Analysis of Musical Timing in Bass Performances for Assessment

|  |
| --- |
| Colm Forkin |
| Masters Graduate, MTG,  Barcelona,Spain.  {c.v.forkin}@gmail.com |

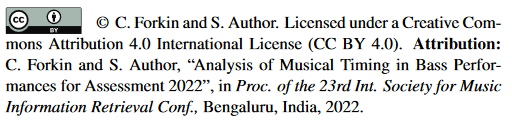
ABSTRACT

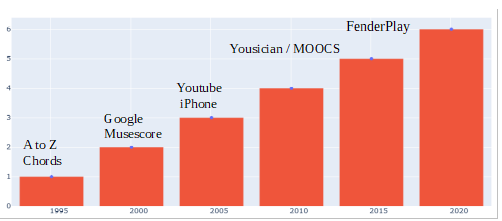
This paper assesses the effectiveness of a commercial education tool for measuring the timing performance of music students in rock and pop music. The focus is on the electric bass instrument and on timing accuracy. The chosen commercial tool, Yousician[[1]](#footnote-2), is used to capture the student performances of basslines from a full length recording of a well known pop song (Yellow by Coldplay) with different timing characteristics in the verse and chorus. To this end, a new Dataset was created consisting of student recordings, covering a wide range of musical performances, from excellent to ones that included deliberate timing mistakes. A reference bass track annotated with onset and offset markings served as a deviation benchmark to measure the student bass performances. Onset Algorithms were applied to the student recordings and the Precision Recall and F-Measure (PRF) accuracies were obtained. Towards measuring the effectiveness of the overall grade given by the commercial tool Yousician, its performance curve was compared to the PRF measurements. It is also aimed to investigate ways and building up the datasets that will give us more insight in how to improve student performances by comparing realtime and non-realtime assessment and finding gaps in the timing feedback information provided.

1. INTRODUCTION

Having researched how people learn popular music using a different paradigm to education methods in classical music, Lucy Green [1] observed how pop musicians were more peer directed in the absence of a central teacher. This then raises the question: how do students direct their learning and receive feedback?

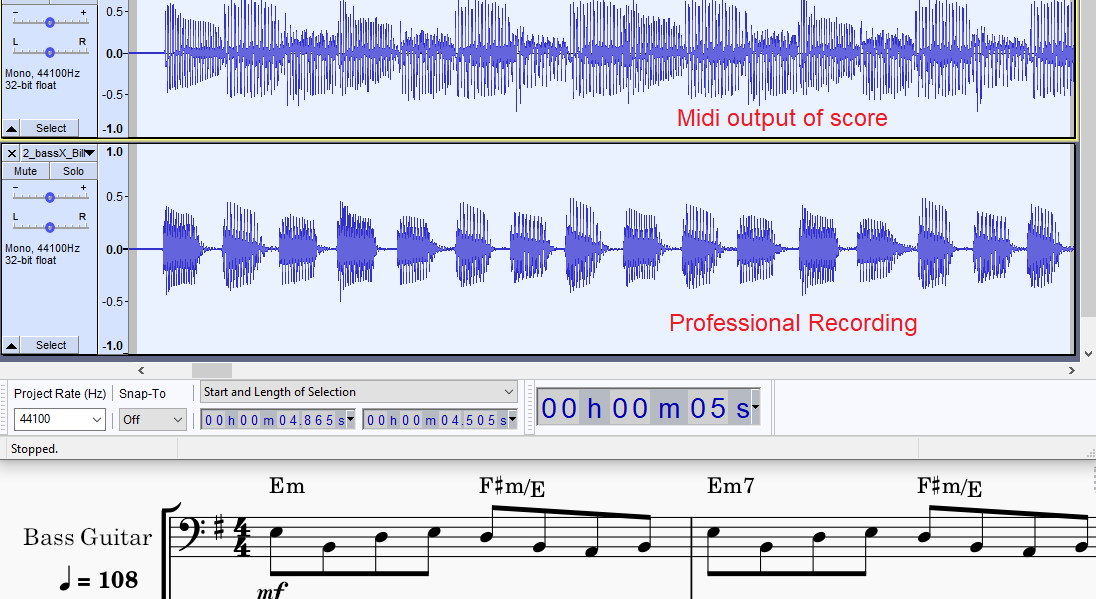
The graph below gives brief overview of the technology background to music education since 1995, progressing from online information and guidance to online feedback.





