Master thesis on Sound and Music Computing

Universitat Pompeu Fabra

DRAFT 0.5

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

Colm Forkin

Supervisors: Vsevolod Eremenko,

Xavier Serra

March 2021



Master thesis on Sound and Music Computing

Universitat Pompeu Fabra

DRAFT 0.7

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

Colm Forkin

Supervisors: Vsevolod Eremenko,

Xavier Serra

2021



Dedication

I would like to thank those near to me personal that made this whole Masters possible. Along this your journey for this past year I have pushed myself to the limit, I suppose because the expected standard are high. I have found myself in an environment where the students are expected to constantly deliver week in week out and always have to bargain hard if an intermediate deadline cannot be met, while teaching staff can deliver feedback whenever it suits them. Where research staff work within their areas of responsibility, with occasional exceptions of some teaching staff making the extra effort. While working on the internship I found the University admin is minimal, especially around the careers service was very bureaucratic and unhelpful I suppose due to lack of resources

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 there are some micro-timing improvements that can be made. Experience in preparing 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 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, Machine Learning, Music Assessment, Music Education, Music Information Retrieval, Music Performance Analysis, offset, onset, playing technique, plucking, position, rhythm, source separation, string detection, style, Automatic Music Transcription, ground -truth, Rhythm, Timing

Contents

[1. Introduction 13](#_Toc77670625)

[2. State of the Art 15](#_Toc77670626)

[2.1. Music Education 15](#_Toc77670627)

[2.2 MIR approaches to onset/offset detection 16](#_Toc77670628)

[2.3 Measuring duration 17](#_Toc77670629)

[2.4 Abessers Research 18](#_Toc77670630)

[2.4 Literature Research Summary 19](#_Toc77670631)

[2.5 Revision of Algorithm Evaluation for Music Pedagogical Purposes 22](#_Toc77670632)

[2.6 Perception of Onsets and Offsets 23](#_Toc77670633)

[2.7 Observations 23](#_Toc77670634)

[3 Datasets 25](#_Toc77670635)

[3.1 IDMT BASS SINGLE TRACKS 25](#_Toc77670636)

[3.2 TCL Dataset 27](#_Toc77670637)

[4 Methodology 31](#_Toc77670638)

[4.1 Core Algorithms 31](#_Toc77670639)

[4.2 Deviation Metrics 37](#_Toc77670640)

[4.3 Correlation with student grades 39](#_Toc77670641)

[5 Experiment 40](#_Toc77670642)

[5.1 Pysimmusic End-to -End test 40](#_Toc77670643)

[5.2 Student Recording Portal 41](#_Toc77670644)

[5.3 Grading 45](#_Toc77670645)

[5.3.2 Teacher Comments 48](#_Toc77670646)

[5.4 Outcome and Lessons Learnt 48](#_Toc77670647)

[6. Results 50](#_Toc77670648)

[6.1 Yellow 51](#_Toc77670649)

[6.2 Billie Jean 52](#_Toc77670650)

[6.3 Just Looking 53](#_Toc77670651)

[6.4 Brown Eyed Girl 54](#_Toc77670652)

[6.5 Roadrunner 54](#_Toc77670653)

[6.6 Walking on the Moon 54](#_Toc77670654)

[6.7 Machine Learning to predict grades 55](#_Toc77670655)

[7. Conclusion 56](#_Toc77670656)

[7 List of figures 58](#_Toc77670657)

[8 List of tables 59](#_Toc77670658)

[9 List of symbols 60](#_Toc77670659)

[10 Bibliography 61](#_Toc77670660)

[APPENDIX A Software Requirements 64](#_Toc77670661)

[APPENDIX B Real Trinity Exam Results 64](#_Toc77670662)

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 etc all [8] 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. These apps[[1]](#footnote-1) offer performance assessment for learning help guide the student without an expert giving feedback on tuning note accuracy and metronome accuracy. However, although they engage the student well with attractive edutainment front ends (e.g. they include scoreboards for highest accuracies) there are other aspects musical endeavour not well covered such as note duration, articulation, good use of dynamics etc.

A music student learning in a formal context can now get daily feedback from an app and this can complement the practicing habits recommended by a teacher. The same apps offer more engagement with the student, “gamifying” the process with score and league tables of performance for songs. 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 a 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 even learnt their instruments. Neill McCormack [2] describes how U2 got together in the early days and difficulties they had in trying to get a good sound with friends and older siblings around giving them

feedback during rehearsal. Their initial success was enough to get them into a studio with producer Steve Lillywhite giving them the feedback they needed for professional quality sound recording on their first album.

But even after achieving success, Adam Clayton sought bass lessons from Patrick Pfeiffer in the mid 1990s, (author of “Bass for Dummies”) and gained new heights in performance of the instrument as a result.

Performance assessment of a particular instrument, in this case the bass guitar, has be placed in the context of the goals that a musician wants to achieve and although it maybe have a role in commercial success, the aims and goals are distinct.

The effectiveness of a learning program, i.e. the goals and purpose has to be considered [1] technology into the music classroom. The Trinity Rock and Pop Bass Syllabus [3] is the syllabus chose in this thesis, since it has a good reputation in 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 in order to produce good quality sound, managing the dynamics.

The SOTA will frequently refer to examples from the Datasets under study,but their formal introduction shall be in Chapter 3.

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 spacing.

* 1. Music Education

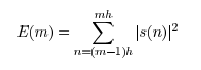
The mastery of any musical instrument without the live presence of teacher is a challenging task. However, there has been interesting developments MIR research that can support Music Education. The algorithms that are used in automated music transcription (AMT) are capable of extracting score information from polyphonic recordings [13] (Salamon/Gomez). In our music education context, we have the advantage of isolated bass stems as input, and we can take advantage of recent success in transcribing bass lines using data driven methods based on the U-Net Architecture [5] and signal processing methods by the same author [6][9]. Abesser has also tested the effectiveness of extracting plucking features (right hand techniques) and expressive features(left hand techniques) in playing bass [7]. AMT methodologies for bass guitar has been seen useful applications in genre classification[10] and sound synthesis and this research aims to apply these techniques to improve the formal context in which bass music students can improve on their rhythmic precision

The effectiveness of a PAT is can only be accurately measured on the readiness level it provides to a student in Assessment of Learning[1] situations as referred to in [1] . This can take various forms: academy entry and final exams, competitions, session job interview, but e The Trinity College London (TCL) Rock and Pop Grades 0,1,2,3 [4] are the quality benchmark used for grading the student. In the syllabus [3] are nine exam or grade levels that can be classified in three groups: Grades 0-2; Grade 3-5 and Grades 6-8. Trained examiners at Trinity perform the evaluation scores using the following guideline: Distinction 87-100; Merit 75-86; Pass 60-74 and below pass. Exam results in the advanced category (Grades 6-8) can award “UCAS points” (Universities and Colleges Admissions Service ) that can be used for higher education entry.

Preparation for a grade 1 exam is estimated to be the following[2]: Guided Learning Hours (12), Independent Learning hours( 48).

* 1. MIR approaches to onset/offset detection

This thesis aims to build on what current technologies can offer. Onset Detection shall be discussed focusing on the properties of the bass which is a Pitch Percussive instrument. The peak picking algorithm introduced by Bello [11] is one of the techniques used for measuring onsets. It requires experimentation to optimise 3 parameters. His subsequent paper on considers the energy in addition to phase for the onsets [14] and it is the equation that considers the local energy of a signal that gives us a technique to measure offsets.



Equation 1 Local Energy

A big question in this research is the following: when does the offset end?

This introduces the topic of signal perception in the human auditory system and this has been researched outside of the context of Music Education by Kopp-Scheinpflug [10]. In the scope of this work, an attempt is made to consider the output of this research in gap-detection and the limits of human auditory temporal acuity which varies from 2/3 ms to 30 ms depending on the level of spectral disparity in the signal

As part of a trade study of current apps [[2]](#endnote-1) that use MIR approaches for onset detection, the performance assessment for singing in Yousician is very different to how bass (and other stringed instruments) are assessed. The piano roll format as used for Vocals training gives better visual feedback to note durations and although no comprehensive test have been carried out on the this, on some songs it has been found to correctly assess note duration but for other songs in the Yousician curriculum (e.g. “Fire” by REM) it has been clearly shown not too 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.

In Bass guitar, we consider two components of onset: the attack part of the rapidly increasing amplitude envelope and the exponentially decaying part. Bello[14] has provided guideline for choosing the right method depending on the requirements. The Wavelet method places focus on precise time localisation, an important aspect of bass rhythm. There are other options that have a high computational cost ranging from complex-domain spectral difference to using training sets and statistical methods. Non-realtime assessments are not constrained by computational cost requirements.

* 1. Measuring duration

As mentioned earlier, experiments with Yousician have shown that duration feedback is missing for some tracks in Vocals Performance assessment, but it is missing for all the tracks for Bass guitar. e g. If you have a half note duration on the score and you play a quarter note duration, you will not be penalized. If you play 4 crotchets, 4 quavers, or 4 semiquavers, the scoring would be the same by all apps, even though there is a clear musical difference in each of the displayed three bars below



*Figure 1: 4 crotchets, 4 quavers and 4 semi-quavers.*

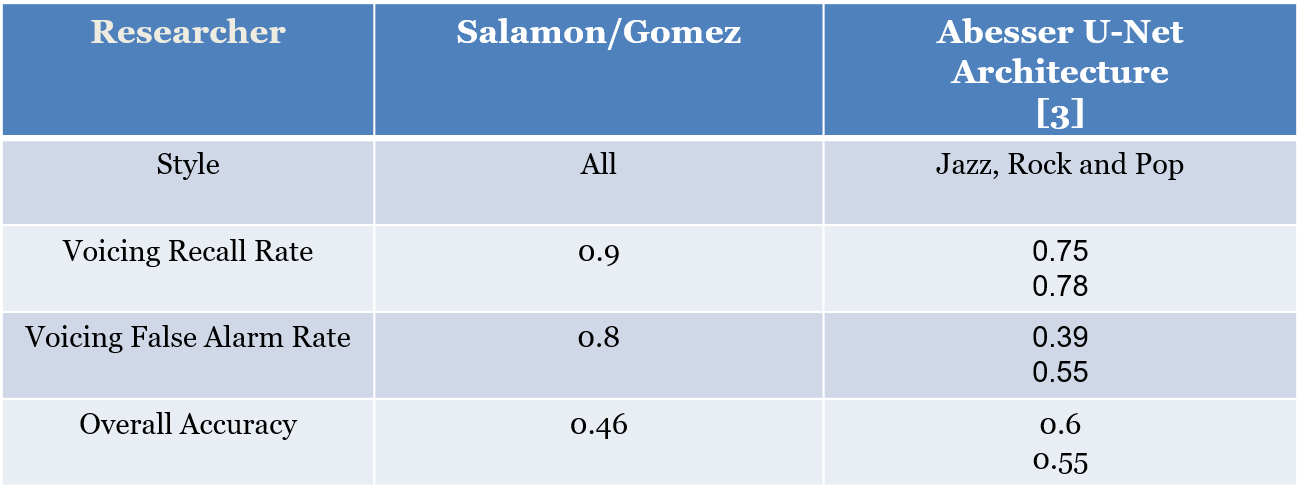
This shortcoming is really surprising when you consider that that Yousician is capable of correctly assessing left hand techniques covered in some of Reboursieres paper [15] such as slides, Hammer-ons and Pull-offs. The challenge of durations measurement is that it is not an instantaneous measurement. 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 the algorithms under study focus on capturing the energy of the bass stem and determining the point of “drop off”, when the bass note is no longer audible. In any given song you may find a variation in the finger style techniques used for playing particular 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. We don’t consider the case where the offset position of a given note exceeds the onset of the subsequent note. 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.

