HARMONIC MODELING OF SINGING VOICE FOR SOURCE SEPARATION

Georgi Dzhambazov
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

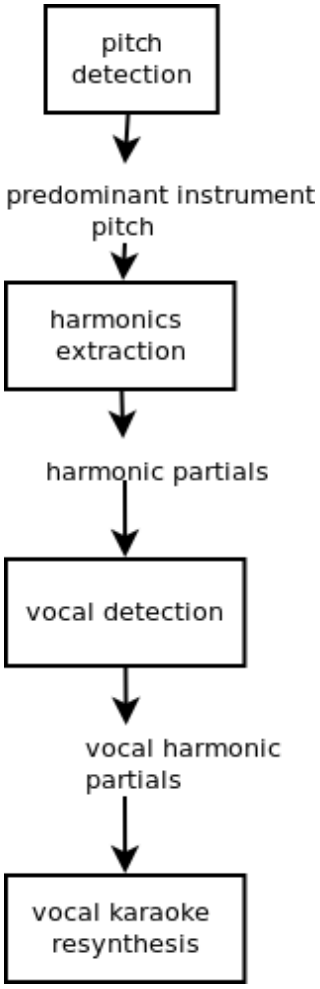
1. INTRODUCTION

In this work we suggest how to adapt an approach, originally suggested for modeling harmonic signals, to the task of singing voice separation. The idea of harmonic modeling is ... Since singing voice is highly harmonic we expect that a such a model is an appropriate choice to capture the spectral content of the main singing voice. However, a pitfall are predominant instruments with harmonic nature.

Recently there has been suggested a vocal detrection algorithm that decreases false positives, caused by instruments that have spectral characteristics similar to human voice [ref].

We suggest to combine the harmonic modeling with a vocal/non-vocal discrimination step to isolate segments, which have their origin in singing voice.

2. APPROACH



2.1 Pitch detection

The harmonic modeling requires as input the pitch of the source. We extract pitch contour regions of the predominant melodic source relying on [ref]. The algorithm performs in the same time a predominant source detection: regions with no predominant melody remain outside the detected contours.

The methodology assumes that there is no predominant melody in time intervals where the peaks of the pitch saliences are below a certain magnitude with respect to the mean of all the peaks. To overcome this we have adjusted some parameters to increase the recall of vocal regions . A side effect is that some instrumental parts increase the false positives rate. We used essentia [ref].

2.2 Harmonic modeling

The harmonic model of [ref] filters the spectral peaks corresponding to the first n harmonic partials of the singing voice.

2.3 Vocal detection

To discard the harmonic regions from background instruments, we have trained a random forest classifier on MFCC-based features, suggested by [ref]. We used 30-MFCCs and vocal variance. Vocal detection on timbral characteristics complements the harmonic-summation-based vocal detection of Melodia. Unlike [ref], we applied the vocal detector on the extracted harmonic spectrum rather than on the recorded mix. This is preferred because it has its origin in mostly one instrument

2.3.1 Training

The classifier is trained on harmonic spectrum as well:

-for each vocal segment or non-vocal segment from annotation features are extracted as if it were standalone audio. This assures that features with wide timespan (e.g. vocal variance) are extracted without interference from adjacent regions.

- the resynthesis is run on vocal track using reference MIDI and features are extracted from it.

2.4 Resynthesis

The vocal source is resynthesized by means of a constant overlapp add resynthesis [ref]

The karaoke is derived by a simple spectral mask [ref].

3. RESULTS

sbmitted:

	t = -70 dB		
nsdr	-0.1693 0.7744		
sir	8.1764 3.7464		
sar	1.5837 5.5095		

--

4. REFERENCES