Aubioonset - performance analysis results

Domenico Stefani
domenico.stefani@unitn.it

This brief report summarizes the results of the performance analysis conducted on the aubioonset onset detector, when applied to individual percussive and pitched sounds recorded with acoustic guitars.
The code used for the analysis can be found at:
`https://github.com/domenicostefani/aubioonset-performanceanalysis`

To this date, aubioonset offers 8 different detection methods:

1. **hfc** (High-Frequency content) (default)
2. **energy** (Energy based distance)
3. **complex** (Complex domain onset detection function)
4. **phase** (Phase based onset detection function)
5. **specdiff** (Spectral difference onset detection function)
6. **kl** (Kulback-Liebler onset detection function)
7. **mkl** (Modified Kulback-Liebler onset detection function)
8. **specflux** (Spectral flux)

All the methods have different latencies and accuracies for different types of sounds.
The parameters that were modified were:

1. Onset method (Method used, from the ones listed before, `-O` option)
2. Buffer size (Size of the aubio analysis window, `-B` option)
3. Hop size (Size of the analysis blocks, `-H` option)
4. Silence threshold (Threshold of the noise floor in dB, `-s` option)
5. Onset threshold (Dynamic thresholding value, `-t` option)
6. Minimum Inter-Onset Interval (Debounce interval in seconds, `-M` option)

Notes on ranges and parameters:

All methods were tested.
The Silence threshold values used range from -55dB to -45dB.
The application intended for my work involves very short latency of a complete product, for this reason the following limitations were applied to the parameter ranges to obtain short latency for onset detection. Maximum buffer size used was 1024.
Almost all studies use HopSize = 64 because higher values affect the latency too much.
Minimum Inter-Onset Interval used was 20ms because the idea for the final product is to be able to identify onsets that are as close as 20ms apart.

Preliminary Observations:

To understand better the results some observations can be done.
Most of the methods show good accuracy and latency for percussive onsets, however pitched onsets seem to be way more complex to detect with high accuracy, precision and recall, without affecting too much the latency.
Low latency (<10-15ms) was considered the main goal, and with that, different methods and parameters were tested to get the best metrics.

Setup:
The aubioonset executable for linux was used for the test. A set of recordings of both pitched and percussive onsets were labeled at the millisecond level with the help of Audacity.
582 percussive onsets were labeled, along with 351 pitched onsets. Labeling of pitched sounds was not always easy, on the contrary percussive onsets were easier to identify.

Evaluation:
All the partial tables are included in the appendix. Here only the best accuracies at different Buffer/Hopsize values are reported. Bear in mind that the best results depend on different settings of the other parameter: for that you can find more info in the tables that are in the appendix. Interesting results are marked in bold.