* 1. Abessers Research

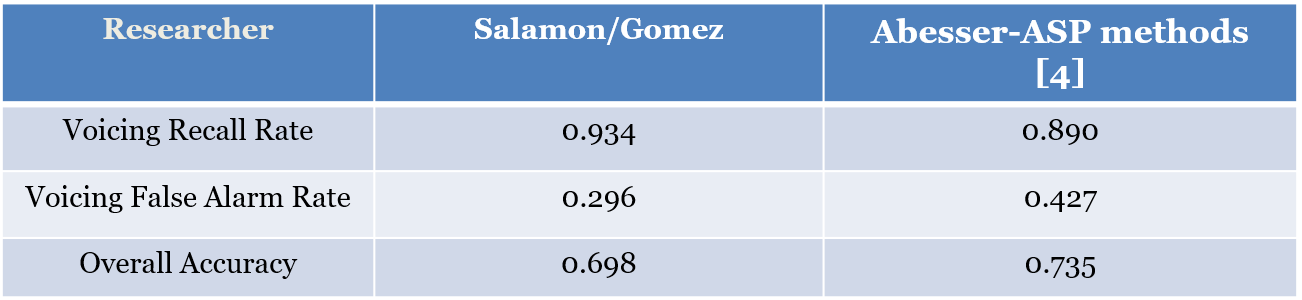
Jacob Abessers applied research in music information retrieval & machine learning / deep learning and audio signal processing is focussed primarily on bass and an evaluation of the results of his research in the context of timing assessment of bass are worth investigation. For timing measurements, the voicing classification (independent of which pitch it is ) is an important metric. The Salamon and E. Gómez [13] techniques have good voicing recall metrics but Abessers-Müller Data-Driven [5] yields lower false alarms

*Table 1: Abessers Data Driven algorithms vs Salamon/Gomez*



The experiment on bass tracks that Abesser performed when comparing the Signal Processing methods with S/G [3] also yielded better overall accuracy (considering pitch) but the on just the Voicing Recall and False alarm rate, the S/G techniques have higher accuracy.

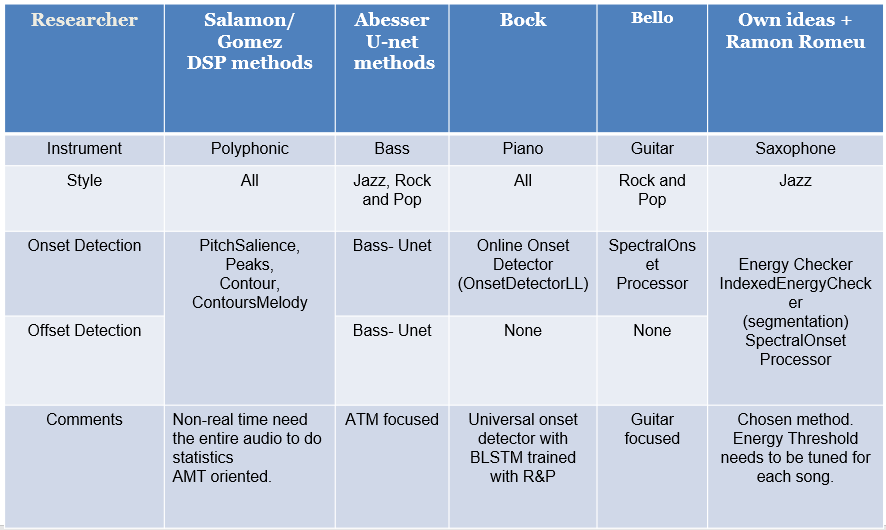
*Table 2: Abesser Signal Processing vs Salamon/Gomez*



* 1. Literature Research Summary

Onset detection is well researched topics with many off-the-shelf library functions available from Madmom[[3]](#footnote-2) and Essentia[[4]](#footnote-3) that can be readily applied to a bass stem. The following table summaries the state of the art of algorithms used for onsets and offsets, many of which use these library functions.

*Table 3: State of the Art algorithms for Onset/Offset detection*



The first columns are based on pitch extraction and have been tested on polyphonic music and work quite well. The main function is the PitchMelodia[[5]](#footnote-4) function, which is designed to extract the predominant melody from polyphonic music and has 18 parameters which can be configured.

The second column is based on a CNN (Convolutional Neural Network) Streamlined Encoder/Decoder Architecture for Melody Extraction. It has been tested on the following Dataset:   
- Real World Computing (RWC)   
- MDB-bass-synth[5],   
- Weimar Jazz Database (WJD)  
It the implementation code for bassunet[[6]](#footnote-5) is open source. In the 2017 [6] paper is based on Onset and Offset detection using Fo Contour Tracking, Abesser used the following annotated Database, which is also used in this Thesis.  
- IDMT-SMT-BASS-SINGLE-TRACKS (Fraunhofer)

The matlab/python code for calculating onsets and offsets is closed source but the Fo Tracking library code is available[[7]](#footnote-6) in the pymus libraries.

Columns 3 to 5 are all based on the Madmom[[8]](#footnote-7) libraries. The Online Onset Detector[[9]](#footnote-8) based on recurrent neural networks by Bock is a universal onset detector with BLSTM and was trained with music mixtures including R&P. The Spectral Onset Processor implements several (up to 11) onset detection functions considering phase and energy information. This algorithm was considered for the first experiment using the guitar focused pysimmusic [[10]](#footnote-9)tools (back end for Music Critic). The same function (madmom.features.onsets.SpectralOnsetProcessor) was considered in the set of algorithms that were developed by Bello [11] based on Peak Picking. This approach forms the core of one of the onset detection methods used in this thesis. The following peak picking formula forms the basis of the [xx]



Equation 2 Peak Picking formula

Ramon Romeu[[11]](#footnote-10) developed a wrapper based on this formula with the following values determined for C\_t, H and delta in sweep experiments

#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*

Initially the values given for C\_t,H, delta shall be tried for chosen Dataset tracks. A sweeping method can then be sued to optimise the detection for a given musical piece.

The final column summaries the method developed that can return an offset for every detected on set, using a “sound island” approach.

* 1. Revision of Algorithm Evaluation for Music Pedagogical Purposes

In the TCL R&P exam[2], 33% of the Music Assessment is directed at Fluency, synchronisation with the backing track, security in notes and rhythm. This is probably the easiest to objectively define and benchmark with MIR solutions currently available. Music Critic[[12]](#footnote-11) is currently developed for Guitar at the MTG (Music Technology Group) and requires adaptation for bass to deal with the importance of timing and locking in with the drum pattern.

The Guitar version has an online version and works well for simple demos. Music Critic gives the student feedback at the end of a performance. It does not flash green or red on each individual note “on the fly” like in the previously mentioned edutainment apps. This leaves some room for discretion in terms of discriminating between the different bars and different notes played for a specific musical piece. Post-Analysis is also more appropriate in preparing the student for an exam and it leaves margin for customizing deviation in timing/tone particular to each song and/or a part. With a defined set of duration formula, an annotated dataset of onset and offset times, the next stage is to gather new annotated data by obtaining real student recording using a Student Portal. (Chapter 5)

The second part of the TCL R&P exam is on Technical Control. This is ability to control the instrument effectively, achieving the various technical demands of the song and sound quality. Of the three songs a student chooses, one song has to be TF (Technical Focus) which means more weigh is given to this section (12 point instead of 9). This prompts, the question, how do we measure and gather data to provide useful feedback on instrument aspects? Is there a Dataset that we can use to train a model to identify good technical focus performances? Abesser addressed this in [6] and the same instrument parameter annotated dataset [17] is used in thesis. Even though, the scope of the thesis is limited to timing and rhythm aspects, the TCL dataset shall be built in a wat so that it can be scaled to consider all technical aspects of playing bass.

The Dataset prepared in this thesis consists of 8 Student recordings of 6 different TCL R&P syllabus and is discussed in Chapter 3. These recordings were passed to a qualified Bass Teacher who assessed the performances using a customised questionnaire. The original Rubik was based on the TCL examiners report of Pass, Merit or Excellent in each of the examination areas of Fluency & Security and Technical Control. With this dataset, a set of histograms of onset deviations in 4 different levels could be developed, like the ones found in [1]. These could be classified according to Trinity scoring criteria as follows: : Level 1 (Fail) Level 2 (Pass) ; Level 3; Merit; Level 4( Excellent). In addition to onset, measurements for offset and note durations could be added.

* 1. Perception of Onsets and Offsets

Onset detection is huge research topic with many alternative formulas for calculation [14], however offset, particularly the perceptual aspects of them, have historically not been given the same level of attention. Perception of offset detection is dealt with very well by Kopp-Scheinpflug [10] in which they researched literature on neurons with sound-offset responses in the auditory system. In sound perception, the concept of “just noticeable difference” is important and what this means for offsets is that a sound will have to be of sufficient duration to allow an offset response. For music education and performance assessment, this means time window needs to be defined that specifies the minimum duration for detecting an onset and an offset. Typically, the onset window [9] has a time value of 50ms being cited as being the minimum time window of detecting onsets, but for Music Education purposes, this is too long and pending further measurements and observations, a time window of about 12 ms would probably be more appropriate.

* 1. Observations

There are a lot of musical aspects to consider in assessment, that are difficult to measure using transcription methods based on score. They are swing ratio, attack displacement (playing exactly, ahead or behind the quantised beat), to mention a few. To capture these characteristics, it is necessary to obtain ideal student performances at the excellent level and also some performances and intermediate and novice level. Ideally these could be obtained with from recording with a microphone capturing the bass performance separate from the backing track. However, if this dataset is limited, Source Separation techniques based on the Spleeter model [18] that have been successfully applied to the Trinity Rock [xx] and Pop Bass syllabus can be applied to previously recorded performances.

The challenge in this project is to find the most suitable machine learning techniques to extract the most relevant parameters from the limited number of annotated student performances. In [6] Abesser used support vector machines to perform not only score transcription for bass but also the extraction of the plucking and expressive styles.

Automatic assessment of musical performance has seen a variety of apps in the marketplace and there are research papers continue to push the boundaries on this [1] achieving similar assessment results to that given by a human-. The area of Technical Control (as understood by the TCL R&P standard) remains a challenge for machines to evaluate and the onset detection and measurement methods as used in Music Critic [1] tools do not sufficiently address the timing needs of the electric bass. The initial studies have shown that Music Critic has problems detecting notes while the tempo is fast and notes are short in guitar, so theses aspects will also need to be measured and addressed for bass.

1. Datasets

This chapter explains the Datasets and how they are used to evaluate algorithm accuracy

The following Datasets shall be used to evaluate the algorithm quality

* 1. IDMT BASS SINGLE TRACKS

The IDMT dataset from Fraunhofer 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 3: Extract from IDMT Dataset. File 002.xml*

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.

