Phoneme Detection in Popular Music

Matthias Gruhne, Konstantin Schmidt and Christian Dittmar
Fraunhofer Institute for Digital Media Technology IDMT, Germany

Introduction
Motivation
Phoneme detection in polyphonic music is an important prerequisite for lyrics synchronization for karaoke applications or browsing in music catalogues.

Challenges
- Finding voiced phonemes in popular music is an ambiguous task due to interference with other instruments playing simultaneously.
- Vocal/Nonvocal detection is disregarded in this paper.

Statement of the problem
- The goal is to automatically identify the sung phonemes of the previously segmented music snippets.

Outline

Database
Labeled Data
- 30 second-snippets from 37 songs have been automatically extracted.
- All phonemes from this pieces have been manually labeled.
- 21 songs have been performed by male singers.
- 16 songs have been performed by female singers.

Evaluation
Settings and Evaluation
- Feature Extraction: 8 LPC, 8 WLPC, 8 PL and 13 MFCC features.
- Windowing: Hamming, 46ms, 23ms overlap.

Results
- The result measures are precision (relevant data to all data), recall (ratio of found data to relevant data) and the value of correct classified instances (ratio between correct classified entities and all entities).
- The value of the correct classified instances (CCI) showed best performance for male singers.
- There are significant differences between phoneme recognition results if harmonics analysis is used as preprocessing and not.
- The average precision and recall are twice as high, if a previous harmonics analysis is performed.

Harmonics Analysis
- Multiresolution Short-Time Fourier Transformation has been performed in order to receive a high resolution spectral representation in short processing time.
- The fundamental frequencies are detected based on [2] and the harmonic structure is extracted thereafter.

Feature Extraction and Classification

Feature Extraction
- A number of different common Low-Level features have been extracted.
- Used features are MFCC, LPC, PLP and Warped LPC.

Feature Classification
- All features have been performed by using different classifiers.
- As classifiers, Gaussian Mixture Models (GMM), Support Vector Machines (SVM) and Multilayer Perceptron (MLP) have been tested.

Table 1. Phonemes recognized in this system

Table 2. Songs in the database

References