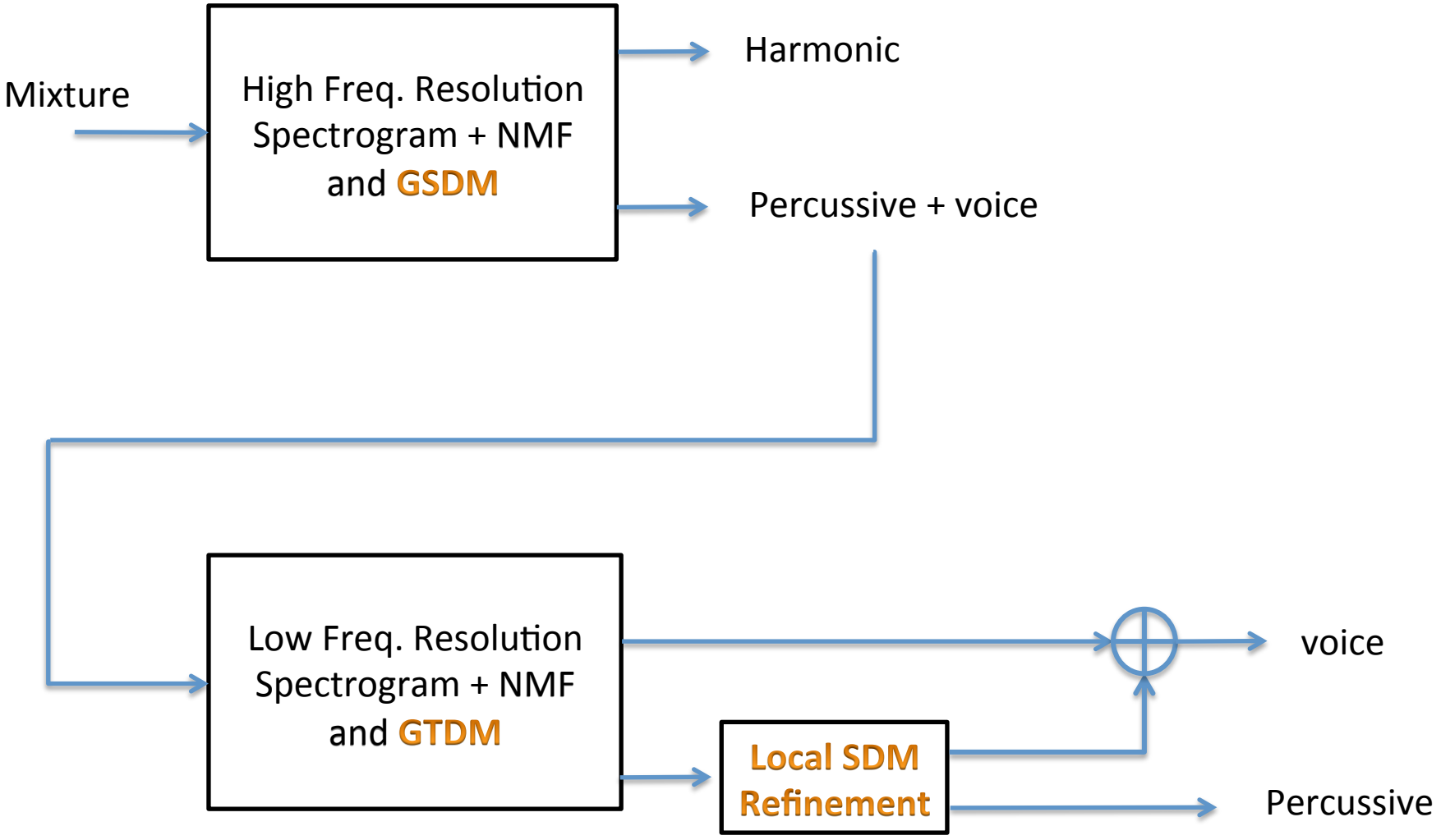
Local Temporal Discontinuity Measures

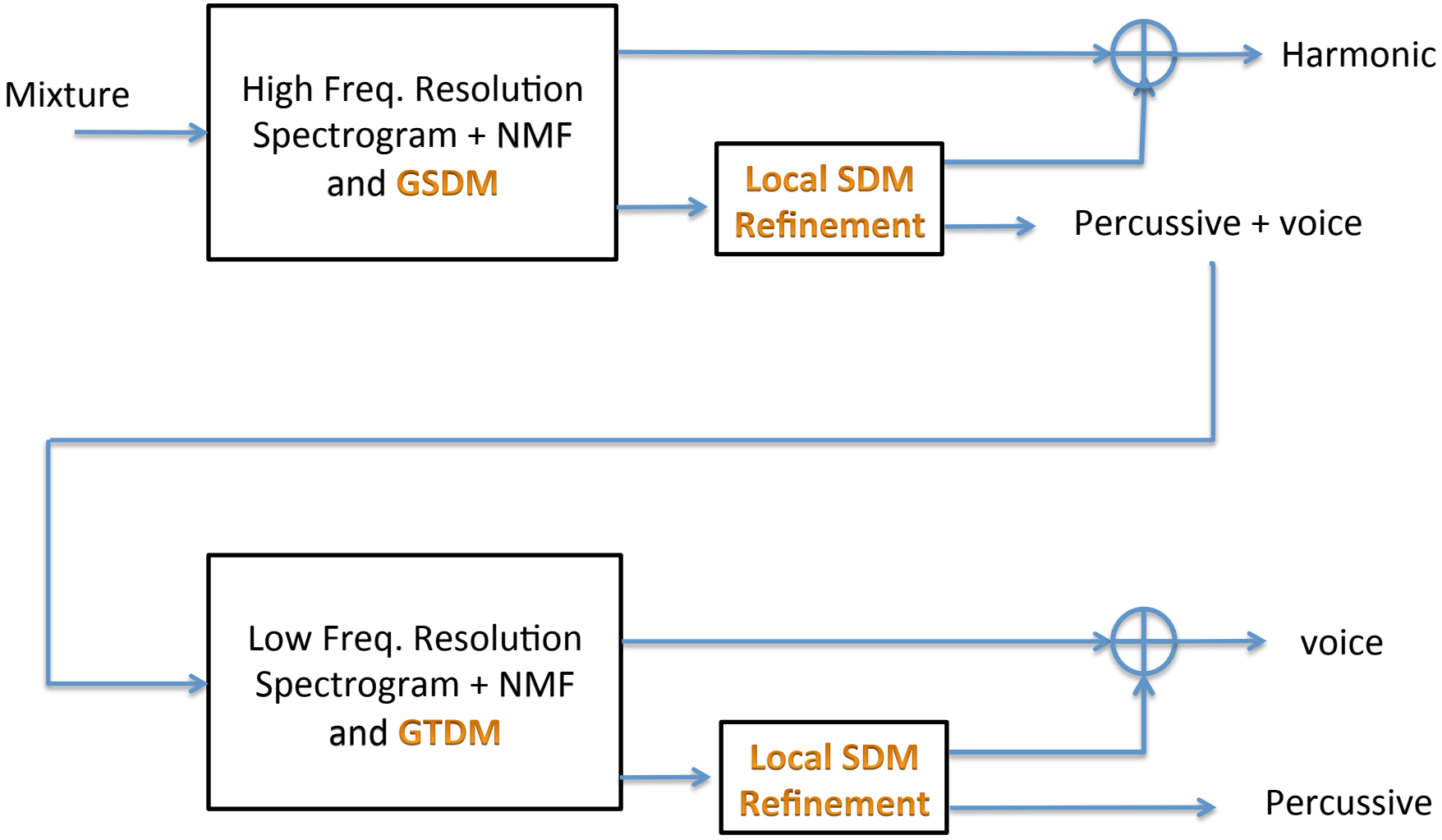



Mixture
High Freq. Resolution Spectrogram + NMF and GSDM
Harmonic
Percussive + voice
Low Freq. Resolution Spectrogram + NMF and GTDM
voice
Percussive