This Dataset is not used for Student performances. It is important the note that the IDMT contain many tracks which have complexity exceeding that which is required for our grading in the TCL dataset and it also has some very particular short notes where the plectrum is used for the style. The objective of this Dataset is to measure the effective ness of the onset/offset algorithms. It is available under a creative commons licence.

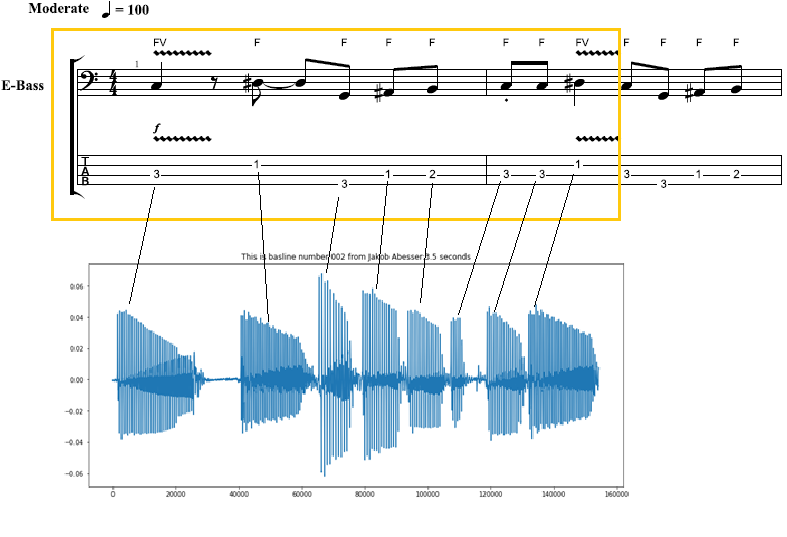
IDMT-Summary

IDMT (Fraunhofer) : 17 tracks

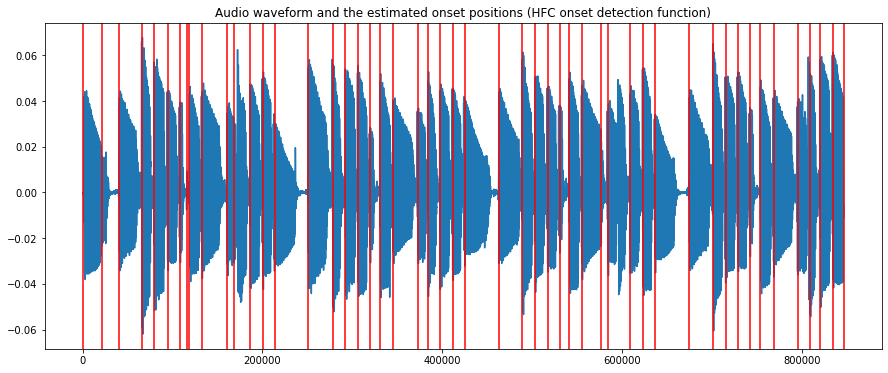
* Varying plucking techniques (Pick, Finger, Muted, Slap)
* PDFs of score and XML of parameters (onsets, offsets, pitch, fret number)
* Expression Style annotated
* 4 tracks within musical and timbral complexity of Trinity G0-3

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 4: Note duration of Staccato notes*



*Figure 5: Audio 002.wav (IDMT DATASET) with onsets*

* 1. TCL Dataset

Talk about full TCL curriculum first and point out this example

If you were to check the TCL R&P criteria for assessing the Grade 1 song “Float On” [4] it says, “*the chorus features some sustained dotted minims, which should be held for their full length*.", so clearly there is a need for correctly measuring not just the start time but also the stop time of musical notes.

The Trinity Dataset consists of Audio of six chosen songs from the bass syllabus ranging from grade 0 to 3. The recordings are cover versions performed by TCL professional musicians. For the Trinity Dataset, the Ground truth is taken to represent the audio of, the isolated bass track of the Trinity recordings .

*Table 4: Six chosen songs from TCL song list*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade** | Artist | Original Artist | Durations | Duration measured |
| **0** | Yellow | Coldplay |  |  |
| **1** | Billie Jean (B. Jean) | Michael Jackson |  |  |
| **1** | Just Looking (J.L) | Stereophonics |  |  |
| **2** | Brown Eyed Girl (B.E.G) | Van Morrison |  |  |
| **3** | Roadrunner (RR) | Junior Walker and the All Stars |  |  |
| **3** | Walking on the Moon (WOTM) | The Police |  |  |

There are two key differences with the previous dataset. First, it is not publicly available and access to the audio and the PDFs is only granted on purchase of the materials from Trinity College London. Secondly, unlike the Fraunhofer stems, these tracks do not come with accompanying Onset/Offset annotations, which were required to be made manually using Sonic Visualizer. The PDF files come accompanied with XML files which provides rich information on the score note duration . There is also technical information e.g.  
 <notations>

<technical>

<fret>2</fret>

<string>2</string>

</technical>

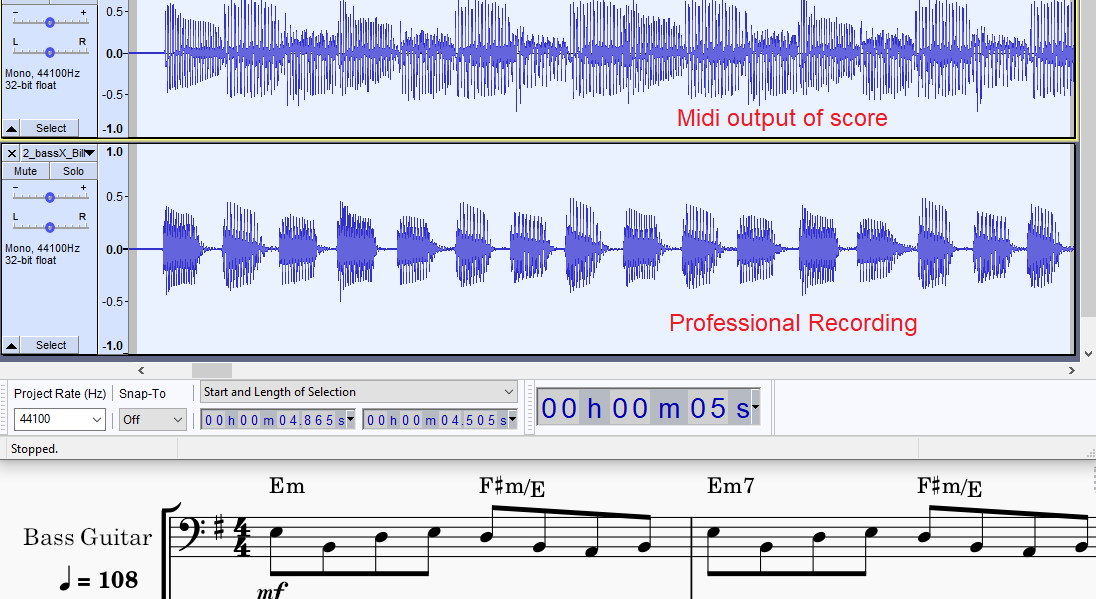
</notations>

And articulation information, e.g the accent in Billie Jean  
 <articulations>

<accent/>

</articulations>

The XML files can be viewed in Musescore and after removing the chord information and export of a WAV file was made. This “mechanical rendering” can provide a useful reference in testing the algorithms. It demonstrates the importance of the incorporating stylistic elements in a real performance especially where duration is concerned



*Figure 6: Billie Jean: Midi Rendering vs Human recording*

The first function of this dataset is to validate the algorithms short listed after testing with the IDMT dataset. The second function is what makes it the core Dataset of this thesis, and that is to serve as a basis to measure student performances.

The term “ground truth” refers to the bass stem of each of the above recordings. This is taken as the watermark to represent a grade of 100% in timing and rhythm.

The following table summarizes the individual musical features of each of the tracks as described in the Grade Books[4]. As stated in the introduction, the objective is to direct the numerical grading and verbal assessment to reflect as much as possible the rhythmic and timing qualities a Trinity Examiner is looking for. Ultimately the goal is to match what are human assessments to objective measurements of extracted audio features. In keeping with the importance of considering Technical Control of the instrument (Chapter 2) , Songs 1-5 were chosen since they were marked “TF” song in the TCL syllabus. The sixth song (WOTM) was chosen because it had widely contrasting note duration features.

The following table summarizes the Technical Control parameters of the 6 chosen audio tracks.

*Table 5: Tech. Control Parameters for 6 TCL songs*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Song | **Coord.** | **Syncop.** | **Repetition** | **Dynamics** | **Articulation** | **Note Len.** |
| Yellow |  | chorus | No rushed feel |  |  | Play evenly (v) |
| B. Jean | intro | Accent just before chorus |  |  | separate jerky quavers+ smooth, melodic material |  |
| J.L |  | Chorus: accented, syncopated motif.,  hard accent |  | Unexpected  subito p at bar 25. |  |  |
| B.E.G |  |  |  |  |  | different note lengths and rests |
| R.R |  |  |  |  | Tenutu (underscore)  loud on beat 1 |  |
| WOTM |  | syncopated  repeated notes |  |  |  | correct  separation |

Yellow was truncated to remove the repeated verse, since no new musical features were introduced.

The song WOTM has been cut to only include the first 50%. The song is symmetrical, so any annotation done on the second part is redundant and only opens the door for introducing more inaccuracies. Moreover, the musical score requests some adlib in the second half, which is beyond the grading requirements of this thesis.

The WOTM verse has a long note duration with zero internote gaps and the bridge has the opposite shorter “reggae” notes with some noticeable internote gaps. The BPM of this song is 146 and for Billie Jean it is 108. Both require very different thresholds. The Grade1 “Billie Jean” song also manifests similar contrasting sections. Evaluations were set up to evaluate the quality of onset and offset measurement algorithms

4 Methodology

Music Education demands more precise timing measurements than state of the art automatic music transcription. Onset detection for electric bass has different requirements for guitar in two key areas. First, as a rhythm section instrument the bass plays a key role in synchronizing with the drum pattern. 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. The main proposal of this thesis is to add duration measurement and to achieve this offset detection is required and this places a constraint on the choice of onset detection algorithm

* Benchmark algorithms for accuracy
* Testing SOTA algorithms to measure student performances
* Using Machine Learning to train new a Model with teacher graded student performances
* Analysis of teacher comments in student performances and correlation with student grades
  1. Core Algorithms

In the algorithms introduced in the literature research in chapter 2 there are two strategies involved: “combined” and “non- combined”. Non combined are the classical approaches to capturing the onset without regard to monitoring the end of the onset. The combined approach is to measure start and stop time with the same audio frame. The non combined approach leaves you without offset captures

The 4 methods introduced in chapter 2 are tested alongside a fifth method to implement combined approach.

* Madmon Online Onset Measurement (Non-paired, data driven)
* Spectral Onset Processor (Non- paired, ASP driven)
* AbesserUNet Algorithm (Paired, Data driven)
* Salamon and Gomez (Paired, ASP driven)
* IndexedEnergyChecker (Paired ASP driven)+ SOP (Non Paired) ASP
  + 1. Indexed Energy Checker Overview

For Onsets, 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, and we will consider the small difference between the actual attack and the perceived attack to be negligible.

