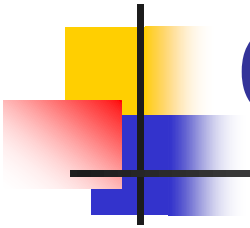
Comparative Study of Singing Voice Detection Methods

Shingchern D. You, Yi-Chung Wu
Department of Computer Science and Information Engineering,
National Taipei University of Technology
Taipei 106, Taiwan



Outline

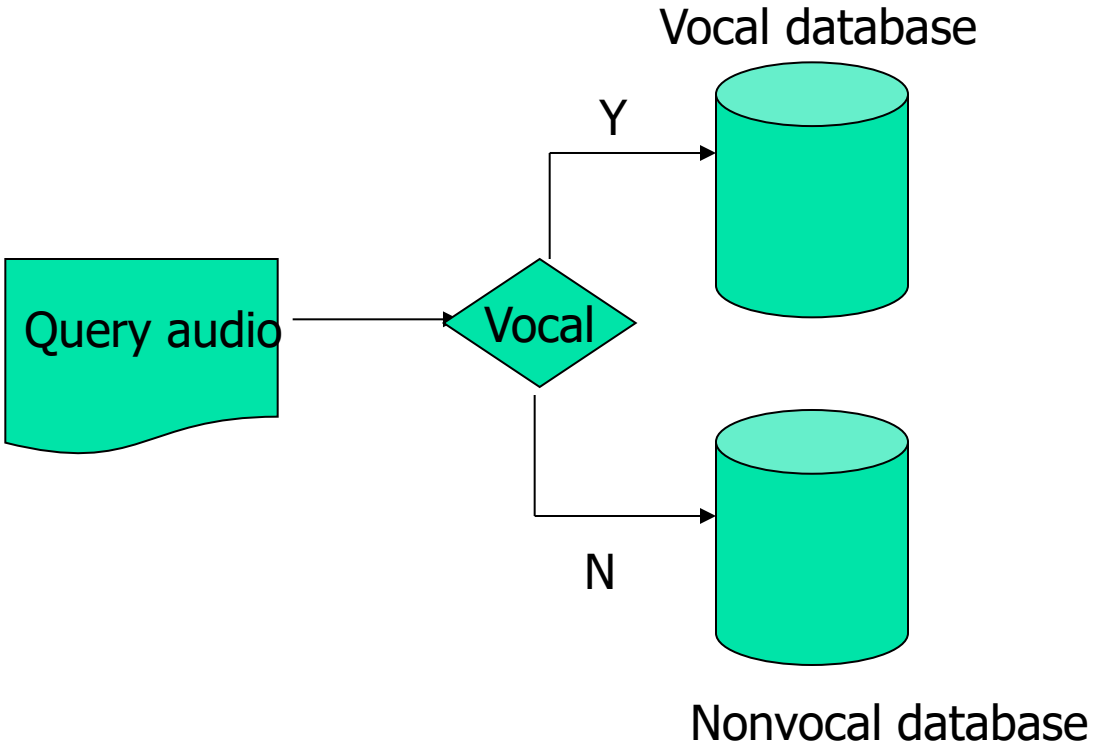
- Motivation
- Approaches
- Experiments & Results
- Conclusions

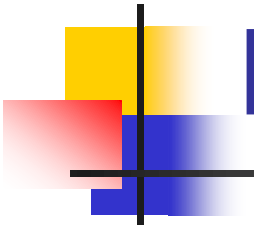
Motivation (1)

- Vocal 
- Nonvocal 

Motivation (2)

- Application scenario:



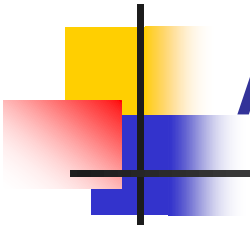


Motivation (3)

- There are many existing papers discussing this problem, but …
- None of them have a unified treatments to two distinct problems (i.e., a survey paper covers all with experiments)

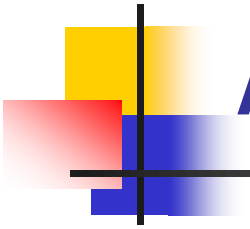
Motivation (4)

- Isolated segment of audio (like query)
  - vocal vs. instrument-only
  - vocal  vs. interlude 
- Identify vocal segments from an entire soundtrack



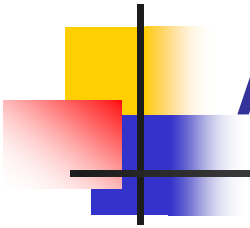
Approaches (1)

- Features examined in the experiments
  - MFCC (Mel Frequency Cepstral Coefficients)
  - LPCC (Linear Predictive Cepstral Coefficients)
  - LPC (Linear Predictive Coefficients)



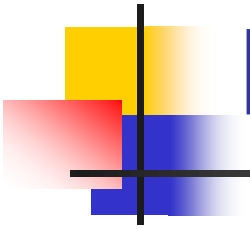
Approaches (2)

