Improving Beat Tracking in the presence of highly predominant vocals using source separation techniques: Preliminary study

José R. Zapata and Emilia Gómez

Music Technology Group
Universitat Pompeu Fabra
{joser.zapata,emilia.gomez}@upf.edu

Abstract. The automatic beat tracking from audio is still an open research task in the Music Information Retrieval (MIR) community. The goal of this paper is to show and discuss a work-in-progress of how audio source separation can be used for improving beat tracking estimations in difficult cases of music audio signal with highly predominant vocals. The audio source separation using FASST (Flexible Audio Source Separation Toolbox) had an average improvement of beat tracking of {14,15%, 17,74%} in the F-measure and {14,21%, 25,70%} in the Amlt of Klapuri and Degara systems respectably in a dataset of 20 songs excerpt.

Keywords: Beat tracking, Source separation, Predominant voice

1 Introduction

The task of Beat tracking is related to the detection of the main pulse beat, defined as "one of a series of regularly recurring, precisely equivalent stimuli" [1]. For Western music, a hierarchical metrical structure is found in different time scales, and the most common ones are: the tatum period, defined as "a regular time division that mostly coincides with all note onsets"; and the tactus period (the perceptually most prominent period), defined as the rate at which most people would regularly tap their feet, hands or finger in time following the music.

Beat is a relevant audio descriptor of a piece of music, which represents the speed of the piece under study. For that reason, much research within the Music Information Retrieval (MIR) community has been devoted to finding ways to automate its extraction and many algorithms have been proposed. Beat tracking algorithms have been used in different application contexts, such as music retrieval, cover detection, playlist generation, and beat synchronization for audio mixing, structural analysis and score alignment. Many approaches for beat tracking have been proposed, and some efforts have been devoted to their quantitative comparisons to find other ways to emphasize and detect the rhythm accents in music, but it's not still clear in which kind of music or interpretations the beat trackers have problems to detect the beats.

A recent study in beat tracking difficulty [2] presented a technique for estimating the degree of difficulty of musical excerpts in beat tracking based on the mutual agreement between a committee of beat tracking algorithms. In this study an audio dataset was built containing 678 excerpts of 40s length from various musical styles such as classical, chanson, jazz, folk and flamenco. In this study difficult cases for beat tracking songs with strong and expressive voice were found. Even with a stable accompaniment, beat trackers encountered problems.

The goal of this paper is to present and discuss a work-in-progress of the improvement of beat tracking estimation in difficult cases with highly predominant vocals, using FASST (Flexible Audio Source Separation Toolbox). Based on the evidence, a discussion of the results and ideas for future work are presented.

This paper is structured as follows. First, we present current challenges for beat tracking, followed by the hypothesis of the experiment. Second, Each part of the evaluated system is briefly explained. Third, we present the results of each beat tracking experiment. Finally, we provide some discussions, limitations, future work and conclusions of this study.

2 Experiment Hypothesis

The hypothesis of this experiment originated from previous research on: automatic beat tracking with percussive/ harmonic separation [3] and tempo estimation that uses source separation [4] or percussive/harmonic separation[5] to improve tempo detection. Based on this research, a source separation technique is proposed to improve beat tracking in difficult cases with highly predominant vocals and quiet accompaniment.

3 Experimental Framework

The main goal of the experiment is to evaluate if audio source separation techniques improve the beat tracking systems. The experiment consists of an evaluation of two beat tracking algorithms on 20 audio song excerpts (highly predominant vocals) before and after a process of source separation.

3.1 Audio Beat Trackers

Two different systems were used for this experiment:

1. The Matlab implementation of the well-known Audio Beat tracking system by Anssi Klapuri [6], which uses the differentials of loudness in 36 frequency subbands as audio features which are then combined in four signals. These signals measure the degree of musical accentuacion over time. The pulse induction block is a bank comb filter. The algorithm estimates the tatum, the beat and the measure through probabilistic modeling the relationships and temporal evolutions.

2. The Matlab implementation of Degara's beat tracker by Norberto Degara [7], analyzes the input musical signal based on complex spectral difference

method and extracts a beat phase and a beat period salience observation signal, with this info estimates the time between consecutive beat events and exploits both beat and non-beat information by explicitly modeling non-beat states. In addition to the beat times, a measure of the expected accuracy of the estimated beats is provided. The quality of the observations used for beat tracking are measured and the reliability of the beats is automatically calculated. The accuracy of the beat estimations are predicted by a k-nearest neighbor regression algorithm.

