Towards Improving Harmonic Sensitivity and Prediction Stability for Singing Melody Extraction

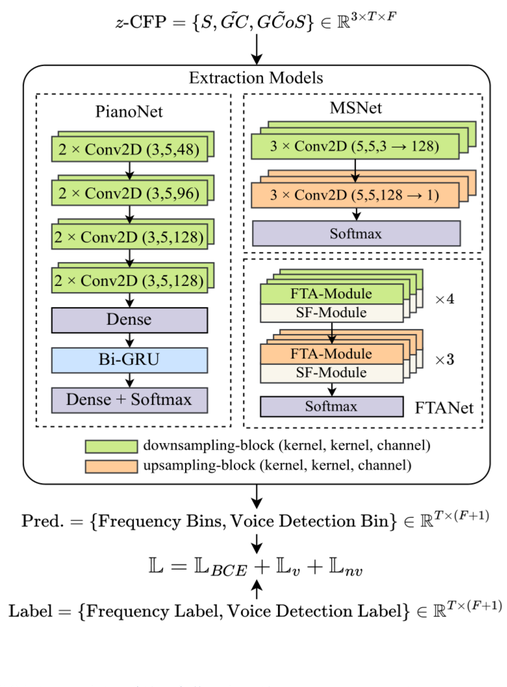
ISMIR 2023 Milan, Italy
Nov. 5-9, 2023

Keren Shao*, Ke Chen*, Taylor Berg-Kirkpatrick, Shlomo Dubnov
University of California San Diego

What is KKNet?

KKNet is a singing melody extraction model with:

- Feature designs on improving the harmonic sensitivity
- Loss designs on improving the prediction stability



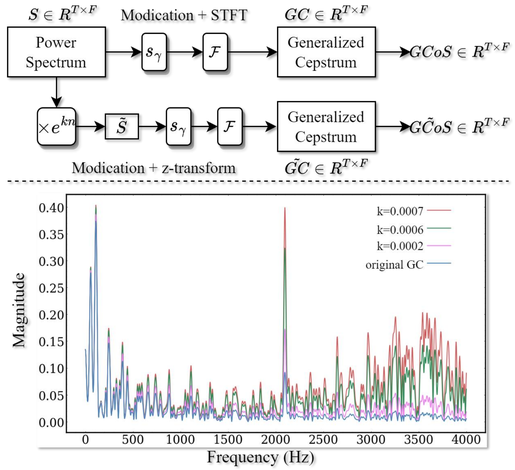
KKNet consists of the following three components:

- Z-CFP as input representation
- PianoNet as core extraction architecture
- Stability Loss Components as key converging objectives

I. Z-CFP

We propose Z-CFP, a revised representation based on CFP, as the model input.

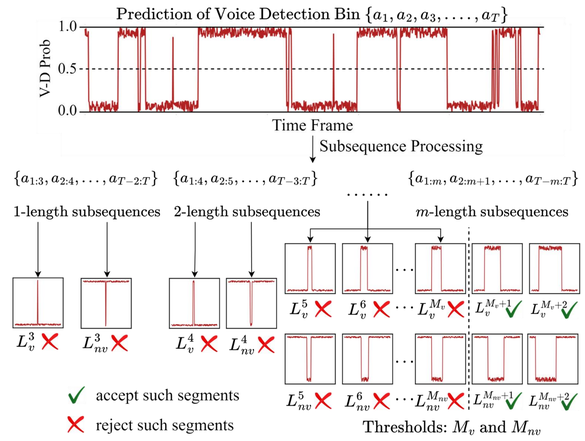
- Z-CFP is obtained by performing the discrete z-transform on the audio power spectrum
- Z-CFP captures the trailing harmonics as important indications for melody extraction



II. Stability Loss Components

We propose additional training objectives to enhance the prediction stability.

- Each component is setup as regularizers for spurious vocal and non-vocal segments
- Each component is design by a set of polynomials to identify these undesirable segments at or below the threshold duration



Experiments & Visualizations

- KKNet yields superior performance on three benchmark datasets of singing melody extractions
- KKNet avoids those spurious peaks whose length falls below the threshold, meanwhile maintains those long spurious segments.

Paper

Code