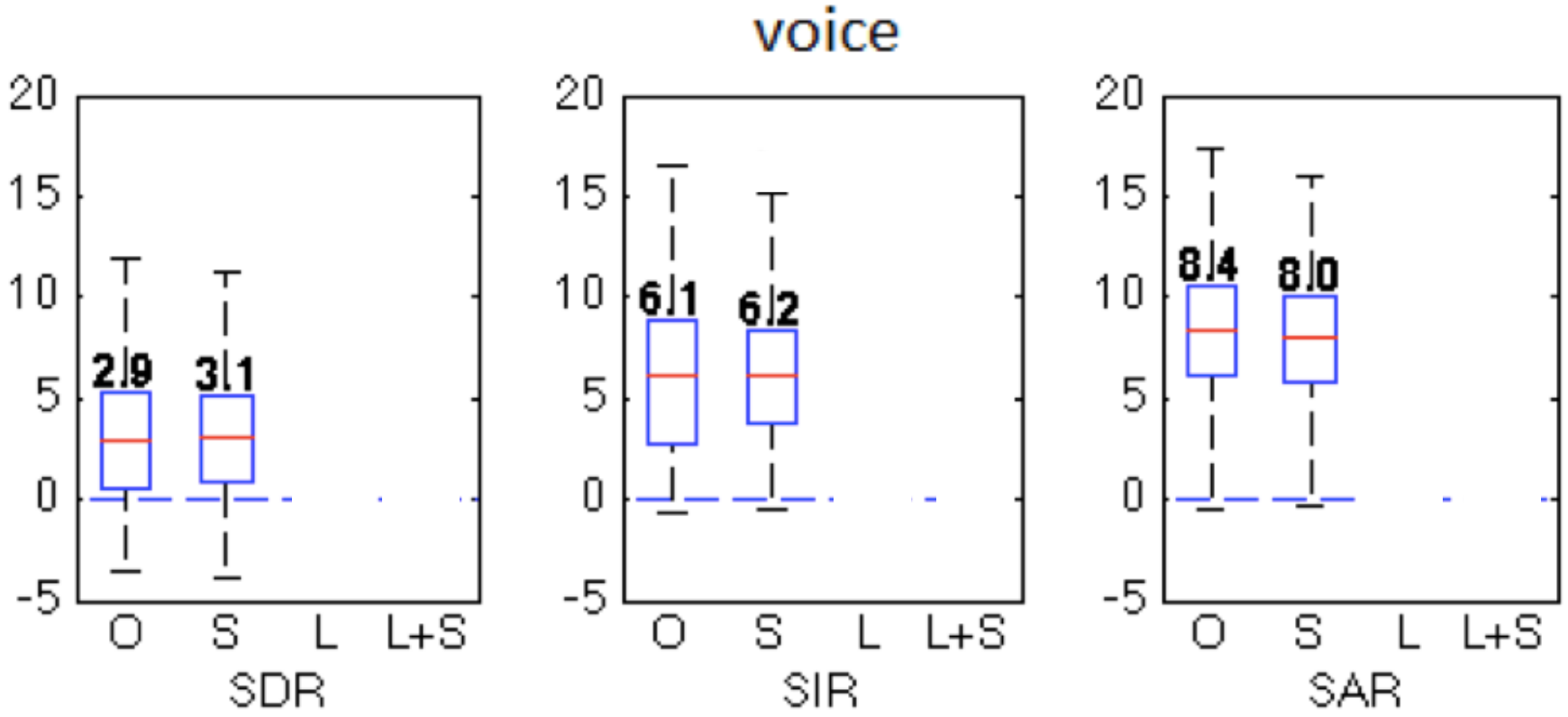

Mixture
High Freq. Resolution Spectrogram + NMF and GSDM
Harmonic
Local SDM Refinement
Percussive + voice
Low Freq. Resolution Spectrogram + NMF and GTDM
voice
Percussive


Mixture
High Freq. Resolution Spectrogram + NMF and GSDM
Harmonic
Percussive + voice
Low Freq. Resolution Spectrogram + NMF and GTDM
voice
Percussive