3.2 Audio Source Separation

The Matlab software tool named Flexible Audio Source Separation Toolbox (FASST) [10] we used as a source separation tool for the experiment. The framework can incorporate prior information about the audio signal. The basic example (EXAMPLE_prof_rec_sep_drums_bass_melody.m) contains information allowing the separation of the following four sources: Bass, Drums, melody (singing voice or leading melodic instrument) and remaining sounds (other).
The Framework FASST is available in `http://bass-db.gforge.inria.fr/fasst/`

3.3 Music Material

The audio files used in the experiment are a subset of 20 excerpts from the databases used in [2]. It consists of difficult song cases of audio beat tracking with highly predominant vocals and the format is the same for all: mono, linear PCM, 44100 Hz sampling frequency, 16 bits resolution. Each excerpt has ground truth annotations of the beats as described in [2]. The artist and the name of each song are in Table 1 and Table 2.

3.4 Evaluation methods

We contrasted the beat trackers output from the original excerpts ans the output of the source separation method. The evaluation techniques considered in this study are:

F-measure [8] : Beats are considered accurate if they fall within a 70ms tolerance window around annotations. Accuracy in a range from 0% to 100% is measured as a function of the number of true positives, false positives and false negatives.

AMLt [9]: A continuity-based method, where beats are accurate when consecutive beats fall within tempo-dependent tolerance windows around successive annotations. Beat sequences are also accurate if the beats occur on the off-beat, or are tapped at double or half the annotated tempo. The range of values for AMLt is 0% to 100%.

It's important to note that F-measure can increase either due to and increase of tru positives or decrease of false positives or negatives. The Amlt measure improvement can be due to the estimation of true positives in different metrical levels, and continuity is not required.

4 Results

Table 1 and Table 2 present the evaluation results of F-measure and Amlt evaluation for Klapuri and Degara beat tracking algorithms respectively from the original excerpts and the source separation output files.

The average result for the original excerpts of Klapuri algorithm is {39,61%, 39,02%} for F-measure and Amlt respectively. Taking only the best beat tracking result from the separated signals per each song, the average resultincreases to {50,43%, 51,97%} for F-measure and Amlt respectively.

For Degara method, the average result for the original excerpts is equal to {33,6%, 28,6%} for F-measure and Amlt respectively. Considering only the best beat tracking result from the separated signals per each song, the average result increases to {45,71%, 47,78%} for F-measure and Amlt respectively.

Results of Klapuri beat tracker using source separation improved 95% on the dataset at least in one measure. F-measure values in 80% of the dataset in a range of {0,3%, 39,67%} (50% on the Bass) and Amlt values in 90% of the dataset in a range of {1,49%, 37,01%} (33,33% on the Bass). Results of Degara beat tracker using source separation improved 85% on the dataset at least in one measure. F-measure values in 75% of the dataset in a range of {1,6%, 46%} (53,33% on the Bass) and Amlt values in 80% of the dataset in a range of {0,3%, 72,95%} (50% on the Bass).

5 Discussion, Limitations and Future work

In the presented experiment we show that, most of the time, beat tracking estimations can be improved by means of source separation techniques in highly predominant vocal songs, although the expressiveness of the voice such as vibrato, rubato, etc, can difficult beat tracking. In future work we will also consider a low latency voice elimination technique (de-soloing) [11] as an alternative option.

5.1 Source Separation

The FASST source separation tools allow source separation without collecting prior information about the input audio signal. One problem is the computational time because it takes more than 20 minutes to process each audio signal. One limitation for source separation is the few implemented and tested systems to use for academic research and implementing low latency algorithms is still a research challenge. For future experiments different source separation systems had to be evaluated to determine the best alternative for our problem.

From the evaluation results Bass output had better results but is not clear which of the four outputs from the source separation is better to use in all the cases, as it depends on the instruments present in the song. A rhythm strength level measure per signal could be used for this purpose, so that we would apply the beat tracking algorithm in the output signal with higher rhythm strength. One open issue is how to combine the beat tracking estimations from the different sources of the same song to improve beat tracking results.

