Master thesis on Sound and Music Computing

Universitat Pompeu Fabra

Automatic Assessment of Timing and Rhythm in Electric Bass for Rock & Pop Repertoire

Colm Forkin

Supervisor: Vsevolod Eremenko

Co-supervisor: Xavier Serra

Aug 2021



Master thesis on Sound and Music Computing

Universitat Pompeu Fabra

Automatic Assessment of Timing and Rhythm in Electric Bass for Rock & Pop Repertoire

Colm Forkin

Supervisor: Vsevolod Eremenko

Co-supervisor: Xavier Serra

Aug 2021



Table of Contents

[1. Introduction 1](#_Toc80989630)

[2. State of the Art 3](#_Toc80989631)

[2.1. Music Education 3](#_Toc80989632)

[2.2. Onset Detection 4](#_Toc80989633)

[2.3. Onset Detection Methods 5](#_Toc80989634)

[2.4. Offset Detection 7](#_Toc80989635)

[2.4.1 Perceptual and musical considerations 7](#_Toc80989636)

[2.4.2 Voicing Classification 8](#_Toc80989637)

[2.4.3 F0 detection and Segmentation 9](#_Toc80989638)

[2.5. Neural Network Strategies 10](#_Toc80989639)

[2.6. Feature selection and Technical Control 12](#_Toc80989640)

[2.6 Summary 13](#_Toc80989641)

[2.7 Observations 14](#_Toc80989642)

[3. Datasets 15](#_Toc80989643)

[3.1. IDMT BASS SINGLE TRACK 15](#_Toc80989644)

[3.2. Rock and Pop repertoire 17](#_Toc80989645)

[3.3. Musical Properties of R&P repertoire 20](#_Toc80989646)

[4. Methodology 21](#_Toc80989647)

[5. Algorithm Development and evaluation 24](#_Toc80989648)

[5.1 Music Critic End-to End test 24](#_Toc80989649)

[5.2 Algorithm Evaluation 25](#_Toc80989650)

[5.3 Indexed Energy Checker ( IEC) 27](#_Toc80989653)

[5.4 Accuracy Metrics : IEC vs SOP 29](#_Toc80989654)

[5.5 Accuracy Metrics: IEC vs MELODIA 30](#_Toc80989655)

[5.5.1 Long Notes 30](#_Toc80989656)

[5.5.2 WOTM Bridge and Verse Comparison 31](#_Toc80989657)

[5.5.3 Classifying Muteness 33](#_Toc80989658)

[5.6 Accuracy Metrics for Neural Network Methods 34](#_Toc80989659)

[5.7 Deviation Metrics 35](#_Toc80989660)

[5.8 Summary 37](#_Toc80989661)

[6. Prediction of Student Performance 39](#_Toc80989662)

[6.1. Student Recording Portal 39](#_Toc80989663)

[6.2. Processing of student recordings 40](#_Toc80989664)

[6.3. Teacher Grading Policy 41](#_Toc80989665)

[6.4. Teacher Grading Results Analysis 43](#_Toc80989666)

[6.5. Yellow 45](#_Toc80989667)

[6.6. Billie Jean 46](#_Toc80989668)

[6.7. Just Looking 48](#_Toc80989669)

[6.8. Brown Eyed Girl 49](#_Toc80989670)

[6.9. Roadrunner 52](#_Toc80989671)

[6.10. Walking on the Moon 52](#_Toc80989672)

[6.11. Technical Control Grades 54](#_Toc80989673)

[The Actual vs Predicted Grade for Technical Control Element 1 for “bjean”, “brown” and “wotm” are shown here. 54](#_Toc80989674)

[7. Conclusion 56](#_Toc80989675)

[7.1. Student Recordings and Teacher Gradings 56](#_Toc80989676)

[7.2. Technical and Musical issues with Experiment 57](#_Toc80989677)

[7.3. Improving analysis and prediction strategies 58](#_Toc80989678)

[7.4. Improving Duration measurements 59](#_Toc80989679)

[7.5. Going forward 60](#_Toc80989680)

[8. List of figures 62](#_Toc80989681)

[9. List of tables 63](#_Toc80989682)

[10. List of symbols 63](#_Toc80989683)

[11. Bibliography 64](#_Toc80989684)

[12. Appendices 67](#_Toc80989685)

[12.1. Methodology Histograms 67](#_Toc80989686)

[BassUnet Accuracies (Billie Jean vs Abbesser 67](#_Toc80989687)

[12.2. Student Portal 67](#_Toc80989688)

[12.3. Recommendations and Instructions for recording. 69](#_Toc80989689)

[12.4. Additional Instructions for recording. 71](#_Toc80989690)

[12.5. Steps for Processing Recording 72](#_Toc80989691)

[12.6. Teacher Portals 73](#_Toc80989692)

[12.7. Teacher Questions for Just Looking 73](#_Toc80989693)

[12.8. Note books for Plots and Code for generating data 76](#_Toc80989694)

[12.9. Billie Jean Student Performances 77](#_Toc80989695)

Dedication

I would like to thank those near to me personal that made this whole Masters program possible, my family in Barcelona. In 2016 I read the biography, “U2 on U2” [1]. This work by Neill McCormack, inspired me to take up the bass again, take lessons, do exams, look at how music technology could help and eventually do the SMC Masters full time. I was fascinated about the musical journey of U2, how the sonic identity developed and how the producers played a big role. I am very grateful for having the opportunity to do research on the electric bass guitar for Music Education purposes with Vsevolod and Xavier as supervisors. Thanks to Marti for diligently correcting the student recordings and to Ramon for sharing his ideas with me.

Acknowledgments

Student Recording Assessments: Marti Benach Obradors, Bass Teacher and Post Graduate Student at the Escuela Superior de Música Moderna i Jazz ,Taller de Músics.

Ideas for the Indexed Energy Checker ( IEC) algorithm.

Ramon Romeu, Researcher in Sound and music computing, Aalborg University Copenhagen, Denmark.

Abstract

Music Education has undergone significant changes in the last twenty years, with a wide array of applications and online tools emerging to help students learn an instrument autonomously offering automatic feedback. Timing and rhythm are crucial in playing good quality electric bass and although tools exist that help measure their synchronization with the metronome, they don’t provide feedback on note length. Experience in having prepared for electric bass music exams and the identification of shortcomings in performance assessment tools have been the motivation of this thesis.

Note length and note rests are two missing measurement criteria in the state-of-the-art tools. The algorithms and technology exist to do this, but their application has been in automatic music transcription where precision requirements are not as high as they are for music education. This thesis evaluates algorithms for onset and offset detection, offers some new suggestions and tests them on songs with different musical properties on the Rock and Pop repertoire

Keywords:

Audio Signal Processing, Automatic Music Transcription, Bass transcription, , durations, electric bass guitar, expression style, expressive performance analysis, fretboard, ground -truth, Indexed Energy Checker (IEC), Machine Learning, Mean Absolute Error, Music Assessment, Music Education, Music Information Retrieval, Music Performance Analysis, offset, onset, playing technique, plucking, position, Rhythm, Sound Archipelagos, Sound Islands, Sound Onset Processor (SOP), source separation, string detection, style, the deviation statistics, Timing.

1. Introduction

Using technology to assist in Music Performance Assessment in the context of Music Educations is the core subject matter of this thesis. Audio Signal Processing and Music Information Retrieval are the technologies used and Dittmar [2] et all gives us a brief history of the role of MIR in Music Education. A key step forward was the transition to digital formats for both recorded and symbolic notation and hence the transition from CDs and score books to today’s smart phone apps. Applications such as “Yousician”[[1]](#footnote-1) offer performance assessment for learning help guide the student on tuning note accuracy, pitch accuracy and metronome accuracy. Despite the engaging front ends (e.g., real time feedback, scoreboards for highest accuracies) there are important musical qualities that are not assessed: duration, articulation, and good use of dynamics.

This thesis aims to bring the push the sound analysis technologies further to better support the strict educational requirements for professional music performance. Typically for aspiring musicians starting out in Rock and Pop Music, the informal context is where all the learning takes place. It was not uncommon for young people starting out to try form a band before they have learnt their instruments. Neill McCormack [1] describes how U2 got together in the early days, how they relied on feedback given by friends and Richard Evans (the Edge’s brother) and how later on, Adam Clayton (U2 Bassist) sought bass lessons from Patrick Pfeiffer, author of the book “Bass for Dummies” [3] after having made a series of successful albums.

So, the performance assessment of a particular instrument, in this case the bass guitar can happen at any stage in one’s journey and the goals maybe different. If the student wants to improve playing the jazz style, there is a separate topic that deals with score deviation and modelling expressive performance [4]. This research focuses on the score adherence that Rock and Pop Music demands and bases the performance assessment on preparing the musicians with the necessary studio, session, and live performance skills in Modern Music. It focuses particularly on micro-rhythmic skills, which can be measured objectively: plucking the string at the correct time, holding the note for the correct length, technical control of the instrument to produce good quality sound and managing the dynamics.

The thesis opens with the State-of-the-Art Chapter and refer to examples from the two Datasets under study: the Fraunhofer IDMT Single Track Dataset[[2]](#footnote-2) and a Rock and Pop Repertoire dataset with six selected popular songs. These two datasets are described in more detail in Chapter 3, focusing on how they helped to achieve the goals of the Thesis. Chapter 4 describes the overall methodology. Chapter 5 then describes the methods used to do the first stage evaluation of the algorithms using the reference bass recording as audio input. Chapter 6 describes the experiments for doing the student and summarizes the results of the analysis of student recordings. Chapter 7 discusses observations, implications and lessons learnt from the research and experiment.

1. State of the Art

The core of this thesis centres around the research and development of a model that can automatically assess a student’s performance of the four-string fretted electric bass guitar and provide them with the useful feedback that can help them improve. The research will focus on the specific qualities that playing the instrument entails for rhythm and timing aspects: onsets, notes duration and technical control.

* 1. Music Education

The mastery of any musical instrument without the live presence of teacher is a challenging task. Taking the application introduced in Chapter 1, “Yousician”, performance assessment for singing is very different to how bass (and other stringed instruments) are assessed. The piano roll format display that it uses for vocals feedback, gives the impression that note length is detected. However, a test on the Yousician-Premium-Plus version of the song “The one I love” by REM, shows that it doesn’t work. The scope of this research is limited to doing non-realtime assessments, i.e., the assessment of a stem recording as opposed to a live performance, so front-end displays are not the priority, but this example does raise awareness of format issue for displaying timing feedback.

The effectiveness of a Performance Assessment Technology (PAT) is the readiness level it provides to a student for auditions or music exams (e.g. TCL R&P) . The Teacher Grading for the experiment is based on the score ranges used by the typical music examiners in Modern Music Academies: Distinction 87-100; Merit 75-86; Pass 60-74 and below pass.

Music Critic[[3]](#footnote-3) is currently developed for Guitar at the MTG[[4]](#footnote-4) (Music Technology Group) and has addressed the rhythm assessment [5] using the SpectralFlux onset detection function. Music Critic is non-realtime: it does not flash green or red on each individual note “on the fly” like in the edutainment apps introduced in Yousician.

* 1. Onset Detection

Before a technical discussion of onset detection can begin, the relevant background concepts (energy and events) need to be described. For a given hop size “h”, hop number “m”, the local energy of the frame of an audio signal “x” can be calculated as follows

Equation 1 Local Energy Equation

The energy “E” is typically measured for a specific energy band (e.g. for bass guitar typically from 25 Hz to 500), so the Energy is calculated from bin index k1 to k2:

Equation 2 Energy in a Frequency Band

where X[k] is the STFT of x(n). An onset detection function (ODF) is used to detect a musically meaning event in an audio signal and returns one value per audio frame. In the case of bass guitar, a pitched percussive instrument, we are concerned with the plucking event. 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)*

Equation 3 Onset Detection Function Equation

A more general strategy for detecting onsets involves various stages

Preprocessing

Onset  
Detection  
Function

Peak  
Picking

Thresholding

Onsets

Signal

*Figure 1: Onset Detection Stages*

The pre-processing is an optional step that emphasises or attenuates aspects of the signal, for example separating the signal into different frequency bands. The ODF reduces the signal to subsampled occurrences of transients. The ODF method is the core stage and is discussed in the next section. The Thresholding stage of the Onset Detection Stages involves eliminating peaks which are not related to the onset event. Bello discussed how a median filter is used to obtain an adaptive threshold curve [6], but the scope of this thesis will be to hold a constant threshold. The Peak Picking stage involves choosing the local maxima after threshold filtering and examples are discussed in the next section on methods.

* 1. Onset Detection Methods

The ODF method can be chosen based on instrument class: pitched/ non-pitched; percussive /non-percussive. The Spectral Onset Processor[[5]](#footnote-5) (SOP) implements several onset detection functions based on magnitude or phase information from the spectrogram of an audio signal [6]

The ODF Methods Spectral Difference and Phase Deviation methods [6]scored the highest accuracy metrics in for pitched percussive instruments. The SOP algorithm in Music Critic tools (back end for Music Critic) for detecting guitar onsets uses the 'superflux' method. For the final stage: “Peak picking” different strategies can be employed here varying from the PitchSalienceFunctionPeaks used in the MELODIA [7] algorithm to the Peak Picking formula in Madmom. This is the formula employed in this thesis and is defined as follows [8]



Equation 4 Peak Picking Formula

Ramon Romeu[[6]](#footnote-6) developed a wrapper based on the above formula with the following values determined for *C\_t, H* and *delta* in sweep experiments

sodf = madmom.features.onsets.SpectralOnsetProcessor(**'superflux'**, diff\_frames=20)

det\_function = sodf(audiofile, fps = fps)  
det\_function\_norm = det\_function/(max(det\_function))

#Dynamic threshold

C\_t = 0.99

H = 100

delta = 0.1

din\_th = np.zeros(len(det\_function\_norm))

for m in range(H, len(det\_function\_norm)):

din\_th[m] = C\_t\*np.median(det\_function\_norm[m-H:m+H])+delta

*Figure 2: dynamic threshold setting*

The values given for C\_t, H, delta can be customised for different Dataset tracks. A sweeping method can then be sued to optimise the detection for a given musical piece.

In addition to Madmom libraries, the Onset Detection function from the Essentia[[7]](#footnote-7) library with examples[[8]](#footnote-8) allows you to choose the following methods:

* hfc
* complex
* complex\_phase
* flux
* melflux
* rms

Another method not listed above is the Wavelet method [9]. It places focus on precise time localisation, an important aspect of bass rhythm, however it did not score as high as the previous too methods in accuracy metrics.

* 1. Offset Detection

While onsets have been well researched by Bello [9]and present a clearly defined points in time for pitched percussive instruments, offsets present more challenges. These include

* identification of stable voiced regions
* identification of the hearing cut-off point

2.4.1 Perceptual and musical considerations

Terhardt [10] showed that with a constant physical duration, the subjective duration of pure tone is reduced when the frequency is reduced when testing with 200Hz and 1kHz tones, but the situation gets more complex when if dominant harmonics are included. Research on the auditory system proposes that we have a temporal energy integrator [11]. The integration time is estimated to be between 50ms and 200ms with some researchers proposing longer values for low frequencies. The relationship between the temporal resolution of the auditory system and onsets is beyond the scope of this thesis. The focus is on “just noticeable difference” or JND when it applies to note length. The question is whether there is sufficient duration to allow an offset response. Kopp-Scheinpflug [12] considers gap-detection and the limits of human auditory temporal acuity which varies from **2/3ms to 30ms** depending on the level of spectral disparity in the signal. This would suggest a more appropriate value of about 12.5ms compared to an onset time window of 50ms [13] in AMT applications

Note duration is the difference between onset and offset and this is the key metric that needs to be identified for student feedback. This section considers a bass example If you play 4 crotchets, 4 quavers, or 4 semiquavers, you get a “green” positive feedback signal in Yousician, even though there is a clear musical difference in each of the displayed three bars below

*Figure 3: 4 crotchets, 4 quavers and 4 semi-quavers.*

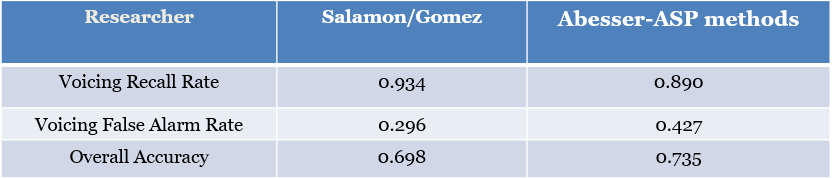
The challenge of duration measurement is that it is not an instantaneous measurement. It’s a bit like the SPECS Speeding camera system where you have start point and an end point. An offset (end point) cannot exist without a start point onset. For offsets, the “exit point” is really an open question for non-muted playing technique.

For offset measurement, energy capture is considered to determine “drop off” as. In any given song you may find a variation in the finger style techniques used for playing bars. A staccato style results in shortening the duration of the notes but also lengthening the inter-note interval. A legato style will result in the offset of a given note to run into the onset of the next note. Clearly a different strategy needs to be applied to each scenario. The case where the offset position of a given note exceeds the onset of the subsequent note is not considered. This would be like holding down the sustain pedal on a piano while playing another note or letting an open string play while you pluck another string.

2.4.2 Voicing Classification

Voicing Classification is an algorithm that determines whether a sound is pitched or not. The voicing classification is a useful measurement instrument that is used in automatic music transcription and can also be useful in performance assessment algorithms that consider timing measurements and exclude pitch accuracy. Jacob Abesser’s applied research in music information retrieval is focussed primarily on bass. The voicing recall statistics for MELODIA [7] techniques compare more favourably against Abessers signal processing methods [14].