A short trade study was conducted comparing a simple RMS measurement approach with a more elaborate “Sound Island” approach (quote github source)

This approach relies on capturing when the energy drops below a certain threshold that marks the offset point. That was my original algorithm which I called “EnergyChecker”. I found an improvement on this by Ramon Romeu []. His algorithm returns a set of start and stop indices representing onset and offset. It is based on the “sound island approach” and uses a threshold input which can be customised for particular songs. In addition to this function, an additional method based on the SpectralOnsetDetection [1] can be used to improve it.

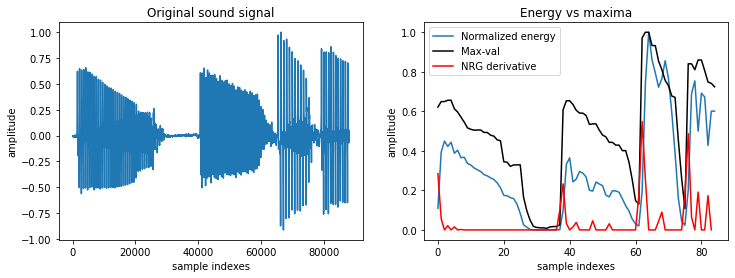
The resulting PRF for onsets for each of IDMT with a 20ms evaluation window averaged over 4 songs that are most similar to TCL Dataset.

*Table 6: Benchmark results of Algorithms*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Researcher | Pysimmusic | Salamon/Gomez | Abesser U-net | Bock | Colm Forkin | Ramon Romeu |
| P |  |  |  |  |  |  |
| R |  |  |  |  |  |  |
| F |  |  |  |  |  |  |

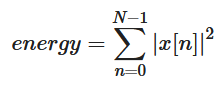
4.1.2 Indexed Energy Checker Detail

Originally a simple Energy Checker function was developed to capture energy drop off points to indicate offset. This was quickly replaced by the IndexedEnergyChecker which optimised the pairing of onsets and offsets, based on an Energy Threshold. This core algorithm 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.



*Figure 7: Normalized Energy of Audio 002.wav sample (IDMT DATASET)*

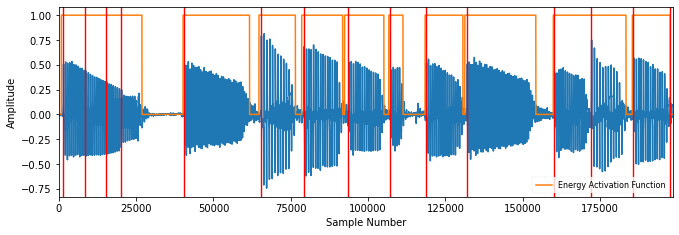
Energy is calculated using the normalise energy function



Equation 2 Normalized Energy Function

The frame size is set to 1024 and the hop size to 512.

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 8: Sound Islands of Audio 002.wav sample (IDMT DATASET)*

This algorithm finds the best matching pairs so in “match\_events” functions

- distance between elements is no greater than matching\_window\_size

- sum of all distances is minimized

4.2 Accuracy Metrics

Accuracy measures for SOP and Energy Island

The resulting PRF for onsets for each of the Trinity RnP ideal bass recordings (stems) with a 20ms evaluation window compare both IndexedEnergyChecker with SOP

*Table 7: Accuracy Metrics for 6 TCL ground truths*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Song | P | | R | | F | |
| IEC | SOP | IEC | SOP | IEC | SOP |
| Yellow | 0.939 | 0.98 | 0.942 | 1.0 | 0.94 | 0.99 |
| Billie Jean | 0.949 | 0.983 | 0.969 | 0.997 | 0.958 | 0.99 |
| Just Looking | 0.682 | 0.9 | 0.828 | 0.869 | 0.748 | 0.884 |
| Brown Eyed Girl | 0.904 | 0.806 | 0.822 | 0.852 | 0.861 | 0.828 |
| Roadrunner | 0.837 | 0.847 | 0.839 | 0.804 | 0.838 | 0.825 |
| Walking on the Moon | 0.732 | 0.952 | 0.975 | 1.0 | 0.836 | 0.975 |

**Table x.y Benchmark results of combined vs non combined**

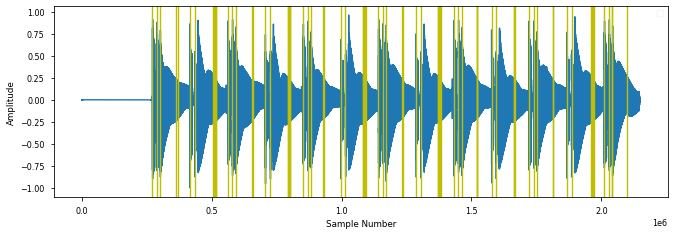
The worst performing EnergyChecker results were then benchmarked against the Salamon /Gomez results

*Table 8: Selected comparison: IEC vs SG*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Song | P | | R | | F | |
| IEC | SG | IEC | SG | IEC | SG |
| Just Looking | 0.682 | 0.78 | 0.828 | 0.745 | 0.748 | 0.761 |
| Walking on the Moon | 0.732 | 0.794 | 0.975 | 0.653 | 0.836 | 0.716 |

For the first half of the track, WOTM , using the IndexedEnergyChecker with Threshold set to 0.06, yielded a lot of false onsets.

0.514 0.949 0.667

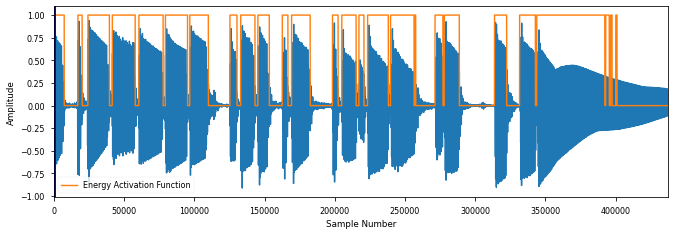


*Figure 9: Calculated IEC onsets for WOTM verse*

For the bridge section of the same track (WOTM), the IndexedEnergyChecker performed better, but again the statistics dropped when considering the last long note.

0.937 0.949 0.943 (with last note)

0.961 0.948 0.954 (Without last note)

****

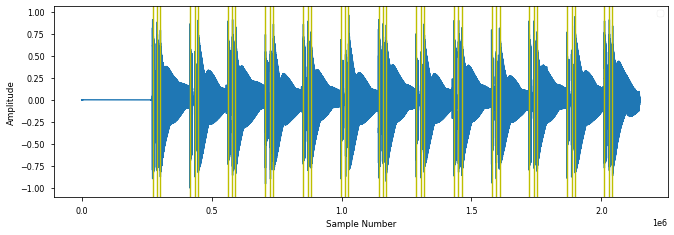
*Figure 10: Calculated IEC Sound Islands for WOTM bridge*

This song has the longest note length of all the tracks. With (thresh = 0.06, we still see a lot of hysteresis on the last sustained note

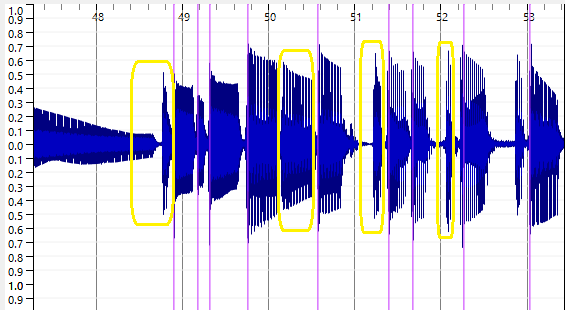
In contrast the Salamon/Gomez algorithm performed a lot better.

First Half S/G

0.974 1.0 0.987



*Figure 11: PitchMelodia derived Onsets for WOTM verse*



*Figure 12: 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 S/G method in the bridge as seen in fig x.y marked in yellow

It is important to clarify some local definition of “mute”. 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 10-10 miliseconds. When there is no gap., there is no offset. The offset in this case is equal to the next onset.

It is important to note that the offset of the ground truths for each of the tracks were “next-onset” adapted. This means that for song sections was no soft mute, were marked appropriately. The CSV file has 3 columns: onset, muted and offset.

Muted is either “Y” or “N”. When it is Y this means an offset exists. When it is ‘N’, the current offset ground truth value is equal to the next onset ground truth value.

* 1. Deviation Metrics

The following plots illustrate the deviation statistics for particular songs for onsets, considering the SOP method and the IEC for offsets, considering only the IEC for offsets.

Fig X.Y IDMT Dataset GT Onset deviations ( SOP ) algorithm

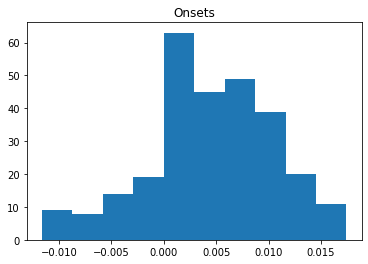
Fig X.Y IDMT Dataset GT Onset deviations ( IEC ) algorithm

Fig X.Y IDMT Dataset GT Offset deviations ( IEC ) algorithm

Fig X.Y Billie Jean GT Onset deviations ( ZZZ ) algorithm

Window 20ms

OnsetsABS Mean: 0.005915, Mean: 0.004029, Dev. from 0: 0.007258



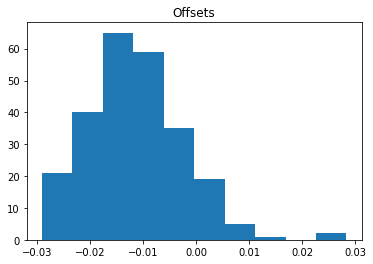
*Figure 13: Billie Jean GT onsets deviation*

misses 9 Percentage Miss 3.147

Fig X.Y Billie Jean GT Offset deviations ( ZZZ ) algorithm

Window 30ms

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



*Figure 14: Billie Jean GT offsets deviation*

misses 39 Percentage Miss 13.636363636363637

The same methods are used to assess the student performances.

* 1. Correlation with student grades

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.

The method used is linear regression and will consider as X inputs the P, R, F results and the mean absolute error and standard deviation of the onset/offset deviations. The Y values shall be 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

5 Experiment

In this chapter, three experiments are described. The first one is an end-to-end test on the chosen dataset against the current State of the Art methods used in the pysimmusic[1] tools. The second experiment is the main experiment and describes a procedure to gather and process electric bass recordings from music students on the six chosen TCL songs. Based on the lessons learnt from the main experiment a minor iteration of the student recordings was performed taking into account lessons learnt from main experiment and focussing on the metrics and songs that yield the most interesting results.