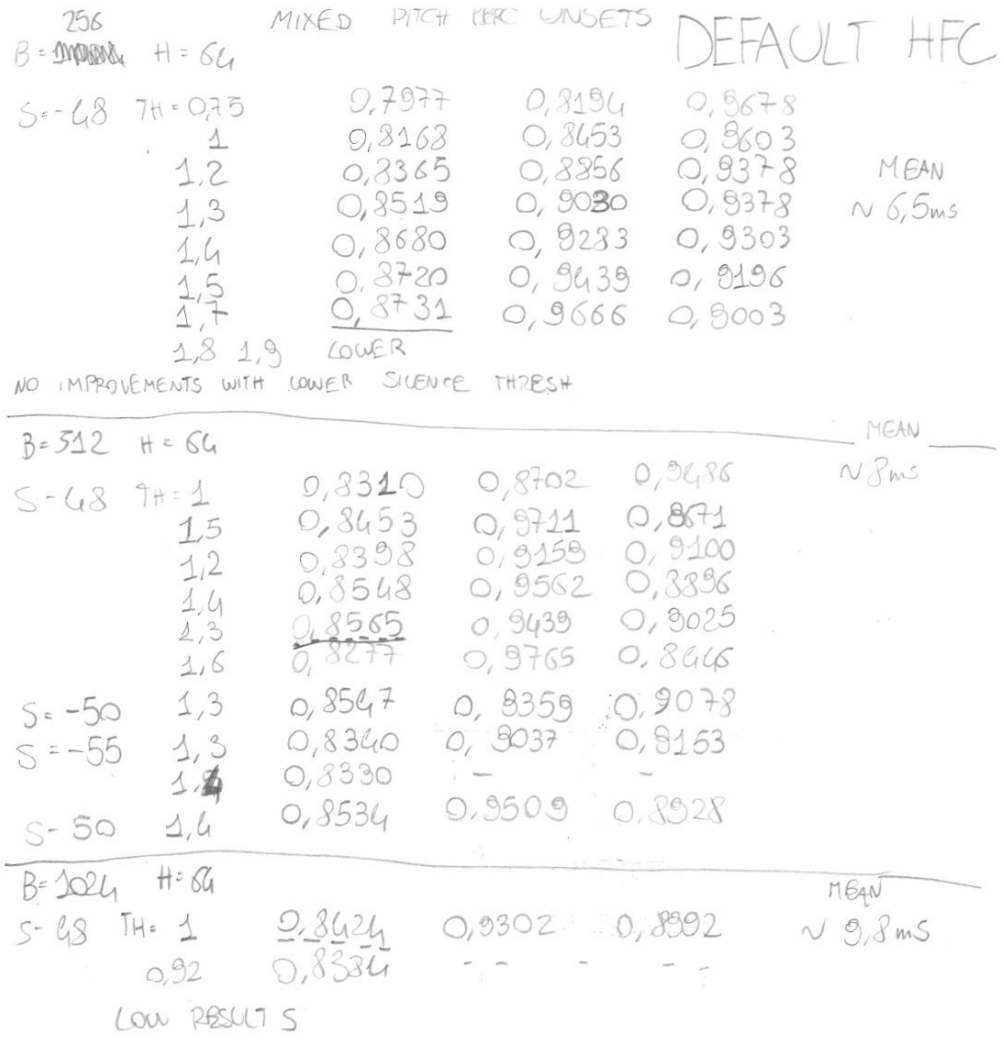
Accuracy Results

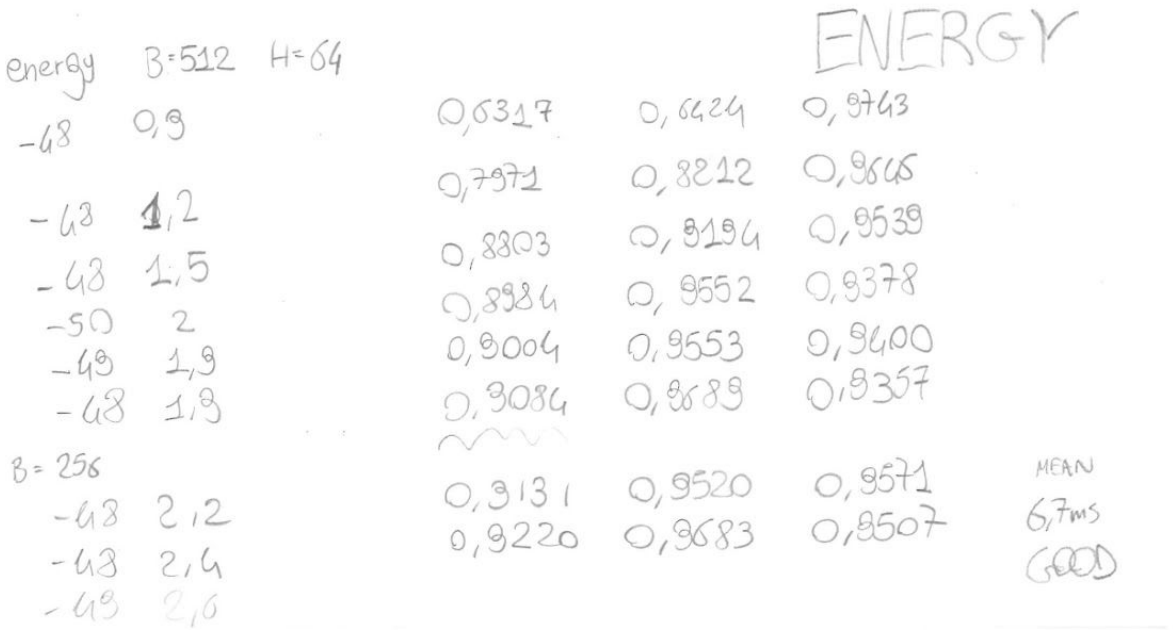
Method	Buffer Size / Hop Size	Buffer Size / Hop Size	Buffer Size / Hop Size
	256 / 64	512 / 64	1024 / 64
hfc	87.31%	85.65%	84.24%
energy	**92.29%**	90.84%	LOW
complex	66.00%	70.00%	72.81%
phase	77.12%	75.00%	LOW
specdiff	90.08%	**93.28%**	**94.39%**
kl	68.00%	69.00%	70.00%
mkl	66.00%	66.00%	67.00%
specflux	86.31%	84.38%	82.53%