*Table 1: Abesser Signal Processing vs Salamon/Gomez[[9]](#footnote-9)*



2.4.3 F0 detection and Segmentation

In this section the relevant background concepts to offsets (fundamental frequency, segmentation, and transcription) are described. For voiced segments of an audio file, the next key component is the estimation of the fundamental frequency (F0), which can be done in Hz or in a more musically meaningful representation in cents. For timing considerations, we are not interested in the actual frequency, but the beginning and end of the segment with a given F0 or fundamental. F0 detection facilitates the segmentation process for music transcription and theoretically it can also be applied to music education applications. One example of F0 detection is the Two Way Mismatch (TWM) algorithm used sms-tools[[10]](#footnote-10), that starts with F0 candidate choices based on peaks of the magnitude spectrum within a specific range of frequencies. Improvements could be improved here to consider sub-harmonics of peak frequencies. The challenge is to also find a vibrato-resistant stable regions. Abessers developed a F0 Tracking function with code available[[11]](#footnote-11) in the pymus libraries. His research on Offset detection [14] considers the problem of Vibrato and is tested with example of this the IDMT dataset [15] in Chapter 3. The full matlab/python code Abesser used for calculating onsets and offsets is closed source and has not been tested in this thesis.

The PitchMelodia[[12]](#footnote-12) function from the Essentia libraries is based on the MELODIA [7] algorithm. A customised version of PitchMelodia is the PredominantPitchMelodia function which can work with polyphonic music signals. Both are based on 4 algorithms that are called in a chain

*Table 2: MELODA Algorithm steps*

|  |  |  |
| --- | --- | --- |
| Function Name | Input | Output |
| PitchSalienceFunction | Spectral Peaks of Audio | Pitch salience function. |
| PitchSalienceFunctionPeaks (like SOP peak picking) | Pitch salience function. | Bins  Saliences |
| PitchContours | Bins Saliences | Contour Bins  Contour Saliences  Contour Start Times  Duration |
| PitchContoursMelody | Contour Bins  Contour Saliences  Contour Start Times  Duration | Pitch  Pitch Confidence |

The tutorial example[[13]](#footnote-13) shows how the individual algorithms and parameters choices (min. and max frequency, voicing tolerance, maximum number of peaks) are employed to estimate the frequency contour of a given audio. The algorithm “PitchContour” has two important parameters for timing:

timeContinuity - the maximum allowed gap duration for a pitch contour and is defaulted to 100ms.

minDuration - the minimum allowed contour duration

For bass segmentation the min. and max. frequencies were chosen as 25 and 340 Hz respectively in the experiments run in this thesis. For the song “wotm” the timeContinuity was set at 14ms and the minDuration set at 112 ms.

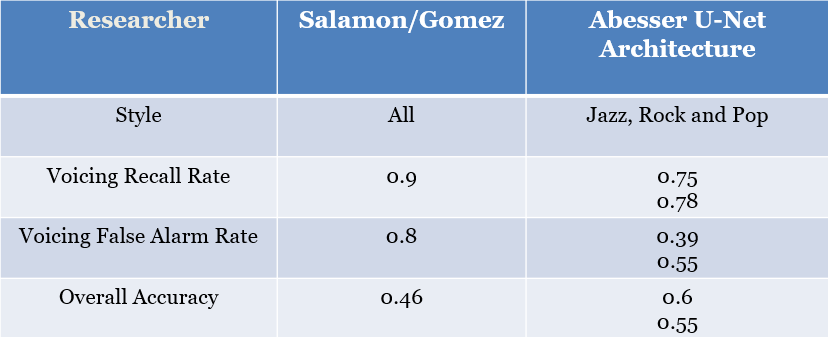
* 1. Neural Network Strategies

Current State of the Art for MIR research focuses more on machine learning / deep learning techniques and onsets/offset detection are no exceptions.

Bass U-Net [[14]](#footnote-14) is based on a Abessers-Müller Data driven algorithms [16]. It uses a CNN (Convolutional Neural Network) Streamlined Encoder/Decoder Architecture for Melody Extraction. It has been tested on the following Datasets:   
- Real World Computing (RWC) [17]  
- MDB-bass-synth [18]  
- Weimar Jazz Database (WJD) [19]

The MELODIA algorithm [7] by Salamon/Gomez as well as producing better voicing recall rate than Abessers [14] F0-detection-based techniques, it also produced better recall rates against as shown in Table 1. However, Bass U-Net[[15]](#footnote-15) produced lower false alarms.

*Table 3: Abessers Data Driven algorithms vs Salamon/Gomez*

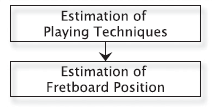


The Voicing accuracy rates still fall short of the requirements for Music Education applications, but there is a trend towards data driven techniques.

Bocks Online Onset Detector[[16]](#footnote-16) is a universal onset detector with BLSTM based on current neural networks and was trained with music mixtures including R&P. Bello has shown that online methods compare favourably with offline methods. which is good news for realtime performance assessment in this area. The main obstacle in all data driven approached in going forward is getting access to useful trained data to drive up accuracy metrics top above 90%.

* 1. Feature selection and Technical Control

In the section on Offset Detection, Abesser’s ASP methods [21] were introduced in terms of their accuracy for identification of voicing section. Besides aiming the research at score level parameters (onset, time pitch and duration) used Feature Selection to extract the plucking and expression style from single note recordings. In the context of R&P performance assessment, this technique could be applied to validate if a student is using the correct technical parameters in a recording. In the single note recordings of the IDMT dataset, you can hear the sonic difference in between plucking styles: e.g. (FS) finger style plucked note and a plectrum picked note (PK). He also introduced five different expression styles: normal (NO), vibrato (VI), bending (BE), harmonics (HA), and dead-note (DN). He also made possible the estimation of the additional instrument-level note parameters string number, fret number. The proposed system achieves accuracy above 0.88 for both left and right hand techniques, thus complimenting similar research on expression styles for guitar by Reboursiere [20].



**Fig. 5. Processing Flow chart of the instrument level extraction process**

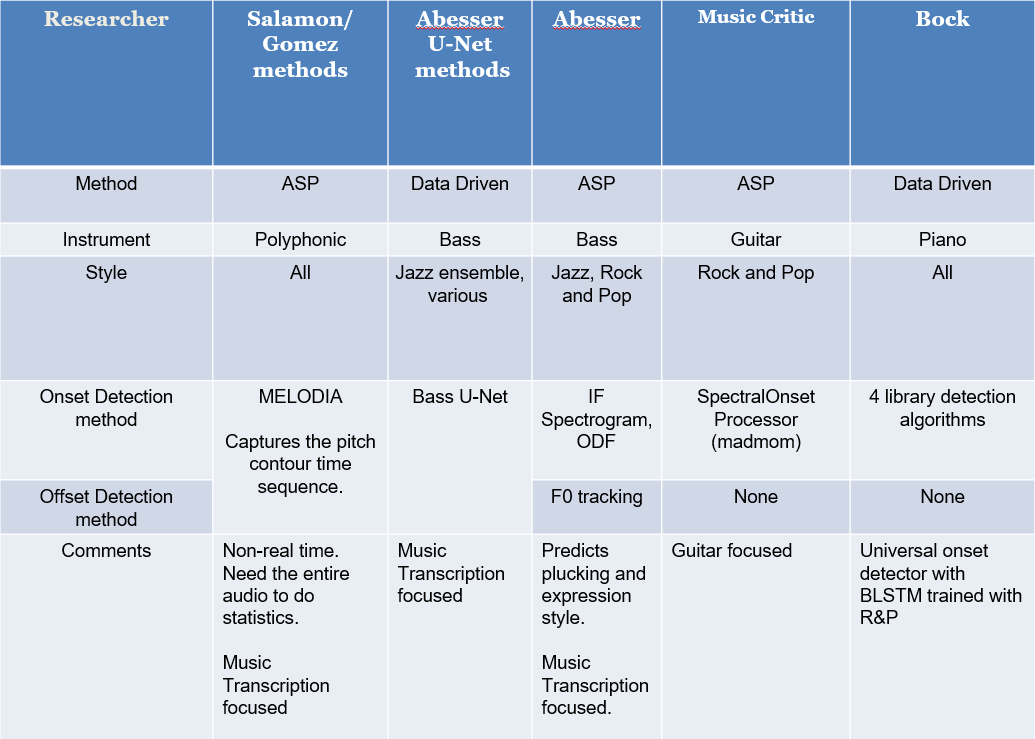
**(borrowed from [5] Instrument-Centred Music Transcription of Solo Bass Guitar Recordings, Jakob Abeßer and Gerald Schuller, Senior Member, IEEE)**

In the context of state of the art tools on the market, Yousician is capable of correctly such as Hammer on (HO), Pull Off (PO), Slide (SL) and also provides real-time recommendations of what finger to use in holding a particular fret position. Trinity Exam examples also publish bass tablature showing recommended string and fret number positions. As an outcome, it is now possible to correctly classify the use or non-use of certain technical parameters while assessing musical performance

* 1. Summary

The following table summarizes the state of the art of algorithms used for onsets and offsets, many of which use these library functions.

*Table 4: Algorithms for Onset/Offset detection*



The final column summaries the central method of the thesis described in the Methodology chapter.

* 1. Observations

The choice of suitable machine learning techniques to extract the most relevant parameter is constrained limited number of annotated student performances. Source Separation techniques based on the Spleeter model [22] have been successfully applied to the songs from Rock and Pop repertoire and could be used to create a large dataset if the recorded examinations were made available. Scaling up would mean wider ML options, like the Support Vector Machines used in the extraction of the plucking and expressive styles [14] .

For music education, the best way forward would be to build an additive Dataset while experimenting. What this means is that a large initial experiment is performed to build a trial performance assessment system based on Linear Regression. At the end of a trial course where a large number of students submit their recordings along with their own “self predicted grades”, this could build new data set that could be used for training a new database for the Bocks Online Onset Detector[[17]](#footnote-17) or Abessers Bassunet[[18]](#footnote-18) .

1. Datasets

This chapter explains the two Datasets that are used to evaluate algorithm accuracy.

* 1. IDMT BASS SINGLE TRACK

The IDMT dataset [15] consists of 17 audio tracks with accompanying score and annotated onsets/offsets with various levels of complexity. Each score is accompanied by a WAV audio file and an XML file with various annotations including MIDI pitch, onset, offset and other instrument characteristics as shown below:

<event>

<pitch>36</pitch>

<onsetSec>2.4</onsetSec>

<offsetSec>2.5552</offsetSec>

<fretNumber>3</fretNumber>

<stringNumber>2</stringNumber>

<excitationStyle>FS</excitationStyle>

<expressionStyle>NO</expressionStyle>

<modulationFrequencyRange>0</modulationFrequencyRange>

<modulationFrequency>0</modulationFrequency>

</event>

<event>

<pitch>36</pitch>

<onsetSec>2.7</onsetSec>

<offsetSec>3</offsetSec>

<fretNumber>3</fretNumber>

<stringNumber>2</stringNumber>

<excitationStyle>FS</excitationStyle>

<expressionStyle>NO</expressionStyle>

<modulationFrequencyRange>0</modulationFrequencyRange>

<modulationFrequency>0</modulationFrequency>

</event>

*Figure 4: Extract from IDMT Dataset. File 002.xml*

This is not used for Student performances. IDMT contains many tracks that have complexity exceeding that required for grading required for this thesis. It also has some very short notes where the plectrum is used for the style. The objective of this Dataset is to measure the effectiveness of the onset/offset algorithms. It is available under a creative commons licence.

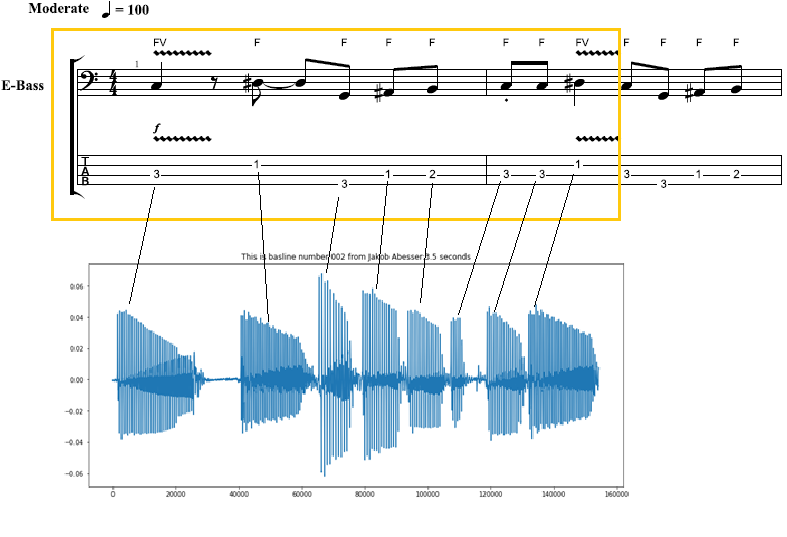
The above annotations show that there is a clear difference in duration when you subtract offset from onset: 159ms for staccato and 300ms seconds for normal.

IDMT (Fraunhofer) : 17 tracks

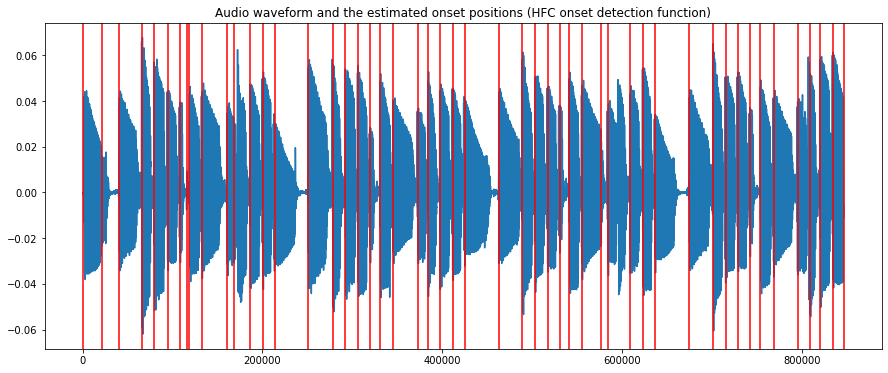
* Varying plucking techniques (Pick, Finger, Muted, Slap)
* PDFs of score and XML of parameters (onsets, offsets, pitch, fret number)
* Expression Style annotated

IDMT-Example on note length

The annotation of a musical note as “staccato” or “legato” can have a big impact on the intended duration. In the figure below you can see that the first C note in the 2nd bar is almost 1/3 the duration of the second C note.



*Figure 5: Note duration of Staccato notes*



*Figure 6: Audio 002.wav (IDMT DATASET) with onsets*

By examining the musical properties of the IDMT dataset of 17 songs, 4 songs were shortlisted for performing the formal evaluations of the algorithms for onset and offset detection. At the beginning when the MIR evaluation window was set high to 50ms, some “stress tests” were made using remaining tracks. However, the lower value of 20ms window made it difficult to achieve good accuracy for difficult tracks ( with slap style for example). The criteria of track selection were based on note the minimal use of semi-quavers, finger style plucking and overall complexity level suitable to lower – medium level grades for music academy exams.

* 1. Rock and Pop repertoire

Six songs with increasing levels of difficulty were chosen for the R&P repertoire with available score (in XML and PDF format) and access to separate stems.

*Table 5: TCL dataset*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Song Nr** | **Song** | **Artist (original)** | **Shortened Name** | **Length** |
| **0** | Yellow | Coldplay | yellow | 93 |
| **1** | Billie Jean | Michael Jackson | bjean | 55 |
| **2** | Just Looking | Stereophonics | just | 86 |
| 3 | Brown Eyed Girl | Van Morrison | brown | 64 |
| 4 | Roadrunner | Junior Walker and the All Stars | road | 82 |
| 5 | Walking on the Moon | The Police | wotm | 120 |
|  | TOTAL |  |  | 500 secs |

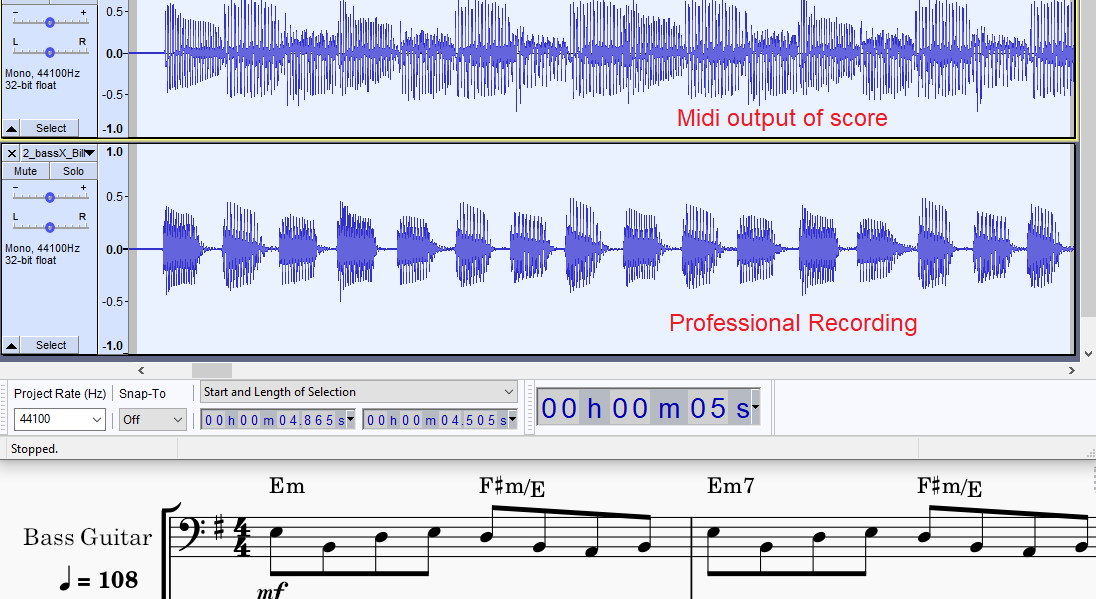
The shortened names in column 4 shall serve as references to the songs. The “Length” columns indicate the duration of the song from the beginning to be considered for grading.