**Fig 1. History of online tools related to Music Education**

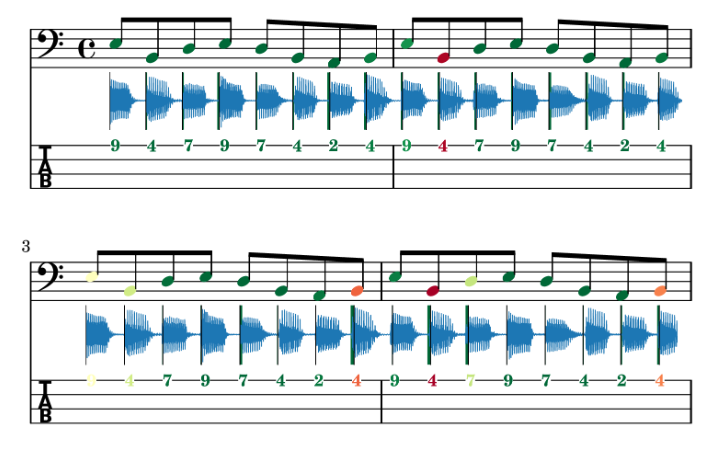
This paper studies one state of the art commercial tool, “Yousician” with focus on timing feedback for bass. Although another tool, FenderPlay[[2]](#footnote-3) has initiated a pilot study of duration feedback for guitar only, duration feedback is absent in Yousician for all instruments.



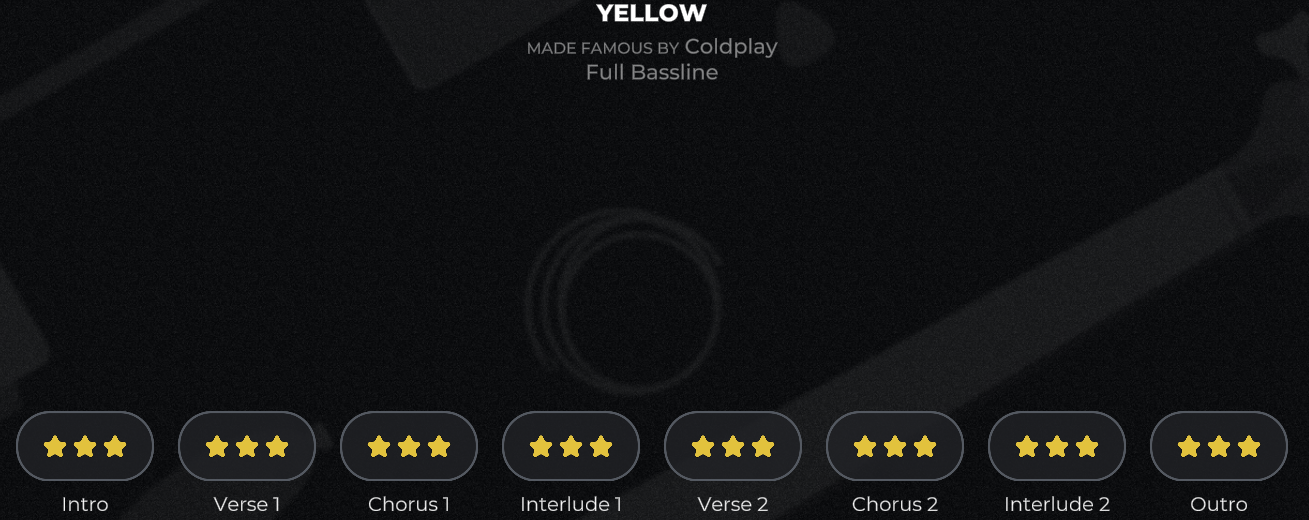
**Fig 2. Billie Jean Sample: Midi vsReal Performance [2]**

Offset curves are highly dependent on playing technique.  
In Billie Jean the performance requires short, jerky plucks. Leaving no gaps between the notes would be a bad interpretation. Measurements on the plucking technique on the violin show a gentle release slope [5] which makes the end point ambiguous. Kehlin/Abesser[4] considered a 200ms margin for the guitar note except, ten times higher than the 10-20ms window typically used used for onsets. The onset algorithms chosen in this paper are based on energy envelope techniques and function[[3]](#footnote-4).