- Classifier: HMM (Hidden Markov Model)
  - Continuous observation densities
  - 10 states
  - Two models, one for vocal and the other for nonvocal



Approaches (3)

- Asking five questions and seeking answers to evaluate the performance of the features for various conditions



Experiments & Results (1)

- Dataset
- Isolated segments: the dataset contains 530 segments of 5-second music
- Entire soundtrack: 100 tracks

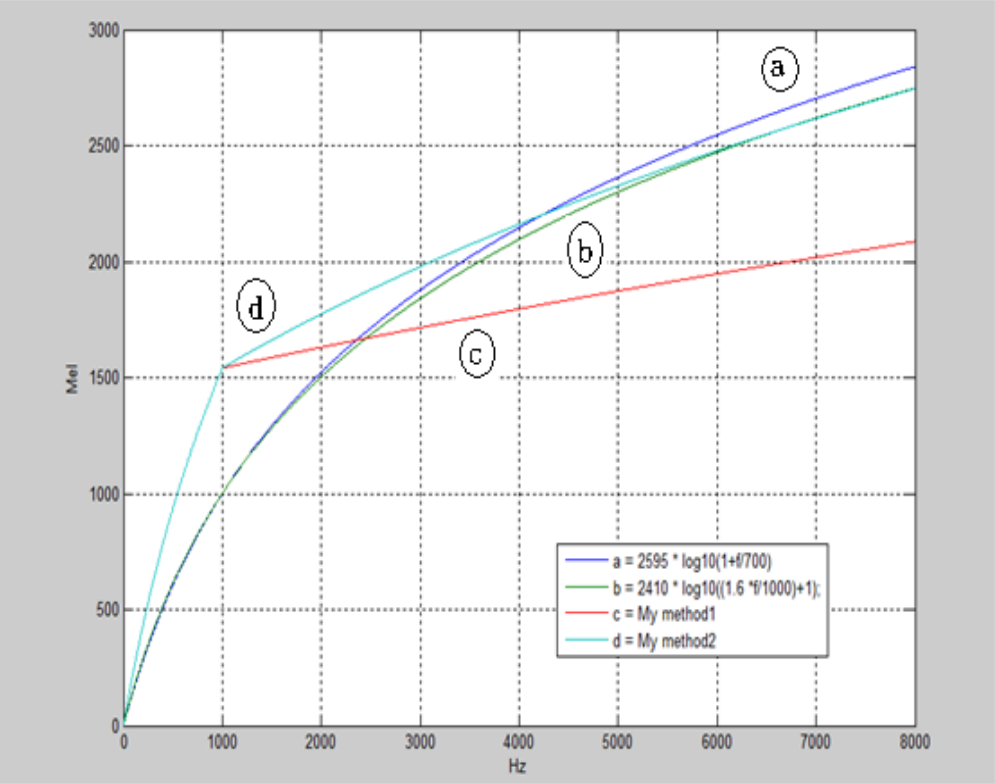
Experiments & Results (2)

- Q1: Between “vocal vs. instrument-only” and “vocal vs. interlude”, which class is more difficult to identify

	Vocal vs. Interlude	vocal vs. instrument-only
Accuracy (%)	89.9	94.2

Experiments & Results (3)

- MFCC curves can be “designed”



Experiments & Results (4)

- Q2: Among MFCC equations, which one yields higher detection accuracy (First two from literature and the other two from us)

	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (4)
Accuracy (%)	89.9	88.0	84.5	89.7

Experiments & Results (5)

- Q3: Among MFCC, LPC, and LPCC, which is more effective?

	MFCC	LPC	LPCC
Accuracy (%)	89.9	62.4	86.7

Experiments & Results (6)

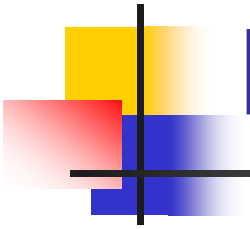
- Q4: Can detection accuracy be improved by combining two features?

	MFCC	MFCC+LPCC
Accuracy (%)	89.9	92.3

Experiments & Results (7)

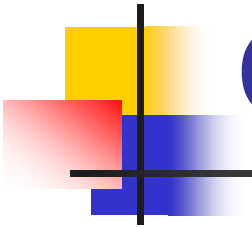
- Q5: Can post-processing improve detection accuracy for entire soundtrack?
- We use bootstrapping technique

	MFCC	Bootstrapping (A)	Bootstrapping (B)
Accuracy (%)	85.6	88.5	64.6



Experiments & Results (8)

- Bootstrapping works, but only apply once
- Placing segments identified by bootstrapping for training for isolated segment detection does not work



Conclusions

- It is a survey/review paper
- Give a brief introduction to vocal/nonvocal problem
- List five problems to investigate performance of various features in various conditions
- Conduct experiments to answer the questions