The Dataset was constructed for all six songs with the following attributes

* Song Descriptions in text.
* Onset/offset annotations of reference bass recording.
* Multiple Student recordings of each song
* Mixes of the Student Recordings with the Minus 1 tracks
* Music Teacher Grades (Numeric and Descriptive )
* Algorithm generated data on the Student and Ground Truth stems.

The Onsets and Offset were initially done by visual inspection in Sonic Visualizer[[19]](#footnote-19). This was improved by running the best performing onset and offset algorithms on the “reference recorded” Stems, removing the false positives and add the visually adding the missing notes. This resulted in a much more consistent set of annotations, so the mean and standard deviations of the onset/offset deviations of ground truths were close to zero.

A “mechanical rendering” of the score can provide a useful reference in understanding the song description elements that are not shown in the score. For “bjean” it is required to pluck with “short/jerky movement” in the verse/chorus, but no dotted notation is used, so the midi output will not reflect desired playing style.



*Figure 7: Billie Jean: Midi Rendering vs Human recording (shorter notes)*

The first function of the R&P dataset is to validate the algorithms short listed state of the art algorithms after testing with the IDMT dataset. The second function is to provide a basis to measure student performances. The reference bass stem is assumed to represent a grade of 100% in timing and rhythm.

Table 6 summarizes the individual musical features of each of the tracks, with their Grade Level (difficulty level) and technical description included [23]. These are the topics of the Grading Sheet questions given to the Bass teacher.

*Table 6: Tech. Control Parameters for 6 TCL songs*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Song** | **Coord.** | **Syncopation** | **Repetition** | **Dynamics** | **Articulation** | **Note Length** |
| Yellow  Grade 0 |  | chorus | No rushed feel |  |  | Play evenly (verse) |
| Bjean  Grade 1 | intro | Accent just before chorus |  | -- | separate jerky quavers+ smooth, melodic material |  |
| Just  Grade 1 |  | Chorus: accented, syncopated motif.,  hard accent |  | Unexpected  subito p at bar 25. |  |  |
| Brown  Grade 2 |  |  |  |  |  | different note lengths and rests |
| Road  Grade 3 |  |  |  |  | Tenuto (underscore)  loud on beat 1 |  |
| Wotm  Grade 3 |  | syncopated  repeated notes |  |  |  | correct  separation |

In the experiment, truncation is applied to the songs, to facilitate grading a section of music once, for example “yellow” was truncated to remove the repeated verse, since no new musical features were introduced.

* 1. Musical Properties of R&P repertoire

The experiment described in section will deal with the tailoring of the teachers questions and grading options for each of the 6 TCL chosen songs based on their musical properties as shown in table 5. The two common grades that apply to all songs are Onset accuracy and Offset accuracy

Two songs, “wotm” and “bjean” “ from the dataset were given more attention because they had interesting note-gap and note length differences between verse and bridge. The “wotm” verse has a long note duration with zero inter-note gaps while the bridge has the opposite: shorter “reggae” notes with some noticeable inter-note gaps. The BPM of this song is 146 and for Billie Jean it is 108. The Grade1 “Billie Jean” song also manifests similar contrasting sections.

In chapter 4 the methodology will be discussed

1. Methodology

Music Education demands more precise timing measurements than state of the art automatic music transcription. Onset detection for bass requires a different approach from current assessment technologies in three key areas. First, as a rhythm section instrument the bass plays a key role in synchronizing with the drum pattern, so a tighter precision is required than the one used for guitar. Secondly, the bass does not require the handing of playing chords that the six-string guitar does. The accuracy levels for bass onset detection should exceed the accuracy achieved by playing guitar melodies. Finally, the choice of onset detection algorithm has to sit with an offset detection algorithm to measure duration.

Abesser used the standard MIR evaluation of 50ms for measuring onset accuracy in Guitar onsets and increased this to 200ms for offsets due to difficulty in handling smoothly decreasing note envelopes [24].

The basic work flow idea to start with was to check the algorithms from the summary table of State of the Art in Chapter 2 with starting with IDMT Dataset then moving onto R&P Repertoire Dataset. The circular workflow would allow continuous improvement of accuracies, by combining algorithms, checking algorithms on different dataset and different song sections in the R&P Dataset.

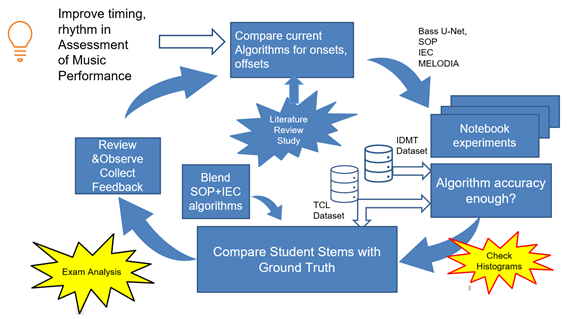
However, some algorithms e.g

**Computational Techniques**

Choice of Frame and Hop Size.

In the end a hybrid algorithm that would apply different threshold parameters to different songs and different offset strategies to different sections of a given song was chosen.

Getting good Precision, Recall and F-Measure metrics for the professionally recorded reference was only half the battle. When gathering and testing the student recordings it proved very difficult to obtain a precision value higher than 60%. In other cases, the performance of an algorithm would deteriorate rapidly when reducing the MIR eval window from 50ms to 20ms. So, another cycle of choosing, parametrizing, and customising an algorithm would have to be repeated for the particular song in the dataset as illustrated in fig. 10.



*Figure 8: Strategy to experiment, test to find best accuracy*

Fig. 10 The accuracy of the algorithms needed continuous improvement to be meaningful for music education standards.

It is important to point out that although PRF figures for Offset detection are measurable and shown in the table, they are not considered as predictor inputs since they are dependant on the onsets

Chapter 5 discusses investigate different onset and offset detection strategies and how to test them with the Datasets discussed in chapter 3. It is expected that songs like “bjean” and “wotm” will present challenges for the analysis techniques because of the diverse audio features within the same song.

1. Algorithm Development and evaluation

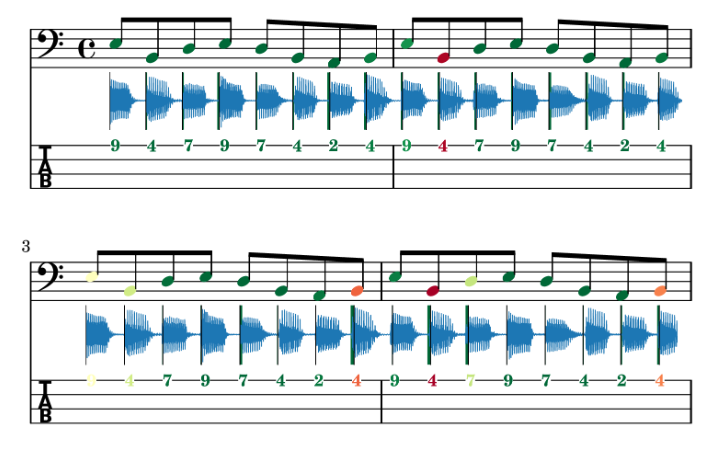
This chapter discusses the following the benchmarking of algorithms for accuracy. It opens with the end-to-end test on a R&P repertoire song using the current State of the Art method used in the Music Critic.

* 1. Music Critic End-to End test

In MUSIC Critic a JSON file is used to mark the overall duration of the songs the beat locations in time. The LY file marks the note pitch, its length and its location in the overall on pattern. To perform the End-to-End test for Billie Jean, the following steps were executed:

* Calculation of Beat positions using Madmom.
* JSON, Lillypond file preparation for Billie Jean.
* Creation of New “Minus-1” track

The diagram below shows the graphical output m of the first experiment



*Figure 9: End to End test of Music Critic with Billie Jean GT*

The alignment of the ground truth stem was not aligned 100%. One possible explanation might be the customisation of the wrapper of the Onset detection function to consider multiple onsets that can occur with guitar chords. The outcome of this experiment is that the Onset detection methods as used currently for guitar need to be improved.

* 1. Algorithm Evaluation



In the algorithms introduced in the literature research in chapter 2 there are two strategies involved: “paired” and “non- paired”. Non paired are the classical approaches to capturing the onset without regard to monitoring the end of the note. The paired approach is to measure start and stop time with the same audio frame. The non-paired approach refers to a method where the onset data on its own.

The 4 methods introduced in table 3 ,chapter 2 were tested alongside a fifth method to implement combined approach.

* Madmom Online Onset Measurement (Non-paired, Data driven)
* SOP: Spectral Onset Processor (Non- paired, ASP driven)
* AbesserUNet Algorithm (Paired, Data driven)
* SG: Salamon and Gomez (Paired, ASP driven)
* IEC: IndexedEnergyChecker (Paired ASP driven)

Precision and Recall, and F-measures were made on Onsets and Durations.  
Duration = Onset-Offset

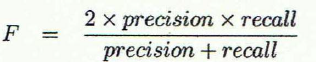
• **Precision**: exactness – How often did the algorithm incorrectly detect an onset/duration? (imposter notes)



• **Recall:** completeness – How often did the algorithm fail to detect an onset/duration? (missing notes)



• ***F* measure (*F1*** or ***F*-score)**: harmonic mean of precision and recall



Deviation measurements shall be made on the difference between the Ground Truth and Measured Value. From this Histograms are derived following methodology used in [5]  
the Mean Absolute Error and Standard Deviation are calculated. We are concerned with two key measurements criteria:

Attack (for onset detection) and Release /Decay (for offset detection).

The attack is the most noticeable for performance assessment. Since for bass we are discounting chords, we do not have to consider multiple attack times. The resulting Precision, Recall and F-measure for onsets for each of IDMT with a 20ms evaluation window averaged over 4 songs that are most like the TCL Dataset.

*Table 8: Benchmarking Onset detection algorithms using IDMT DS*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Researcher** | **Simple Energy Checker** | **Indexed Energy Checker** | **Bock (SOP)** | **MELODIA** | **Abesser U-net** |
| Precision | 0.425 | 1.0 | 0.894 | 0.933 | 0.67 |
| Recall | 0.518 | 0.649 | 1.0 | 0.927 | 0.797 |
| F-measure | 0.465 | 0.783 | 0.943 | 0.93 | 0.728 |

The experiment also extracted Precision, Recall and F-measure for offset, for the IEC, Melodia and Abesser U-net algorithms, but they are not shown, since offsets depend on onsets.

* 1. Indexed Energy Checker ( IEC)

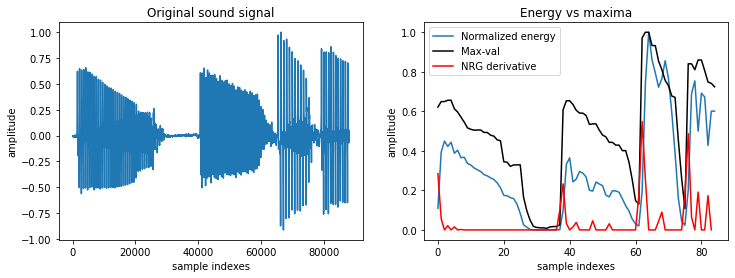
Originally a simple Energy Checker function called “calculateOffsetOnset()” was developed to capture RMS band values of an input signal. Subsequently those RMS band values would then be checked in sequence to see if they dropped below a certain threshold, in which case they would be added to a new offset array and a flag would be set to indicate an offset was detected. The flag would have to be cleared before threshold drop off detection could start again.

index= 0  
array\_of\_time\_offsets= []  
flag = False  
last\_index=0  
while index < len(rms\_bands):  
 if (abs(rms\_bands[index])<threshold) and flag == False and (index!=0):  
 *# Skip very first* array\_of\_time\_offsets.append(index)  
 flag = True  
 last\_index=index  
 index+=1  
 *#We set flag back to false after determined time period* increment\_factor= int(hopSize/hopSizeScaleFactor)  
 if index > last\_index+increment\_factor:  
 flag = False

*Figure 10: Extract Original Energy Checker for offset detection*

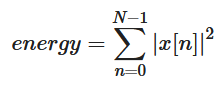
The main problem with this algorithm was that it did not calculate corresponding onsets within the same frame. The Onsets were calculated separately using standard HFC method, but these values were independent of the offset values. Too much effort would be needed to align them and fix this “un-paired” problem.

The IndexedEnergyChecker[[20]](#footnote-20) (IEC) algorithm, derived from combining functions in the song assessment python notebooks, pairs the onsets and offsets, based on an Energy Threshold. It is based on the concept of a Sound Island. Depending on the energy threshold level set, it decides on how to split the wave boundaries. The splitting is represented by set of start and stop indices representing onset and offset. This threshold parameter was hard coded to 0.05 in his experiments with the saxophone. For the TCL Dataset, it was configured uniquely for each song.



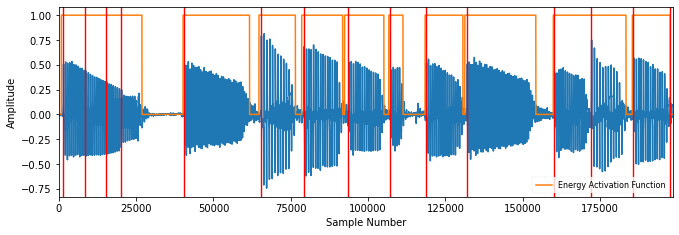
*Figure 11: Normalized Energy of Audio 002.wav sample (IDMT DATASET)*

Energy is calculated using the normalise energy function



Equation 5 Normalized Energy Function

The returned parameter “split\_decision\_func” function is an array of 1s and 0s that can be plotted as an overlay to the sound wave to give a graphical view of start and stop times of each voiced section



*Figure 12: Sound Islands of Audio 002.wav (IDMT). FS = 1024, HS = 512.*

This algorithm finds the best matching pairs with the following criteria  
 - distance between elements is no greater than matching\_window\_size

- sum of all distances is minimized

The red lines in figure 12 show that the onsets detected by the SOP algorithm throw more false positives for long notes.

* 1. Accuracy Metrics : IEC vs SOP

**5.4.1 Overall Performance**

To calculate Accuracy measures for SOP and IEC algorithms, PRFs were captured for onsets for each of the TCL bass recordings (stems) with a 20ms evaluation window compare both IEC with SOP.

*Table* *8: Accuracy Metrics for 6 TCL ground truths (combined vs non combined)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Song | P | | R | | F | |
| IEC | SOP | IEC | SOP | IEC | SOP |
| yellow | 0.939 | 0.98 | 0.942 | 1.0 | 0.94 | 0.99 |
| bjean | 0.949 | 0.983 | 0.969 | 0.997 | 0.958 | 0.99 |
| just | 0.682 | 0.9 | 0.828 | 0.869 | 0.748 | 0.884 |
| brown | 0.904 | 0.806 | 0.822 | 0.852 | 0.861 | 0.828 |
| road | 0.837 | 0.847 | 0.839 | 0.804 | 0.838 | 0.825 |
| wotm | 0.732 | 0.952 | 0.975 | 1.0 | 0.836 | 0.975 |

**5.4.2 Billie Jean**

The methodology in this chapter was concentrating on one particular song, “bjean”. It was a great candidate for testing the “hybrid algorithm”, which combined SOP and IEC techniques. This song used customized time tags for switching to the IEC to the verse and the SOP to the bridge and the result was as follows:

*Table 9: Accuracy Metrics for Billie Jean Hybrid Algorithm*

|  |  |  |  |
| --- | --- | --- | --- |
| Song | P | R | F |
| Hybrid | Hybrid | Hybrid |
| bjean | 0.9893 | 0.97555 | 0.9824 |

* 1. Accuracy Metrics: IEC vs MELODIA
     1. Long Notes

The songs “just” and “wotm” were interesting candidates for containing long notes in different sections of their songs. The relationship between the “long note-short gap” pattern and the low PRF score for IEC was observed. They were tested with SG [7] to see there were better accuracy alternatives for algorithms that could detect both onset and offset.

