AOD TASK MIREX 2018 SUBMISSION – KFLUX2

Ricardo V. Koscky
IFES Serra, ES, Brazil
ricardokoscky@gmail.com

Flávio Garcia Pereira
IFES Serra, ES, Brazil
flavio.garcia@ifes.edu.br

Karin S. Komati
IFES Serra, ES, Brazil
kkomati@ifes.edu.br

ABSTRACT

KFlux is an algorithm to detect onsets on audio tracks. It is based in spectral dissimilarity, as other good detectors, but with some important differences. Firstly, the Onset Detection Function (ODF) is calculated using a convoluted dissimilarity kernel over a very coarse spectrogram, followed by an optional mid or low-range frequency filter. Secondly, the peak-picking stage uses a prominence based adaptive thresholding and a clustering sub-block. Finally, by varying its parameters, KFlux has access to a family of ODFs, so a fusion with a few of these ODFs was done, in order to achieve yet better results. In KFlux dataset, the algorithm has reached F_1=0.9209.

1. INTRODUCTION

KFlux is entirely based in [1], a spectral dissimilarity-based audio onset detector [2]. However, instead of using original formula [2], KFlux uses a dissimilarity kernel to improve soft onsets detection.

KFlux uses the same architecture proposed by the authors of [3]: pre-processing, reduction and peak-picking, but important details and some differences are considered, in order to achieve good and meaningful outcomes.

2. ALGORITHM

Figure 1 shows the architecture and the internal KFlux details.

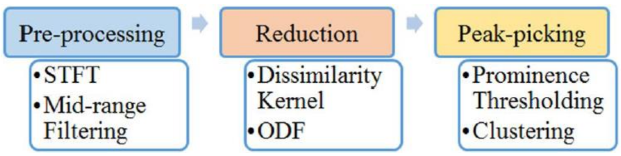

Figure 1. KFlux architecture: blocks and internal details.

2.1 Pre-processing

Using short-term Fourier transform (STFT), KFlux computes a very coarse spectrogram, with only 30 frequency bins linearly distributed in a Mel scale. Parameters α and β control window size and overlap, respectively (Table 1). The window type chosen was Hann and spectrogram magnitudes are calculated in dB.

Symbol	Description (unit)	Values
α	Window size (ms)	5, 7, 10, 14, 19, 25, 32, 40, 50, 60, 70, 80
β	Overlap (%)	0, 20, 40, 60, 80
N	Kernel size (frames)	1-30, in 1 steps
φ	Spectrogram filter	'F', 'M', 'L'
γ	Threshold (dB)	2-5.5, in 0.5 steps

Table 1. KFlux parameters: 43200 combinations.

Then, KFlux applies an optional mid or low-range frequency filter (Table 1), controlled by parameter φ. Should the filter be applied, KFlux eliminates frequency bins outside the 4 kHz mid-range band around 160-4160 kHz (φ='M') or outside the 500 Hz low-range band around 160–660 Hz (φ='L'). Otherwise (φ='F'), no filtering is applied, so the full spectrogram is used.

2.2 Reduction

KFlux initiates reduction block by calculating an intermediary processed spectrum, convolving a dissimilarity kernel [1] over the post-filtered spectrogram. Parameter N controls the kernel size (Table 1).

Then, KFlux calculates the average from each column (i.e. frame) from the processed spectrum. By taking all these averages, KFlux obtains an ODF.

2.3 Peak-picking

KFlux doesn't need to do any post-processing such as suggested by [3] in this block, not even calculate ODF moving averages, standard deviations or variances. Instead, KFlux uses peak prominences [1] to select onset candidates. This provides a new adaptive thresholding mechanism. Prominences are calculated in dB, providing a physically meaningful value for onset candidates. Parameter γ ensures selection of only the best candidates (Table 1).

Finally, KFlux includes a new sub-block for grouping peaks that are too close [1] from each other. Together with the prominence-based adaptive thresholding, this makes more sense to the human auditory system due to temporal masking phenomena.

2.4 ODF Family Fusion

By varying the first 4 parameters from Table 1, KFlux has 12 x 5 x 30 x 3 = 5400 possible combinations in its ODF family. Each combination is called a pODF (parameterized ODF). During KFlux development, some pODFs excel detecting hard (i.e. percussive) onsets while others were better in detecting soft (i.e. non-percussive) ones.

Since MIREX AOD dataset has both hard and soft onset types, a fusion of some few pODFs seems to be the logical choice for aiming better F_1 performance.

In this submitted version, KFlux has 19 pODFs that were picked and γ tweaked with a more sophisticated, though not extensive, optimization algorithm to achieve the best F_1 in KFlux dataset.

Also, in this year, the function that implements KFlux has received an additional external parameter, a threshold offset that shifts γ of all pODFs by the same amount. A sweep in this new parameter allows the generation of a precision-recall operator characteristic curve.

3. RESULTS

3.1 KFlux Dataset

The KFlux dataset, used during development, has 602 onsets annotated manually by 3 experts. 298 are percussive onsets and 304 are non-percussive onsets. This dataset has 61 excerpts with different kinds of instruments.

KFlux follows the same MIREX 2018 criteria for considering true positives (TP), false negatives (FN) and false positives (FP). Also, the resulting F_1 is calculated in the same way.

Scenario	TP	FN	FP	F_1
Percussive Dataset	297	1	25	0.9581
Non-Percussive Dataset	256	48	21	0.8812
Complete Dataset	553	49	46	0.9209

Table 2. Results on KFlux dataset.

KFlux has reached an overall F_1=0.9209, correctly detecting almost all (FN=1) percussive onsets. The partial results for the percussive and non-percussive datasets were F_{1p}=0.9581 and F_{1np}=0.8812, respectively.

3.2 MIREX 2018 Dataset

The MIREX 2018 dataset is not publicly available, but if it were the same from 2005 (which was used till 2017), also not publicly available, it has 85 excerpts and more than 2500 onsets distributed over 9 track classes: complex, poly pitched, solo bars and bells, solo brass, solo drum, solo plucked strings, solo singing voice, solo sustained strings and solo winds. These onsets were also manually annotated by 2 to 5 people.

<TO DO in camera-ready version: include MIREX 2018 results here and comment about it.>

4. CONCLUSION

<TO DO in camera-ready version: include conclusions.>

<TO DO in camera-ready version: second paragraph.>

5. REFERENCES

[1] A. Anonymous: "Towards Audio Onset Detection with KFlux," *Journal of Audio Engineering Society*, under submission.

[2] P. Masri, "Computer modeling of sound for transformation and synthesis of musical signal," Ph.D. dissertation, University of Bristol, UK, 1996.

[3] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in musical signals," *IEEE Trans. Speech and Audio Proc.*, Vol. 13, No. 5, pp. 1035–1047, 2005.