Considerations:
Pitched and percussive sounds seem to have distribution of result peaks with regard to the parameter used. These results are an average of both the sound types so they might be biased toward percussive ones since there are slightly more examples. Anyway the new analysis script reports metrics also for the 2 sound types and the different intensities: experiments of interest can be repeated with the scripts in the github repository.
Latency results are reported in the appendix: latency seems to depend mostly on buffersize and hop-size so the results are not written when these do not change.
Calling the computeLatency.py script and feeding the right parameters will produce boxplots of the distribution of latency across all onsets and barplots for the relevance metrics.
Some of the methods perform really well with percussive sounds while their performance with pitched ones are very low.

Appendix:
Annotated tables for each method, excuse my bad handwriting.

On the left of the tables, S and TH represent respectively the silence threshold and the onset threshold. When values are not reported it means that they didn't change from the last value used.
The following three rows of numbers (on the right) represent respectively Accuracy, Precision and Recall.
On the last row sometimes the average latency is reported, however this is quite stable when the buffer size and hop size are set. Horizontal lines separate experiments with different values of these two.


256
MIXED PITCH COMPLEX ONSETS
DEFAULT HFC
B = [illegible] H = 64
S = -48 TH = 0,75 0,7977 0,8194 0,8678
1 0,8168 0,8453 0,8603
1,2 0,8365 0,8856 0,9378
1,3 0,8519 0,9030 0,9378
1,4 0,8680 0,9283 0,9303
1,5 0,8720 0,9439 0,9196
1,7 0,8731 0,9666 0,9003
1,8 1,9 LOWER
MEAN ~ 6,5ms
NO IMPROVEMENTS WITH LOWER SILENCE THRESH
B = 512 H = 64
MEAN ~ 8ms
S = -48 TH = 1 0,8310 0,8702 0,9486
1,5 0,8453 0,9711 0,871
1,2 0,8398 0,9159 0,9100
1,4 0,8548 0,9562 0,8896
1,3 0,8565 0,9439 0,9025
1,6 0,8277 0,9765 0,8646
S = -50 1,3 0,8547 0,8359 0,9078
S = -55 1,3 0,8340 0,9037 0,8163
1,4 0,8330 - -
S = 50 1,4 0,8534 0,9509 0,8928
B = 1024 H = 64
MEAN ~ 9,8ms
S = 48 TH = 1 0,8424 0,9302 0,8992
0,92 0,8334 - - - -
LOW RESULTS


ENERGY
energy B=512 H=64
-48 0,9
-48 1,2
-48 1,5
-50 2
-48 1,9
-48 1,8
B=256
-48 2,2
-48 2,4
-48 2,6
0,6317 0,6424 0,9743
0,7971 0,8212 0,9545
0,8803 0,9194 0,9539
0,8984 0,9552 0,9378
0,9004 0,9553 0,9400
0,9034 0,9589 0,9357
0,9131 0,9520 0,9571
0,9220 0,9683 0,9507
MEAN
6,7ms
GOOD

COMPLEX (PITCHED + PERCUSSIVE)

B = 256	H = 64	ACC	PRE	REC	MEAN
> -48	TH = 0,75	0,6627	0,7635	0,8339	6,5ms
-55	0,75	0,6140			
-50	0,9	0,6325			
-48	1,5	0,58			
"	1,1	0,64	0,80	0,76	
B = 512	H = 64				
-48	1	0,7052	0,7996	0,8564	6,2ms
	1,2	0,66	0,83	0,76	
B = 1024	H = 64				
-50	1	0,7281	0,7921	0,8014	