Experiments on Yousican[2] show that “imposter notes” are also not penalized. The scope of this research does not attempt to solve the latter problem. Research examples of offline assessment tools, such as MusicCritic[[4]](#footnote-5) offer a post mortem feedback to students without realtime indicators. This allows the student to go back over the score and review the onset timing accuracies of the notes with traffic light (green, orange, red) indicators.

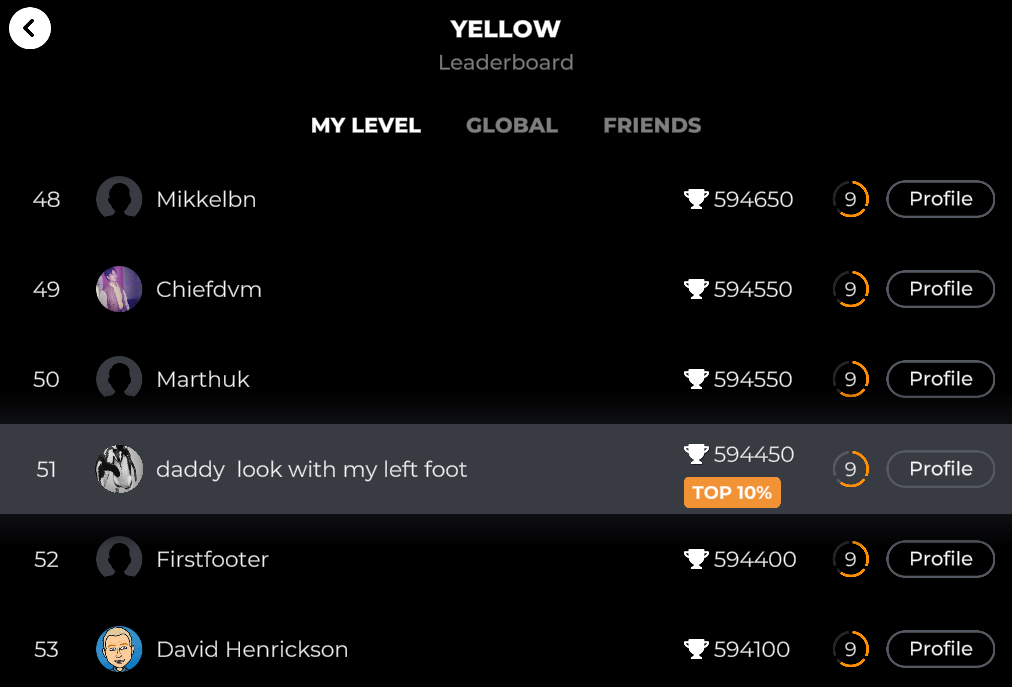
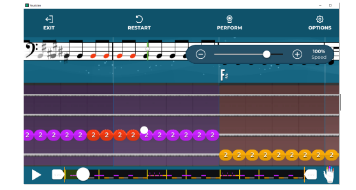


**Fig 3. Music Critic Feedback of Billie Jean bass**

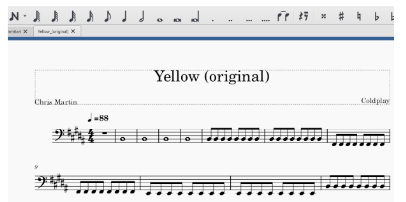
Yousician reports three gold stars on every section that has no note misses or deviations outside the timing limit.

**Fig 4. Yousician Feedback of Yellow**

Depending on how many errors occur, the feedback given is either three, two or one silver stars. Currently there is only a recording of both score and audio performance for Vocals. Having this feature available for Bass would open more opportunities for analysis.

