Automatic Chord Recognition using Neural Networks

Audio Signal Processing

Ayoub Ghriss

MVA, 2017

Contents

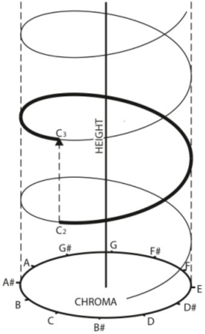
Introduction

A bit of music and history

- Automatic Chord Recognition, first start in 1991
- Classic methods : KNN, Logistic, HMM
- Leap in precision with the introduction of Neural Networks

Chords and Pitch

- Pitch : subjective perception of a note's height
- The pitch system is composed of 12 classes,
- An octave means a doubling of the frequency, each octave : $f_n = 2^{\frac{1}{12}} f_{n-1}$
- Define an equivalence relation between notes
- A chord is combination of 2 or more pitches



	Octave				
Note	2	3	4	5	6
C	66 Hz	131 Hz	262 Hz	523 Hz	1046 Hz
C♯/D♭	70 Hz	139 Hz	277 Hz	554 Hz	1109 Hz
D	74 Hz	147 Hz	294 Hz	587 Hz	1175 Hz
D♯/E♭	78 Hz	156 Hz	311 Hz	622 Hz	1245 Hz
E	83 Hz	165 Hz	330 Hz	659 Hz	1319 Hz
F	88 Hz	175 Hz	349 Hz	698 Hz	1397 Hz
F♯/G♭	93 Hz	185 Hz	370 Hz	740 Hz	1480 Hz
G	98 Hz	196 Hz	392 Hz	784 Hz	1568 Hz
G♯/A♭	104 Hz	208 Hz	415 Hz	831 Hz	1661 Hz
A	110 Hz	220 Hz	440 Hz	880 Hz	1760 Hz
A♯/B♭	117 Hz	233 Hz	466 Hz	932 Hz	1865 Hz
B	124 Hz	247 Hz	494 Hz	988 Hz	1976 Hz

Features Extraction

Short-Time Fourier Transform

- Discrete FT on equally spaced segments of the song
- Frequency mapping can be scaled (Mel, Logarithmic)

Drawbacks :

- Constant resolution and frequency difference
- Not suited to represent the pitch class concept

Constant Q Transform

$$f_k = f_{min}.2^{\frac{k}{B}} \tag{1}$$

where k is the frequency index,B is the number of bins per octave.

$$X(k;r) = \frac{1}{N_k} \sum_n x(nr)w(n)e^{j2\pi nQ/N_k} \tag{2}$$

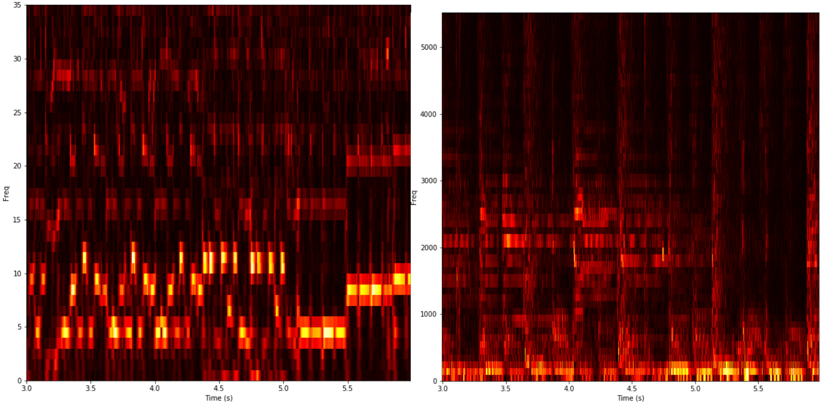
$Q = \frac{1}{2^B - 1}$, and the resolution$^{-1} N_k = [Q\frac{f_s}{f_k}]$

PCP, the first ACR System

- Can be seen as 1-bin Constant Q Transform
- Uses pitches pattern matching using NN and hand crafted score

Takuya Fujishima, Realtime Chord Recognition of Musical Sound

CQT vs STFT : who's the best ?