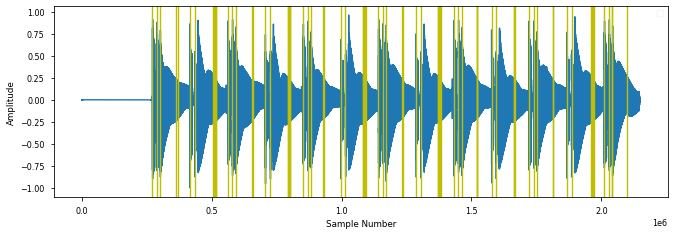
*Table 10: Selected comparison: IEC vs SG*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Song | P | | R | | F | |
| IEC | SG | IEC | SG | IEC | SG |
| just | 0.682 | 0.78 | 0.828 | 0.745 | 0.748 | 0.761 |
| wotm | 0.732 | 0.794 | 0.975 | 0.653 | 0.836 | 0.716 |

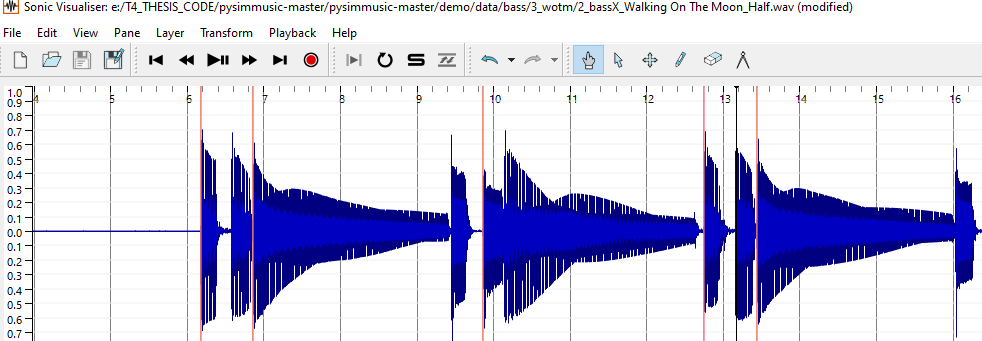
* + 1. WOTM Bridge and Verse Comparison

**WOTM Verse-IEC Algorithm**

Since there was no improvement, I broke the song “wotm” into components to see which song sections could be used for the paired-technique (detecting both onsets and offsets). For the first half of the track, “wotm” , using the IEC with Threshold set to 0.06, yielded a lot of false onsets, hence the low precision in the PRF reading: P= 0.514, R = 0.949, F = 0.667



*Figure 13: IEC onsets for “wotm” verse*



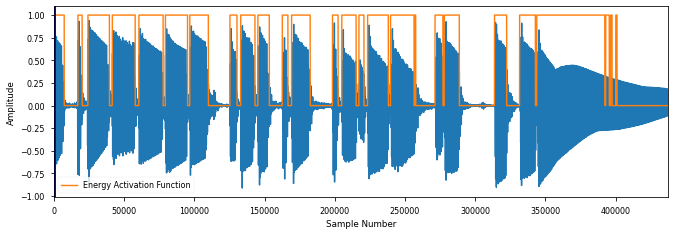
*Figure 14: Section of “wotm”verse showing missed IEC onsets.*

**WOTM Bridge -IEC Algorithm**

For the bridge section of the same track “wotm”, the IEC performed better, but again the PRF accuracies improved when omitting the last long note.

P =0.937 R= 0.949 F=0.943 (With last note)

P =0.961 R= 0.948 F=0.954 (Without last note)

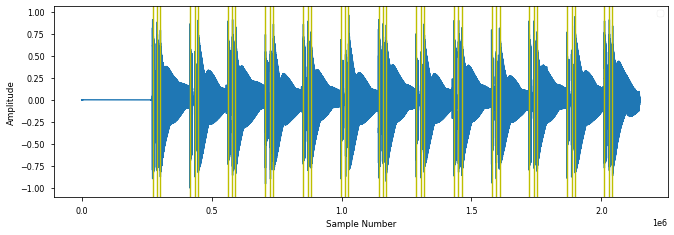
****

*Figure 15: Calculated IEC Sound Islands for WOTM bridge, P = 0.937*

**WOTM Verse -Melodia Algorithm**

This song has the longest note length of all the tracks. With Threshold = 0.06, we still see a lot of hysteresis on the last sustained note in Figure 17 and this kicks 4 percentage points off the precision.

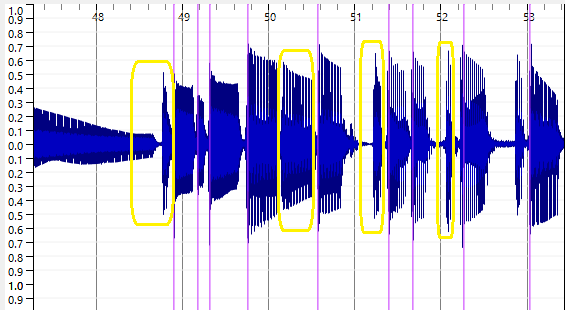
In contrast the SG [7] algorithm performed a lot better in the “wotm” verse with PRF measures of P = 0.974, R = 1.0 and F = 0.987.



*Figure 16: PitchMelodia derived Onsets for WOTM verse , P = 0.97*

**WOTM Bridge -Melodia Algorithm**

But SG [7] missed a lot of onsets in the bridge section of “wotm”.

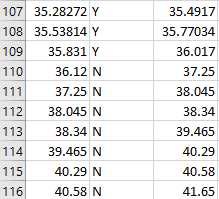
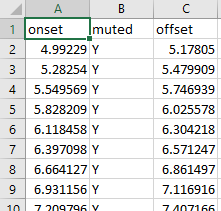


*Figure 17: PitchMelodia derived Onsets for WOTM bridge*

The same behaviour was observed for “Just Looking” which also had long sustained notes. There are a lot of “misses” for the short notes in SG [7] method in the bridge as seen in fig 19 marked in yellow.

* + 1. Classifying Muteness

It is important to clarify some local definition of “mute”, that it does not relate to the concept of the left-hand muting technique described by Abesser [14]. Normally muting strings on the bass means damping them with left or right hand to short sounds with rapid decays. The term “soft mute” is introduced here to signify a gap of at least 10-20 milliseconds. When there is no gap., there is no offset. The offset in this case is equal to the next onset. This introduced the concept of an annotation file called the “”<song>\_rhythnm.csv” file with 3 columns: onset, muted and offset. Example extracts of “bjean\_rhythm.csv” is shown below for the verse and the bridge sections



*Figure 18: Sample of bjean\_rhythm.csv verse (left) and bridge(right)*

The middle column is marked Y when a slight muting of the string occurs. IF the srting is allowed to sound right up until the next onset then the middle column is marked “N”. You can see in row 110 that the offset is equal to the next onset value in row 111.

* 1. Accuracy Metrics for Neural Network Methods

Bass U-Net was tested using the Mixed Genre and Jazz Training models ('bassunet\_mixed.h5', 'bassunet\_jazz.h5)

The IDMT Dataset was tested for tested for PRF accuracies alongside “bjean”. The tables in the Appendix show that the IDMT Dataset precisions range from 0.22 to 0.65. In contrast, “bjean” returned the following: precision, recall, f\_measure\_value values:

P = 0.815, R = 0.969, F= 0.885.

The high recall is interesting, in that it is higher than precision. Overall, these figures make it suitable for Student comparison.

Following the Trend shown for “bjean” the same song was tested with Bocks Online Detector are summarized here.

The Bock Online Onset Detector returned the following results, with noticeable high recall.

yellow

precision, recal, f\_measure\_value 0.733 1.0 0.846

bjean

precision, recal, f\_measure\_value 0.713 0.997 0.831

* 1. Deviation Metrics

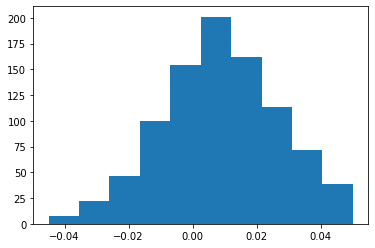
The following plots illustrate the deviation statistics for particular songs for onsets, considering the SOP method and the IEC for onset deviations, considering only the IEC for offset deviations. The MIR evaluation window for measuring the accuracy was set at 50ms to get a better understanding of the histograms shapes for onset and offsets. This tolerance was further extended for the offsets. The EnergyChecker Threshold was set to 0.05 The PRF accuracy measures are shown for the IEC and SOP algorithms using IDMT Dataset.

**IEC Measurements:**

Average precision: 0.947, recall: 0.632, f-measure: 0.75

|  |  |
| --- | --- |
|  |  |
| Onsets  Mean: 0.003599  Deviation from zero: 0.019611 | Offsets  Mean: 0.109678  Deviation from zero: 0.282235 |

*Figure 19: IDMT Dataset GT Onset + Duration deviations (IEC) algorithm*

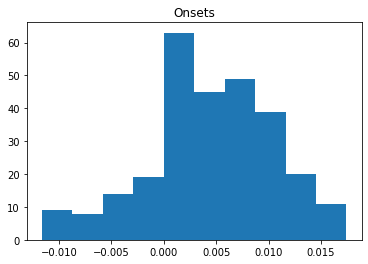
**

*Figure 20: IDMT Dataset GT Onset deviations ( SOP) algorithm*

**SOP Measurements:**

Average precision: 0.921, recall: 0.967, f-measure: 0.943

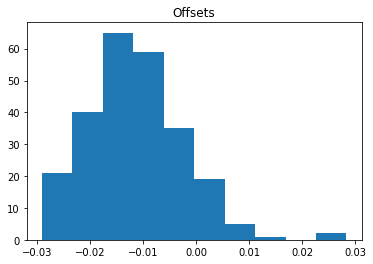
Doing a deviation measurement for Onsets for “bjean” with the MIR eval window set to 20ms yielded the following:



*Figure 21: Billie Jean GT onsets deviation*

SOP Onset Deviations  
ABS Mean: 0.005915, Mean: 0.004029,   
Dev. from 0: 0.007258

Doing a deviation measurement for Onsets for “bjean” with the MIR eval window set to 30ms yielded the following:



*Figure 22: Billie Jean GT offsets deviation (? Algorithm)*

Offsets ABS Mean: 0.012817, Mean: -0.011807, Dev. from 0: 0.014876

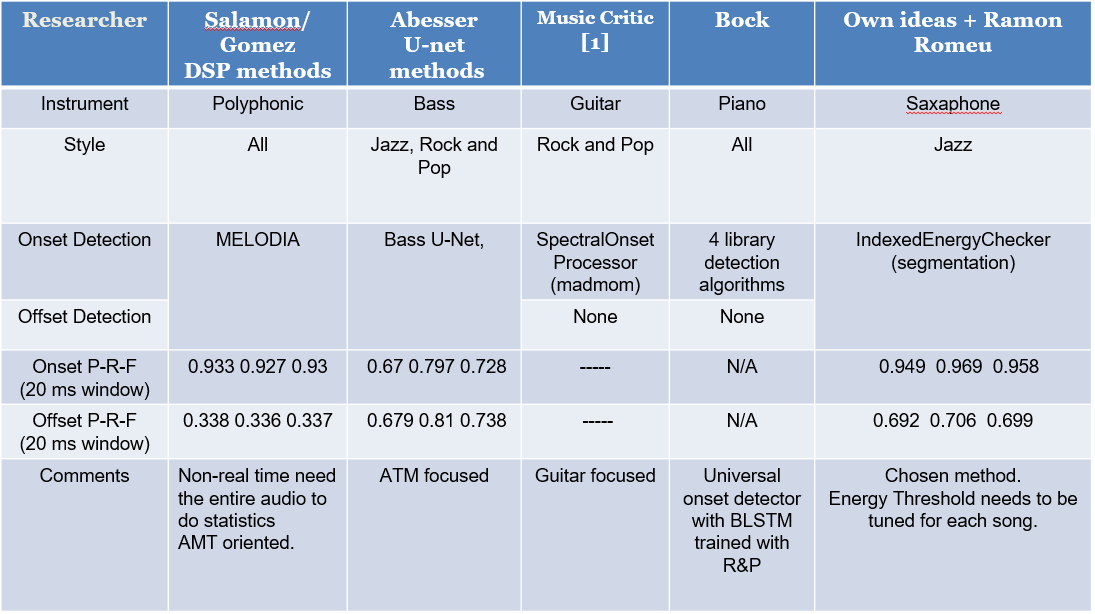
The same methodology for calculating deviation metrics were used alongside PRF metrics to assess the student performances.

* 1. Summary

The contrasting Precision, Recall and F-Measure readings for the “bjean” reference stem with the Indexed Energy Checker (Average precision: 0.921, recall: 0.967, f-measure: 0.943) and the SOP algorithm (Average precision: 0.947, recall: 0.632, f-measure: 0.75) gives initial results which prompt further investigation into finding out which part of the songs yield better accuracies for different algorithms.

The Neural Network strategies were not tested on the Students since there was not enough scope in the project to choose both ASP and Data Driven methods. The ASP path was chosen since there was no easy way to prepare H5 models using the R&P Dataset. The next chapter describes the main experiment of the thesis in which real teacher grades are obtained from real student recordings with the aim of allow us to predict grades on a set of test recordings.

*Table 11: Benchmarking Algorithms*



1. Prediction of Student Performance

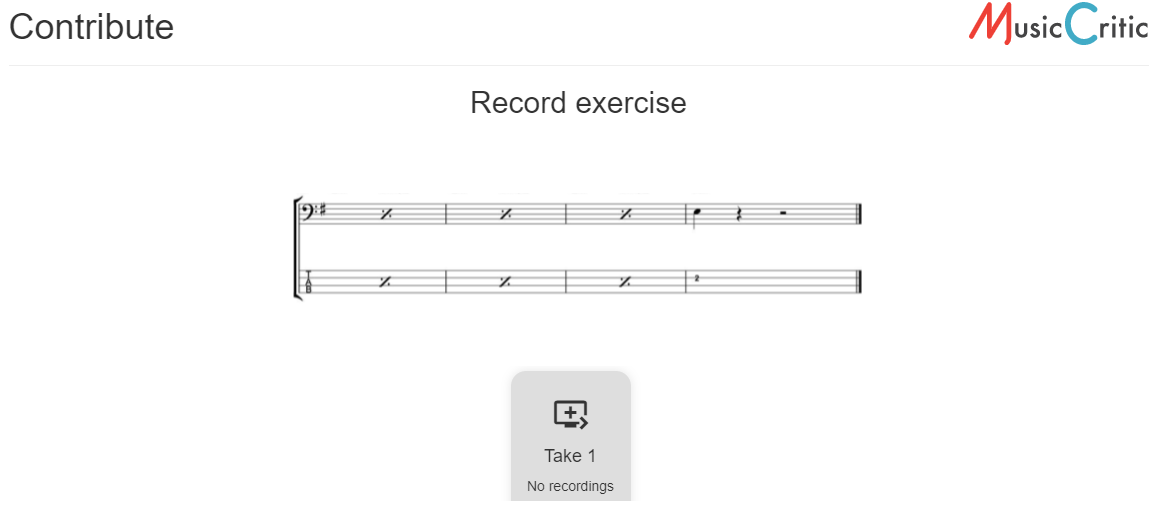
This chapter discusses the following:

* Student Portal and Teacher Grading
* The testing algorithms to measure student performances
* Using a Multi-variable Linear Regression to train new a model with teacher graded student performances

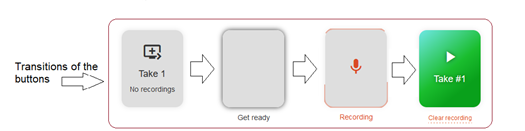
This chapter is dedicated to gathering and collecting data: onset/offset annotations, the student stem recordings, the mixes, the grades & comments on the mixes and using this to predict Teacher Grades. The method used is linear regression and will consider as X inputs the PRF results and the mean absolute error and standard deviation of the onset/offset deviations. The Y values are the specific grades the teacher assigns for metronome accuracy, for note length and the specific demands that each son requires for Technical Control. There has been a slightly higher deviation noted in the offsets. This is to be expected since there is no clear end point for long sustained notes.

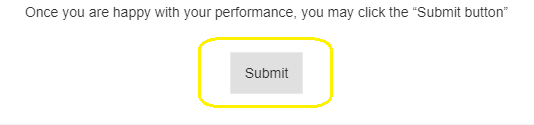
* 1. Student Recording Portal

The largest experiment was the campaign to find eight students to do at least one recording of each of the six chosen bass tracks. For this reason, each song has a Google Form containing instructions and links to a song portal (sample portal shown below) to perform the live recording with a backing track.



*Figure 23: Student Portal (copyrighted score removed)*

**

**

*Figure 24: Transition states on student portal*

The student presses the play button, listens and plays along with the option to listen back on the stem before pressing “Submit”.More Details of this Portal can be found in Appendix 12.1.

* 1. Processing of student recordings

After removing the latency from the recordings, it was noticed that the JSON calculated value still did not align with the first onset, so the student recording stems were aligned manually with the first onset. Apart from aligning the student stem like this, some amplitude boost was given to equalize the loudness of the bass stem with the backing track. This was achieved using as greater Signal Boost ( Audacity 2.4) to increase the volume, to be in line amplitude of the Minus-1 track.

A copy of the post processed student stem was then stored for analysis. A copy mixed with the minus 1 track was provided to the teacher for grading. In the initial recordings, the microphone on the headset was used, there was some noticeable background noise which affected the teachers sound quality grading. However, the algorithms were still capable of onset detection for noisy stems. Full details of all the steps involved in going from having a bass stem recording on a server, to a clean bass stem ready for analysis and a clean mix ready can be found in Appendix 12.1.

