

# “Polyphonic Music Transcription by Non-negative Sparse Coding of Power Spectra”

by Samer Abdallah and Mark Plumbley

presented by Yu Shiu  
ISE 575: 02/17/2005

## Review Paper

- Title:
  - “Polyphonic Music Transcription by Non-negative Sparse Coding of Power Spectra”
- Authors:
  - Samer A. Abdallah and Mark D. Plumbley
- Goal
  - Music transcription
    - It is difficult, if not impossible
  - Sparse coding
    - Certain kind of basis decomposition
  - Power spectra
    - Feature (observation)

## Basic Ideas

- Training phase
  - Decompose the training data (one or more music pieces) into a collection of atomic features
- Transcription phase
  - The features of a testing music piece are represented as
    - A linear combination of the atomic features with appropriate weight

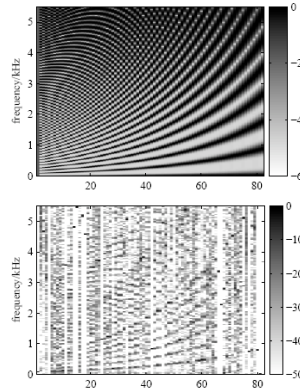
## Training Phase

- $X = A S$ 
  - $X$ :  $n$  by 1 column vector, Fourier spectra of music signals
  - $A$ :  $n$  by  $m$  matrix, each column represents an “atomic feature”
  - $S$ :  $m$  by 1 column vector, the weight of column vectors
    - Activity of atomic features

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} = [A_1, A_2, \dots, A_m] \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} = s_1 A_1 + s_2 A_2 + \dots + s_m A_m$$

## Training Phase (Cont)

- Before Training:  
Structured initial dictionary
  - Ordered collection of roughly pitched spectra on a quarter-tone spacing
  - The higher the frequency is, the broader the width is.
- After optimization of A & s
  - Alternate applications of (14) & (17)
  - (14): update s
  - (17): update A



## Training Phase (Cont) Mysterious (14) & (17)

- Independent Component Analysis (ICA)
  - Reducing the dependence among components, that is, atomic features in the dictionary matrix
- Sparse coding
  - Most observations can be encoded with only “a few” non-zero weights, “s”.
  - Only a few atomic features are needed to account for a note
  - Assuming a noisy generative model

## Training Phase (Cont) Estimation of Variance

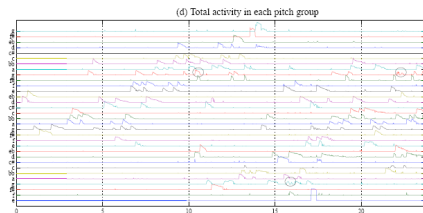
- Power Spectral Density (PSD)
  - Fourier Transform is used as the diagonalizing transform in  $\langle YY^T \rangle = UDU^T$
  - Pairs of eigenvectors have the same eigenvalues and span 2D subspaces
- Problem of spectral estimation is equivalent to that of ...
  - estimating variances of subspaces

## Training Phase (Cont)

- Pitch labeling on atomic features
  - Pitched or Non-pitched
  - Assigning “pitch” to each pitched spectra
- “Pitch group”
  - Many pitches are represented by more than one atomic feature

## Testing Phase

- Testing Data
  - J. S. Bach's Fugue in G minor No.16
- Activity of certain pitch
  - Summation of activities of elements (atomic features) in certain pitch group



## Discussion

- Unfinished work
  - End results: show activities of each pitch group
  - Need algorithm to pick the "pitched" atomic features
  - Needs robust peak-picking algorithm to assign certain segment to certain pitch values
  - Needs beat or measure detection to do note segmentation
- Mathematical?
  - Bright side: improve our math
- Piano only?

## Questions

- Plenty on math
- ???