Master thesis on Sound and Music Computing

Universitat Pompeu Fabra

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

Colm Forkin

Supervisor: Vsevolod Eremenko

Co-supervisor: Xavier Serra

Aug 2021



Master thesis on Sound and Music Computing

Universitat Pompeu Fabra

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

Colm Forkin

Supervisor: Vsevolod Eremenko

Co-supervisor: Xavier Serra

Aug 2021



Table of Contents

[1. Introduction 1](#_Toc78887551)

[2. State of the Art 3](#_Toc78887552)

[2.1. Music Education 3](#_Toc78887553)

[2.2 MIR approaches to onset/offset detection 3](#_Toc78887554)

[2.3 Measuring duration 5](#_Toc78887555)

[2.4 Abessers Research 6](#_Toc78887556)

[2.4 Literature Research Summary 7](#_Toc78887557)

[2.5 Revision of Algorithm Evaluation for Music Pedagogical Purposes 10](#_Toc78887558)

[2.6 Perception of Onsets and Offsets 12](#_Toc78887559)

[2.7 Observations 12](#_Toc78887560)

[3. Datasets 14](#_Toc78887561)

[3.1. IDMT BASS SINGLE TRACK 14](#_Toc78887562)

[3.2. TCL Dataset 16](#_Toc78887563)

[4. Methodology 22](#_Toc78887564)

[4.1 Algorithm Evaluation 23](#_Toc78887565)

[4.2 Indexed Energy Checker 24](#_Toc78887566)

[4.3 Accuracy Metrics 26](#_Toc78887567)

[4.4 Deviation Metrics 30](#_Toc78887568)

[4.5 Correlation with Student Grades Metrics 32](#_Toc78887569)

[5. Experiment 34](#_Toc78887570)

[5.1. Pysimmusic End-to -End test 34](#_Toc78887571)

[5.2. Student Recording Portal 35](#_Toc78887572)

[5.2.1. Latency Test 35](#_Toc78887573)

[5.3. Grading 41](#_Toc78887574)

[5.4. Teacher Comments 45](#_Toc78887575)

[5.5. Outcomes and Lessons Learnt 45](#_Toc78887576)

[6. Results 48](#_Toc78887577)

[6.1. Yellow 49](#_Toc78887578)

[6.2. Billie Jean 51](#_Toc78887579)

[6.3. Just Looking 53](#_Toc78887580)

[6.4. Brown Eyed Girl 54](#_Toc78887581)

[6.5. Roadrunner 57](#_Toc78887582)

[6.6. Walking on the Moon 58](#_Toc78887583)

[6.7. Technical Control Grades 59](#_Toc78887584)

[6.8. Teacher Comments 60](#_Toc78887585)

[7. Conclusion 62](#_Toc78887586)

[8. List of figures 66](#_Toc78887587)

[9. List of tables 67](#_Toc78887588)

[10. List of symbols 68](#_Toc78887589)

[11. Bibliography 69](#_Toc78887590)

[12. Appendices 70](#_Toc78887591)

[12.1. Appendix 1 70](#_Toc78887592)

[12.2. Appendix 2 73](#_Toc78887593)

Dedication

I would like to thank those near to me personal that made this whole Masters possible, my family in Barcelona. One of the music biographies, I read around 2016, U2 on U2 [2] was the beginning of a long journey which led me to choosing to do the SMC Masters full time. I was fascinated about the musical journey their careers took and hwo the producers played a big role. Fast forward to November 2020 and my thesis topic was chosen, I was know faced with the burden of getting this one chance to really unlock what makes up a great bass sound. In terms of heavy project workload, writing a Thesis is like recording an album. But I felt fortunate, I had Vsevelod and Xavier as supervisors in the same way Adam Clayton felt fortunate in having Brian Eno and Danny Lanois as producers for U2.

Along the way there were a lot of challenging assignments and the one that definitely stick out was the MPC exercise for P. Herrero. I learnt a lot about music and emotions and I found myself getting better at choosing music to listen to help my mood at different occasions ( I really hope that nobody noticed I had Phil Collins on Spotify when I was preparing my presentation).

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

Acknowledgments

(Optional, if used placed on a right page next to an empty left page)

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

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.

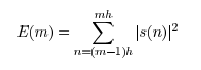
* 1. Music Education

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 scenario which can be a final exam, competitions or audition. This thesis uses the final exam approach with the Trinity College London (TCL) Rock and Pop from Initial level to Grade 3 [4] as the quality standard. The Teacher Grading for the experiment is based on the score ranges used by the Trinity Examiners: Distinction 87-100; Merit 75-86; Pass 60-74 and below pass.