PHASE — PITCHED + PERCUSSIVE

B= 256 H= 64

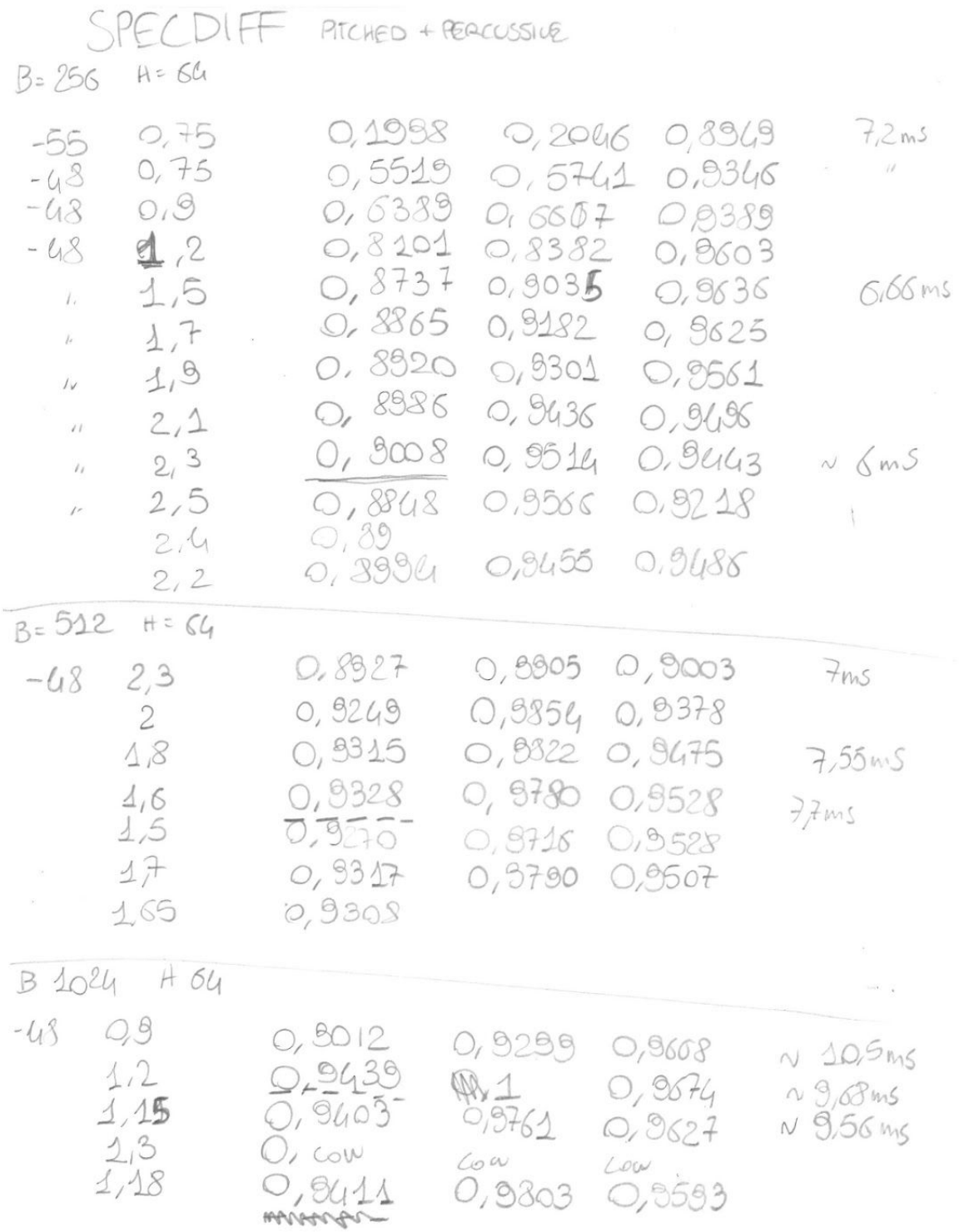
S	TH	ACC	PRE	REC	TIME
-55	0,75	0,1	0,1	0,8810	
	1,2	0,26	0,27	0,9228	
-50		0,4151	0,4289	0,9281	~6ms
-48	1,5	0,5549	0,5828	0,9206	
	2	0,6807	0,7347	0,8024	~5ms
	2,5	0,7318	0,8005	0,8850	
-46	2,5	0,7712	0,8568	0,8853	
-45	2,6				