* 1. Teacher Grading Policy

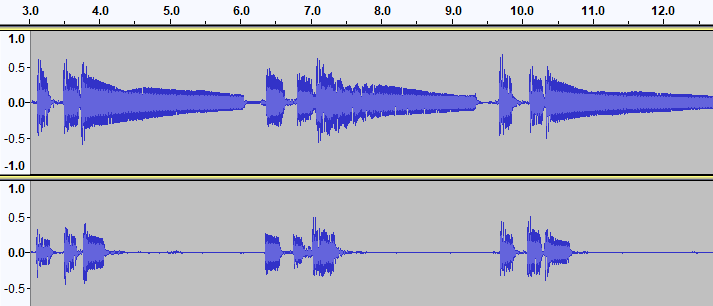
For each song there were grade given each of the following categories:  
Onset, Duration, Technical Control and Sound Quality.

Technical Control (TC) is song dependent, e.g. for “bjean” the main TF is “Articulation and Coordination”. Exam Grades were chosen to fit within the typical ranges of a music academy system: Excellent (100%), Very good (88%), Good with occasional lapses.(80%), Generally reliable.(63%), Unreliable (37%).  
The Onset and Offset Grade sections are common to all songs, with slight variations on how the questions was phrased. On the Technical Control side, the table below summarizes the topics covered by each question.

*Table 11: Technical Control Question Topics*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Question 1 | Question 2 | Question 3 |
| Yellow | Repeated Notes | Syncopation | Sound Quality |
| Bjean | Articulation | Dynamics | Sound Quality |
| Just | Accented syncopation | Dynamics | Sound Quality |
| Brown | Groove | Sound Quality | ----- |
| Road | Syncopation | Articulation | Sound Quality |
| wotm | Syncopation | Sound Quality | ----- |

Comment sections on each question were added to reflect how Music Academies correct exams.



*En esta cancion las notas deben ser mas largas al prinicipio y màs cortas en el "bridge".*

**Duration: 2 (Fair)  
Overall: 2**

*Muy buena "note duration". Buena diferenciaciòn entre las notas largas de la "intro“ y el "verse" y las cortas del "bridge".*

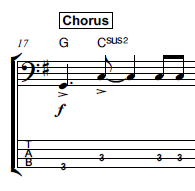
Duration: **5 (Excellent)**

Overall: **4**

*Figure 25: Contrasting comments of “wotm” performances*

Comments also proved useful to understand the rationale behind why a particular recording was graded a certain way. An additional global grade was added to help do consistency checks in the combination of the other grades allocated.

As mentioned in the State of the Art, it is difficult to map technical focus skills such as syncopation, dynamics to numeric measurements in audio features. Each song has customized assessment policy that depend on the Technical Control Questions in Table 10. For example, table 10 shows that Just Looking examines “Accented Syncopation”. The score shows that the emphasis is off the beat in the chorus, i.e. on the “and” beat after 2, when counting 1+ 2**+** 3+4.



*Figure 26: Syncopation example Just Looking*

To measure its effectiveness, the amplitude of the G and first C note would have to be greater than the last two C notes. One way would be to add a column in the “rhythm” CSV file that indicates Syncopation, and this could flag the onset detection algorithm to check for greater energy. The same musical property occurs in the chorus of “yellow”. The Teacher gives specific grades on specific technical demands of each song and the big challenge it to study their relationship with measured onsets and offsets.

* 1. Teacher Grading Results Analysis

After grading was complete on the Teacher Google Forms the CSV files were downloaded. Details in the Appendix are provided on the Google Form Histograms, the python scripts developed to parse the text output of the CSV files to map Radio Box Choices to the numeric values (100%, 88%, 80%, 63 %, 37%) that can be loaded into the Data Tables to be used for training the Model.

The student recordings were analyzed for their onsets and offset measurements and compared against the reference stems. The appendix section xyz gives details on how to reproduce the data with python scripts that is used for generating the accuracy results contained in the “StudentStatistics\_<song>.csv file (e.g. StudentStatistics\_yellow. Each student performance has an associated Deviations file for Onsets and Offsets, (yellow\_devs\_student1.csv… yellow\_devs\_student8) and raw data on the detected onsets and offsets (yellow1\_onsets.csv… yellow2\_onsets.csv )

Linear Regression Machine Learning algorithms were used to train the models with multiple variables. The Cross Validation strategy was that 30% of the total cases are used as Test Cases. Table 12 and 13 gives an overview of how the different songs responded to predicting the main grades: Onset and Duration. N represents the number of Student recordings used. The Algorithm columns indicates choice of algorithm (IEC or SOP or a combination). Onset and Duration. The core feature selection was the PRF accuracy measures. There are three Deviations Statistics values available as input features are: (i) Mean Deviation, (ii) Absolute Mean Deviation and (iii) Standard Deviation

*Table 12: Onset Grade Prediction Errors*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Song.** | **N** | **Algorithm** | **P** | **R** | **F** | **Mean Deviation** | **MAE[[21]](#footnote-21) Onset** | **RMS Error Onset** |
| yellow | 8 | SOP |  |  |  | Mean | 8.12 | 9.87 |
| bjean | 15 | IEC/ SOP |  |  |  | - | 8.62 | 12.25 |
| just | 11 | SOP |  |  |  | - | 7.84 | 11.7 |
| brown | 8 | IEC |  |  |  | - | 3 | 3.5 |
| road | 8 | IEC |  |  |  |  | 7.87 | 12.17 |
| wotm | 11 | SOP/IEC |  |  |  | - | 5.92 | 6.38 |

With the exception of the song “yellow”, the minimum Mean Absolute Error was achieved by only selecting P,R F as feature inputs.

*Table 13: Duration Grade Prediction Errors*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Song.** | **N** | **Algorithm.** | **P** | **R** | **F** | **ABS Mean Deviation** | **Mean Deviation** | **MAE**  **Duration** | **RMS Error Duration** |
| yellow | 8 | SOP |  |  |  | ABS Mean | -- | 11.97 | 12.2 |
| just | 11 | SOP |  |  |  | -- | Mean | 4.67 | 6.38 |
| road | 8 | IEC |  |  |  | -- | Mean | 7.95 | 9.5 |
| wotm | 11 | SOP/IEC |  |  |  | -- | Mean | 5.173 | 6.249 |

For duration grade prediction, the same PRF accuracy features were also selected for the above 4 songs. The Absolute Mean Deviation was also considered for “yellow” the Mean Deviation for the remining three songs.

The remainder of this chapter discusses the relationship between the predicted grades and the nuances and musical properties of the audio tracks. The final section discusses the Technical Control grade predictions followed by a review of the role played by the Teachers Comments in the grade prediction.

* 1. Yellow

While the PRF score of the reference stem was close to 100% the highest graded Student 5 only scored a 50% precision while 2nd highest graded Student 4 for a P.R.F score with a 59% precision. Student 4 produced lower deviation statistics than Student 5 but this was not reflected in the grades allocated by the Teacher.

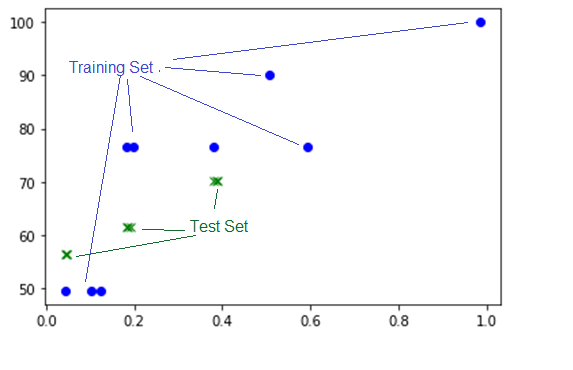
*Table 14:Best two “yellow” students*

|  |  |
| --- | --- |
|  |  |
| Onset Grade = 90.0 Duration Grade = 90.0 Onset ABS Mean: 0.010412 Onset Mean: 0.002990, Dev. from 0: 0.011992 Offset Mean: 0.014845, Dev. from 0: 0.055124 Articulation Grade = 90.0 Sound Control Grade = 90.0  Volume Control Grade = 90.0 Final Mark = 4.5 Precision = 0.508 | Onset Grade = 76.5 Duration Grade = 76.5 Onset ABS Mean: 0.009035 Onset Mean: -0.000789, Dev. from 0: 0.010740 Offset Mean: 0.000439, Dev. from 0: 0.069981 Articulation Grade = 76.5 Sound Control Grade = 76.5  Volume Control Grade = 49.5 Final Mark = 1.8 Precision = 0.594 |

**Sound Quality**

One explanation for the higher grade given by the teacher for Student 5 was because it was a different student with different instrument, with an overall superior sound quality. Student 1 was an outlier. Apart from a slight lack of consistency in hitting the A notes,

[Add git hub reference] when you listen to the Audio you can hear some clicks which cannot be located on the audio waveform. These noise sources were probably due to the settings or the environment of the Sound Card. For Student1, it was noticed that the IEC algorithm returned a 12 % rather than a 4% precision from SOP. It may be that the sound island approach is less sensitive to noise.



*Figure 27: Plot of Onset Grades vs Precision for Yellow*

Green plot: Test Set  
Blue Dots: Training Set

Yellow musically speaking is the most basic song and had the least number of outliers of all the songs the above plot shows.

* 1. Billie Jean

The song was divided into three sections: 1st verse (Muted), Bridge (Non-Muted), Chorus (Muted). Full details of the StudentStatistics are provide in the appendix. An analysis of the Overall Grades is shown here to highlight an error.

*Table 15:Billie Jean: Actual vs Predicted Grades Overall*

X=dataset['precision','recall','f\_measure\_value','Onset ABS Mean','Duration ABS Mean']

|  |  |  |
| --- | --- | --- |
| Test Case | Actual Grade | Predicted Grade |
| 0 | 3.600 | 2.458553 |
| 1 | 3.645 | 1.818748 |
| 2 | 2.700 | 1.369066 |
| 3 | 0.900 | 4.263822 |
| 4 | 3.150 | 3.510796 |
| Mean Absolute Error (M.A.E) | 1.6% | |
| Root Mean Squared Error (RMS Error) | 1.89% | |

Despite a low Mean Absolute Error, Test Case 3 shows a huge difference between predicted grade and actual grade, which raises some questions on grading policy.

Figure 36 shows 4 overgraded ( yellow) and one undergraded (red) outliers considering onset precision. The offsets have 4 over graded students for large duration deviations and 1 undergraded student for low deviation, using Absolute Mean as an indicator.

|  |  |
| --- | --- |
| *Precision (X)*  *Onset Grades (Y)* | *Duration ABS Mean (X)*  *Duration Grades (Y)* |
|  |  |

*Figure 28: Plot of Precision (X) vs Onset Grades (Y) Billie Jean*

**Bad Mixing**

Let us examine the 50% grades with precision value above 0.5. Listening back on the stem and looking at the waveform there are no major differences with the ground truth, but the problem was the synchronization with the bass drum.[add git hub reference] The teacher’s comment was

*Although the bass follows the song quite well, there is a tendency to play behind the "beat". Try to match the bass and kick drum.*

Therefore this outlier could be explained by the fact that the step was incorrectly synchronized with the Minus 1 track.

**Imposter Notes**

Student 6 (appendix 12.8) has a much higher onset grade than Student 8 despite having the same P value. The reason for this is that Student 6 did not follow the score in the A chord parts and played no rest notes. Nevertheless, the teacher overlooked this and gave a high grade. Subsequently this grade was further scaled down by 90% to compensate this score deviation.

You can see that the best (Student 14) and worst (Student 10) in figure 37 student performance show similar statistics, partly because of the “filter effect” [[22]](#footnote-22)of the minimum window size.

|  |  |
| --- | --- |
| High Grade: Student 14 157 deviations | Worst Grade: Student 10  62 deviations |
|  |  |

*Figure 29:Best vs Worst bjean students*

* 1. Just Looking

**Poor Duration results**

Even though there are muted notes in the chorus, the SOP algorithm is used for the entire song, given its superior accuracy on the reference stem. This grade 1 song has wider outliers than “yellow” (grade 0).

|  |  |
| --- | --- |
| Onsets | Duration |
|  |  |

*Figure 30: Just Looking: Grades (Y) Precision (X)*

The left plot shows the Onset grade curve against precision. The right plot shows the Duration grade against precision.

Student 4 had best accuracy with a precision of around 0.43, suggesting high sensitivity of algorithm or an overall distinct performance characteristic from ground truth. This student was also the highest graded student summing all the grades together.

* 1. Brown Eyed Girl

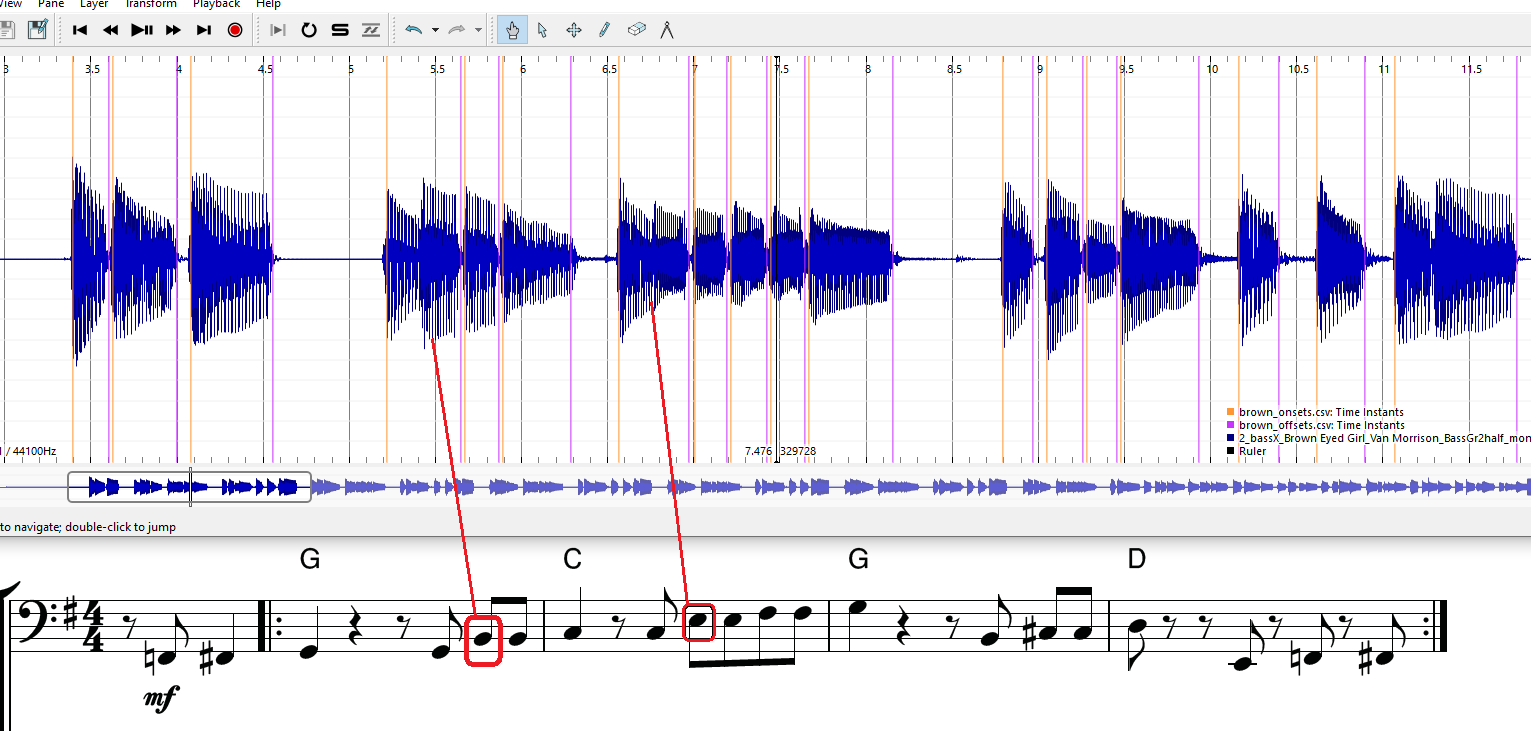
For “brown”, there was no way to improve accuracy by blending the algorithms. The Recall metric for the IEC methods was only 82 % and the SOP not much better at 85%. Although IEC had higher precision the closeness of the notes generated “Sound Archipelagos” instead of sound islands. These reference figures make Student Evaluations very unreliable.

**Short Difference between Notes**

This song tests the limitations of the MIR evaluation window of 20ms, when you consider the Onsets for the first 8 notes of the song, 5 of the gaps are less than 25ms.

*Table 16:First 8 onsets of “brown”*

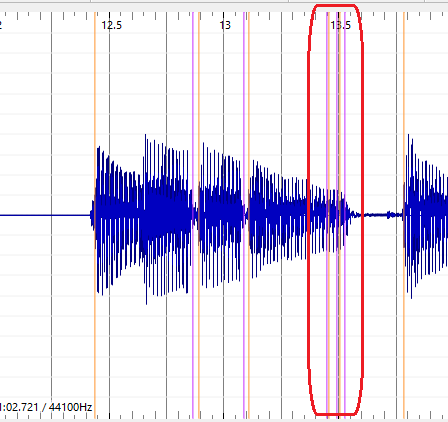
|  |  |
| --- | --- |
| **Onset Mark** | **Difference** |
| 3.385 |  |
| 3.62 | 0.235 |
| 3.96 | 0.34 |
| 4.08 | 0.12 |
| 5.215 | 1.135 |
| 5.42 | 0.205 |
| 5.66 | 0.24 |
| 5.88 | 0.22 |



*Figure 31: Brown: Missed Onset pattern*

**Sound Archipelagos**

In bar 2 (the first G chord bar) the four notes B, B, C# and D are all detected as 4 sound islands. In another occurrence of the exact same note sequence audio blow, not only is there a missing onset of the B note but there are two false alarm sound islands (13.4 seconds). This particular note pattern tends to throw false alarms, thus reducing precision below an acceptable amount.



