Fast Cover Song Retrieval in Advanced Audio Coding Domain based on Deep Learning Technique

Jiunn-Tsair Fang, Yu-Ruey Chang, and Pao-Chi Chang

Communication Engineering
National Central University
Taoyuan, Taiwan

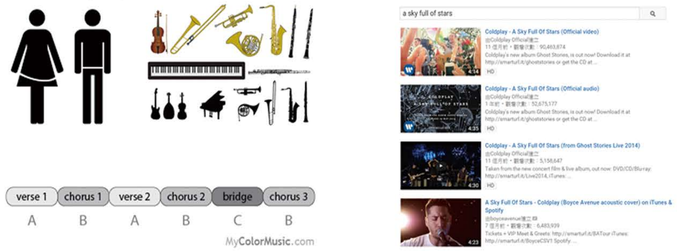
Abstract

- In this study, we proposed a novel deep neural network based **cover song retrieval** method, which can significantly reduce the computational complexity.
- The **MDCT coefficients** from the **Advanced Audio Coding (AAC)** were extracted and mapped into the **12-dimensional chroma features**.
- Chroma features were further segmented to preserve the melody of music.
- Each segment of chroma features was trained and learned to reduce its dimension by using an **autoencoder**, used for **deep learning of artificial neural networks**.

1. Introduction

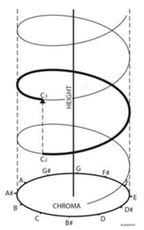
1.1 Cover Song Identification

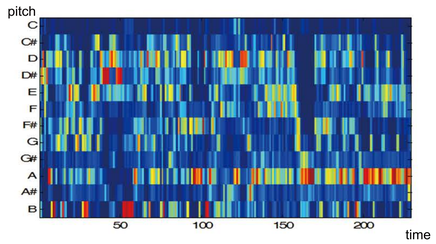
- Cover song is the novel rendition of the original work,
  which may differ from original song in performer, instruments or even the musical structure.
- Cover song may have lots of versions on the multimedia web.



Songs can be represented by the Pitch Class Profile (so called Chroma feature).

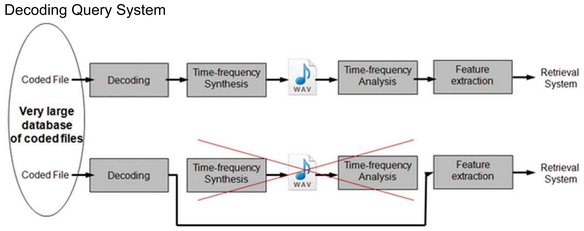
- The octave is divided into 12 pitch classes, which is motivated by 12-tone equal-temperament.
- Frequency of MDCT is mapped to 12-bin chroma
- Chroma features are the basis for most Music Information Retrieval (MIR) approaches to tonality.





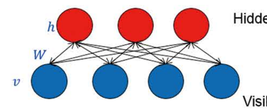
1.2 Partial Decoding in AAC domain

- Partial decoding saves the computation of transformation from MP3 to WAV.
- The Partial Decoding Query System

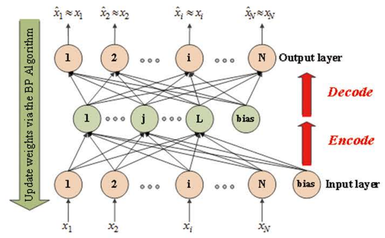


1.3 Deep Learning

- Restricted Boltzmann machines (RBMs)

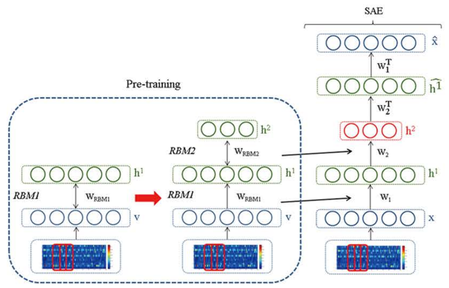


- Autoencoder (Nonlinear dimensionlity reduction)