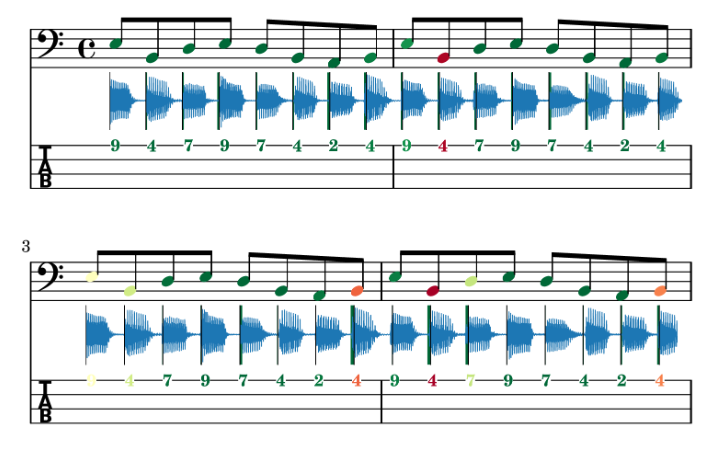
5.1 Pysimmusic End-to -End test

In Performance Assessment for Pysimmusic it is required to pre-program a song with BPM, meter and beat location. In the pysimmusic program[1] 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 with pysimmusic software for Billie Jean, the following steps were executed:

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

The diagram below shows the graphical output



*Figure 15: End to End test of Music Critic with Billie Jean GT*

The alignment of the ground truth stem was not aligned 100% which tells us the onset detection methods need improvement. 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.

5.2 Student Recording Portal

The second part of build the Bass Trinity Dataset involved a large campaign to find eight students to do at least one recording of each of the six chosen bass tracks. The name “Bass Critic” was given to the Student Portal” which was implemented as a Google Form containing instructions and links to each of the six songs to perform the live recording with a backing track and then submit. Details of this Portal can be found in the appendix. The Student was expected to use good headphones to isolate the backing track from the microphone and preferably use an direct audio interface for the bass instead of relying on positioning the microphone close to the bass amp.

5.2.1 Latency Test

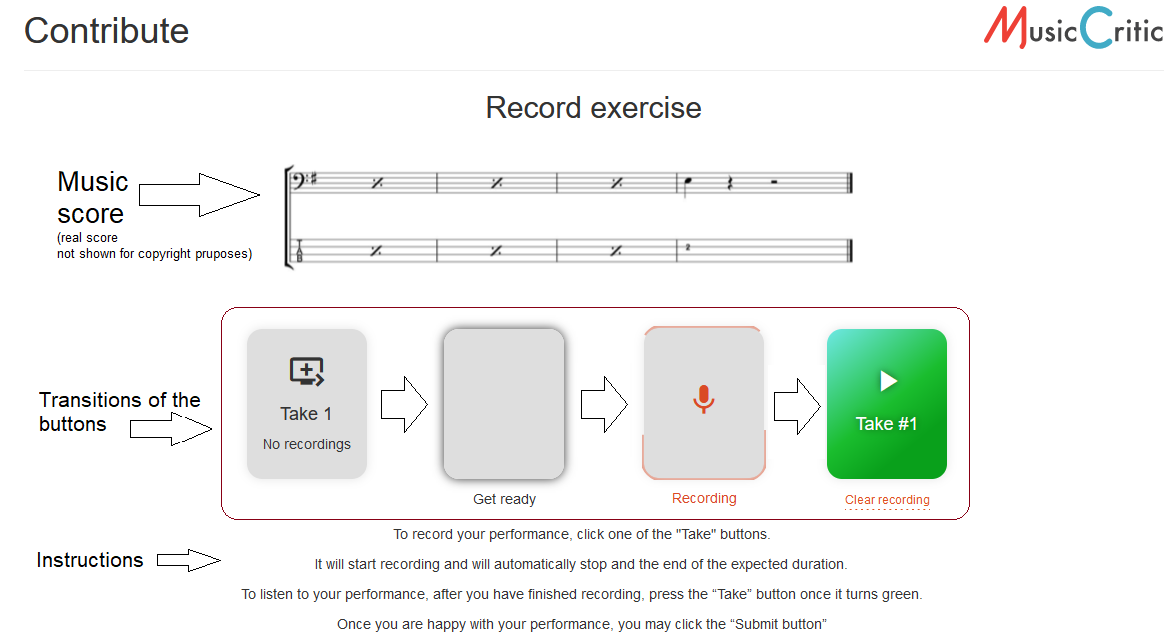
A pre-requisite before recording an attempt is the Latency Test. This involves placing the microphone close to the headphone speaker while recording the Click track. This allows to calculate a latency value which is then stored in the submissions.json file.

Ideally the microphone should go through the audio interface for the latency test however the hardware adapter components were not available at the time so the regular headset connection with the laptop was used.

As with the first experiment with Billie Jean in section 5.1, the “minus 1” tracks for the remaining five were created from the individual stems.

In the portal the student is requested to provide 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)

Refer to Appendix D1 and D2 for full details on Student Instructions. The middle section of the diagram below illustrates the transitions on the web portal for capturing a student recording.



*Figure 16: Workflow on Student Portal*

After the recordings were completed, the stems were downloaded from the server and the latency was removed from the track. In the JSON file an approximation for the latency was give, the snipbit below shows an example of the JSON file contents:

{

"path": "submissions/1814\_00463b7fd67144c2ba1be416c80afb70.wav",

"exercise\_id": 203,

"session\_id": "hyajrui9tpaivk52hqaiano3m56ktv01",

"latency": 0.056,

"user\_agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/90.0.4430.93 Safari/537.36",

"id": 1814,

"duration": 104.36160997732426,

"grades": [],

"created": "2021-04-27T15:17:20.481839Z"

},

*Figure 17: Latency for a given submission*

After removing the latency from the recordings, it was noticed that the json calculate value still did not align with the first onset. As a result, the student recording stems were aligned manually with the first onset. This assumption was reasonable since it reflected the reality of the “best effort approach” [[13]](#footnote-12)of the student recording. 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 and also copy mixed with the minus 1 track to be provided to the teacher for grading.

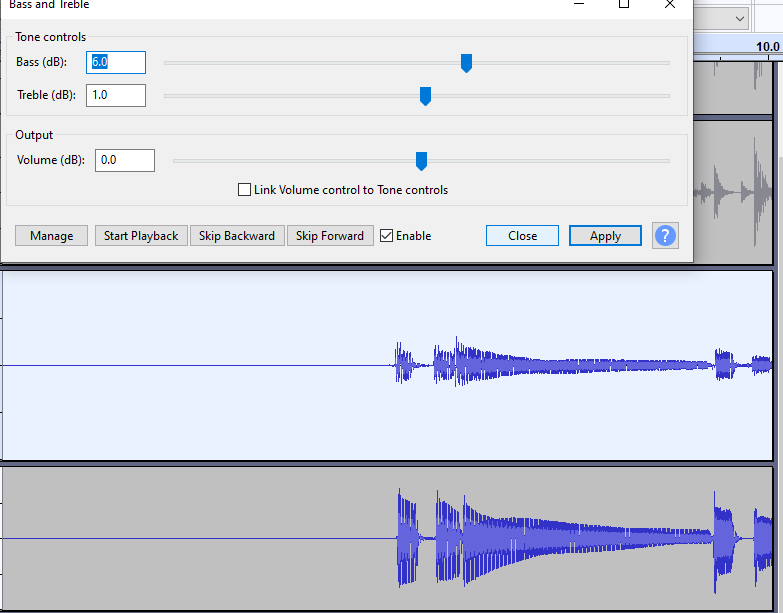
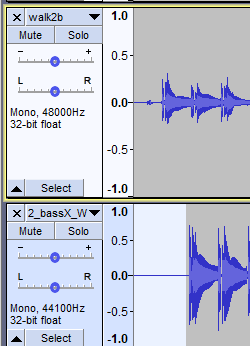
In the initial recordings, the microphone on the headset was used, there was some noticeable background noise which adversely affected the teachers sound quality grading. However, the algorithms were still capable of onset detection for noisy stems.

Steps:

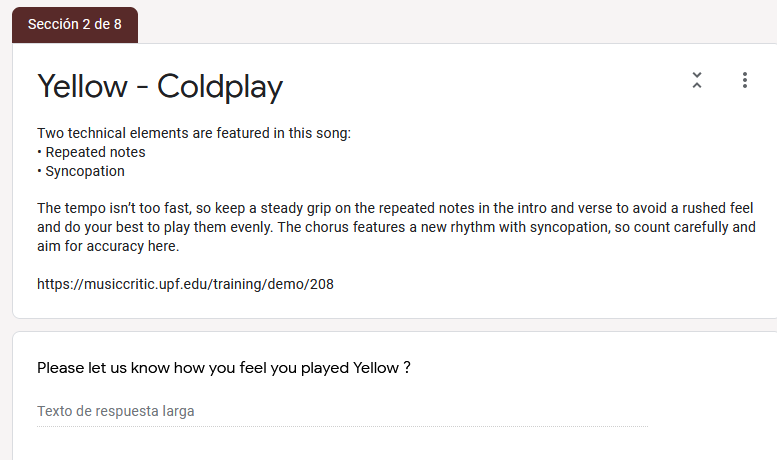
1. Download the student recording from the server.  
   IT has a name like this: e-e-g- "submissions/1805\_52a4886b326c4301b2760c8df6404c96.wav"
2. Rename it to the Student name.
3. Check that the playback rate is 44100Hs If its 48000Hz, left click on the audio file in Audacity and choose the rate 44100Hz, 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



Here is an example of how the link to student recording looks. (Full details of the instructions are provided in Appendix D )



*Figure 18: Technical instruction and link to recording portal*

5.2.2 Criteria for grading

Criteria for Grading

Since it was initially perceived that the investment in building the infrastructure for gathering quality data on Student performances for further research a set of question on the Teacher Portal Design were prepared to deal with musical aspects of the TCL syllabus that go beyond just physical onset and offset measurement.

For each song there were grade given each of the following categories:

Onset, Duration, Technical Focus, Dynamics and Sound Quality

Technical Focus (TF) is song dependent.

Billie Jean the TF is “Articulation and Coordination”  
Exam Grades are given on a scale of 5 to 1. (radio button).

Comment required for each exam grade.  
Overall grade also given(5-1)

Refer to Appendix D 3 for more details on Google Form sections

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.

