Aubio Study Results

Domenico Stefani

27 January 2021

Abstract

This brief report presents the results of a comparison study on all the methods of the Aubio suite for onset detection.

Some choices are strongly affected by the target application for the onset detector.

1 Introduction

The study is performed on sounds recorded from 6 combinations of different acoustic guitars and experienced guitarists. 8 different techniques were used to play the sounds, 4 of them being conventional techniques played with a pluck/pick on the strings, while the remaining 4 are percussive sounds produced by hitting the body of the guitar.

Aubio onset offers 8 different detection methods and different parameters. The methods are:

1. hfc.
2. energy.
3. complex.
4. phase.
5. specdiff.
6. kl.
7. mkl.
8. specflux.

The parameters are:

- *Hop size:* the value used was **64 samples** since this is a value for the audio block size for the vst plugin that offers a good trade off between latency and computational resources requirements on Elk Audio Os (Running on Raspberry Pi 4).
- *Minimum Inter-Onset Interval*: Debounce parameter which was set to **20ms** as it is the target latency for the whole classification pipeline in mind for this study.
- *Buffer Size*: This represents the size of the analysis window for which the onset detection function is computed. Different values were tested: **64,128,256,512,1024,2048 samples**. This affects in minor way the detection latency.
- *Silence Threshold*: It is used to cut low amplitude onsets. It's expressed in dB and values between **-60dB** and **-40dB** were used.
- *Onset Threshold*: A value used in dynamic thresholding to select onsets. Values between **0** and 4 were used.

2 Test

For each method, different combinations of the parameters were tested to optimize the f1-score. Since the cardinality of examples for each class does not reflect the one of the full dataset and it is not equal between classes, the metric optimized was the **macro average f1-score**, which is the mean of the f1-scores for each individual class. Similarly, all the latency metrics reported are averaged across all classes. For latency, upper and lower Tukey fences are reported (k = 1.5) along with the sample mean of the distribution.

The 2 objectives to optimize, in order to choose the combination of method and parameters to use, are the f1-score (to maximize) and the variability of the latency, represented by the Interquartile Range (to minimize). A flexible constraint can be posed on the upper fence which should not exceed a set value. In this regard, the pareto front is found with the 2 main objectives first, and then with the f1-score and upper fence as objectives.

3 Results

The best f1-score results along with the latency metrics connected to them are presented in tables 1 and 2.

Table 1: The best f1-score avg. values are shown. Different combinations of Buffer size and Method produce different latency values, which are reported in the following tables. Bold values represent the points in the Pareto front defined by the points in the space of 2 objectives: the f1-score (to maximize) and the Inter Quartile Range (to minimize). More info in fig. 1 and table 3.

		Buffer size					
		64	128	256	512	1024	2048
Method	hfc	0.9330	0.9229	0.9118	0.8883	0.8902	0.8724
	energy	0.9364	0.9419	**0.9470**	**0.9444**	**0.9533**	0.6447
	complex	0.8351	0.8422	0.8579	0.8720	0.8623	0.7909
	phase	0.6967	0.8138	0.8507	0.7857	0.7331	0.6849
	specdiff	0.8524	**0.9330**	**0.9511**	**0.9523**	0.9528	0.9294
	kl	0.8532	0.8610	0.8658	0.8702	0.8919	0.8053
	mkl	0.8482	0.8522	0.8718	0.8661	0.8690	0.8706
	specflux	**0.8436**	0.9185	**0.9239**	0.9067	0.8799	0.8659

Table 2: The results of the latency recorded on the examples which f1-score is reported in table 1 are shown here. Each cell contains 3 values: the first and the last are the lower and upper Tukey fences with k = 1.5, which are defined starting from the Interquartile range and are commonly used to define outliers of a distribution, while the central value is the sample mean of the latency distribution.