Artist - Song title	Measure	Original	Melody	Bass	Drums	Other
Joss Stone	**F-measure**	26,51	31,71	**34,04**	29,27	32,10
Dirty Man	**Amlt**	3,08	2,04	**13,85**	2,04	4,17
Edith Piaf	**F-measure**	47,80	42,70	50,91	**53,41**	44,32
La Foule	**Amlt**	22,41	35,48	44,83	**56,67**	56,67
Joss Stone	**F-measure**	22,86	19,13	14,58	**23,16**	23,16
The Chokin' Kind	**Amlt**	9,88	**20,99**	9,09	12,96	11,11
Diana Krall	**F-measure**	18,18	9,26	**32,65**	16,82	8,00
Just The Way You Are	**Amlt**	8,00	8,00	17,33	**22,67**	4,00
Tomwaits	**F-measure**	17,48	40,38	29,03	34,86	**57,14**
The Piano Has Been Drinking	**Amlt**	38,46	41,51	12,68	33,93	**75,47**
Tomwaits	**F-measure**	31,07	30,91	32,65	20,00	**38,46**
Foreign Affair.wav	**Amlt**	18,99	25,32	20,69	8,33	18,99
Joss Stone	**F-measure**	8,33	8,33	15,22	**22,50**	8,33
Understand	**Amlt**	67,35	63,27	0,00	24,56	**75,51**
Tomwaits	**F-measure**	44,44	24,24	**54,35**	14,58	20,45
The One That Got Away	**Amlt**	65,00	26,09	**90,32**	21,21	42,37
Edith Piaf	**F-measure**	28,32	**40,35**	18,18	20,34	21,43
L'Accordeoniste	**Amlt**	13,56	**23,33**	13,43	17,19	8,62
Edith Piaf	**F-measure**	50,00	26,80	**79,12**	28,83	21,05
Correqu' Et Reguyer	**Amlt**	56,63	21,82	**67,35**	31,33	26,42
Edith Piaf	**F-measure**	27,87	19,67	**42,59**	32,73	31,67
Prisonnier De La Tour	**Amlt**	11,34	4,11	**35,59**	16,39	12,37
Edith Piaf	**F-measure**	14,81	22,43	24,30	29,36	**33,64**
Il Pleut	**Amlt**	7,69	14,06	4,71	9,41	**18,75**
Diana Krall	**F-measure**	36,17	15,53	31,11	34,34	31,11
Abandoned Masquerade	**Amlt**	40,00	17,57	**45,90**	30,00	36,07
ABBA	**F-measure**	80,65	77,42	47,62	**93,55**	75,41
The Winner Takes It All	**Amlt**	83,87	87,10	43,75	**96,77**	80,65
Tony Bennett	**F-measure**	21,74	18,60	**42,55**	31,11	24,39
i used to be colourblind	**Amlt**	35,48	6,90	**56,25**	33,33	27,59
Ivor Novello	**F-measure**	17,54	29,51	**32,65**	3,70	18,87
I Can Give You	**Amlt**	14,29	**21,88**	20,00	17,86	13,79
Joe Cocker	**F-measure**	80,28	77,14	28,57	52,35	68,57
That's the way her love is	**Amlt**	85,92	**90,14**	14,44	44,87	94,37
Roberto Goyeneche	**F-measure**	74,29	38,46	67,29	51,92	**78,10**
Ventanita florida	**Amlt**	81,13	40,38	67,27	48,08	81,13
Bruce Springsteen	**F-measure**	87,34	11,45	28,00	82,82	86,34
Thunder Road	**Amlt**	85,34	73,68	9,20	79,82	**86,84**
Meat Loaf	**F-measure**	56,60	39,75	41,10	36,76	52,56
Bat out of hell	**Amlt**	31,97	30,61	25,00	30,65	26,53