  
  
  
  
  
  
  
  
 **Fig 5 Yousician Grade Table**  
 In addition your total score is given and your relative position with percentile information is also given in a global table.   
 

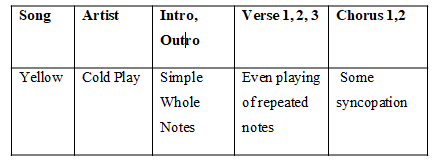
**Fig 6. “Yellow” score in Yousician**



**Fig 7. Imported into Musescore[[5]](#footnote-6)**  
  
The exported audio file from this bass clef rendering provides the generated bass reference from which ground truth onset and offsets are obtained.

1. Methods

Finger style plucking and normal expression style had accuracy rates of 96% in Abessers research[4].The song chosen “Yellow” was one that fits this musical style. It also has a low difficulty level for correct pitch, an important factor when focusing on timing. The cover version of Yellow used in Trinity College London Rock and Pop Syllabus[[6]](#footnote-7) has a difficulty level equivalent to the initial grade band 0. However, the original version used in Yousician has a chorus which is more in line with a grade 1 or grade 2 level. The generated bass reference was split into seven sections: Intro, Verse1, Chorus 1, Verse 2, Chorus 2, Verse3, Outro. These sections do not correspond to the Yousician diagram in Fig.4. The sections for the experiments were created to identify three groups of musical properties summarized in table 1.



**Table 1: Song Characteristics**

Theses sections form the basis of supporting detailed song analysis using onset and offset detection algorithms explained in the next section.

**2.3 Onset Algorithms**

An Onset is detected by calculating the energy in a frequency band. where X[k] is the STFT of x(n). A simple ODF calculation can be made by taking the energy difference of two consecutive audio frames:

*ODF(frame index) = E(frame\_index)-E(frame\_index -1)*

In the case of bass guitar, a pitched percussive instrument, we are concerned with the plucking event. A more general strategy for detecting onsets involves various stages:

• pre-processing is an optional step that emphasizes or attenuates aspects of the signal into different frequency bands.

• ODF reduces the signal to sub-sampled occurrences of transients using “ODF Methods”.

• The “Thresholding” eliminates peaks which are not related to the onset event.

• Peak Picking involves choosing the local maxima after threshold filtering.

The chosen ODF method for this paper is Superflux method implemented in madmom libraries.[[7]](#footnote-8)

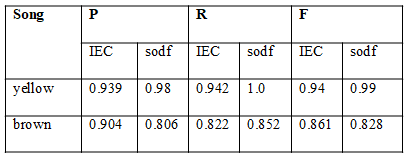
The peak picking strategy in this paper is based on the formula given in [2]. In the same research, another onset method based on sound islands was also used for the more legato sections of the songs and is described next.

**2.4 Offset Algorithms**

Offsets can only be detected with in an onset aware. The Indexed Energy Checker method described in [2] “sound island concept and makes a decision when a voiced section begins and ends. The Essentia[[8]](#footnote-9) algorithm “Effective Duration” can also be used to measure inter-note intervals as an alternative method for measuring offset. Rather than take 100% of the distance between two onsets, perceptual aspects are considered alongside with the attenuation of the signal in determining the true duration.

**2.4 Annotating**  
After running the reference bassline through a script [[9]](#footnote-10)running the ODF / Superflux method the positions of the Onsets were checked visually using Sonic Visualizer. The offsets were manually annotated in the same tool and exported as CSV files. For the Verses, no offsets were captured given the short inter-note interval of all repeated notes. The same procedure was applied to all of the seven subsections of the song (Intro, Verse 1, etc)

In the paper [2], the following accuracy metrics obtained using the Superflux(sodf) and IEC (Indexed Energy Checker [2]) algorithm on a Trinity cover version version of an ideal recording of Yellow, one verse and one “less complex chorus” with just simple quarter notes instead the tied eighth notes used in the original. Also included are the results obtained for the song Brown Eyed Girl, which is a Grade 2 song. This approach is different to using a score generated audio which yields 100% PRF results.