* 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.*

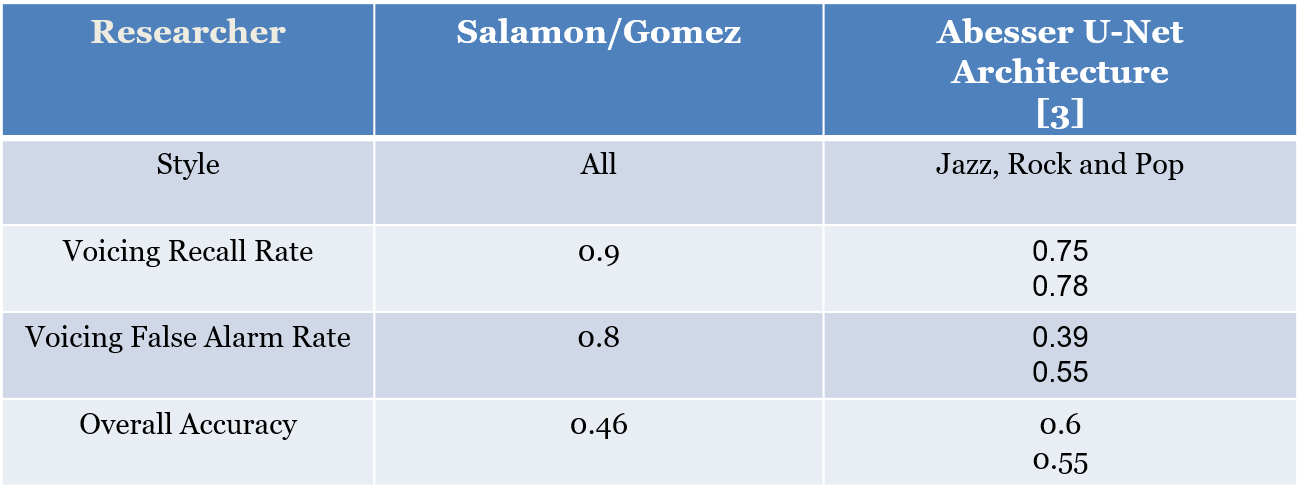
This shortcoming is really surprising when you consider that that Yousician is capable of correctly assessing left hand techniques covered in some of Reboursieres paper [15] such as slides, Hammer-ons and Pull-offs. The challenge of durations measurement is that it is not an instantaneous measurement. 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

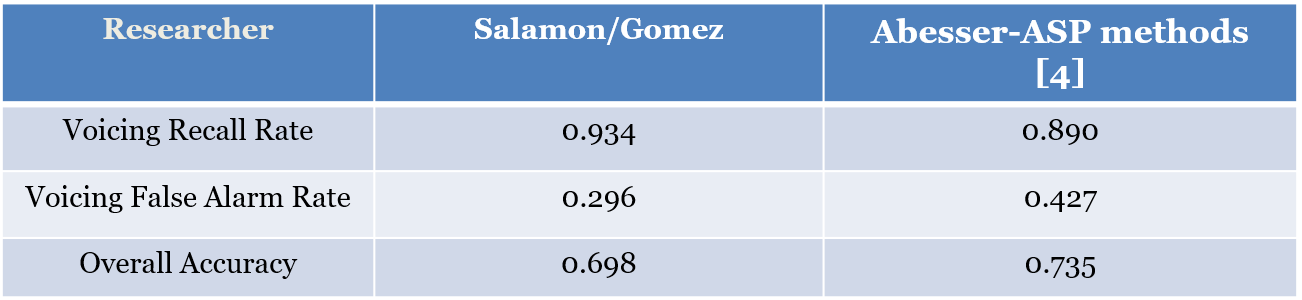
Jacob Abessers applied research in music information retrieval & machine learning / deep learning and audio signal processing is focussed primarily on bass. For timing measurements, the voicing classification (independent of which pitch it is ) is an important metric. The Salamon and E. Gómez [13] techniques 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.

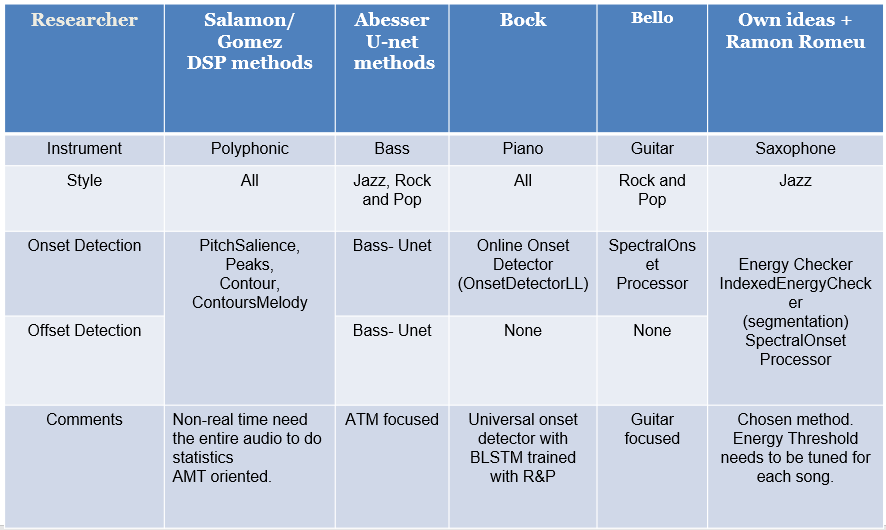
*Table 2: Abesser Signal Processing vs Salamon/Gomez*