*Figure 32: Brown: False Onsets*

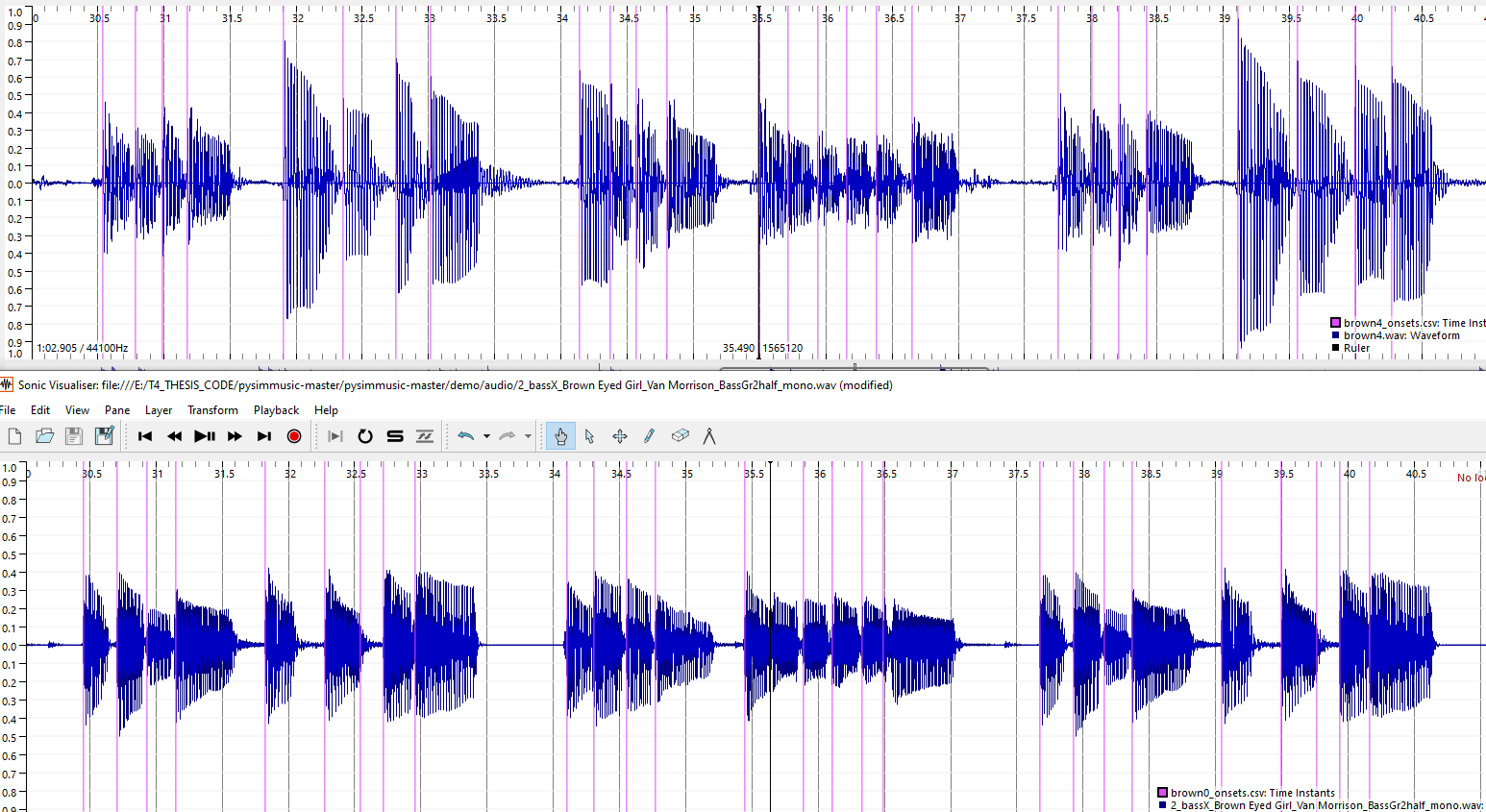
**Mismatch on Grade Prediction**

It is not surprising therefore that big outliers can happen: Student 4 who scored 90% grade for onsets, but a precision around 10%.

|  |  |
| --- | --- |
| *Onset Grades (Y)  Precision (X)* | *Duration Grades (Y)  Duration Mean Deviation (X)* |
|  |  |

*Figure 33: Brown: Grade Plots*

To try to understand this we look at the waveforms. At 36.5 secs Student 4 correctly hits 6 offsets but the GT only hits 5 onsets. There is almost a 10ms drift at 39.5 second timestamp.



*Figure 34: Brown: GT and Student 4 stems with onsets detected*

As a result of having a very low PRF, the IEC algorithm returned zero offset deviations for Student 4) because the evaluation window was too narrow to measure the difference.

* 1. Roadrunner

**Improved after “Tenuto” section removed.**

The tenuto notes for the song Roadrunner, were too weak to be measured and this ambiguity could explain the overall low PRF of the IEC and SOP methods. The cutoff point was made at 34.5 seconds to remove them all. This resulted in getting a better PRF score using the IEC algorithm. The Grade Prediction for Onset and Duration worked optimally when considering the Mean Onset and Mean Duration respectively alongside the PRF as feature inputs. As expected for a grade 3 song, the number of outliers was higher compared to lower grades.

|  |  |
| --- | --- |
| *Onset Grades (Y)  Precision (X)* | *Duration Grades (Y)  Duration Mean Deviation (X)* |
|  |  |

*Figure 35: Roadrunner: Onset & Duration Grades*

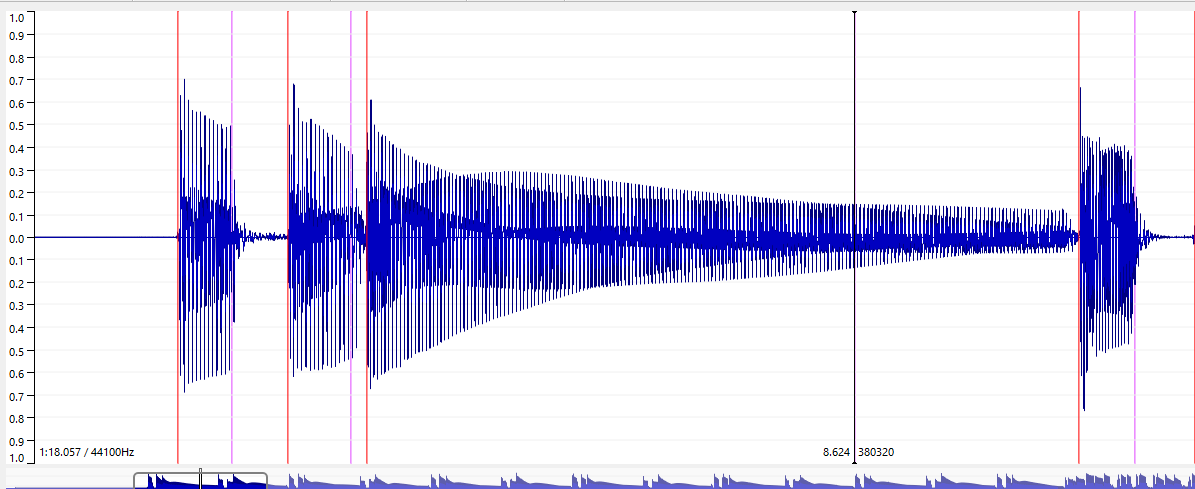
* 1. Walking on the Moon

The PRF using the IEC algorithm alone for WOTM was 0.701, 0.932, 0.8. After segmenting the verse into the SOP algorithm and the bridge into the IEC algorithm, the new PRF was as follows. 0.983, 0.991, 0.987.

It is important to point out a tradeoff made in using SOP for the first part. The very first note of WOTM has a subtle gap, which is ignored if you use the SOP, “offset=next onset” algorithm. Given that the Energy Checker didn’t return a high Recall, how would you determine if this subtle gap is respected in a student recording?

**Effective Duration**

The Essentia library function “Effective Duration” considers perceptual factors [25]when measuring note length. For the first three notes of “wotm”, the threshold setting of 0.05 returns offset points that approximately align with the human hearing threshold



*Figure 36: WOTM: Offsets using Effective Duration (T= 0.05)*

The orange line are the onsets, and the purple lines are the offset (the black cursor aligns with 3rd offset at 8.624 secs). Any future improvement in automating or semi-automating the assessment of duration should consider these adjusted offsets. This should improve the value that the Mean Duration Deviation has for predicting Duration Grades.

|  |  |
| --- | --- |
| *Onset Grades (Y)  Precision (X)* | *Duration Grades (Y)  Duration ABS Mean Deviation (X)* |
|  |  |

*Figure 37: Wotm: Onset & Duration Grades*

* 1. Technical Control Grades

The Actual vs Predicted Grade for Technical Control Element 1 for “bjean”, “brown” and “wotm” are shown here.

*Table 17: Tech Control Grades Actual vs Predicted*

|  |  |  |
| --- | --- | --- |
|  | Technical Focus 1 | |
| Song. | Actual | Predicted |
| bjean | 72.0  90.0  79.2  33.3  79.2 | 75.6  76.3  70.3  67.7  86.3 |
| brown | 79.2  79.2  72.0 | 76.8  77.2  76.9 |
| wotm | 72.0  72.0  79.2  56.7 | 80.0  82.6  74.9  78.8 |

*Table 18: Technical Grade Prediction Errors*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Song. | # S | Algo. | MAE Onset TF1 | RMS Error Onset TF1 | Extra Inputs  TF1 |
| bjean | 15 | IEC/SOP | 13.56 | 18.15 |  |
| brown | 8 | IEC | 3 | 3.3 |  |
| wotm | 11 | SOP/IEC | 11.26 | 13.07 | - |

“bjean” having the most recordings is discussed here under the two technical control. A Mean Absolute Error of 13.56 % in predicting the Articulation Grades (TF1) using the PRF and Mean Duration Deviation as inputs.

“Brown” yielded a low error prediction, but caution has to be exercised in drawing a conclusion, considering the onset and duration grade error were over 10%. However, it does highlight the case that maybe the teacher should grade the Technical Focus Areas 1 and 2 in one joint grade.

1. Conclusion
   1. Student Recordings and Teacher Gradings

It wasn’t possible to find music students to do the experiment. Five submissions were collected by researchers at the MTG, and I submitted the remaining 48 recordings. Some of the bassists who declined to participate had issues with music notation literacy and others were not comfortable in having to strictly follow the score and technical control details as laid out.

For future experiments the bass students should be sourced from undergraduate music programs at different third level music schools. This would avoid music literacy problems and any other issues related to following the musical techniques that are evaluated on each song, on the other hand it could exclude bassists who learn in informal contexts.

A comprehensive experiment would require compensating the participating students sufficiently to motivate them to complete the questionnaire, do the latency test correctly and optimize their technical setup to produce the best sound possible on the playback. A compensation scheme could be economic one (e.g., a fixed amount per song for the “informal context” students), academic (e.g., negotiating allocating credit as part of a subject in an undergraduate program or possible prize for best recording.). If an agreement could be established with a Music school, an agreement to allow music-undergrads and post grads to participate as students and teachers in a grading experiment and allocate credits for participation could be mutually beneficial. The benefits of collecting different performances and sounds from different bassists are clear from the experiences in this experiment, especially for the songs like “Roadrunner” (“road”) that were difficult to execute. The most interesting observations came from the recordings that were not my own. One student managed to obtain the best PRF score with a guitar and another student scored highest on “bjean”, despite having deviated from the score. Having had to do over 40 recordings (and more attempts), I noticed improvements each time I recorded the difficult songs “brown” and “road”, but then the “law of diminishing returns” began, and subsequent performances don’t improve.

Due to time constraints, the Bass Teacher got 43 out of 48 gradings done. One mistake that was pointed out by participating student, was that the Bass Teacher was not given the ground truth under the guise of a student attempt. This would have helped benchmark the real attempts. The subsequent “downgrading compensation” that I applied was to add more validity to the grades rather than push for lower prediction errors.

* 1. Technical and Musical issues with Experiment

Another restriction was that Music Critic is not supported on Tablets and Smart Phones, and one professional bassist reported that he had no sound card on his PC, indicating he uses his Tablet for recording. One workaround for this problem would be to set up a recording workshop on a fixed date, with a dedicated laptop running Audacity (or another DAWS), connected with an audio interface and headphone. The students could come in record without any latency test. This would be like an exam simulation and may suit some students who like to time-limit their attempts and are driven by “getting it right first time”. On the other hand, there might be students who prefer to prepare the recordings in the comfort of their own environment, taking as many attempts as they wish.

One mistake made in the experiment was not performing the truncation of the backing track and score at a very clear point in time or section of the music. The idea of truncation was to reduce the student and teacher effort. It was also necessary to truncate some songs further, e.g., “road”, to make the algorithms work properly for subsequent experiments. The score that gets uploaded should clearly show the stop point for student evaluation and the backing track should either fade out or stop accordingly.

The quality of the mix is also a factor in facilitating the Bass Teacher allocate a fair score. Perhaps a more thorough grading policy would be to allow the teacher listen to the recorded stem as well as the mix. Another weakness in the grading technique is linearly “going through the students one-by one”. In real exam correcting situations a teacher may prefer to have multiple iterations of grading student. Finally, the onset and offset grade should be mark between 1-8 and the Technical Control grades should be discarded, until such time that significant progress is made on accurately measuring duration.

* 1. Improving analysis and prediction strategies

The semi-automatic strategy of customizing the algorithms according to the songs was a compromise and not the original intention. It was proposed to use as input a Three-Column “rhythm” csv-format annotation files. The threes columns are: Onset Time, Muted, Offset Time”. The Onset and Offset time represent the annotations of the reference stem recording while the “Muted” property is set to Y to denoted “muted” or “N” to denote “not muted”. The idea was that the IEC algorithm would choose the “next onset” as the offset, but it did not yield significant improvements. An improvement could be made to switch between the calling of one algorithm or another depending on the CSV column contents. More columns could be added also to handle other musical properties.

Although I made additional submitted recordings and gradings, the overall quantity of data meant a limited choice of Machine Learning algorithms. The feature extraction papers by Abesser the data sets used were larger and that support vector machines were used for prediction. I found that in some cases, thatadding the deviation statistics information did not help in improving the predicted grade. For the Onset Grade Prediction, it was the Mean rather than Absolute Mean deviation which added more value. A wider set of tests, with Mean deviation calculated separately for the early and late onset would help to get a better understanding of this relationship. For offsets, absolute mean rather than mean deviation, helped reduce the predicted grade error metrics. Apart for Duration Grade prediction in “bjean”, adding the Standard Deviation statistic made the prediction a lot worse. Musically speaking , the song style would dictate the teacher sensitivity is to lateness/vs earliness and note longitude.

No further research was added to Abesser’s classification of plucking styles [14], but the three column “rhythm” csv file, would be a good continuation point for adding more functionality and parameters to the algorithms. For example, the ground truth input files would not only contain annotated onsets and offsets, but also the plucking style on each note which a detector could compare against. Adding a Tenuto marking in the Rhythm Files, could be an extension to the existing methodology to customize algorithms for certain technical focus elements.

For the sound quality grade, no grade prediction was attempted as the audio features that the algorithms extract doesn’t relate directly to sound quality . This requires other audio features to be extracted, such as background noise level, stability of the dynamics which are not the subject of this thesis. Apart from collecting valuable data for future .  
research, having the recordings labelled with Sound Quality is useful in trying to understand how robust the onset detection is in the presence of noise.

* 1. Improving Duration measurements

The main challenge was to do a score accurate check on long notes, typically composed of tied quarter, eight notes. If a particular song yielded better results for the full extract using SOP, then this meant that there was effectively no strict validity check on whether a long note was properly held. The scope of the project was limited to blending two algorithms., IEC and SOP and the tolerance window for Offset detection was double that of Onset detection. A more rigorous measurement approach would incorporate the Effective Duration as discussed in section 6.10 into the algorithms and narrow the matching window size.

The repeated quavers in the verse for “yellow” leaves no gaps for duration so it is meaningless here. This applies to many of the TCL songs, apart from the ones mentioned here (e.g. AC/DCs grade 1 song). Theoretically it would be possible to play shorter notes than quavers, but that is actually more difficult to do, so it is not worth checking for a minimum note length.

There are some areas that were too complex to analyze, e.g. The Tenuto musical property in “road”. A starting point for looking further into it would be to examine the two extreme cases: Max Case: All notes equal, Min Case the third note is absent.  
Between these two extremes: an ideal energy statistic should be chosen for tenuto, considering the limit of human perception.

* 1. Going forward

To end I list suggested future paths for scaling up future experiments to continue the research:

* Further customization of algorithm to include the PitchMelodia (SG [7]) instead of the SOP algorithm in “wotm” for the verse.
* A pilot project that would involve a selection of up to 20 students who are studying bass guitar in private schools and conservatories and partitioning into different “tendencies”: late/early onsets and short/long duration playing styles.
* Continually annotating more Ground Truth songs.
* Source Separation to obtain more annotated data.
* Consideration of Rockschool[[23]](#footnote-23).Music Syllabus

Finally, the results have shown us that the precision, recall and f-measure metrics are good indicators of good performance, but they need more robustness to close the gap between the Student Performance and the ground truth. It is only possible to compare the Student performance deviation statistic data with another Student performance of similar PRF measure. If the overall mean value of the onset deviation is negative that indicates early tendency and likewise if the mean of duration deviation is negative this indicates tendency to play shorter notes. Earliness and lateness (and duration deviation) has to be looked at in the context of the musical properties of a song. Workarounds to the limits of PRF accuracy were found by applying different algorithms to different song sections. There is more scope for further segmentation using threshold parameters for the IEC algorithm and setting for the SOP. Data driven methods were not investigated in the student assessments.