These percentage points (88,80,63,37) were chosen to fit within the ranges of the 4 categories of the TCL syllabus (ref SOTA chapter excellent, merit, pass, below pass)

Bass Teacher Selection: Marti Brenach Obradors

Song length limit to reduce load on Teachers and Students(8)

*Table 9:Duration of 6 TCL song extracts*

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **Seconds** |
| **Grade 0** | 1 | Yellow | 93 |
| **Grade 1** | 2 | Billie Jean | 55 |
| **Grade 1** | 3 | Just Looking | 86 |
| **Grade 2** | 4 | Brown Eyed Girl | 64 |
| **Grade 3** | 5 | Roadrunner | 82 |
| **Grade 3** | 6 | Walking on the Moon | 120 |
|  |  | TOTAL | 500 secs |
|  |  |  | 8m 20 secs |

Teacher Grading done on 89% of recordings

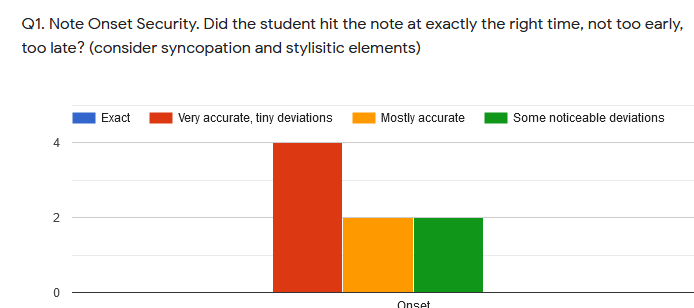
Teacher Assessment Sheet

* Explain rational behind question choice
* Explain rational for global and non sectioned questions
* Explain rational for teacher comments

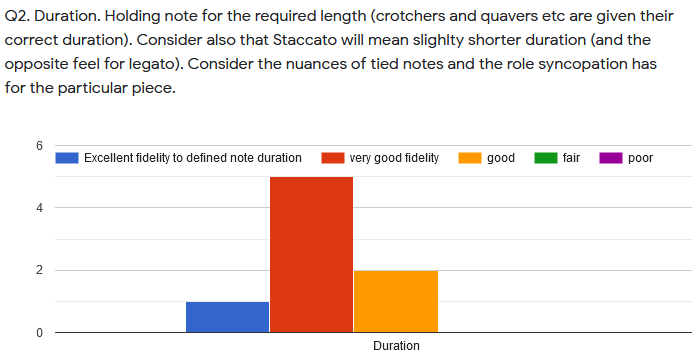
Student Instructions

* To play the first 60-90 seconds of the song
* Teacher should only correct this section
* Student question was required, (technical set up, feedback on experience, feedback on their own performance)
  1. Grading

As mentioned in the State of the art the TCL R&P exam serves as the syllabus that we will use as a guide for teacher grading. Only the timing criteria in the fluency section is considered, splitting the evaluation to match onset and duration accuracy.



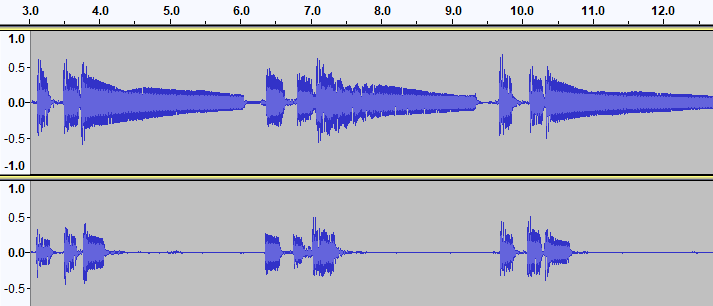
*Figure 19: Grading Histogram Onsets: Billie Jean*



*Figure 20: Grading Histogram Offsets: Billie Jean*

Five of the six songs are “technical focus” songs and the grading here is considered to add more insight into the rhythmic and timing aspects of the students’ performance and perhaps insight into volume handing and dynamics.

Three comment sections and have been added to emulate the comments that are given on a per song basis in the trinity exam.



*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".*

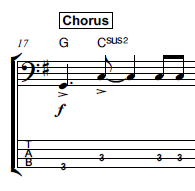
Duration: **5 (Excellent)**

Overall: **4**

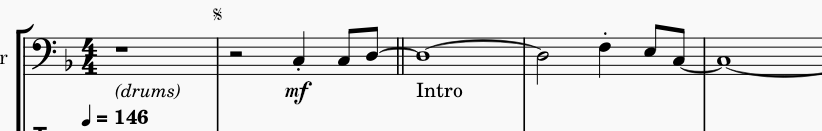
*Figure 21: Audio extracts of very different WOTM student performances*

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. The TCL R&P template is the chosen benchmark reference for performance assessment, so this requires the collection of Technical Control Grades. This also requires that each song has customized assessment policy depending on the Technical Control parameters as summarized in table X.Y in chapter 3. Even though these T.F teacher grades were collected, without knowledge to direct mapping to an extracted audio feature that we can offer some suggestions. But first it is important to understand its meaning in the musical context of a particular song. In Just Looking the score shows that the emphasis is off the beat in the chorus, i.e. o the “and” beat when counting 1+ 2+ 3+4.



*Figure 22: 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 field int eh 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”. For WOTM the syncopation manifests itself inn the audio, with the first bass note occurring half a beat after where it is indicated in the score.  


*Figure 23: Syncopation example: WOTM*

The dot is used to indicate this and it also rendered to have half the duration of a regular crotchet not only in the ground truth audio but also in the WAV file generated from the XML file in Musecore. So the syncopation is characterised by both the note position and note length both of which are annotated in the ground truth onsets and offsets. There fore the Syncopation grade should correlate with the combination of these two measurements.

So in the end how was it possible to make good use of the Technical Control (e.g. Syncopation or Dynamics) and Sound Quality Information? This question is dealt with on a song by song basis in the Results chapter.

5.3.2 Teacher Comments

In the TCL R&P exam the student performs 3 songs and is graded in the previously mentioned 3 areas (Fluency, TF and Communication). A number between 1 and 8 is given and comments between 10 and 40 words are given. In the thesis experiment, separate comments were allocated for each of the five metrics that are graded: onset, duration, technical focus, sound quality and dynamics. The text in these comments can be classified into three areas:

Technical terms related to rhythm and timing

Positive descriptions/adjectives related to the terms

Negative descriptions/adjectives related to the terms

Appendix B shows the extract from a real Grade 4 and Grade 5 TCL Bass exam that I made. In 2018.

5.4 Outcome and Lessons Learnt

Getting Students  
It wasn’t not possible to find students outside the MTG 90% recordings done by myself and the remaining 10% by researchers at the MTG. Some students had issues with music notation literacy and therefore were not comfortable in having to strictly follow the score and technical control details as laid out. There was as preference for adding ones own interpretation which is outside the scope of R & P Syllabus.

Portal Complexity

The portal had a lot of instructions, especially around the latency test and entering details on Sound Card used, which was off putting to students expecting an easier just “plug in and play approach”. This raises the question as to whether a wider survey should be carried out in the diverse bass players community about music literacy, knowledge of assessment apps, knowledge of sound cards, but how would you identify a bassist community. One approach would be to collect data on all registered students in third level academies and in the private music academies, but this would exclude a huge number of bassists who learn in informal contexts.

In the end the Student Portal was not used at all. Since practically all the recordings were done by myself on my laptop or with other researchers using my laptop, there was no real diversity of platforms to consider and the instructions were already known. Nevertheless, it is hoped that the current Student Portal can be resued, improved for further recording campaign in other Thesis projects.

Platforms

Another restriction was that Music Critic is not supported on Tablets and Smart Phones, and one professional bassist reported no having a sound card on his PC, indicating he uses his Tablet for recording.

One precaution taken in the experiment set up was to limit the section of the songs that come under scrutiny for grading and this is shown in table X.Y.Z

Song Length  
However, for convenience the songs were not truncated to reflect this request, in order not to disturb the flow of finishing the section and this caused confusion to some students and in nearly all cases they just kept playing. The fact that it felt natural for the student to continue playing, this in turn prompted the teacher to assume that the full song was to be grades.

One observation was that since I was doing almost all the recordings, the natural tendency was to improve on the previous recording, but this didn’t always happen. Fatigue set in and it was noticed that a fresh one-off performance, but another student yielded the best teacher grade.

There were some grading errors by the teacher, since some Imposter Notes are not penalized in Billie Jean (Student 6)

Teacher: Time constraints 43 out of 48 gradings done. Comments in Spanish.

Follow Up:

Interpolating and creating the ideal deviation scale

Cutting existing recordings and re-grading segments to create more data. (Walking…)

On of the mistakes in providing the student recordings to the teacher was not giving him an unlabeled mix of the original 100% version. This would have helped benchmark the other performances better.

The gradings were requested to be applied to first non-repeating sections of the songs.

6. Results

This chapter presents the most significant observations when doing the analysis.

Using Table X.Y in chapter as a guide, the student recordings were analyzed for their onsets and offset measurements and compared against the ground truth stems. In each case the P,R, F metrics of the onsets shall be taken as well as an analysis of the onset/offset deviations using both the IEC and SOP algorithms

The CSV files were checked for valid offsets: no offset could overtake the next onset. Also the stem waveforms were studied carefully to manually discriminate between a muted and non-muted string. The first annotated rhythm files were prepared using the “inspect and click” method in Sonic Visualizer. The problem with this approach apart from being tedious was that it was difficult to apply a consistent strategy. To improve this the output onsets and offsets of the ground truths were exported to files and then imported into Sonic Visualizer. The False Positive annotations were removed and any missing onsets or offsets were added. This gives us a better PRF starting point for the ground truth so any subsequent student performance becomes more of a similarity check with the Ground Truth. From an objective measurement strategy aspect, this is good news but it holds the assumption that only 1 ground truth exists. For higher grades where more interpretation is allowed, this assumption may not hold.

The software included in this thesis allows you to generate onset /offset analysis with the chosen algorithm and chosen song. It then invites you to run the analysis for the student recording which can generated the output data files with Onset and Offset Deviations, the associated statistics and also P, R, F measures for each student recording. Since the data is very wide it is not possible to display in one table.

In each table the “Student 0” represents the ground truth

The first group of X values for the Student performance of Yellow is as follows:

*Table 10:Accuracy results: Yellow*

|  |  |  |  |
| --- | --- | --- | --- |
| Student | precision | recall | f\_measure\_value |
| 0 | 0.938 | 0.938 | 0.938 |
| 1 | 0.4 | 0.364 | 0.381 |
| 2 | 0.527 | 0.545 | 0.536 |
| 3 | 0.21 | 0.199 | 0.205 |
| 4 | 0.163 | 0.164 | 0.164 |
| 5 | 0.223 | 0.22 | 0.221 |
| 6 | 0.334 | 0.336 | 0.335 |
| 7 | 0.088 | 0.091 | 0.089 |
| 8 | 0.238 | 0.252 | 0.244 |

The second group of X values to be considered are:

with the teacher grades merged. It then Runs a predictor for a given test student are as follows:

*Table 11:Deviation results: Yellow*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Student | Onset ABS  Mean | Onset  Mean | Onset  Std | Duration ABS  Mean | Duration  Mean | Duration  Std |
| 0 | 0.004 | 0.000 | 0.005 | 0.054 | -0.015 | 0.109 |
| 1 | 0.01 | 0 | 0.011 | 0.054 | 0.052 | 0.189 |
| 2 | 0.009 | -0.002 | 0.01 | 0.019 | -0.005 | 0.031 |
| 3 | 0.011 | -0.004 | 0.012 | 0.03 | -0.005 | 0.049 |
| 4 | 0.011 | 0.004 | 0.012 | 0.043 | 0.038 | 0.104 |
| 5 | 0.01 | -0.001 | 0.011 | 0.024 | 0.003 | 0.037 |
| 6 | 0.009 | 0.006 | 0.011 | 0.058 | 0.041 | 0.186 |
| 7 | 0.011 | -0.006 | 0.012 | 0.226 | -0.226 | 0.301 |
| 8 | 0.009 | 0 | 0.01 | 0.167 | -0.151 | 0.379 |

All the above X values are to be considered for correlation with selected Y values from the following table:

*Table 12: Teacher Grades Yellow*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Student | onset Mark | Duration Mark | articulation Mark | soundQuality Mark | volumeControl Mark | final Mark |
| 0 | 100.000 | 100.000 | 100.000 | 100.000 | 100.000 | 5.000 |
| 1 | 85 | 85 | 70 | 85 | 85 | 4 |
| 2 | 55 | 70 | 70 | 70 | 70 | 3 |
| 3 | 70 | 85 | 85 | 85 | 100 | 4 |
| 4 | 70 | 85 | 85 | 85 | 85 | 4 |
| 5 | 85 | 85 | 70 | 85 | 85 | 3 |
| 6 | 85 | 100 | 100 | 85 | 100 | 4.5 |
| 7 | 85 | 85 | 85 | 55 | 70 | 2 |
| 8 | 55 | 70 | 85 | 70 | 85 | 3 |

6.1 Yellow

It has to be remembered that the PRF should give the main measure of onset accuracy.