		Buffer size					
		64	128	256	512	1024	2048
Method	hfc	2.4/4.6/6.6	2.7/5.0/7.1	3.7/6.0/8.1	5.2/7.5/9.7	8.3/10.9/13.6	11.3/14.0/16.5
	energy	1.7/3.9/6.1	2.7/4.9/7.0	3.9/6.0/8.0	5.7/7.8/9.8	9.0/11.6/14.1	16.0/18.6/21.4
	complex	2.6/5.0/7.2	3.1/5.4/7.5	3.5/6.0/8.3	4.0/6.6/9.0	4.6/7.5/10.3	5.9/8.7/11.4
	phase	1.1/4.0/6.6	1.9/4.1/6.1	2.8/4.8/6.7	3.4/5.3/7.1	3.9/5.9/7.9	4.3/7.6/10.8
	specdiff	2.4/4.5/6.4	2.8/4.7/6.6	3.8/5.9/7.9	4.9/7.1/9.2	6.6/9.1/11.8	7.4/12.6/18.1
	kl	1.9/4.2/6.3	2.4/4.6/6.7	3.3/5.6/7.8	5.2/8.0/10.6	8.3/12.2/16.4	12.8/16.1/19.9
	mkl	2.3/4.6/6.7	2.9/5.3/7.5	3.8/6.4/8.8	5.4/8.2/11.0	8.5/11.0/13.6	11.4/14.5/18.1
	specflux	2.1/3.8/5.3	2.5/4.3/5.8	3.1/4.8/6.3	3.8/5.6/7.3	4.4/6.3/8.1	5.3/7.5/9.5

Pareto front results are shown in table 3 and fig. 1 for the first analysis, and in table 4 and fig. 2 for the second one.

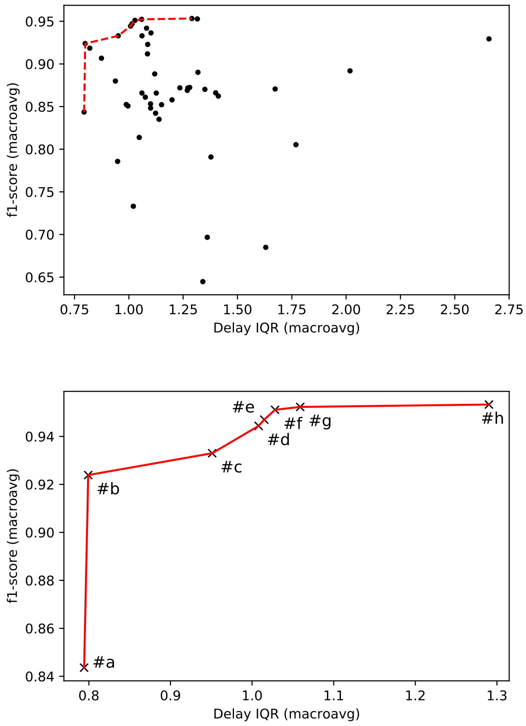

Figure 1: Pareto front computed for f1-score and the Interquartile Range of the latency distribution. The upper plot shows all the solution while the lower plot represents only the points in the front. The labels refer to the detailed information that can be found in table 3.

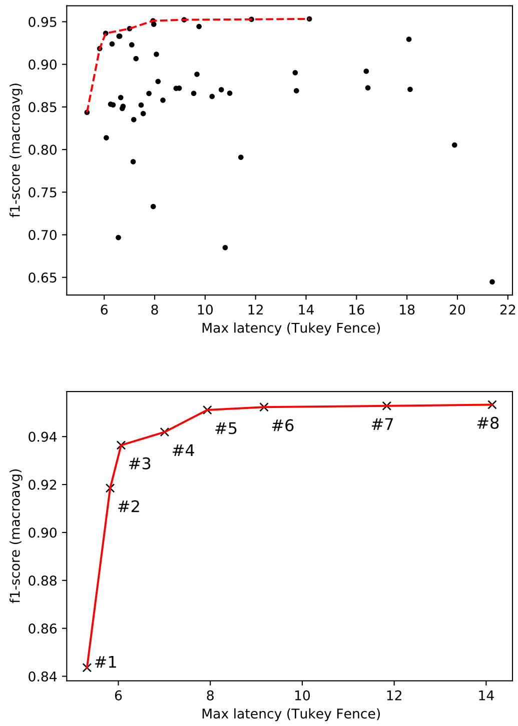

Figure 2: Pareto front computed for f1-score and upper Tukey fence. The upper plot shows all the solution while the lower plot represents only the points in the front. The labels refer to the detailed information that can be found in table 4.