Table 1. F-measure and Amlt results for Klapuri beat tracking algorithm

Artist - Song title	Measure	Original	Melody	Bass	Drums	Other
Joss Stone	**F-measure**	36,70	23,93	**46,15**	32,97	26,83
Dirty Man	**Amlt**	38,16	14,29	0,00	**38,46**	3,08
Edith Piaf	**F-measure**	44,32	40,82	40,41	40,21	29,32
La Foule	**Amlt**	30,43	3,75	1,30	30,14	6,67
Joss Stone	**F-measure**	13,46	17,58	**41,07**	32,20	28,57
The Chokin' Kind	**Amlt**	14,29	20,00	**46,91**	35,80	32,94
Diana Krall	**F-measure**	17,02	14,29	**39,25**	20,00	22,86
Just The Way You Are	**Amlt**	7,14	16,67	**46,67**	17,33	21,33
Tomwaits	**F-measure**	34,11	21,24	22,61	33,33	**35,71**
The Piano Has Been Drinking	**Amlt**	10,48	23,33	24,19	26,23	**40,68**
Tomwaits	**F-measure**	36,04	29,63	23,85	21,36	24,00
Foreign Affair.wav	**Amlt**	36,71	32,91	18,99	5,06	17,72
Joss Stone	**F-measure**	17,78	7,84	14,74	5,48	**25,32**
Understand	**Amlt**	17,91	0,00	0,00	28,00	**28,57**
Tomwaits	**F-measure**	27,72	24,49	**52,75**	9,88	25,26
The One That Got Away	**Amlt**	28,17	30,88	**83,61**	6,78	44,62
Edith Piaf	**F-measure**	29,06	21,05	11,97	21,85	14,68
L'Accordeoniste	**Amlt**	15,87	16,67	14,29	**20,00**	16,36
Edith Piaf	**F-measure**	32,08	**38,33**	36,36	20,00	18,00
Correqu' Et Reguyer	**Amlt**	13,25	38,55	**49,40**	14,46	8,62
Edith Piaf	**F-measure**	34,38	32,06	**54,17**	43,56	35,29
Prisonnier De La Tour	**Amlt**	23,71	25,77	**73,47**	46,15	30,19
Edith Piaf	**F-measure**	19,64	18,69	23,21	27,35	**28,57**
Il Pleut	**Amlt**	7,06	4,71	10,59	**21,18**	16,36
Diana Krall	**F-measure**	28,30	17,65	21,95	24,14	24,49
Abandoned Masquerade	**Amlt**	15,58	5,48	**24,56**	0,00	20,29
ABBA	**F-measure**	31,43	32,88	16,67	**77,42**	27,45
The Winner Takes It All	**Amlt**	7,69	0,00	29,41	**80,65**	0,00
Tony Bennett	**F-measure**	20,00	32,65	**38,10**	16,00	17,02
i used to be colourblind	**Amlt**	31,43	44,12	**44,83**	34,29	28,13
Ivor Novello	**F-measure**	57,14	35,29	25,00	**64,52**	34,62
I Can Give You	**Amlt**	44,12	3,45	25,00	**54,55**	4,35
Joe Cocker	**F-measure**	59,15	46,81	**69,01**	32,43	41,42
That's the way her love is	**Amlt**	84,51	71,83	81,69	27,66	36,73
Roberto Goyeneche	**F-measure**	16,36	12,84	37,84	31,48	**59,62**
Ventanita florida	**Amlt**	44,83	52,63	32,20	33,93	**67,31**
Bruce Springsteen	**F-measure**	76,39	39,60	34,04	29,95	55,70
Thunder Road	**Amlt**	70,83	14,16	37,70	13,51	50,00
Meat Loaf	**F-measure**	40,94	43,02	**71,74**	52,24	42,86
Bat out of hell	**Amlt**	29,93	30,61	**40,14**	31,67	31,29

Table 2. F-measure and Amlt results from Degara beat tracking algorithm

5.2 Data

It's important to note that this evaluation has been specifically carried out for difficult beat tracking cases with highly predominant vocals in the audio signal and one limitation is found with these kinds of cases from the beat tracking databases that exist right now with ground truth. For future evaluation, more data with these issues could be collected using an automatic identification system of difficult examples for beat tracking[2] and manually classifying highly predominant vocals cases, or by using an automatic highly predominant vocals detection system.

Most of the source separation algorithms use the spatial information to improve the separation. In this evaluation the datasets are mono audio signals. For future evaluations, it would be good to add some stereo song excepts.

5.3 Beat Tracking

The song excerpt with best improvement of F-measure (13,46% to 41,07%) with Degara algorithm is the sameas the Klapuri has the lowest improvement (22,86% to 23,16%), but the Klapuri algorithm reach better F-measure result for this song excerpt. One limitation of the beat tracking evaluation is the use of different measures to determinate the good performance of the systems. There is no consensus on how to measure with a single value, or which evaluation measure is more reliable for beat tracking proposes.