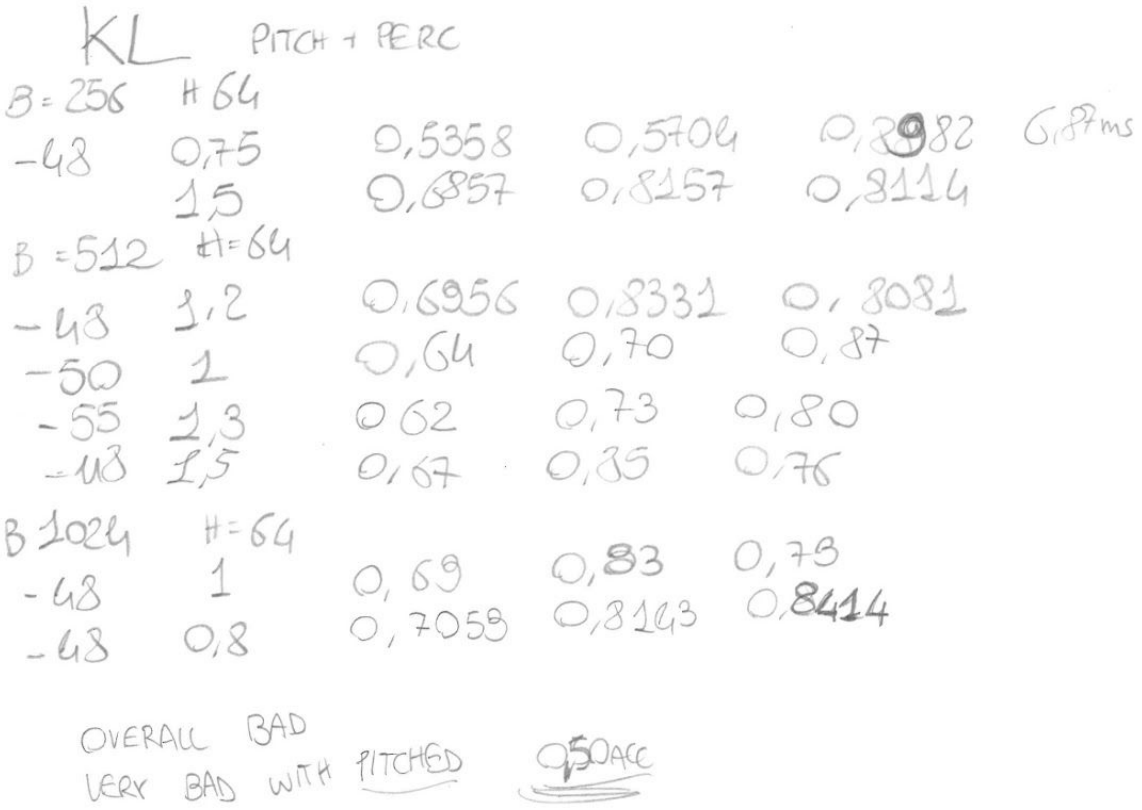
B = 512 H = 64

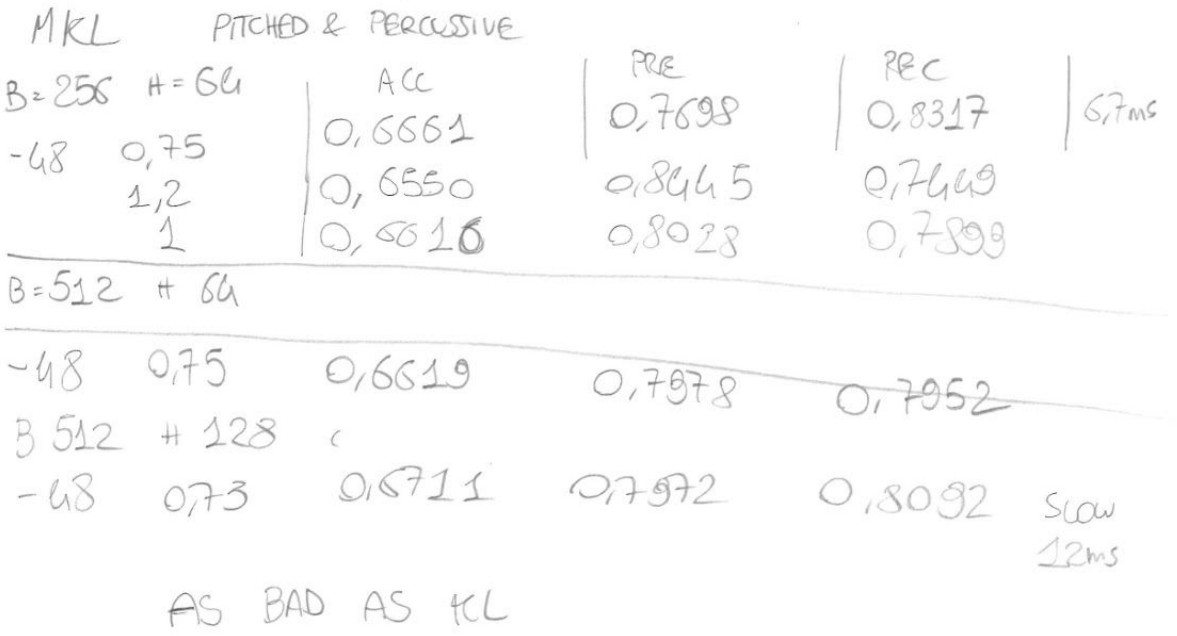
-48	1	0,19	0,20	0,93
	2	0,48	0,50	0,94
	3	0,75	0,88	0,83
-50	3	0,71	0,82	0,83