* 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 summarizes 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. PitchMelodia consists of 4 algorithms that are called in a chain: PitchSalienceFunction, PitchSalienceFunctionPeaks, PitchContours and PitchContoursMelody. Using these library functions for testsing offers more configuration flexibility than the combined PitchMelody function and that was the approach used in this Thesis

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

The peak picking strategy used here is the equivalent of the one contained in the S/G Essentia algorithm PitchSalienceFunctionPeaks. Ramon Romeu[[11]](#footnote-10) developed a wrapper based on the above 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. A more rigorous approach to measuring onsets would distinguish between perceived and actual onsets, but this distinction is not considered in this Thesis.

* 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 [3] 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. Support vector machines to perform not only score transcription for bass [6] but also the extraction of the plucking and expressive styles. In this AMT-domain cases the Dataset consisted of small single note extracts for the style and short 15 sec tracks for the score. Music Education requires larger audio segments and more widely ranging quality performances.

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 two Datasets that are used to evaluate algorithm accuracy

* 1. IDMT BASS SINGLE TRACK

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.

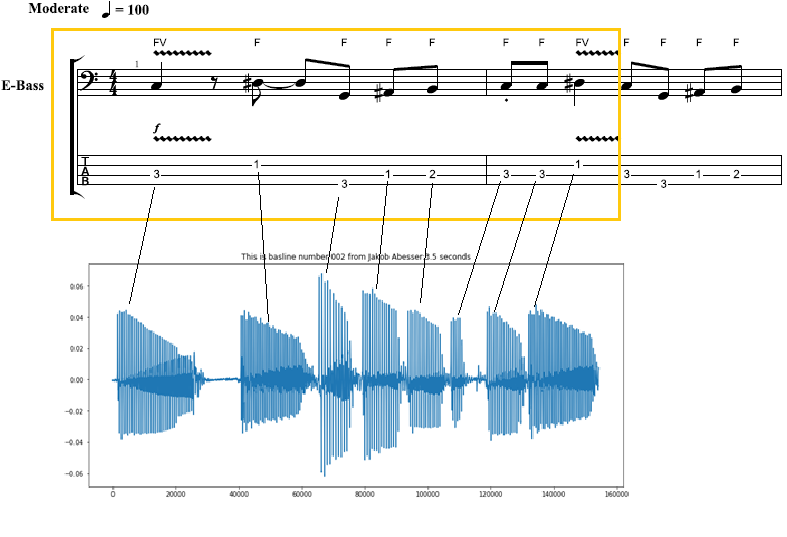
IDMT-Summary

IDMT (Fraunhofer) : 17 tracks

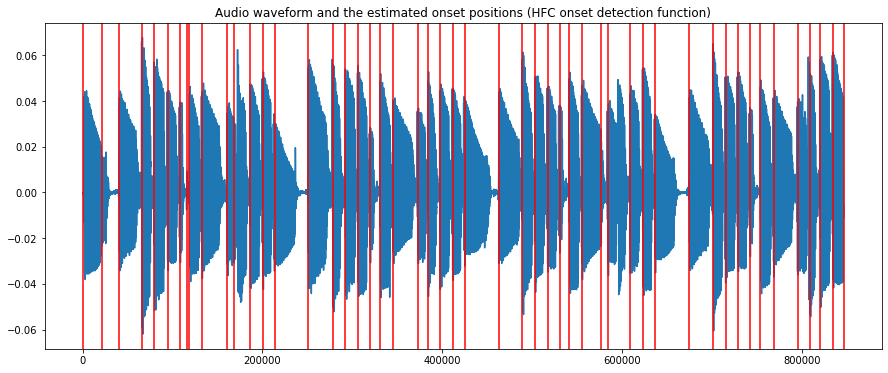
* Varying plucking techniques (Pick, Finger, Muted, Slap)
* PDFs of score and XML of parameters (onsets, offsets, pitch, fret number)
* Expression Style annotated
* 4 tracks within musical and timbral complexity of Trinity G0-3

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

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. It was necessary to make these manually using Sonic Visualizer.