THE PRF of the stem is high as expected (see table X)

|  |  |  |
| --- | --- | --- |
| 0.938 | 0.938 | 0.938 |

But it doesn’t have the lowest Std Devs of all.

Student 2 has a low PRF but it also has a lower Duration Mean and Duration Absolute Mean are lower than in the stem. The reason this happens is that less onset/offset pairs are considered (through higher missed notes) for calculating the deviations.

After that the deviations that are included depending on how many deviations are included in the measurement window

One of this significant observations in experimenting with this song was to set all notes in the “muted” column in the “”rhythm csv files” to “N”. This means all onset took the next onset

6.2 Billie Jean

The PRF measures were considers with bridge section set to non muted and the results were as follows…

Onset and Duration Histograms

The following are the results of the above algorithms to the Histograms of two sample students performing Billie Jean.



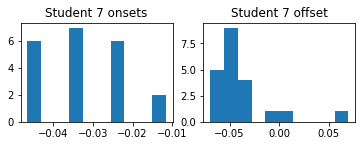
Onset Grade = 85 Duration Grade = 85

Onset ABS Mean: 0.021678,Onset Mean: 0.014059, Dev. from 0: 0.026423

Offset Mean: 0.020862, Dev. from 0: 0.035131

Articulation Grade = 70   
Sound Control Grade = 85   
Volume Control Grade = 85

Final Mark = 4.0



Onset Grade = 85 Duration Grade = 85

Onset ABS Mean: 0.032619,Onset Mean: -0.032619, Dev. from 0: 0.034459

Offset Mean: -0.038147, Dev. from 0: 0.048403

Articulation Grade = 85   
Sound Control Grade = 55   
Volume Control Grade = 85

Final Mark = 2.0

*Figure 24: Billie Jean Deviation Histograms*

The following tables summarise the X measures (PRF, Durations) against the Teacher Onset Grades (for the song Billie Jean)

*Table 13:Billie Jean: Onset accuracy vs Grades*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student | P | R | F | Onset ABS Mean | Onset Mean | Onset Std | Grade | Overall |
| *0* | *0.983* | *0.9965* | *0.9896* |  |  |  | *100* | *5* |
| *1* | *0.381* | *0.399* | *0.39* |  |  |  | *85* | *4* |
| *2* | *0.565* | *0.535* | *0.549* |  |  |  | *55* | *3* |
| *3* | *0.224* | *0.231* | *0.227* |  |  |  | *70* | *4* |
| *4* | *0.195* | *0.185* | *0.19* |  |  |  | *70* | *4* |
| *5* | *0.201* | *0.206* | *0.203* |  |  |  | *85* | *3* |
| *6* | *0.286* | *0.304* | *0.295* |  |  |  | *85* | *4.5* |
| *7* | *0.031* | *0.031* | *0.031* |  |  |  | *85* | *2* |
| *8* | *0.231* | *0.255* | *0.243* |  |  |  | *55* | *3* |

The following are the results of the first train/test Linear regression.

 test\_size=0.3

*Table 14:Billie Jean: Actual vs Predicted Grades*

|  | **Actual** | **Predicted** |
| --- | --- | --- |
| **0** | 85.0 | 51.454149 |
| **1** | 55.0 | 55.239781 |
| **2** | 85.0 | 64.189365 |

Mean Absolute Error: 40.09932968521929

Root Mean Squared Error: 45.26950911827377

6.3 Just Looking

Just Looking had the lowest precision for the Energy Checker. 0.682. This very low precision means this song cannot use Energy Checker and must be confined to using SOP method. It turns that nearly all next onsets can be considered as offsets.

The best student accuracy was a precision of around 40%, suggesting high sensitivity of algorithm or poor student performance sensitivity was very high, even after the truncation of the end of the bridge at 84 seconds.

The modified SOP algorithm is used to calculate this for these Non-Muted notes.

6.4 Brown Eyed Girl

This track had a better onset performance with Energy Checker so we can discard the SOP method as a onset improvement strategy. One way to reduce this could be to partition the song to so that no repeated errors are overly accumulated.

The performances were well distributed.

*Table 15: Brown Eyed Girl: student accuracies*

|  |  |  |  |
| --- | --- | --- | --- |
| Stem | precision | recall | f\_measure\_value |
| 0 | 0.903654 | 0.821752 | 0.860759 |
| 1 | 0.177 | 0.224 | 0.197 |
| 2 | 0.253 | 0.29 | 0.27 |
| 3 | 0.225 | 0.329 | 0.267 |
| 4 | 0.212 | 0.242 | 0.226 |
| 5 | 0.113 | 0.133 | 0.122 |
| 6 | 0.288 | 0.347 | 0.315 |
| 7 | 0.177 | 0.224 | 0.197 |
| 8 | 0.253 | 0.29 | 0.27 |

One major outlier is that the best graded student did not correspond with the best PRF results, and they were very low in comparison with the stem. This highlighted the mistake in not letting the teacher consider the stem as a “blind” student performance.

6.5 Roadrunner

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 Student grades were in a very range, which suggest a high difficulty with the song. The original truncation was set at 82 seconds. This was further truncated to 68 seconds.

Regarding the Tenuto scenario there are two extremes:Max Case: All notes equal , Min Case the third note is absent.

6.6 Walking on the Moon

The PRF before applying this muting classification for WOTM was

0.701 0.932 0.8.

The precision was low for the stem and the range of student precisions was also very low in comparison.

After doing the classification to the bridge section to be Y and the other part N the result PRF were as follows;

An analysis of the text can be performed on the 48 student performances by correlating the comments with the grades given. A classification algorithm can then be applied to help predict a grade with a given set of comments. We don’t aim to generate text for a given audio recording. Personal experience has shown that low grades (e.g. 5/8 for Day Tripper Grade 5 Exam, Appendix C) tend to be a bit longer on words. A good insight into the teacher comments can allow micro adjustments of the final grade that was given.

A new set of plots can be made to then check if this improves the correlation between audio measurements and the grades given.

* 1. Machine Learning to predict grades

Here we present the Machine Learning algorithms used to train the models

Cross Validation ( Leave 1 out)

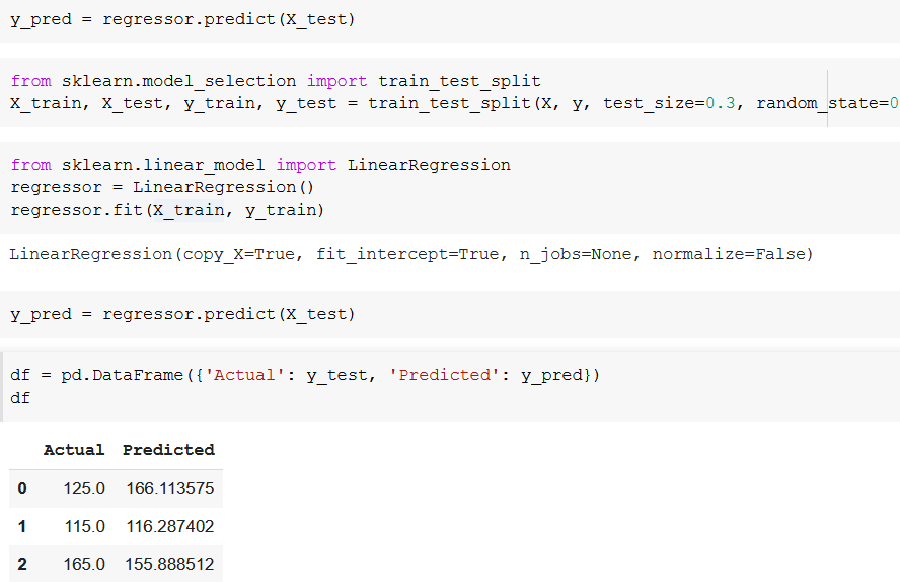
Matrix of different X inputs

1. Linear Regression with multiple variables
2. Classification for handling the text processing of the examiners report.

The different X inputs are the P,R,F measures, Mean-Absolute Error ,Standard Deviation of Onset and Duration Deviations

The Y output is the Final Grade combined with the different grades given for Onset and Duration. There are other Y outputs to consider, and this can be seen in section 4.3

Onsets/Durations vs Student Grades



*Figure 25: Billie Jean Predicted Grades*

1. Conclusion

Despite having the parameter for the minimum window size included, there were some outliers in the duration measurements. There is room for improvement in capturing accurate onset by blending different algorithms together (e.g. using the Essentia algorithm EffectiveDuration() to check if the duration returned by the EnergyChecker algorithm is valid.

We found that the teacher may overlook some “imposter notes” and this causes mismatch between mean absolute error and associated grade. A typical audience member may not detect these errors, but a TCL examiner should.

Another significant Teacher characteristic is their ability to hear the required bass duration in the presence of the mix. The first experiment did not have enough bass boost to help the teacher discriminate and this may help explain some of the outliers. This brings us back to the perception topic by Kopp-Scheinpflug [10], discussed in the State of the Art.

Another weakness in the grading is that the teacher used the “go through the students once”, so grade was given an absolute scale rather than a relative scale. To ameliorate this I added new student recordings and after grading them re-calibrated the previous grades so that they were all aligned on a relative to the ground trutgh grade of 100%.

The method for classifying notes can improve the deviation statistics and it opens the door for more sophisticated methods to consider technical control parameters. The Duration grade only has meaning depending on the context. For example, it is only worth measuring duration in the bridge in Yellow. The repeated quavers in the verse leaves no gaps for duration. Theoretically it would be possible to play shorter notes than quavers, but that is actually more difficult to do, so its not worth checking for a minimum note length. Overall properties such as detecting adequate energy levels after a certain time for long notes, is more effective than trying to accurately determine the exact offset point. The testing of the algorithms with of section of the songs taken from midi translations of score highlight the importance of considering articulation as a parameter to be grades.

I made a second iteration of the g

Summary of remaining actions

* Re-check the prediction accuracy by after adjusting final grade from comment analysis
* Improve accuracy of the algorithms with hybrid methods and “next onset” approach
* Perform annotation for the remaining 4 songs (Yellow, Just Looking, Brown Eyed Girl, Roadrunner)
* Record more with deliberate errors (like in the Police song) for better curve alignment
* Perform the grades objectively against the actual score and mark down missing or extra notes
* Perform grading only on Onset and Offset and give a mark between 1-8 and just these two metrics and make song sections shorter.

The last action is the key to opening the door to getting more recordings graded.

Suggested future paths for scaling up the experiment as a follow on to the thesis.

