Singer Identification

Bertrand SCHERRER

McGill University

March 15, 2007
Outline

1. Introduction
   - Applications
   - Challenges

2. Feature Extraction

3. Vocal/NonVocal Region Segmentation
   - GMM-based methods

4. Classification
   - GMM

5. Results

6. Conclusion
Singer Identification is to be (has been) applied on pop music mainly
Automatically label data for which no/or not much information is available ⇒ recognize the singer

- Distinguish between original version of a song and cover songs
- Copyright enforcement: recording companies could scan bootleg sites on the internet to check if there are any unauthorized recorded versions of a concert [Kim, 2002 and Tsai and Wang, 2006]
- Music recommendation systems could use singer identification to group singers with same voice characteristics.
- Automatically label data for which no/or not much information is available ⇒ recognize the singer
- Distinguish between original version of a song and cover songs
  - Copyright enforcement: recording companies could scan bootleg sites on the internet to check if there are any unauthorized recorded versions of a concert [Kim, 2002 and Tsai and Wang, 2006]
- Music recommendation systems could use singer identification to group singers with same voice characteristics.
- Automatically label data for which no/or not much information is available ⇒ recognize the singer
- Distinguish between original version of a song and cover songs
- Copyright enforcement: recording companies could scan bootleg sites on the internet to check if there are any unauthorized recorded versions of a concert [Kim, 2002 and Tsai and Wang, 2006]
- Music recommendation systems could use singer identification to group singers with same voice characteristics.
Automatically label data for which no/or not much information is available ⇒ recognize the singer

Distinguish between original version of a song and cover songs

Copyright enforcement: recording companies could scan bootleg sites on the internet to check if there are any unauthorized recorded versions of a concert [Kim, 2002 and Tsai and Wang, 2006]

Music recommendation systems could use singer identification to group singers with same voice characteristics.
Singing Voice = hybrid btw speech and musical instrument ⇒ create specific methods of analysis.

In pop music, voice is never heard alone: presence of accompaniment
• Singing Voice = hybrid btw *speech* and *musical instrument* ⇒ create specific methods of analysis.

• In pop music, voice is never heard alone: presence of *accompaniment*
Outline

1. Introduction
   - Applications
   - Challenges

2. Feature Extraction

3. Vocal/NonVocal Region Segmentation
   - GMM-based methods

4. Classification
   - GMM

5. Results

6. Conclusion
As seen in the previous diagrams: need to extract some features from the sounds.

Features used:
- MFCC (Mel-Frequency Cepstral Coefficient)
- MDCT (Modified Discrete Cosine Transform)
- LPCC (Linear Predictive Coding Coefficients)
- WLPCC (Warped ...)
- Cepstral Coefficients of the LPC spectrum
- LPMFCC (MFCC of the LPC spectrum)
As seen in the previous diagrams: need to extract some features from the sounds.

Features used:
- MFCC (Mel-Frequency Cepstral Coefficient)
- MDCT (Modified Discrete Cosine Transform)
- LPCC (Linear Predictive Coding Coefficients)
- WLPCC (Warped ...)
- Cepstral Coefficients of the LPC spectrum
- LPMFCC (MFCC of the LPC spectrum)
Outline

1. Introduction
   - Applications
   - Challenges

2. Feature Extraction

3. Vocal/NonVocal Region Segmentation
   - GMM-based methods

4. Classification
   - GMM

5. Results

6. Conclusion
Difference in spectrum between voiced regions and accompaniment-only: **harmonicITY** of the voice.
Fig. 1 [Tsai and Wang, 2006]
Tsai’s Approach

Fig. 1 [Tsai, 2004]
This method is supposed to yield 82.3% accuracy [Tsai and Wang, 2006]
Vocal/NonVocal Region Segmentation

GMM-based methods

Fujihara’s Approach

from Fig.1 [Fujihara 2005]
The GMM classification between Vocal and Non Vocal is done on the resynthesized signal.
Outline

1. Introduction
   - Applications
   - Challenges

2. Feature Extraction

3. Vocal/NonVocal Region Segmentation
   - GMM-based methods

4. Classification
   - GMM

5. Results

6. Conclusion
3 main strategies

- GMM
- SVM
- $k$-NN
GMM Method with Solo Voice Modeling

Fig.3 [Tsai and Wang, 2006]
Performance

- Kim and Whitman 2002 ⇒ 45%
- Liu and Huang, 2002 ⇒ 80 %
- Tsai and Wang, 2006, Fujihara et al., 2005 ⇒ 95%
Outline

1. Introduction
   - Applications
   - Challenges

2. Feature Extraction

3. Vocal/NonVocal Region Segmentation
   - GMM-based methods

4. Classification
   - GMM

5. Results

6. Conclusion
Singer identification yields satisfactory results.
But ...

- Only one article tackles Target Singer Detection or Target Singer Tracking: [Tsai and Wang 2006]. ⇒ results are not perfect for duet but are better than doing GMM without solo modeling.
- Specific to pop music ⇒ what happens with *a cappella* singers?
- Specific to geographical area (Asia) ⇒ important because of voice mix
But ...

- Only one article tackles Target Singer Detection or Target Singer Tracking: [Tsai and Wang 2006]. ⇒ results are not perfect for duet but are better than doing GMM without solo modeling.

- Specific to pop music ⇒ what happens with a cappella singers?

- Specific to on geographical area (Asia) ⇒ important because of voice mix
But ...

- Only one article tackles Target Singer Detection or Target Singer Tracking: [Tsai and Wang 2006]. ⇒ results are not perfect for duet but are better than doing GMM without solo modeling.
- Specific to pop music ⇒ what happens with **a cappella** singers?
- Specific to on geographical area (Asia) ⇒ important because of voice mix
But ...

- Only one article tackles Target Singer Detection or Target Singer Tracking: [Tsai and Wang 2006]. ⇒ results are not perfect for duet but are better than doing GMM without solo modeling.
- Specific to pop music ⇒ what happens with **a cappella** singers?
- Specific to on geographical area (Asia) ⇒ important because of voice mix


Questions ?