

MIREX 2016: Singing Voice Separation by Harmonic Modeling

Georgi Dzhambazov, Xavier Serra

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

contact: georgi.dzhambazov@upf.edu

Idea

1. Model voice as a harmonic source

- Distinguish distorted harmonic partials

2. Mute non-vocal regions:

vocal detection (VD) based on combination of

- Saliency of estimated pitch
- Timbre of harmonic partials

MELODIA
(Salomon and Gómez, 2012)

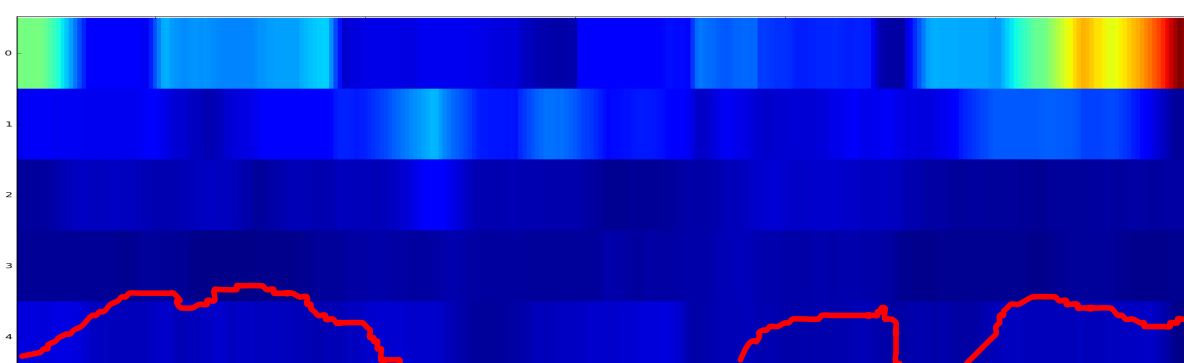
3. Resynthesis

- Vocal by extracted harmonics
- Background by spectral residual

$$Yh[k] = \sum_{r=1}^R A_r W[k - r\hat{f}_0]$$

(Serra, 1989)

MFCC variance



random forest classifier trained on

- MFCCs
- MFCCs variance
- spectral flatness

(Lehner et al, 2014)

Results

VD evaluation:

	before VD	after VD
recall	0.83	0.76
false alarms	0.35	0.22

Source separation evaluation:

- Normalized Signal-to-Distortion Ratio (NSRD)
- Signal-to-Interference Ratio (SIR)
- Signal-to-Artifacts Ratio (SAR)

	voice		accompaniment	
	mean	st dev	mean	st dev
NSDR	-2.281	3.534	0.395	1.470
SIR	6.562	9.778	1.984	9.805
SAR	2.394	4.562	2.708	2.661

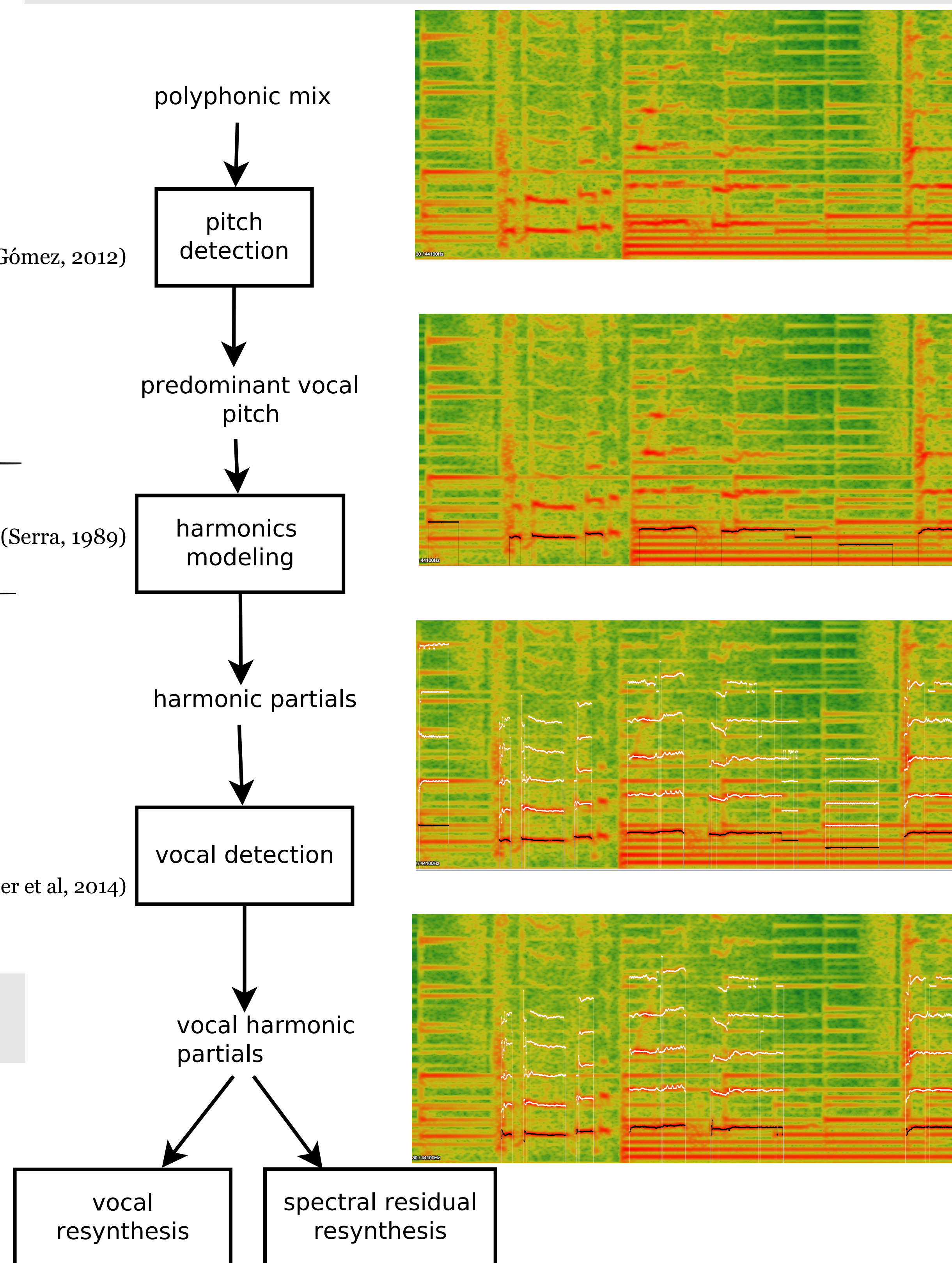
References

Salomon J. and Gómez, E. Melody extraction from polyphonic music signals using pitch contour characteristics. IEEE TSALP, 20(6):1759–1770, 2012.

Lehner B., Widmer G. and Sonnleitner. R. On the reduction of false positives in singing voice detection. In 2014 IEEE ICASSP, pages 7480–7484. IEEE, 2014.

Serra X., A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition. Technical report, 1989

Method overview



Conclusion

- Potential for improvement by refining VD
- Careful harmonic modeling results in high voice SIR
- Non-voiced consonants are not resynthesized

available in python at:

<https://github.com/georgid/vocal-detection>