Melody Extraction and Tempo Estimation: MIREX 2011

Justin Salamon, Jose R. Zapata and Emilia Gómez

Music Technology Group, Univeristat Pompeu Fabra, Barcelona, Spain

Contact:
{justin.salamon,joser.zapata}@upf.edu
http://mtg.upf.edu/~jsalamon

Two algorithms are presented in this poster, the first for melody extraction from polyphonic music (Justin Salamon and Emilia Gómez) and the second for tempo estimation (Jose R. Zapata and Emilia Gómez).

Melody Extraction from Polyphonic Music

Introduction

The system presented here is an updated version of the system submitted to last year's campaign. Following a detailed analysis of the first two blocks of the system [1], system parameters have been optimised for melody extraction and the implementation is now more efficient. Two variants of the system have been submitted (SG1 & SG2), each using a different spectral transform.

Sinusoid Extraction

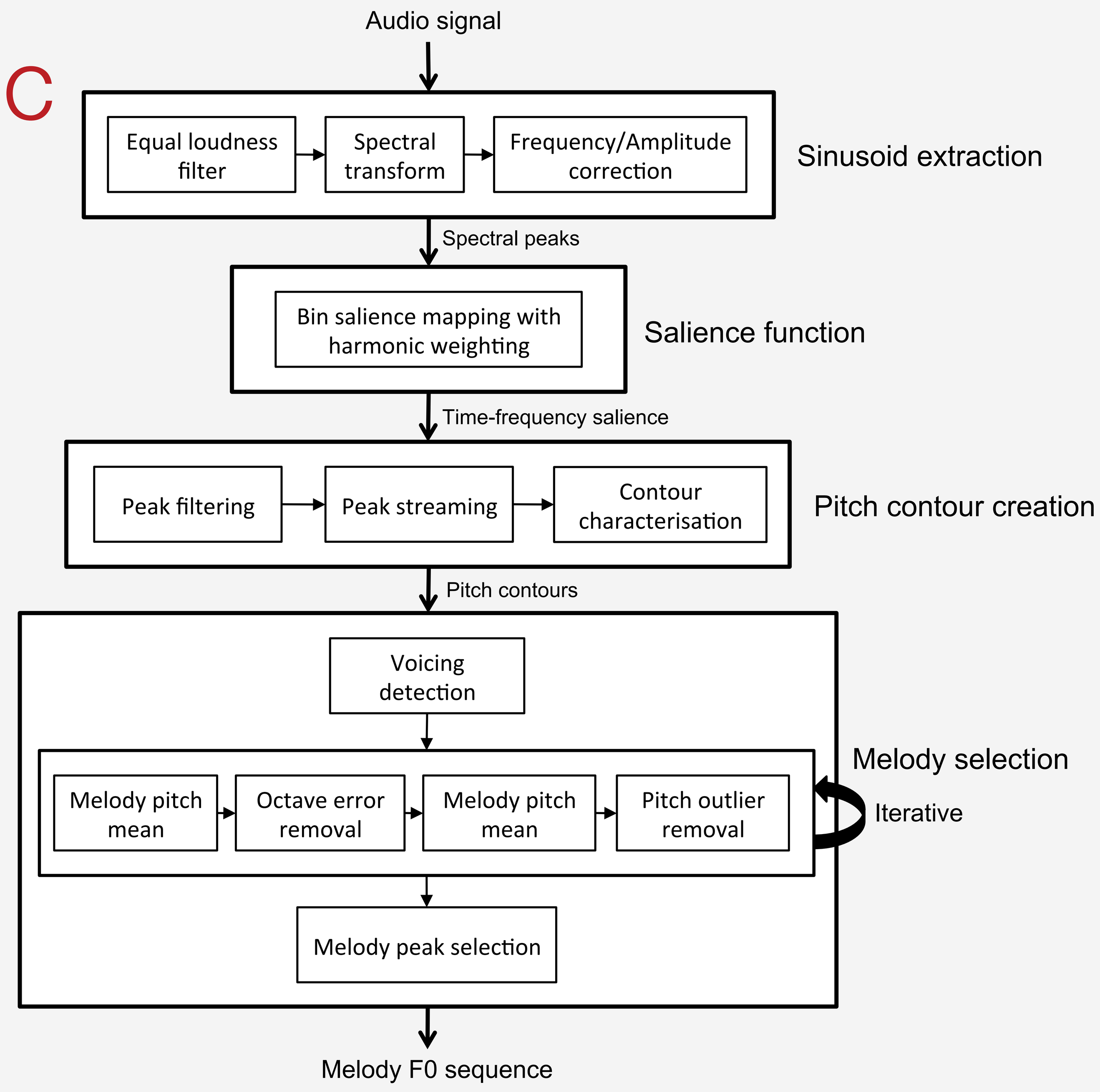
- Pre-filtering: Equal Loudness filter.
- Transform: STFT (**SG1**), MRFFT [2] (**SG2**).
- Frequency/Amplitude correction: Instantaneous frequency using phase vocoder method.

Salience Function

Based on harmonic summation with magnitude weighting, and spans a range of almost five octaves from 55Hz to 1760Hz. Further details are provided in [1].