The Beat tracking in the source separated signals fail when the accompaniment had pauses, tempo changes and the principal metrical level is a musical combination between of all the instruments and the voice (e.g Diana Krall - Abandoned Masquerade).

Another limitation is the lack of methodology to combine the beat tracking results from different algorithms. For future work this evaluation can be performed with more beat trackers to extend the results of the experiment and establish more accurate statements of the advantage of use source separation for improve beat tracking. The evaluation and research of this method can be applied like a pre-process stage in beat tracking.

6 Conclusions

The audio source separation made by FASST algorithm had an average improvement of beat tracking of {14,15%, 17,74%} in the F-measure and {14,21%, 25,70%} in Amlt of Klapuri and Degara systems.

Comparing only the best result from each separated signals per song with the original beat tracking result, the Klapuri and Degara algorithms enhanced the average results in {10,81%, 12,1%} for F-measure and {12,96, 19,18%} for Amlt value respectively.

The Bass output from the source separation enhanced the beat tracking results in the dataset more than the other outputs at least in 50% on F-measure and 33% on the Amlt for Klapuri and Degara Beat trackers. This is the clearest and common instrument output in most of the songs on the dataset.

Audio source separation could then be used as a pre-process stage for improving beat tracking estimation in difficult songs with highly predominant vocals, without changing the beat tracking algorithm.

Acknowledgments. Thanks to Anssi Klapuri, Norberto Degara and A. Ozerov, E. Vincent and F. Bimbot, the authors of the algorithms of beat tracking and source separation respectively for making their algorithms available for research topics. Matthew Davies, Andre Holzapfel and Fabien Gouyon for the intership support in INESC in Porto. Thanks to Colciencias and Universidad Pontificia Bolivariana (Colombia), Music Technology Group at Universitat Pompeu Fabra, Classical Planet and DRIMS project for the financial support. Robin Motheral for the paper review and Justin Salamon for your helpful recommendations.

References

1. Cooper, G., Meyer, L. B.: The rhythmic structure of music. University Of Chicago Press,Chicago (1960)
2. Holzapfel, A., Davies, M.E.P., Zapata, J.,Oliveira, J.L., Gouyon, F.: On the automatic identification of difficult examples for beat tracking: towards building new evaluation datasets. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP. IEEE Press, Kyoto, Japan (2012)
3. Gkiokas, A., Katsouros, V., Carayannis, G.: ILSP Audio Beat Tracking Algorithm for MIREX 2011. 6th Music Information Retrieval Evaluation eXchange (MIREX). Miami (2011)
4. Chordia, P., Rae, A.: Using source separation to improve tempo detection. In: Proceedings of 10th International Conference on Music Information Retrieval ISMIR, pp. 183–188 (2009)
5. Gkiokas, A., Katsouros, V., Carayannis, G.: ILSP Audio Tempo Estimation Algorithm for MIREX 2011. 6th Music Information Retrieval Evaluation eXchange (MIREX). Miami (2011)
6. Klapuri, A. P., Eronen, A. J., Astola, J. T.: Analysis of the meter of acoustic musical signals. In: IEEE Trans. on Audio, Speech, and Language Processing, vol. 14, no. 1, pp. 342–355 (2006)
7. Degara N., Argones, E., Pena, A., Torres-guijarro, S.,Davies, M.E.P., Plumbley, Mark, D.: Reliability-Informed Beat Tracking of Musical Signals. IEEE Transactions on Audio, Speech and Language Processing, Vol. 20, pp. 290–301 (2012)
8. Dixon,S.: Evaluation of the audio beat tracking system BeatRoot. Journal of New Music Research, vol. 36, pp.39–50 (2007)
9. Hainsworth,S.W. and Macleod,M.D.:Particle filtering applied to musical tempo tracking. Journal of Advances in Signal Processing, vol. 15, pp. 2385–2395 (2004)
10. Ozerov, A., Vincent, E., Bimbot, F.: A General Flexible Framework for the Handling of Prior Information in Audio Source Separation. IEEE Transactions on Audio, Speech, and Language Processing. vol. 19, no. 8. (2011)
11. Marxer, R., Janer, J., Bonada, J.: Low-latency Instrument Separation in Polyphonic Audio Using Timbre Models. In: 10th International Conference on Latent Variable Analysis and Source Separation, LVA/ICA 2012, Tel-aviv, Israel (2012)