Data-driven alternatives to ASP methods were not tested in depth in this thesis. The advantage of non-real-time performance assessments is that they are not constrained by the computational cost requirements that training sets and statistical methods demand. The evaluation results given by the pre-trained models in the U-Net architecture were low [16] but perhaps new models based on the training the TCL dataset would yield better results. There are also other ASP methods that were mentioned in the State of the Art that may merit further consideration, e.g., the Wavelet [9] method by Bello. difference to using training sets and statistical methods.

1. List of figures

(Optional)

[*Figure 1: Onset Detection Stages* 2](#_Toc80906098)

[*Figure 2: dynamic threshold setting* 2](#_Toc80906099)

[*Figure 3: 4 crotchets, 4 quavers and 4 semi-quavers.* 2](#_Toc80906100)

[*Figure 4: Extract from IDMT Dataset. File 002.xml* 2](#_Toc80906101)

[*Figure 5: Note duration of Staccato notes* 2](#_Toc80906102)

[*Figure 6: Audio 002.wav (IDMT DATASET) with onsets* 2](#_Toc80906103)

[*Figure 7: Billie Jean: Midi Rendering vs Human recording (shorter notes)* 2](#_Toc80906104)

[*Figure 8: Strategy to experiment, test to find best accuracy* 2](#_Toc80906105)

[*Figure 9: End to End test of Music Critic with Billie Jean GT* 2](#_Toc80906106)

[*Figure 10: Original Energy Checker for offset detection* 2](#_Toc80906107)

[*Figure 11: Normalized Energy of Audio 002.wav sample (IDMT DATASET)* 2](#_Toc80906108)

[*Figure 12: Sound Islands of Audio 002.wav (IDMT). FS = 1024, HS = 512.* 2](#_Toc80906109)

[*Figure 13: IEC onsets for “wotm” verse* 2](#_Toc80906110)

[*Figure 14: Section of “wotm”verse showing missed onsets.* 2](#_Toc80906111)

[*Figure 15: Calculated IEC Sound Islands for WOTM bridge, P = 0.937* 2](#_Toc80906112)

[*Figure 16: PitchMelodia derived Onsets for WOTM verse , P = 0.97* 2](#_Toc80906113)

[*Figure 17: PitchMelodia derived Onsets for WOTM bridge* 2](#_Toc80906114)

[*Figure 18: Sample of bjean\_rhythm.csv verse (left) and bridge(right)* 2](#_Toc80906115)

[*Figure 19: IDMT Dataset GT Onset + Duration deviations (IEC) algorithm* 2](#_Toc80906116)

[*Figure 20: IDMT Dataset GT Onset deviations ( SOP) algorithm* 2](#_Toc80906117)

[*Figure 21: Billie Jean GT onsets deviation* 2](#_Toc80906118)

[*Figure 22: Billie Jean GT offsets deviation (? Algorithm)* 2](#_Toc80906119)

[*Figure 23: Student Portal (copyrighted score removed)* 2](#_Toc80906120)

[*Figure 24: Transition states on student portal* 2](#_Toc80906121)

[*Figure 25: Contrasting comments of “wotm” performances* 2](#_Toc80906122)

[*Figure 26: Syncopation example Just Looking* 2](#_Toc80906123)

[*Figure 27: Plot of Onset Grades vs Precision for Yellow* 2](#_Toc80906124)

[*Figure 28: Plot of Precision (X) vs Onset Grades (Y) Billie Jean* 2](#_Toc80906125)

[*Figure 29:Best vs Worst bjean students* 2](#_Toc80906126)

[*Figure 30: Just Looking: Grades (Y) Precision (X)* 2](#_Toc80906127)

[*Figure 31: Brown: Missed Onset pattern* 2](#_Toc80906128)

[*Figure 32: Brown: False Onsets* 2](#_Toc80906129)

[*Figure 33: Brown: Grade Plots* 2](#_Toc80906130)

[*Figure 34: Brown: GT and Student 4 stems with onsets detected* 2](#_Toc80906131)

[*Figure 35: Roadrunner: Onset & Duration Grades* 2](#_Toc80906132)

[*Figure 36: WOTM: Offsets using Effective Duration (T= 0.05)* 2](#_Toc80906133)

[*Figure 37: Wotm: Onset & Duration Grades* 2](#_Toc80906134)

[*Figure 38: Technical instruction and link to recording portal* 2](#_Toc80906135)

[*Figure 39: Extract of guide given to teacher for grading* 2](#_Toc80906136)

[*Figure 40: Grading Histogram Onsets: Billie Jean* 2](#_Toc80906137)

[*Figure 41: Grading Histogram Offsets: Billie Jean* 2](#_Toc80906138)

1. List of tables

(Optional)

[*Table 1: Abesser Signal Processing vs Salamon/Gomez* 2](#_Toc80906139)

[*Table 2: MELODA Algorithm steps* 2](#_Toc80906140)

[*Table 3: Abessers Data Driven algorithms vs Salamon/Gomez* 2](#_Toc80906141)

[*Table 4: S Algorithms for Onset/Offset detection* 2](#_Toc80906142)

[*Table 5: TCL dataset* 2](#_Toc80906143)

[*Table 6: Tech. Control Parameters for 6 TCL songs* 2](#_Toc80906144)

[*Table 7: Benchmarking Onset detection algorithms using IDMT DS* 2](#_Toc80906145)

[*Table 8: Accuracy Metrics for 6 TCL ground truths (combined vs non combined)* 2](#_Toc80906146)

[*Table 9: Accuracy Metrics for Billie Jean Hybrid Algorithm* 2](#_Toc80906147)

[*Table 10: Selected comparison: IEC vs SG* 2](#_Toc80906148)

[*Table 11: Technical Control Question Topics* 2](#_Toc80906149)

[*Table 12: Onset Grade Prediction Errors* 2](#_Toc80906150)

[*Table 13: Duration Grade Prediction Errors* 2](#_Toc80906151)

[*Table 14:Best two “yellow” students* 2](#_Toc80906152)

[*Table 15:Billie Jean: Actual vs Predicted Grades Overall* 2](#_Toc80906153)

[*Table 16:First 8 onsets of “brown”* 2](#_Toc80906154)

[*Table 17: Tech Control Grades Actual vs Predicted* 2](#_Toc80906155)

[*Table 18: Technical Grade Prediction Errors* 2](#_Toc80906156)

[*Table 19: Billie Jean Grades Actual vs Predicted* 2](#_Toc80906157)

[*Table 20:* PRF, Deviations (inputs) vs Billie Jean Onset Grade (outputs) 2](#_Toc80906158)

[*Table 21:* PRF, Deviations (inputs) vs Billie Jean Duration Grade (outputs) 2](#_Toc80906159)

1. List of symbols

[Equation 1 Local Energy Equation 2](#_Toc80906160)

[Equation 2 Energy in a Frequency Band 2](#_Toc80906161)

[Equation 3 Onset Detection Function Equation 2](#_Toc80906162)

[Equation 4 Peak Picking Formula 2](#_Toc80906163)

[Equation 5 Normalized Energy Function 2](#_Toc80906164)

1. Bibliography

|  |  |
| --- | --- |
| [1] | N. McCormick, U2 by U2: Bono, The Edge, Adam Clayton, Larry Mullen Jr, New York: HarperCollins, (2009). |
| [2] | C. Dittmar, E. Cano, J. Abeßer and S. Grollmisch, "Music Information Retrieval Meets Music Education," in *Multimodal Music Processing*, Dagstuhl, Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik, p95-120, (2012).. |
| [3] | P. Pfeiffer, Bass Guitar For Dummies, 3rd edition, Hoboken, NJ.: John Wiley and Sons, (2014). |
| [4] | B. H., S. Giraldo and R. Ramirez, "Jazz ensemble expressive performance," in *17th International Society for Music Information*, (2016). |
| [5] | 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). |
| [6] | J. Bello, C. Duxbury, M. Davies and M. S. M. I. Sandler, "On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain," in *IEEE Signal Processing Letters*, (2004). |
| [7] | E. Gómez and J. Salamon, "Melody extraction from polyphonic music signals using pitch contour characteristics," in *IEEE Transactions on Audio, Speech, and Language Processing,*, (2012). |
| [8] | J. P. Bello and M. B. Sandler, "Phase-based note onset detection for music signals," in *ICASSP-88*, 2003. |
| [9] | J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies and M. B. Sandler, "IEEE A Tutorial on Onset Detection in Music Signals," in *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING,*, 2005. |
| [10] | E. Terhardt, "Psychoacoustic evaluation of musical sounds," *Perception and Psychoacoustics,* vol. 23 (6), 483-492, (1978). |
| [11] | F. E. MUSIEK1 and G. D. CHERMAK2, "Psychophysical and behavioral peripheral and central auditory tests," in *Handbook of Clinical Neurology, Chapter 18*, vol. 129 (3rd series), Elsevier, (2015). |
| [12] | C. Kopp-Scheinpflug, J. L. Sinclair and Jennifer F. Linden., "When Sound Stops: Offset Responses in the Auditory System," *Trends in Neurosciences,* no. Special Issue: Time in the Brain, (2018). |
| [13] | Mirex, "music-ir.org," MIREX, [Online]. Available: https://www.music-ir.org/mirex/wiki/2020:Singing\_Transcription\_from\_Polyphonic\_Music. |
| [14] | J. Abeßer and G. Schuller, "Instrument-centered music transcription of solo bass guitar recordings," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing 25 (9), 1741*, (2017). |
| [15] | Fraunhofer. [Online]. Available: https://www.idmt.fraunhofer.de/en/business\_units/m2d/smt/bass.html. |
| [16] | J. Abeßer and M. Müller, "Jazz Bass Transcription Using a U-Net Architecture," *Electronics,* vol. 10, 12 March (2021). |
| [17] | M. Goto, "Predominant-F0 estimation for detecting melody and bass lines in," in *Speech Communication*, (2004). |
| [18] | J. Salamon, R. Bittner, J. Bonada, J. Bosch, E. Gómez and J. Bello, "An Analysis/Synthesis Framework for Automatic F0 Annotation of Multitrack Datasets," in *International Society for Music Information Retrieval Conference*, Suzhou, China, (2017). |
| [19] | J. Abeßer, K. Frieler, M. Pfleiderer and W.-G. Zaddach, "Introducing the Jazzomat project - Jazz solo analysis using Music Inf. Retrieval methods," in *Int. Symposium on Computer Music Multidisciplinary Research (CMMR) Sound, Music and Motion*, Marseille., (2013). |
| [20] | J. Abeßer, H. Lukashevich and G. Schuller, "Feature-based extraction of plucking and expression styles of the electric bass guitar.," in *IEEE International Conference on Acoustics,Speech and Signal Processing*, (2010). |
| [21] | L. Reboursiere, O. Lahdeoja, T. Drugman, S. Dupont, C. Picard and N. Riche, "Left and right-hand guitar playing techniques detection," in *New Interfaces for Musical Expression*, (2012). |
| [22] | R. Hennequin, A. Khlif, F. Voituret and M. Moussallam, "Spleeter: a fast and efficient music source separation tool with pre-trained models," *The Journal of Open Source Software,* (2020). |
| [23] | Trinity College London, "R&P Bass Grade 1," in *R&P Bass Grade 1*, Londoin, (2017), p. 9. |
| [24] | 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). |
| [25] | G. Peeters, ""A large set of audio features for sound description (similarity and classification) in the CUIDADO project," CUIDADO I.S.T., (2004). |
|  |  |
|  |  |

1. Appendices
   1. Methodology Histograms

BassUnet Accuracies (Billie Jean vs Abbesser

* 1. Student Portal

Recommendations:

* good headphones, capable of isolating the backing track from the microphone
* A direct audio interface for the Bass is preferred over a headset microphone

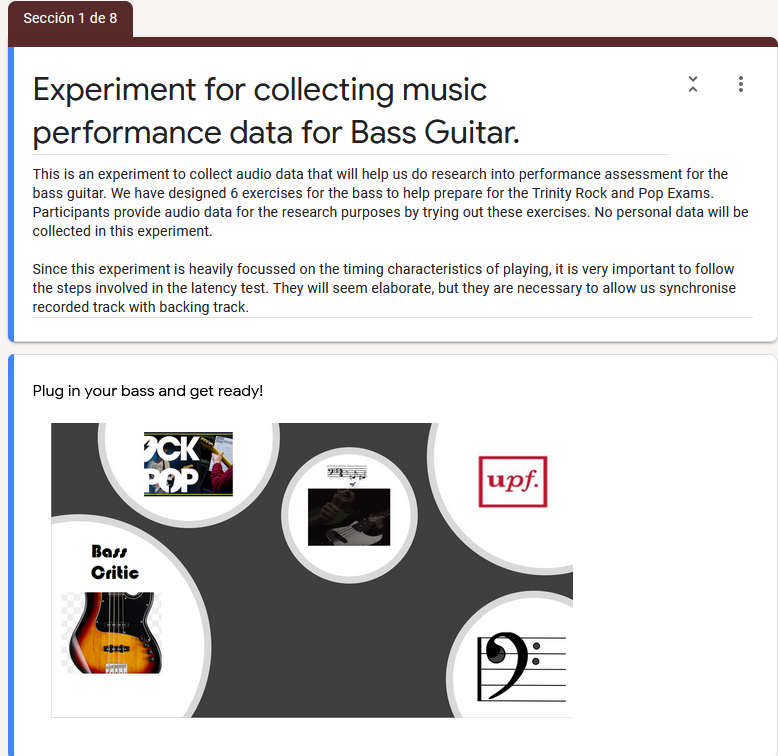
In the portal the student is requested to provide the following:

* a description of hardware (external or internal microphone, Audio Interface), soundcard and driver (e.g. Realtek Audio),
* Browser (e.g. Chrome) and Operating System (Windows/ Mac / Linux).

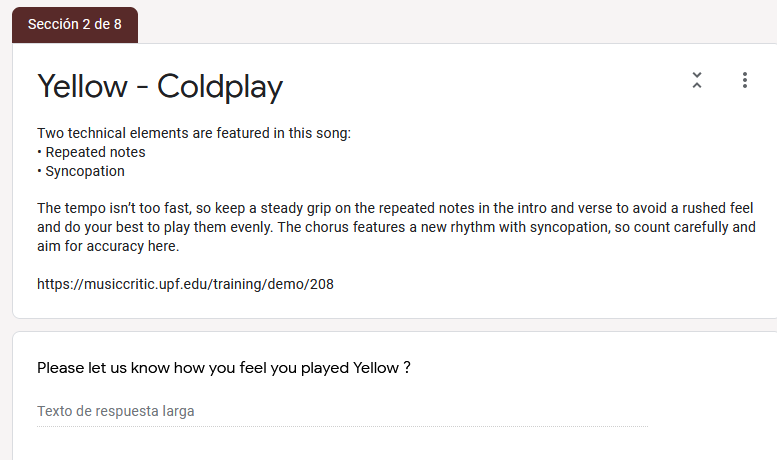
Here is the link to the Bass Critic Landing Page (restricted access)

<https://docs.google.com/forms/d/16AAuiEyNhDUJNg6Rs3cQORj8LozZAWuBETrHv_vcP-M/edit>

Here is how it appears:



* Sections on links to Minus-Tracks for recording



*Figure 38: Technical instruction and link to recording portal*

* 1. Recommendations and Instructions for recording.

===========================

1. It is recommended to close all other programs on your computer to minimise load.

2. The recommended browser is Chrome and Apple Mac is preferred but not mandatory.

3. Use a Laptop or Computer. Tablets/smart phones are not supported.

4. microphone should go through the audio interface for the latency test

The latency calibration test for your recording setup is Mandatory. You do not need your bass for this part.

Please follow these steps:

===========================

1. Place your headphones near the microphone.

2. Click on link https://musiccritic.upf.edu/training/demo/182

3. Follow the instructions on the above link carefully.

4. The click track you will hear is approximately 25 seconds in duration.

5. Listen to your recording by clicking "Play" control. You should hear the click sound clearly.

6. Submit, if the recording is ok. If you don't hear the click, try again ("Clear" and then "Record"). If you can't achieve good quality, try with another setup (browser or computer).

7. A few seconds later you will see the message "Submission Feedback: Overall performance is accurate."

8. Record and submit another click track recording. Below is an illustration of workflow.

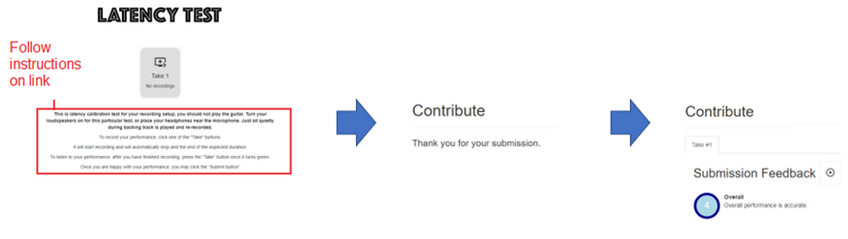
On the latency test page, the following instructions appear:

**This is latency calibration test for your recording setup, you should not play the guitar. Turn your loudspeakers on for this particular test, or place your headphones near the microphone. Just sit quietly during backing track is played and re-recorded.**

To record your performance, click one of the "Take" buttons.   
It will start recording and will automatically stop and the end of the expected duration.  
To listen to your performance, after you have finished recording, press the “Take” button once

Once you are happy with your performance, you may click the “Submit button”

The Latency Test workflow steps can be summarised as follows



The submission of the click track is treated the same way as the submission of any recording.