Pitch Contour Creation and Melody Selection based on Contour Characteristics

Peaks of the salience function are grouped over time using heuristics based on auditory streaming cues [3], resulting in a set of **pitch contours**. The contours are automatically analysed and a set of **contour characteristics** is computed and used to filter out non-melody contours: we use contour **feature distributions** to remove contours in segments where there is no melody present (voicing detection). The remaining contours are used to iteratively calculate an overall melody pitch trajectory, which is used to minimise octave errors and remove pitch outliers. Finally, contour salience features are used to select the melody F0 at each frame from the remaining contours.



Overall Accuracy Results: MIREX 2011

Algorithm	ADC 2004	MRX 2005	MRX 2008	MRX 09 (0dB)	MRX 09 (-5dB)	MRX 09 (+5dB)	Mean (Unweighted)	Mean (Weighted*)
TY3	0.47	0.51	0.70	0.52	0.41	0.56	0.53	0.50
TY4	0.47	0.51	0.70	0.52	0.41	0.56	0.53	0.50
TOS1	0.59	0.57	0.72	0.74	**0.62**	0.82	0.68	0.72
LYRS1	0.73	0.59	0.72	0.47	0.36	0.54	0.57	0.47
HCCPH1	0.44	0.45	0.64	0.50	0.39	0.59	0.50	0.50
CWJ1	0.73	0.57	0.69	0.53	0.40	0.62	0.59	0.52
YSLP1	**0.85**	0.65	0.73	0.52	0.39	0.66	0.63	0.53
PJY1	0.81	0.65	0.71	0.74	0.54	0.83	0.71	0.70
SG1	0.74	0.66	0.83	**0.78**	0.61	**0.85**	0.74	**0.75**
SG2	0.74	**0.68**	**0.84**	**0.78**	0.61	**0.85**	**0.75**	**0.75**

* Weighted by data-set total playtime.

References

[1] J. Salamon, E. Gómez, and J. Bonada. Sinusoid extraction and salience function design for predominant melody estimation. In Proc. 14th Int. Conf. on Digital Audio Effects (DAFX-11), Paris, France, September 2011.

[2] K. Dressler. Sinusoidal Extraction using an Efficient Implementation of a Multi-resolution FFT. In Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06), pages 247–252, Montreal, Quebec, Canada, Sept. 2006.

[3] A. Bregman. Auditory scene analysis. MIT Press, Cambridge, Massachussetts, 1990.

Combination of Audio Tempo Estimation Approaches

Introduction

Using the tempo estimation results from four different approaches (BeatIt, Ellis, Davies and MPEG7-XM), we use a heuristic strategy to obtain the slow and fast tempo from audio music combinig their results, based on the hypothesis that the tatum and tactus tempo hierarchal levels can be obtained from these estimations analyzing their relations.

Combination Method

1. Sort all the tempo estimation values and eliminate the repeated ones.
2. Cluster tempo values with differences of 4% (e.g: 127.6, 125.8 , 128.4) and calculated the median value of each cluster.
3. Check if each value has a relation of (½, 2, 1/3, 3) and eliminated the not related ones.
4. Heuristic analysis

- If only two values are obtained, the lowest value is the slow tempo (T1) and the highest value is the fast tempo (T2).
- If there are 3 values with a binary relation between them, the lowest value is the slow tempo (T1) and the double value is the fast tempo (T2).
- If there are 3 values and two of these ones had a ternary relation between them we take this two numbers and the lowest value is the slow tempo (T1) and the double value is the fast tempo (T2).
- If there are more than 3 values we choose the values related with the BeatIt estimation, and the heuristic analysis is checked, to obtain the slow tempo (T1), the fast tempo (T2).
- The strength of T1 relative to T2 is taked from the Ellis Results.

Algorithm	BeatIt	Ellis	Davies	Mpeg7-xm
Author	Bonada and Gouyon	Ellis	Davies and Plumbley	Rohden
Infrastructure	Windows Binary	Matlab	Matlab	Matlab
Output	One Bpm	slower and faster tempo, relation between them	Beats of the song	BPM value each time the estimated tempo changes
Feature list	Energy envelope differences for 8 bands	Onsets obtained from the Mel spectrogra m	Spectral Flux	Energy envelopes for 6 bands
Pulse induction	ACF	ACF	ACF	ACF

Results for the train data-set

- **At least one tempo correct: 100%**
- **Both tempi correct: 85%**

MIREX 2011 Results

Audio Tempo Extraction			
Algorithm	**Tempo P-Score**	**At least One Tempo Correct**	**Both Tempi Correct**
SP1	0.7105	0.9286	0.3857
GKC6	0.6777	0.8214	0.4286
GKC3	0.8290	0.9429	0.6214
SB5	0.6559	0.8429	0.3500
FW2	0.7385	0.8357	0.5429
ZG1	0.7275	0.8214	0.5714

References

[1] D. Ellis: "Beat Tracking by Dynamic Programming," Journal of New Music Research, Vol. 36, No. 1, pp. 51-60, 2007.

[2] J. Zapata and E. Gómez: "Comparative Evaluation and Combination of Audio Tempo Estimation Approaches," Audio Engineering Society Conference: 42nd International Conference: Semantic Audio, Ilmenau - Germany, 2011.

[3] M.E.P. Davies, and M. D. Plumbley: "Context-dependent beat tracking of musical audio," Audio,Speech,and Language Processing, IEEE Transactions, Vol. 15, No. 3, pp. 1009-1020, 2007.

§ This work was funded by the Programa de Formación del Profesorado Universitario (FPU) of the Ministerio de Educación de España.

§§ A big thanks to the IMIRSEL team at the University of Illinois at Urbana-Champaign for running MIREX!