 **Table 2: Onset Accuracies with IEC and sodf methods  
on the TCL version of Yellow**

1. **Recording**

Thirty Student isolated bass recordings of Yellow were made. The Yousician grade of each recording was logged and the “.wav” files were imported into Audacity to synchronize the first onset with the reference audio. The initial noise was removed and the amplification was boosted by by 5db to align the volume with the reference.

**3.1 Recording setup**

A set of Marshal headphones with was connected directly to the laptop running Yousician Premium with Yellow. The inbuilt microphone of the headphones captured the notes for Yousician. The instrument used was a Fender Jazz bass connected to an Ashdown amplifier. During performances 0-14 a Shure dynamic microphone was pointing to the amp and connected to a smart phone capturing the recording that would be processed for onsets and offsets. For performances 20-24, the inbuilt microphone on a smart phone was used, since the dynamic microphone was picking up too much high frequency glitches that were ignored by the microphone in the Marshal headphone connected to the Yousician app. For some of the recordings there was clipping which could not be later smoothed out.

**3.2 Grading Method**

The Yousician score was logged for each recording. The scoring is based on pitch and timing accuracy. Assuming the correct note is hit, the scoring is determined on five timing bands: “early, slightly early, perfect, slightly late, late”. Anything outside this band is zero points. We dont have visibility on how much weighting is applied to each band, but at the end of the song we can see the accumulated number of points. Some reverse engineering of the first four notes shows that 600 points are allocated for a perfect whole note played on time.

**3.3 Song Partitions**

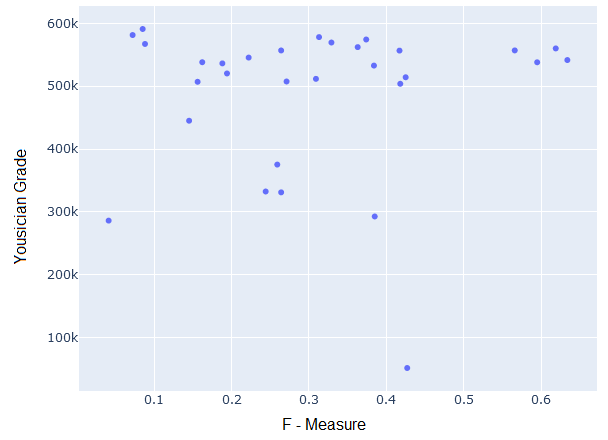
In order to facilitate PRF score on specific audio characteristics, the song was partitioned into 7 sections according to Table 1.These sections do not correspond exactly to the sections of the Song as described in Yousician, but there is a some overlap. For instance our Intro considers the first four notes, whereas the official intro also adds one verse cycle without vocals.

1. Tests

One of the most sensitive aspects of this experiments was aligning the first offset with the reference which assumed perfect timing on the first note, which was not always feasible. In that cases, the general alignment of the first four whole notes where checked and also the last two whole notes were checked manually that no drift was introduced during synchronization.

These recorded bass stems were processed using the algorithms described in section 2 to obtain the PRF scores for onsets The results with high precision score were further processed to obtain deviation statistics for both onsets and offsets. The recordings that yielded low precision scores were studied to see if there were any quality issues in the recording that could be fixed with some post processing.

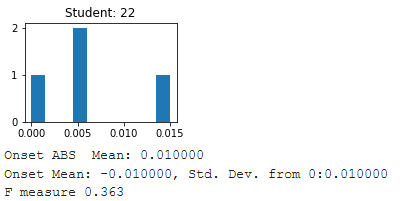
The plot below show the F score vs Yousician score. This plot was obtained running the Jupyter Notebook in the ISMIR 2022 folder in my github link[[10]](#footnote-11).