Table 3: Pareto front solution with f1-score (macro average over all techniques) as the first objective and Interquartile Range of latency as the second.

#	Method	F1-score	Low Tukey fence (ms)	Delay mean (ms)	High Tukey fence (ms)	Onsets inside fences (%)
a	specflux	0.8436	2.1412	3.7592	5.3180	95.54
b	specflux	0.9239	3.1157	4.8102	6.3123	93.44
c	specdiff	0.9330	2.7828	4.7029	6.5870	96.92
d	energy	0.9444	5.7263	7.7971	9.7582	97.30
e	energy	0.9470	3.9054	5.9803	7.9649	97.94
f	specdiff	0.9511	3.8230	5.8531	7.9353	95.84
g	specdiff	0.9523	4.9329	7.0556	9.1687	96.60
h	energy	0.9533	8.9710	11.5858	14.1319	97.29

4 Solution Choice

At this point, all the solutions in the Pareto front computed for IQR and f1-score (fig. 1 and table 3) are viable non-dominated solutions which offer different trade-offs between f1 performance and latency variance. In particular, points between #b and #g form a knee-like shape in the front and are particularly good. Solutions #f and #g are

Table 4: Pareto front solution with f1-score (macro average over all techniques) as the first objective and maximum latency as the second, in the form of upper Tukey fence.

#	Method	F1-score	Low Tukey fence (ms)	Delay mean (ms)	High Tukey fence (ms)	Onsets inside fences (%)
1	specflux	0.8436	2.1412	3.7592	5.3180	95.54
2	specflux	0.9185	2.5403	4.2609	5.8217	95.41
3	energy	0.9364	1.6527	3.9025	6.0597	97.73
4	energy	0.9419	2.6799	4.8975	7.0059	97.49
5	specdiff	0.9511	3.8230	5.8531	7.9353	95.84
6	specdiff	0.9523	4.9329	7.0556	9.1687	96.60
7	specdiff	0.9528	6.5762	9.0782	11.8381	94.38
8	energy	0.9533	8.9710	11.5858	14.1319	97.29

shared with the second pareto front (respectively #5 and #6) and have IQR values that induce a range of respectively *4.1 ms* and *4.2 ms* between the upper and lower Tukey fences. Moreover, solution is a good solution with reasonably low IQR, resulting in a Inter-fence range which is about *1 ms* lower than the 2 aforementioned solutions (*3.2 ms*).

More details about the solutions in table 5.

Table 5: Non dominated solutions of choice.

#	Study Id	Buffer Size	Hop size	Min IOI (s)	Silence Thresh. (dB)	Onset Thresh.	F1-score	Low Tukey fence (ms)	Delay mean (ms)	High Tukey fence (ms)	IQR (ms)	Stdev (ms)
b	542	256	64	0.02	-48	1.80	**0.9239**	3.1	4.8	6.3	0.80	0.86
f	259	256	64	0.02	-45	1.80	**0.9511**	3.8	5.9	7.9	1.03	0.99
g	273	512	64	0.02	-48	1.75	**0.9523**	4.9	7.1	9.2	1.06	1.00

5 Results on dynamics

Table 6: Results on dynamics

#	Piano Accuracy	Piano Precision	Piano Recall	Mezzoforte Accuracy	Mezzoforte Precision	Mezzoforte Recall	Forte Accuracy	Forte Precision	Forte Recall
b	0.7382	0.8982	0.8056	0.8219	0.9022	0.9022	0.9167	0.9167	1.0000
f	0.8188	0.9236	0.8782	0.8584	0.9615	0.8889	0.9719	0.9740	0.9978
g	0.8506	0.9182	0.9204	0.8553	0.9682	0.8800	0.9450	0.9531	0.9911