* A pilot project that would involve a selection of up to 20 students who are studying bass guitar in private schools and conservatories.
* A custom portal that would allow them
* Continually annotating recorded data for future training
* Source Separation to obtain more annotated data.

A per-song solution has been found as a work around to the limits encountered in overlapping onsets, offsets. For “Yellow, a song with eight nots ats xxx bpm

* 1. Suggested Improvements

Adding a Tenuto marking in the Rhythm Files.

1. List of figures

(Optional)

[*Figure 1: 4 crotchets, 4 quavers and 4 semi-quavers.* 17](#_Toc77670450)

[*Figure 2: dynamic threshold setting* 21](#_Toc77670451)

[*Figure 3: Extract from IDMT Dataset. File 002.xml* 25](#_Toc77670452)

[*Figure 4: Note duration of Staccato notes* 26](#_Toc77670453)

[*Figure 5: Audio 002.wav (IDMT DATASET) with onsets* 27](#_Toc77670454)

[*Figure 6: Billie Jean: Midi Rendering vs Human recording* 29](#_Toc77670455)

[*Figure 7: Normalized Energy of Audio 002.wav sample (IDMT DATASET)* 33](#_Toc77670456)

[*Figure 8: Sound Islands of Audio 002.wav sample (IDMT DATASET)* 34](#_Toc77670457)

[*Figure 9: Calculated IEC onsets for WOTM verse* 35](#_Toc77670458)

[*Figure 10: Calculated IEC Sound Islands for WOTM bridge* 36](#_Toc77670459)

[*Figure 11: PitchMelodia derived Onsets for WOTM verse* 36](#_Toc77670460)

[*Figure 12: PitchMelodia derived Onsets for WOTM bridge* 37](#_Toc77670461)

[*Figure 13: Billie Jean GT onsets deviation* 38](#_Toc77670462)

[*Figure 14: Billie Jean GT offsets deviation* 39](#_Toc77670463)

[*Figure 15: End to End test of Music Critic with Billie Jean GT* 40](#_Toc77670464)

[*Figure 16: Workflow on Student Portal* 42](#_Toc77670465)

[*Figure 17: Latency for a given submission* 42](#_Toc77670466)

[*Figure 18: Technical instruction and link to recording portal* 43](#_Toc77670467)

[*Figure 19: Grading Histogram Onsets: Billie Jean* 45](#_Toc77670468)

[*Figure 20: Grading Histogram Offsets: Billie Jean* 46](#_Toc77670469)

[*Figure 21: Audio extracts of very different WOTM student performances* 46](#_Toc77670470)

[*Figure 22: Syncopation example Just Looking* 47](#_Toc77670471)

[*Figure 23: Syncopation example: WOTM* 47](#_Toc77670472)

[*Figure 24: Billie Jean Deviation Histograms* 53](#_Toc77670473)

[*Figure 25: Billie Jean Predicted Grades* 56](#_Toc77670474)

1. List of tables

(Optional)

[*Table 1: Abessers Data Driven algorithms vs Salamon/Gomez* 18](#_Toc77670486)

[*Table 2: Abesser Signal Processing vs Salamon/Gomez* 19](#_Toc77670487)

[*Table 3: State of the Art algorithms for Onset/Offset detection* 20](#_Toc77670488)

[*Table 4: Six chosen songs from TCL song list* 27](#_Toc77670489)

[*Table 5: Tech. Control Parameters for 6 TCL songs* 30](#_Toc77670490)

[*Table 6: Benchmark results of Algorithms* 33](#_Toc77670491)

[The resulting PRF for onsets for each of the Trinity RnP ideal bass recordings (stems) with a 20ms evaluation window compare both IndexedEnergyChecker with SOP *Table 7: Accuracy Metrics for 6 TCL ground truths* 34](#_Toc77670492)

[*Table 8: Selected comparison: IEC vs SG* 35](#_Toc77670493)

[*Table 9:Duration of 6 TCL song extracts* 44](#_Toc77670494)

[*Table 10:Accuracy results: Yellow* 50](#_Toc77670495)

[*Table 11:Deviation results: Yellow* 51](#_Toc77670496)

[*Table 12: Teacher Grades Yellow* 51](#_Toc77670497)

[*Table 13:Billie Jean: Onset accuracy vs Grades* 53](#_Toc77670498)

[*Table 14:Billie Jean: Actual vs Predicted Grades* 53](#_Toc77670499)

[*Table 15: Brown Eyed Girl: student accuracies* 54](#_Toc77670500)

1. List of symbols
2. Bibliography

1. Eremenko V, Morsi A, Narang J, Serra X. Performance assessment technologies for the support of musical instrument learning. Paper presented at: CSEDU 2020 The 12th International Conference on Computer Supported Education; (2020)

https://musiccritic.upf.edu/ (2021)

OK

2. McCormick, Neil (2009). U2 by U2: Bono, The Edge, Adam Clayton, Larry Mullen Jr (1st It Books ed.). New York: HarperCollins. ISBN 978-0-06-190385-4. OCLC 313657117.

ok

3. Trinity Rock and Pop Bass Syllabus. Qualification specifications for graded exams from 2018. https://www.trinityrock.com/instruments/bass (2018)

OK

4. Trinity Rock and Pop Bass Grades 0,1,2,3. Contemporary & classic songs for Trinity College London Rock & Pop exams from 2018..Trinity College London Press.

ok

5. Jazz Bass Transcription Using a U-Net Architecture by Jakob Abeßer and Meinard Müller.Electronics 2021, 10(6), 670   
https://doi.org/10.3390/electronics 10060670

ok

6. J Abeßer, G Schuller. Instrument-centered music transcription of solo bass guitar recordings IEEE/ACM Transactions on Audio, Speech, and Language Processing 25 (9), 1741. DOI: 10.1109/TASLP.2017.2702384 (2017)

ok

7. Jakob Abeßer; Hanna Lukashevich; Gerald Schuller, Feature-based extraction of plucking and expression styles of the electric bass guitar, IEEE International Conference on Acoustics, Speech and Signal Processing, DOI: 10.1109/ICASSP.2010.5495945 (2010)

ok

8. Christian Dittmar, Estefanía Cano, Jakob Abeßer, and Sascha Grollmisch, Music Information Retrieval Meets Music Education, DOI: 10.4230/DFU.Vol3.11041.95 (2012)

ok

9. Abesser, Jakob; Schuller, Gerald. Instrument-Centered Music Transcription of Bass Guitar Tracks, 53rd AES International Conference: Semantic Audio (2014)

ok

10. Prof. Dr.-Ing. Gerald Schuller, Prof. Dr. Meinard Müller, Dr. Tech. Anssi Klapuri, Automatic Transcription of Bass Guitar Tracks applied for Music Genre Classification and Sound Synthesis, urn:nbn:de:gbv:ilm1-2014000294, Dissertation, Technischen Universität Ilmenau.(2014)

ok

11. Phase-based note onset detection for music signals, Juan Pablo Bello, Mark Brian Sandler, Conference Paper in Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on · November 2003 DOI: 10.1109/ASPAA.2003.1285811 · Source: IEEE Xplore

ok

14. On the use of phase and energy for musical onset detection in the complex domain. J.P. Bello; C. Duxbury; M. Davies; M. Sandler, Published in: IEEE Signal Processing Letters ( Volume: 11, Issue: 6, June 2004) Page(s): 553 - 556, Date of Publication: 24 May 2004,Print ISSN: 1070-9908, Electronic ISSN: 1558-2361, INSPEC Accession Number: 7972894 DOI: 10.1109/LSP.2004.827951

ok

12. Conny Kopp-Scheinpflug, James L. Sinclair and Jennifer F. Linden., When Sound Stops: Offset Responses in the Auditory System, Trends in Neurosciences, Special Issue: Time in the Brain (2018)

ok

13. J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 6, pp. 1759–1770, 2012

[ok]

14 Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, Senior Member, IEEE A Tutorial on Onset Detection in Music Signals, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 1063-6676 (2005).

15 Reboursi`ere, L., L¨ahdeoja, O., Drugman, T., Dupont, S., Picard,

C., & Riche, N. (2012). Left and right-hand

guitar playing techniques detection, In Proc. of

the Int. Conf. on New Interfaces for Musical Expression

(pp. 1–4).

Company Sites

Coursera Sites

Github links

11. Youscian, Oy (2021), Company that provides application for feedback on playing bass. https://yousician.com/

12. Today’s Music Industry, Darko Velichkovski, West Virginia University.  
https://www.coursera.org/learn/music-industry.

13. Scotts Bass Lessons, https://scottsbasslessons.com/ (2021)

16. https://github.com/RamoonRoomeu/AutomaticAssessmentSax

Homeless

On the Use of Phase and Energy for Musical Onset

Detection in the Complex Domain

Juan P. Bello, Chris Duxbury, Mike Davies, and Mark Sandler*, Senior Member, IEEE*

Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, Senior Member, IEEE A Tutorial on Onset Detection in Music Signals, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 1063-6676 (2005).

APPENDIX A Student Portal

A.1 Recommendations and Instructions for recording.

Please follow recommendations:

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

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.

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."

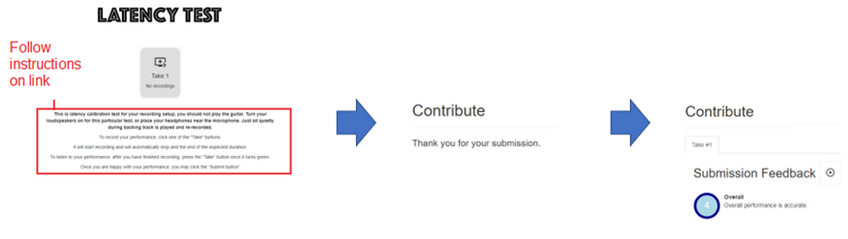
8. Record and submit another click track recording. Below is an illustration of workflow.

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 it turns green.  
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.

A.2 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.

B Teacher Portal

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 stylisitic 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 (ocassion 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.

Appendix C Note books for Plots

Appendix D: Code for generating data

1. https://yousician.com/ [↑](#footnote-ref-1)
2. www.yousician.com, www.fretello.com, www.songs2see.com [↑](#endnote-ref-1)
3. https://github.com/CPJKU/madmom [↑](#footnote-ref-2)
4. https://essentia.upf.edu/ [↑](#footnote-ref-3)
5. https://essentia.upf.edu/reference/std\_PitchMelodia.html [↑](#footnote-ref-4)
6. <https://github.com/jakobabesser/bassunet> [↑](#footnote-ref-5)
7. https://github.com/jakobabesser/pymus/tree/master/pymus/sisa/f0\_tracking [↑](#footnote-ref-6)
8. https://github.com/CPJKU/madmom [↑](#footnote-ref-7)
9. https://github.com/CPJKU/madmom/blob/master/bin/OnsetDetectorLL [↑](#footnote-ref-8)
10. https://github.com/MTG/pysimmusic-experiments [↑](#footnote-ref-9)
11. https://github.com/RamoonRoomeu/ToneExperiments [↑](#footnote-ref-10)
12. https://musiccritic.upf.edu/ [↑](#footnote-ref-11)
13. The student tries to give best recording possible, improving on each take with no deliberate mistakes. [↑](#footnote-ref-12)