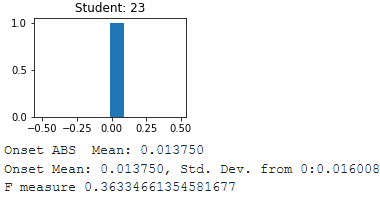
**Fig 8. Onset Accuracy vs Yousician Score**

The left outliers were recordings with some sound glitches caused by the student hitting the pickup with their finger. The right outlier is a low yousician score but the PRF accuracy is high, which suggests that Yousician evaluation window is lower than the 25 ms window we used for detecting onsets. Plots were also made of the MAR (Mean absolute error) of both onset and offset deviation

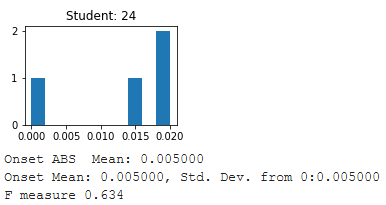
As was discovered in [2], gathering reliable deviation data only makes sense for Precision scores above a reasonable threshold, which in normal cases should be higher than around 75%, however, given the low quality of the results the threshold was dropped to 33%. The deviation plots are based a method used in Ermenkos research [3]. The plot below shows Onset Deviations for the best three student performances:



**Fig 9. Onset Deviation Student 22**

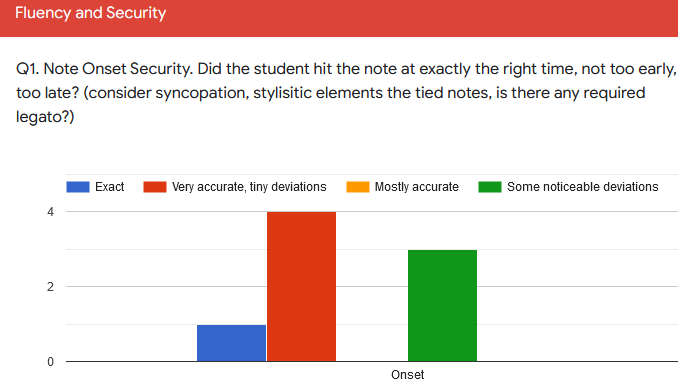
****

**Fig 10. Onset Deviation Student 23**

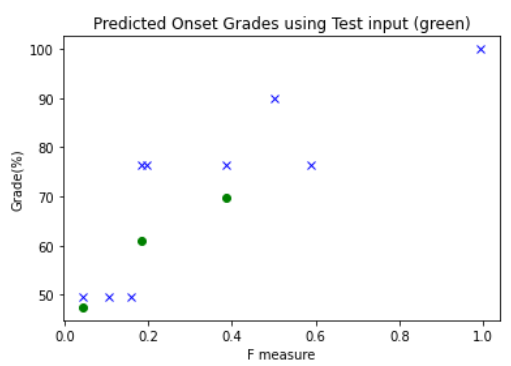
****

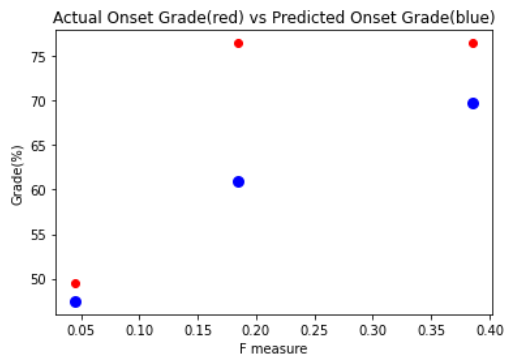
**Fig 11. Onset Deviation Student 24**

Students 22 and 24 have no early onsets. Student 23 has a misleading Histogram. It has a low F – Measure so a wider eval window would be needed to see more deviations.   
Instead of using an edutainment tool, a teacher can graded the recordings and the following histogram gives an overview of the research performed to construct the Rock and Pop Repertoire Dataset [2]



**Fig 12. Teacher Grading Histogram of Students**Having split the dataset into training and test sections the following Predicted Grades were obtained:





Actual Predicted

0 76.5 60.924

1 76.5 69.761

2 49.5 47.442

Mean Absolute Error: 8.124

Root Mean Squared Error: 9.87

**Fig 13. Yellow (cover) predicted Grades[2]**

Offsets are trickier to measure, since they have no clear end point, so a MIR evaluation window is made longer depending on the notes. Unfortunately the PRF measures are too low in this experiment to warrant investigation into the offsets.

Compared with the linear regression curve of the shorted Yellow cover version above, the Yousician scoring system has produced more outliers in fig. 5. Given he 4m20sec duration of the original song, it was always going to be a challenge to get a low a global PRF score.