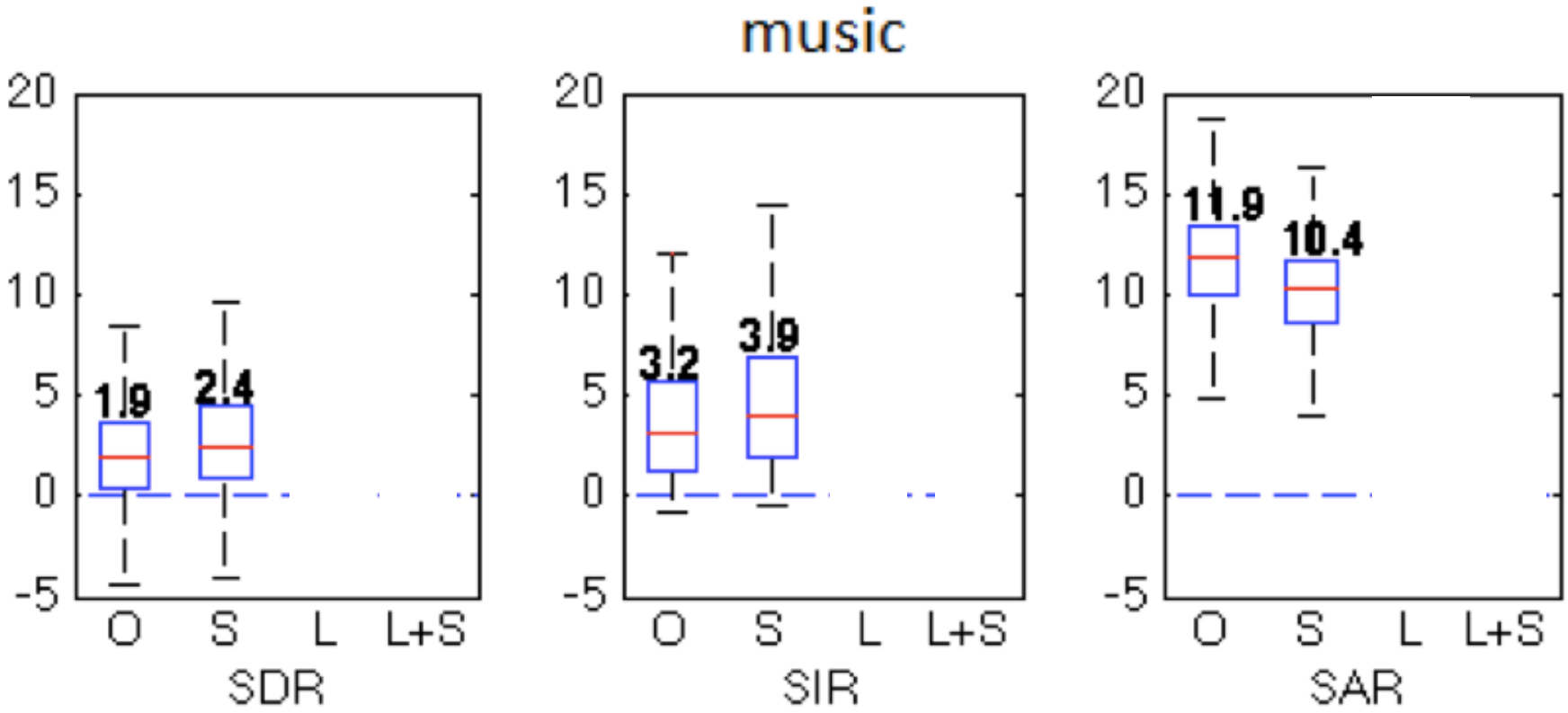

Mixture
High Freq. Resolution Spectrogram + NMF and GSDM
Harmonic
Percussive + voice
Low Freq. Resolution Spectrogram + NMF and GTDM
voice
Local SDM Refinement
Percussive


Mixture
High Freq. Resolution Spectrogram + NMF and GSDM
Harmonic
Local SDM Refinement
Percussive + voice
Low Freq. Resolution Spectrogram + NMF and GTDM
voice
Local SDM Refinement
Percussive

Results



Results



Thank you