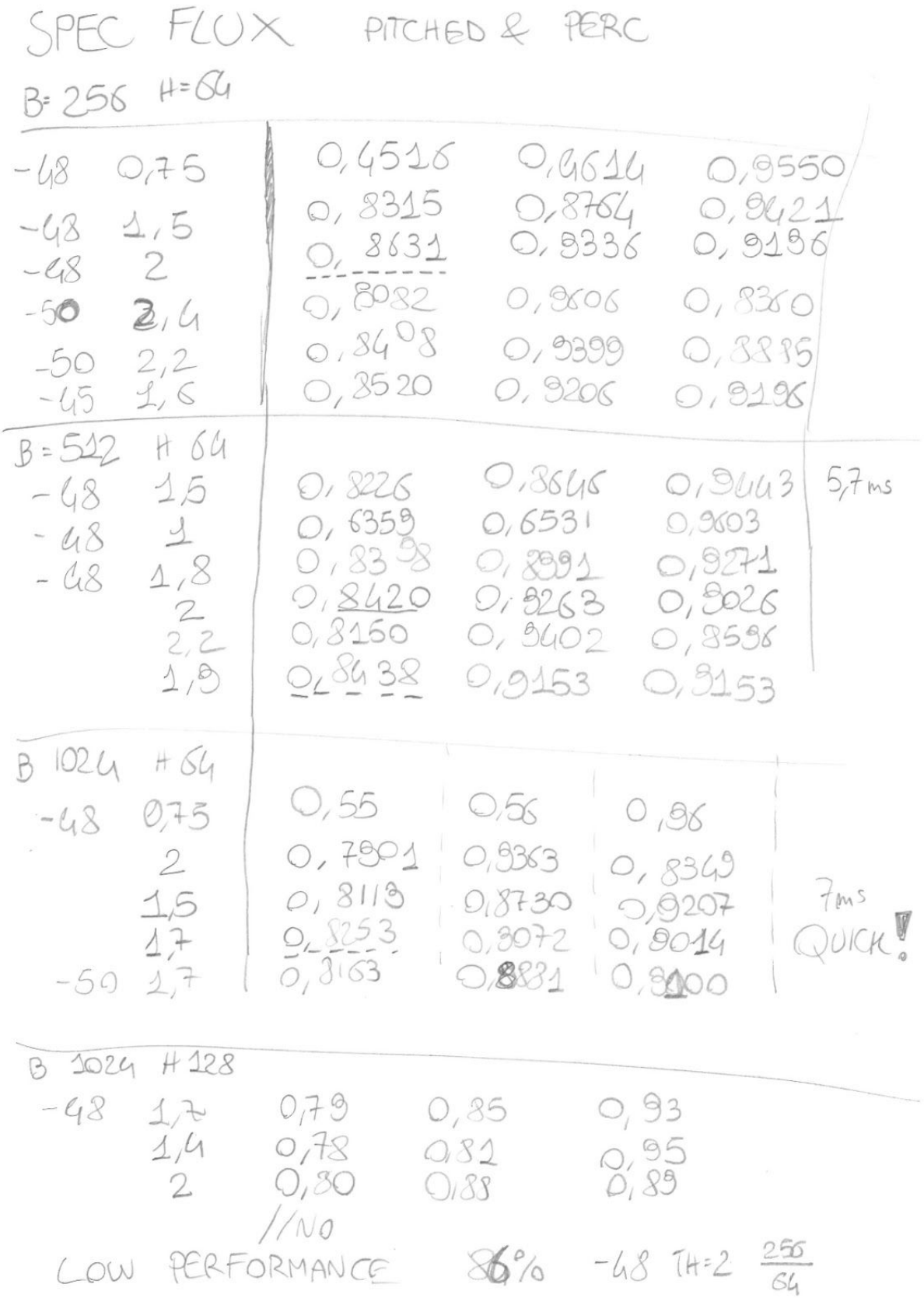
GOOD FOR PERCUSSIVE BUT VERY BAD FOR PITCHED