1. DISCUSSION

The results from show that the teacher provides a slighlty more accurate predictor of grades of than commercial tool. Regarding experimenting with Yousician, not having a section by section grade, makes the evaluation of the local F scores very difficult. Imposter notes are also an important factor to consider when grading, but the emphasis on the research is getting useful feedback to the student on areas that they may not be aware of or notice while playing. Realtime feedback can guide a student but it needs to be removed to reflect the reality of an exam environment.,

The timing algorithm showed more sensitivity to sound glitches than yousician. There are other instrument aspects that must be considered, e.g. playing a fuzzy sound by holding too close to fret. One noisy recording in number, resulted in a low onset accuracy, that was not reflected in the Yousician grade.

A more rigorous test would require Bass Students already familiar with playing the track to deliberately play short eighth notes in the Intro and Outro and very short stacatto notes in the chorus that have correct onsets and pitch but incorrect duration. The Dataset has already got two recordings of this Use Case, so its a question of improving algorithm performance on Onsets, before analysis of Offset can be done.   
  
The main objective was to configure an experimental framework in which you could plug-in different onset and offset detection algorithms to find the optimum combination for performance feedback for a given repertoire of songs. Choosing and configuring an external arbitrator (human teacher or edutainment tool) for grading the performances. For a Music Teacher it requires designing an optimal survey from where they grade manually the onset and duration of each section and Fig 9 gives one example of how this can be doen with Google Forms.

Getting back to Lucy Greens paper [1], this research has encourages autonomous learning with technological feedback. The experience of using an external tool with a group of students showed that it is possible for other students to learn from each other while they perform with visible real time feedback and can then study the Audio waveforms afterwords, thus combining the best of both worlds that real-time and post-mortem feedback in a group environment offer.

1. REFERENCES

[1] Lucy Green, What Can Music Educators Learn from Popular Musicians?: Evaluation and assessment in the contemporary classroom Authors Lucy Green Publication date2017/7/5 Book Music Education as Critical Theory and Practice, Pages 211-226, Publisher Routledge.

[2] Colm Forkin "Automatic Assessment of Timing and Rhythm in Electric Bass for Rock & Pop Repertoire”,Thesis supervisor(s), Vsevolod Eremenko; Xavier Serra, Publication date: August 31, 2021, DOI: 10.5281/zenodo.5553959

[3] V. Eremenko, A. Morsi, J. Narang and X. Serra, "Performance assessment technologies for the support of musical instrument learning," in 12th International Conference on Computer Supported Education, (2020).

[4] C. Kehling, J. Abeßer, C. Dittmar and G. Schuller, "Automatic Tablature Transcription of Electric Guitar Recordings by estimation of score and instrument-related parameters," in DAFx, (2014).

[5]Che-Yuan Liang; Li Su; Yi-Hsuan Yang, Musical Onset Detection Using Constrained Linear Reconstruction Page(s): 2142 – 2146, Date of Publication: 11 August 2015, ISSN Information:,INSPEC Accession Number: 15357036, DOI: 10.1109/LSP.2015.2466447 Publisher: IEEE

1. www.yousician.com [↑](#footnote-ref-2)
2. https://try.fender.com/play [↑](#footnote-ref-3)
3. https://github.com/CPJKU/madmom [↑](#footnote-ref-4)
4. https://musiccritic.upf.edu/ [↑](#footnote-ref-5)
5. https://musescore.org/en [↑](#footnote-ref-6)
6. https://www.trinityrock.com/instruments/bass [↑](#footnote-ref-7)
7. https://github.com/CPJKU/madmom/blob/7e560700987d2800f1c78fdb2a5faba77ecece0d/madmom/features/onsets.py#L585 [↑](#footnote-ref-8)
8. https://essentia.upf.edu/documentation.html [↑](#footnote-ref-9)
9. Scripts can be found in ISMIR folder in https://github.com/cvf-bcn-gituser/bass-critic [↑](#footnote-ref-10)
10. https://github.com/cvf-bcn-gituser/bass-critic [↑](#footnote-ref-11)