Muhammad Huzaifah, Comparison of Time-Frequency Representations.

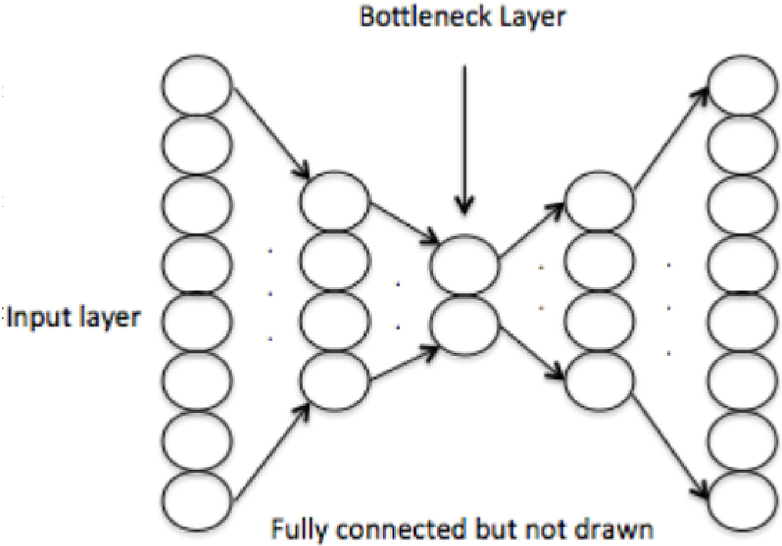
Features Processing

- Time slicing : concatenating adjacent frames
- First low pass filter : $y_n = \alpha y_{n-1} + (1-\alpha)x_n$
- A pair of low pass filters : exponentially weighted mean

$$\sum_{i=-r} ra^{-|i|}x[.+r]$$

- Other papers : Geometric mean/ Median filter

Learning Architecture


Bottleneck Layer
Input layer
Fully connected but not drawn

Deep Belief Networks

- Layers of fully connected layers of Restricted Boltzmann Machines
- A stochastic neural network :
- One layer of visible units : chords
- One layer of hidden units : latent variables
- the hidden units of layer i are the visible for the layer $i + 1$
- Seen as an out-dated model in the Deep community

Deep architectures

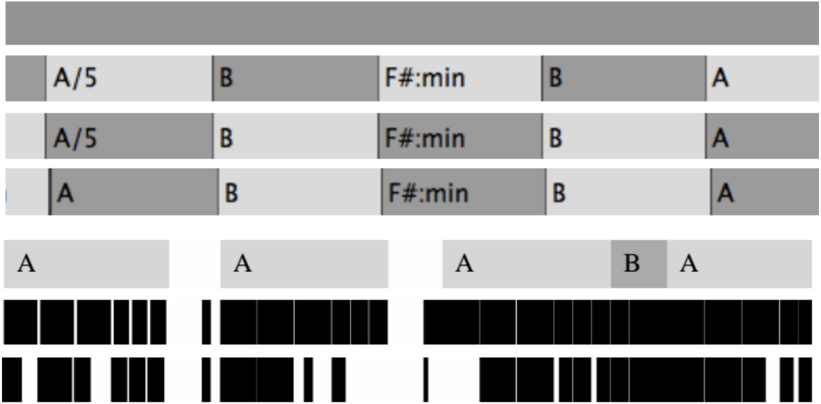
- Recurrent Neural Networks
  - Nicolas Boulanger-Lewandowski, Audio Chord Recogntion with Recurent Neural Networks.
- Convolutional NN:
  - Anis Rojb al., Music Transcription by Deep Learning with Data and "Artificial Semantic" Augmentation

Conclusion

Conclusions

- Improvement with the introduction with neural networks
- little difference in performance between different structures
- Smoothness of solutions to be improved

Matthias Mauch & Katy Noland & Simon Dixon (2009) Using Musical Structure to Enhance Automatic Chord Transcription.


A/5
B
F#:min
B
A
A/5
B
F#:min
B
A
A
B
F#:min
B
A
A
A
A
B
A