B = 1024
BAD RES


SPECDIFF PITCHED + PERCUSSIVE
B=256 H=64
-55 0,75 0,1988 0,2046 0,8949 7,2ms
-48 0,75 0,5519 0,5741 0,9346 "
-48 0,9 0,6389 0,6607 0,9389
-48 1,2 0,8101 0,8382 0,9603
" 1,5 0,8737 0,9035 0,9636 6,66ms
" 1,7 0,8865 0,9182 0,9625
" 1,9 0,8920 0,9301 0,9561
" 2,1 0,8986 0,9436 0,9498
" 2,3 0,9008 0,9514 0,9443 ~6ms
" 2,5 0,8848 0,9566 0,9218
2,4 0,89
2,2 0,8994 0,9455 0,9488
B=512 H=64
-48 2,3 0,8927 0,9905 0,9003 7ms
2 0,9249 0,9854 0,9378
1,8 0,9315 0,9822 0,9475 7,55ms
1,6 0,9328 0,9780 0,9528 7,7ms
1,5 0,9270 0,9716 0,9528
1,7 0,9317 0,9790 0,9507
1,65 0,9308
B 1024 H 64
-48 0,9 0,8012 0,9299 0,9608 ~10,5ms
1,2 0,9439 1 0,9674 ~9,68ms
1,15 0,9403 0,9761 0,9627 ~9,56ms
1,3 0,cow cow cow
1,18 0,8411 0,9803 0,9593


KL PITCH + PERC
B=256 H 64
-48 0,75 0,5358 0,5704 0,8982 6,87ms
1,5 0,6857 0,8157 0,8114
B=512 H=64
-48 1,2 0,6956 0,8331 0,8081
-50 1 0,64 0,70 0,87
-55 1,3 0 62 0,73 0,80
-48 1,5 0,67 0,85 0,76
B 1024 H=64
-48 1 0,69 0,83 0,78
-48 0,8 0,7059 0,8163 0,8414
OVERALL BAD
VERY BAD WITH PITCHED
50 ACC


MKL PITCHED & PERCUSSIVE
B=256 H=64
-48 0,75 ACC 0,6661 PRE 0,7698 REC 0,8317 6,7ms
1,2 0,6550 0,8445 0,7649
1 0,6616 0,8028 0,7899
B=512 H 64
-48 0,75 0,6619 0,7978 0,7952
B 512 H 128
-48 0,73 0,6711 0,7972 0,8092 slow 12ms
AS BAD AS KL


SPEC FLUX PITCHED & PERC
B=256 H=64
-48 0,75 0,4516 0,6614 0,9550
-48 1,5 0,8315 0,8764 0,9421
-48 2 0,8631 0,9336 0,9196
-50 2,4 0,8082 0,9606 0,8360
-50 2,2 0,8408 0,9399 0,8895
-45 1,6 0,8520 0,9206 0,9196
B=512 H 64
-48 1,5 0,8226 0,8646 0,9443 5,7ms
-48 1 0,6359 0,6531 0,9603
-48 1,8 0,8398 0,8991 0,9271
2 0,8420 0,9263 0,9026
2,2 0,8150 0,9402 0,8598
1,9 0,8438 0,9153 0,9153
B 1024 H 64
-48 0,75 0,55 0,56 0,98
2 0,7901 0,9363 0,8349
1,5 0,8113 0,8730 0,9207 7ms
1,7 0,8253 0,8072 0,9014 QUICK!
-50 1,7 0,8163 0,8831 0,9100
B 1024 H 128
-48 1,7 0,79 0,85 0,93
1,4 0,78 0,81 0,95
2 0,80 0,88 0,89
//NO
LOW PERFORMANCE 86% -48 TH=2 256/64