AN ITERATIVE APPROACH TO MONAURAL MUSICAL MIXTURE DE-SOLOING

Jean-Louis Durrieu, Gaël Richard and Bertrand David

Institut Télécom, Télécom Paristech, CNRS LTCI

Introduction

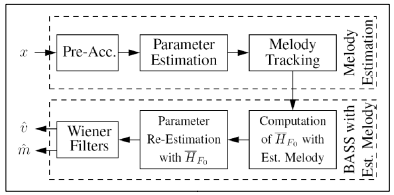
Separation (BASS) / indexing (MIR) hybrid method for mono-channel solo (or main instrument) / accompaniment separation in 2 steps:

- Main melody extraction [Durrieu08],
- Estimation of solo/accompaniment with melody.

Separation system characteristics:

- **NMF-based** power spectrum models,
- **Unsupervised** technique

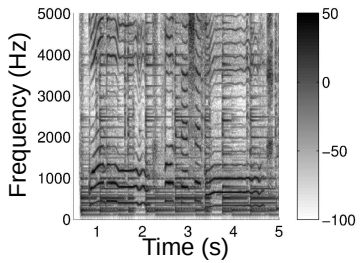
Separation system outline



MIXTURE X = Solo V + Accompaniment M

- Modeling power spectrum $S=|X|^2$ such that:

$$S \approx (W_K H_K) .* (W_{F_0} H_{F_0}) + W_R H_R$$

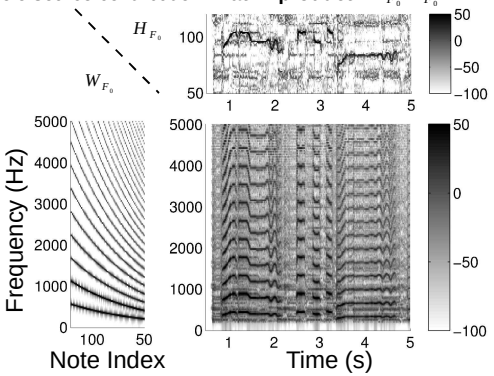


Tamy – *Que Pena Tanto Faz*

Solo V source/filter model:

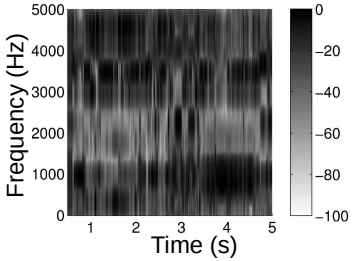
Solo source model:

- **Fixed** dictionary matrix W_{F_0} (spectral combs from glottal source model)
- **Estimated** amplitudes matrix H_{F_0}
- Solo source contribution: **matrix product** $W_{F_0} H_{F_0}$



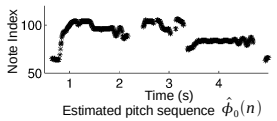
Solo filter model:

- **Estimated** dictionary of **smooth filters** W_K
- **Estimated** amplitudes H_K
- Filter contribution: $W_K H_K$



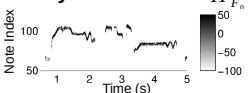
Melody estimation:

Viterbi smoothing on H_{F_0}



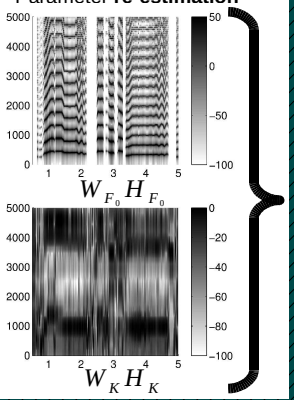
Solo re-estimation

- **Melody constrained** $\overline{H}_{F_0}$



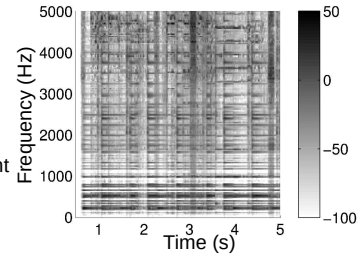
$$|f_0 - \hat{\phi}_0(n)| < 1/4\,tone \Rightarrow \overline{H}_{F_0}(f_0,n) = H_{F_0}(f_0,n),$$
$$otherwise \Rightarrow \overline{H}_{F_0}(f_0,n) = 0.$$

- Parameter **re-estimation**

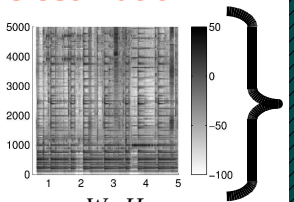


Accompaniment M model:

- **Estimated** dictionary W_R
- **Estimated** amplitudes H_R
- Accompaniment contribution: $W_R H_R$

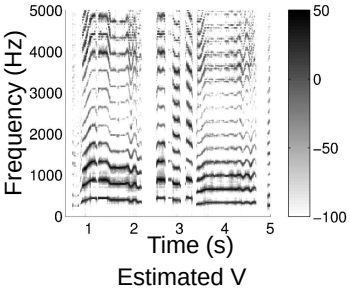


Accompaniment re-estimation



Results:

- **Database:**
  - **A:** SiSEC08 "Professionally Produced Music Recordings" development set,
  - **B:** excerpts from DB used in [Ozerov] and [Lagrange08],
  - **C:** pop songs with female singer (ccmixter.org/shannon-hurley)



- **Main Instrument / Accompaniment** separation:

Dataset	SDR	SIR	SAR
All (1)	1,6	5,4	5,2
A (2)	8,2	17,4	8,9
B (2)	2,4	6,6	5,2
C (2)	2,7	9,2	5,0
All (2)	2,7	8,1	5,2

Main instrument separation performances: (1) after 1st round and (2) after 2nd round

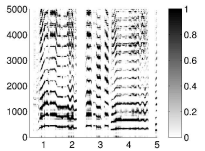
Dataset	SDR	SIR	SAR
All (1)	7,7	14,6	10,7
A (2)	10,8	15,4	12,9
B (2)	8,5	14,6	11,7
C (2)	9,1	14,1	12,7
All (2)	8,8	14,4	12,1

Accompaniment separation performances

Wiener Filters:

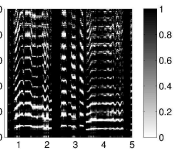
Time-Frequency masks

- **Main instrument:**



$$\frac{(W_K H_K) .* (W_{F_0} H_{F_0})}{(W_K H_K) .* (W_{F_0} H_{F_0}) + W_R H_R}$$

- **Accompaniment:**



$$\frac{W_R H_R}{(W_K H_K) .* (W_{F_0} H_{F_0}) + W_R H_R}$$

Conclusion:

- Bayesian framework successfully integrating **BASS and MIR** aspects
- **State-of-the-art** separation of the main instrument and of the accompaniment

Perspectives:

- Modeling **unvoiced parts** of the main instrument
- Extension to **multi-channel** signals (stereophonic recordings)
- Use this technique as a **pre-processing for MIR applications** (e.g. vocals recognition, chords recognition,...)

International Conference on Audio, Speech and Signal Processing, 2009, Taipei, Taiwan

References: [Durrieu08] J.-L. Durrieu *et al.*, "Singer melody extraction in polyphonic signals using source separation methods", ICASSP, 2008.
[Lagrange08] M. Lagrange *et al.*, "Normalized cuts for the predominant melodic source separation", IEEE Trans. on ASLP, 2008.
[Ozerov07] A. Ozerov *et al.*, "Adaptation of bayesian models for single-channel source separation in popular songs", IEEE Trans. on ASLP, 2007.