* 1. Additional Instructions for recording.

IMPORTANT: IF YOU DO THE RECORDINGS IN DIFFERENT SESSIONS, YOU WILL NEED TO REPEAT THE LATENCY TESTS FOR EACH SESSION.

For each song you can attempt as many recordings as you wish.

The backing tracks used are cover versions of original. It is a good idea to listen to the original to get a better feeling of the groove. Links to originals are not provided here, but you can search for them on the internet.

The total playing time for all the backing tracks is approximately 12 mins 30 secs. The displayed sheet music does not cover the full song length, so you are only required to play the bass up to the last bar of the score.

There are six songs in total, it would be nice to get recordings from all songs, but you can perform the ones you wish.

On each take or attempt you can hear the playback of your bass part and you can decide to submit (save on the server) or not.

If the volume of the playback is low, try increasing the gain on the microphone through the control panel.

Try to avoid boosting it too much or else it will distort.

There is no limit to the number of takes or submissions you can do.

Remember we are trying to collect all kinds of performances, so dont be shy of submitting something with a few blemishes.

Please note, for some songs we do not have the full score available. When you get to the end of the score you can stop playing (or you wish just keep playing along. Only the displayed score will get assessed.)

To start recording a particular song, click the song link in each of the provided sections.

When you have completed the first link (Yellow-Cold Play), place your comments and proceed to next song and so on.

The songs are ordered below in terms of complexity (easiest songs first).

Each song has a backing track with bass track removed, sheet music and bass tabs.

The Teachers focus will be focussed mainly on timing, rhythm and dynamics. Read technical focus advise given for each song (in the text at the bottom of score). The expectation is to follow the score (no creative embellishments).

This is an experiment, so no results will be posted on the performances. The audio data and text answer to this form are used for research.

* 1. Steps for Processing Recording

|  |
| --- |
| Steps:   1. Download the student recording from the server. It has a name like this: "submissions/1805\_52a4886b326c4301b2760c8df6404c96.wav" 2. Rename it to the Student name. 3. Check that the playback rate is 44100Hs If its 48000Hz, left click on the audio file in Audacity and choose the rate 44100Hz, make sure it is also this rate in the Project settings, then go to Tracks menu and choose “Resample”. 4. Import Isolated Student Stem to Audacity 5. Import Ground Truth and make a split track to Mono 6. Zoom in on initial onsets and align them manually. Align the first onsets. 7. Boost the Bass 6db and the Treble 1db and after words add another 1edb to align amplitude with stem. 8. Add other tracks to audacity playback and check synchronisation and volume mix. 9. Boost the bass volume so you can hear it clearly. You may need to attenuate the other tracks, particularly vocals. 10. When you are happy with the mix so that you can grade the bass as a teacher export as WAV file. 11. Remove the other tracks and export the student stem also as a WAV file. It is recommended to have a good naming system to distinguish Bass stems from mixes.   Use the “\_m” suffix to signify a mix (wotm1\_m.wav)and leave bass stem with student name (e.g. wotm1.wav)   1. Don’t wait too long before you having the mixes graded |

Any future development that would require the collection of recordings on a large scale would require automating some or preferably all these steps.

* 1. Teacher Portals

|  |
| --- |
| Fluency and Security  Trinity classifies according to the following scale for Fluency, Synchronisation & Security  -----------------------------------------------------  Excellent sense of fluency and synchronisation (100%) Very good of fluency, synchronisation with only momentary lapses. (88%) Good sense of fluency and synchronisation though with occasional lapses. (80%) Generally reliable level of fluency and synchronisation though with some lapses.(63%) Unreliable fluency, synchronisation (37%)  We want to focus on two aspects of fluency and that is on the 2 key timing aspects:  (i) Note Onset (hitting at the right time) (ii) Note Duration (holding it for correct length)  Please Note, the Song is organised as follows. (Ignore the NO BASS INTRO)  Part 1: Verse. Bars 5-16 Part 2: Chorus. Bars 17-24 Part 3: Bridge. Bars 25-32  Please refer to particular sections or bars of the song when making comments. |

*Figure 39: Extract of guide given to teacher for grading*

Link to Teacher Grading Google Form for Yellow

https://docs.google.com/forms/d/1B8AyKgHRtxJdv4LXE23plPZTIpF5IWRU\_RvHKKBApiY/edit#responses

* 1. Teacher Questions for Just Looking

Teacher Portal Design:

There are three sections in the Teacher  
Section 1: Title of songs summary of grading scheme, name of student track

Section 2 Fluency and Security section customise for onset and offset measurement.

Section 3 Technical Focus section with customized question for the song (e.g. Accented Syncopation for Just Looking) and questions on Dynamics and Sound Quality.  
Final sections: Contains following general question:  
  
Please add additional comments on stylistic understanding (e.g. mood and character), musical detail (e.g. dynamics and articulation), audience engagement. Finally write a number between 5 (highly convincing) and 1 (unreliable) classify the overall impact of the song.

Q1. Note Onset Security. Did the student hit the note at exactly the right time, not too early, too late? (consider syncopation and stylistic elements)

Q2. Duration. Holding note for the required length, consider tied note,etc.

Technical focus

Technical Control: Classify each aspect between 5 and 1 as follows:

-----------------------------------------------------

5 Fulfilled to a very high degree

4 Fullfilled

3 Largely fulfilled (occasional lapse)

2 Generally fulfilled

1 Often not fulfilled

Q1. Accented syncopation In chorus bass plays an accented, syncopated motif. (5-1)

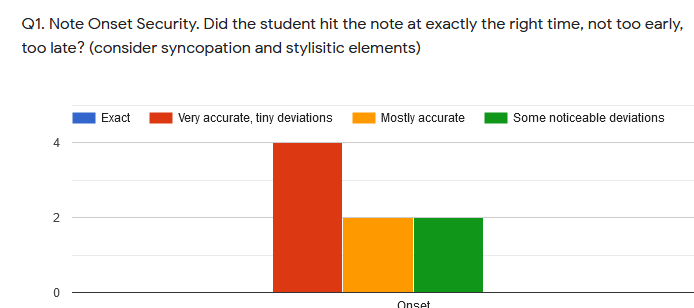
Q2. Dynamics, subito, contrast (5-1)

Q3. Sound Quality (5-1)

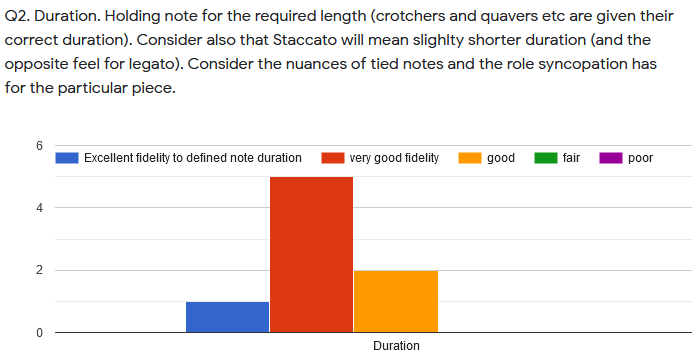
Please add additional comments on the above Technical Control aspects of the song, to justify your choice.

Please add additional comments on stylistic understanding (eg mood and character), musical detail (e.g. dynamics and articulation), audience engagement. Finally write a number between 5 (highly convincing) and 1 (unreliable) classify the overall impact of the song.

The original “bjean” Google Form has Q2 covering “Sound Quality” and Q3 covering Dynamics. In order for the parsing programs to work these columns had to be switched to read Q2 -Dynamics and Q3 Sound Quality as shown in table 10



*Figure 40: Grading Histogram Onsets: Billie Jean*



*Figure 41: Grading Histogram Offsets: Billie Jean*

* 1. Note books for Plots and Code for generating data

<https://github.com/cvf-bcn-gituser/bass-critic>

* 1. Billie Jean Student Performances

Table 11 also shows that two actual grades of 49.5 produce two different grades of 42.868 and 63.165 respectively. The failed grades of 36 are from Students 9 and 10 have very low PRF scores as shown in the table. However, if these two rows are removed and the Linear Prediction of the Duration Grade is applied on the reduced set , the MAE jumps to 63%, but removing the Duration Standard Deviation would then bring it back down to 14%. Billie Jean is the song with most Student recordings; however they are more divergent in quality than other songs.

*Table 19: Billie Jean Grades Actual vs Predicted*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **ONSET** | | **DURATION** | |
| **Song**. | **Actual** | **Predicted** | **Actual** | **Predicted** |
| yellow | 76.5  76.5  49.5 | 60.92  69.76 47.44 | 90 63 76.5 | 79.76 78.3 66.13 |
| bjean | 76.50  68.85  49.50  49.50  76.50 | 68.76  55.13  42.87  63.16  78.93 | 76.5  81.0  63.0  36.0  76.5 | 64.39  72.57  67.02  64.31  74.54 |
| just | 79.19  72.00  90.00  56.70 | 72.86  73.87  90.70  79.14 | 90.00  79.2  90.00  79.2 | 78.19  78.90  92.32  83.46 |
| Brown | 79.2  79.0  72.0 | 77.53  81.1  77.53 | 72  90  90 | 79.56  86.55  79.56 |
| Road | 79.2  79.2  72 | 76.87 79.54  92.96 | 79.2  90  79.2 | 94.5  86  74.6 |
| Wotm | 72  79.2  72  72 | 77  69.7  69.07  65.68 | 72.0  79.2  79.2  72 | 82.99  74.50  81.06  75.15 |

The Mean Absolute Error increased to 11% when including 'Onset Mean', 21%, when including ‘Absolute Mean’ and 11% and when including Onset Standard Deviation. Low F-measure results in a lower deviation count because “bad onsets” are filtered out. If you have two student recordings with very similar Precision, Recall , F Measure Values, then you can do a fair comparison of the Statistics

*Table 20:* PRF, Deviations (inputs) vs Billie Jean Onset Grade (outputs)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stud. | P | R | F | Abs. Mean | Mean | Std. Dev | ONSET GRADE | OVERALL GRADE |
| **0** | **1** | **0.99** | **0.995** | **0** | **0** | **0** | **100** | **5** |
| 1 | 0.328 | 0.315 | 0.321 | 0.008 | 0.002 | 0.009 | 76.5 | 3.6 |
| 2 | 0.519 | 0.531 | 0.525 | 0.006 | 0 | 0.009 | 49.5 | 2.7 |
| 3 | 0.189 | 0.185 | 0.187 | 0.008 | -0.004 | 0.009 | 63 | 3.6 |
| 4 | 0.102 | 0.098 | 0.1 | 0.009 | 0.002 | 0.01 | 63 | 3.6 |
| 5 | 0.206 | 0.21 | 0.208 | 0.007 | -0.001 | 0.009 | 76.5 | 2.7 |
| 6 | 0.201 | 0.206 | 0.203 | 0.009 | 0.006 | 0.01 | 68.85 | 3.645 |
| 7 | 0.107 | 0.108 | 0.108 | 0.009 | -0.004 | 0.011 | 76.5 | 1.8 |
| 8 | 0.239 | 0.259 | 0.248 | 0.008 | 0 | 0.01 | 49.5 | 2.7 |
| 9 | 0.088 | 0.091 | 0.09 | 0.005 | -0.001 | 0.008 | 49.5 | 0.9 |
| 10 | 0.18 | 0.154 | 0.166 | 0.009 | 0 | 0.01 | 49.5 | 0 |
| 11 | 0.568 | 0.455 | 0.505 | 0.007 | -0.001 | 0.009 | 76.5 | 3.6 |
| 12 | 0.525 | 0.469 | 0.495 | 0.007 | -0.002 | 0.009 | 63 | 2.7 |
| 13 | 0.437 | 0.374 | 0.403 | 0.007 | -0.001 | 0.009 | 76.5 | 3.15 |
| 14 | 0.512 | 0.448 | 0.478 | 0.008 | -0.003 | 0.01 | 90 | 4.5 |
| 15 | 0.468 | 0.43 | 0.448 | 0.008 | -0.003 | 0.01 | 90 | 1.98 |

*Table 21:* PRF, Deviations (inputs) vs Billie Jean Duration Grade (outputs)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stud. | P | R | F | A. Mean | Mean | Std. D | Acc. | DUR | OVERALL |
| **0** | **1** | **0.99** | **0.995** | **0** | **0** | **0** | **1** | **100** | **5** |
| 1 | 0.328 | 0.315 | 0.321 | 0.046 | -0.017 | 0.14 | 0.48 | 76.5 | 3.6 |
| 2 | 0.519 | 0.531 | 0.525 | 0.023 | -0.009 | 0.036 | 0.57 | 63 | 2.7 |
| 3 | 0.189 | 0.185 | 0.187 | 0.046 | -0.031 | 0.124 | 0.48 | 76.5 | 3.6 |
| 4 | 0.102 | 0.098 | 0.1 | 0.013 | 0.001 | 0.018 | 0.33 | 76.5 | 3.6 |
| 5 | 0.206 | 0.21 | 0.208 | 0.04 | -0.033 | 0.133 | 0.39 | 76.5 | 2.7 |
| 6 | 0.201 | 0.206 | 0.203 | 0.023 | 0.005 | 0.026 | 0.68 | 81 | 3.645 |
| 7 | 0.107 | 0.108 | 0.108 | 0.326 | -0.326 | 0.411 | 0.7 | 76.5 | 1.8 |
| 8 | 0.239 | 0.259 | 0.248 | 0.149 | -0.115 | 0.326 | 0.65 | 63 | 2.7 |
| 9 | 0.088 | 0.091 | 0.09 | 0.014 | -0.001 | 0.018 | 0.94 | 36 | 0.9 |
| 10 | 0.18 | 0.154 | 0.166 | 0.042 | -0.02 | 0.103 | 0.52 | 36 | 0 |
| 11 | 0.568 | 0.455 | 0.505 | 0.045 | 0.03 | 0.181 | 0.4 | 76.5 | 3.6 |
| 12 | 0.525 | 0.469 | 0.495 | 0.026 | -0.005 | 0.078 | 0.57 | 63 | 2.7 |
| 13 | 0.437 | 0.374 | 0.403 | 0.019 | 0.011 | 0.022 | 0.67 | 76.5 | 3.15 |
| 14 | 0.512 | 0.448 | 0.478 | 0.017 | -0.002 | 0.021 | 0.22 | 90 | 4.5 |
| 15 | 0.468 | 0.43 | 0.448 | 0.029 | -0.017 | 0.087 | 0.39 | 49.5 | 1.98 |

1. Examples include <https://yousician.com/>, https://fretello.com/ [↑](#footnote-ref-1)
2. https://www.idmt.fraunhofer.de/en/business\_units/m2d/smt/bass.html [↑](#footnote-ref-2)
3. https://musiccritic.upf.edu/ [↑](#footnote-ref-3)
4. https://www.upf.edu/web/mtg [↑](#footnote-ref-4)
5. https://github.com/CPJKU/madmom/blob/7e560700987d2800f1c78fdb2a5faba77ecece0d/madmom/features/onsets.py#L585 [↑](#footnote-ref-5)
6. https://github.com/RamoonRoomeu/ToneExperiments [↑](#footnote-ref-6)
7. https://essentia.upf.edu/reference/std\_OnsetDetection.html [↑](#footnote-ref-7)
8. https://github.com/MTG/essentia/blob/master/src/examples/tutorial/example\_onsetdetection.py [↑](#footnote-ref-8)
9. Abessers overall accuracy is higher at 0.735, but that metric considers pitch which is not of concern here. [↑](#footnote-ref-9)
10. https://github.com/MTG/sms-tools/blob/master/software/models/harmonicModel.py [↑](#footnote-ref-10)
11. https://github.com/jakobabesser/pymus/tree/master/pymus/sisa/f0\_tracking [↑](#footnote-ref-11)
12. https://essentia.upf.edu/reference/std\_PitchMelodia.html [↑](#footnote-ref-12)
13. https://github.com/MTG/essentia/blob/master/src/examples/tutorial/example\_predominantmelody\_by\_steps.py [↑](#footnote-ref-13)
14. <https://github.com/jakobabesser/bassunet> [↑](#footnote-ref-14)
15. [↑](#footnote-ref-15)
16. https://github.com/CPJKU/madmom/blob/master/bin/OnsetDetectorLL [↑](#footnote-ref-16)
17. https://github.com/CPJKU/madmom/blob/master/bin/OnsetDetectorLL [↑](#footnote-ref-17)
18. <https://github.com/jakobabesser/bassunet> [↑](#footnote-ref-18)
19. https://www.sonicvisualiser.org/ [↑](#footnote-ref-19)
20. https://github.com/RamoonRoomeu/AutomaticAssessmentSax/tree/main/Training%20and%20Experiments [↑](#footnote-ref-20)
21. MAE = Mean Absolute Error [↑](#footnote-ref-21)
22. Student performances with low accuracies will mean that less onset and offset measurements get inside the 20ms window for checking, so less deviation calculations are made. [↑](#footnote-ref-22)
23. https://www.rockschoolespana.com/ [↑](#footnote-ref-23)