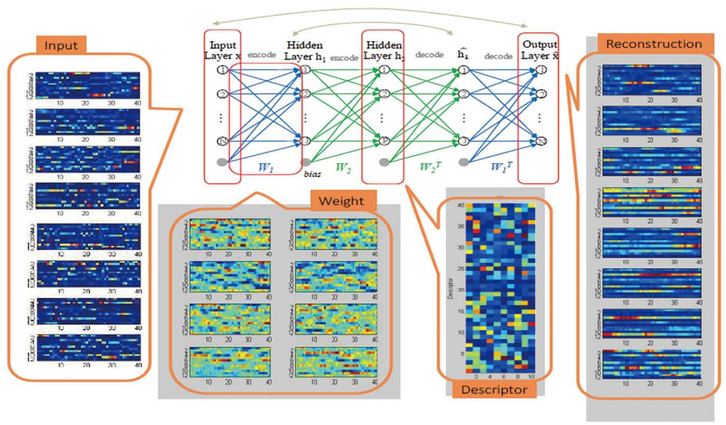


Deep learning model

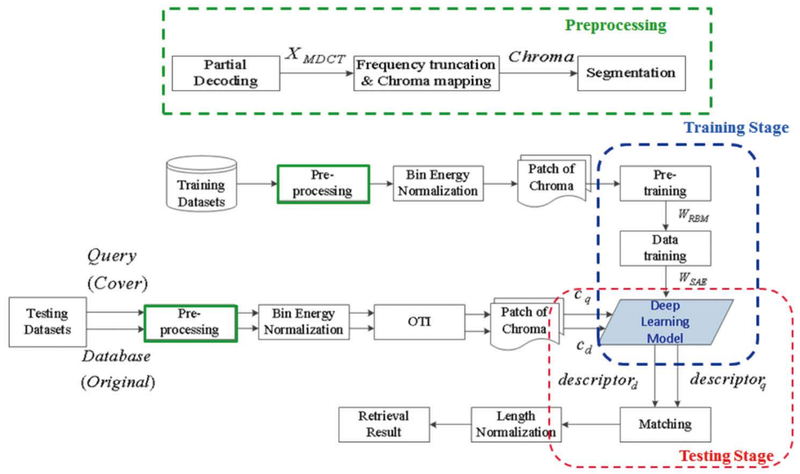
- Stacked RBMs to pretrain the weight.
- Sparse autoencoders to reduce dimension.
- The Stack of RBMs constitudes a SAE training model



- An example to illustrate the chroma features transformation by SAE.



2. Proposed Cover Song Retrieval System Diagram



3. Experimental Results

Experimental Environment

- **Hardware environment**
  CPU : Intel Core i7@3.4GHz
  RAM : 8GB Ram DDR3-1666MHz
- **Software environment**
  DeepLearnToolbox
  MATLAB R2013a
- **Training Dataset**
  Western and eastern pop music: 698 songs
- **Testing Dataset**
  Covers80 [3]: Western pop music stored in the MP3 format
  80 song sets: Original and Cover versions
  All audio files are converted to the AAC format at a sampling rate of 16 kHz for compressed issue

- **Performance measurement**
  - Mean Reciprocal Rank (MRR)

Query	Top 3 Returns	Rank	Reciprocal Rank
●	■ ▲ ●	3	1/3
▲	● ▲ ■	2	1/2
■	■ ▲ ●	1	1/1

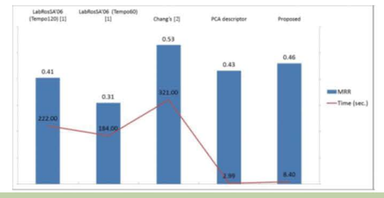
Top 3 MRR=(1/3+1/2+1/1)/3=0.61

- Performance of proposed method under different segment of frames

		Testing Data (frame/segment)									
		8		12		16		20		24	
Training (frame/segment)	Descriptor	MRR	Time	MRR	Time	MRR	Time	MRR	Time	MRR	Time
8	8	0.430	19.165	0.426	8.214	0.420	4.503	0.408	2.840	0.371	1.964
12	12	0.430	19.274	0.426	8.294	0.420	4.532	0.408	2.833	0.371	1.941
16	16	0.438	20.025	0.422	8.585	0.422	4.678	0.386	2.951	0.397	2.000
20	20	0.442	19.638	**0.463**	**8.401**	0.449	4.612	0.412	2.886	0.420	1.978
24	24	0.407	19.409	0.401	8.305	0.419	4.583	0.389	2.874	0.428	1.978

- Performance comparison with other systems from cover80

	MRR	Matching Time (sec.)
LabROSA'06 (Tempo120) [1]	0.4051	222
LabROSA'06 (Tempo60) [1]	0.3095	184
C.B.Hsiehs [3]	0.53	321
PCA descriptor	0.432	2.993
Proposed SAE descriptor	**0.463**	**8.4**



Conclusion

- In our work, we used the MDCT coefficients from AAC compressed files to represent chroma feature **without full decoding**.
- We used the RBM to **pre-train** the weight of SAE, and used SAE to **reduce dimensionality** in our cover song retrieval.
- Experimental results showed that the proposed method increased **MRR** to be **0.46**, and reduced approximately **95%** of the **matching time** as compared with traditional approaches.

References

[1] D. P. W. Ellis and G. E. Poliner, "Identifying cover songs with chroma features and dynamic programming beat tracking," in Proc. Int. Conf. Acoustic, Speech and Signal Processing, Honolulu, HI, Apr. 2007, pp. 1429 –1432.

[2] T. M. Chang, E. T. Chen, C. B. Hsieh, and P. C. Chang, "Cover Song Identification with Direct Chroma Feature Extraction From AAC Files," in Proc. of Global Conference on Consumer Electronics, Tokyo, Japan, pp. 55-56, Oct. 2013.

[3] The Covers80 cover song data set, [Online]. Available: http://labrosa.ee.columbia.edu/projects/coversongs/covers80