Trinity College London published specifications how the grading for the on the 2018 Rock and Pop Bass syllabus is graded[3]. Originally twelve songs were identified to cover as wide as possible, the syllabus topics that are important in grading. Finally Six songs have been chosen from the Grade books [4], ranging from grade 0 to grade 3. The recordings are cover versions performed by TCL professional musicians.

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

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

As part of a recent collaboration between the Music Technology Group (MTG) and Trinity College London (TCL), the separate stems of these recordings have been made available for the analysis purposes. The TCL recorded bass stem is to serve as the Ground Truth. The remaining stems are used to build up “minus 1” tracks for mixing the with the student performances. Trinity have also made available the Scores and Grade book data in the form of PDF publications and XML files (for loading into Musescore). The description of how each song is grades is given before the Tablature is shown, and this parts is key part of the dataset as it describes how the song is graded.

Throughout the Thesis a Dataset was built up consisting of the following:

* Onset/offset annotations of ground truth (bass stems of TCL recordings)
* Multiple Student recordings of each of the six songs
* Mixes of the Student Recordings with the Minus 1 tracks
* Music Teacher Numeric and Descriptive grades allocated to each Mix.
* Algorithm generated data on the Student and Ground Truth stems.

For the purposes of brevity, the term “TCL Dataset” is used to describe this 6 song dataset. In summary the objective of the Thesis is to generate full generated all the metadeta and result data for this dataset and to discuss those results and their implications related to timing and rhythm.

Duration plays an important role in 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*.” Table 4, list the songs chosen for analysis in this Thesis and the shortened name/acronym to be used hereafter in the thesis.

The PDF files of the Grade books [4] containing the 6 songs are accompanied with XML files which provides rich information on the score note duration . There is also technical information related to the fret position 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>

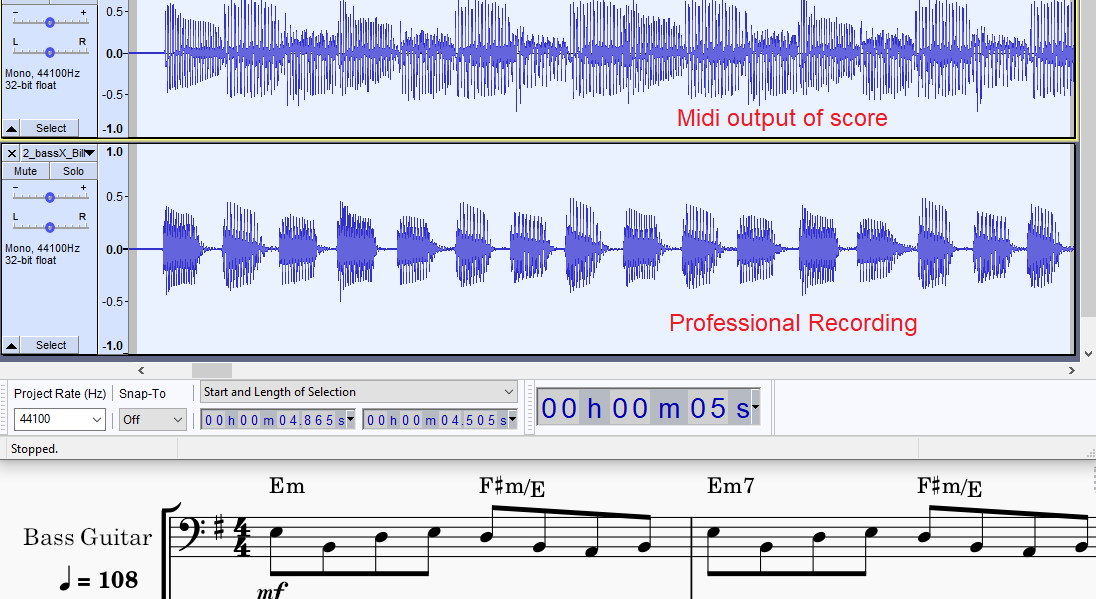
Annotating the Onsets and Offsets

An iterative approach was taken into defining the onsets and offsets. The first pass, was done by visual inspection in Sonic Visualizer. The problem with this approach is that ones eyes are not reliable enough to apply a consistent rule throughout. The second pass was to run the best performing onset and offset algorithms to the Stems and then remove the false positives and add the visually add the missing notes. This resulted in a much more consistent set of annotations, so the mean and standard deviations of the onset/offset deviations of ground truthswere close to zero.

It is important to remember that the Ground Truths Annotations were updated to be aligned with the measured onset/offsets of the Trinity Recordings

From an objective measurement strategy aspect, this is good design and it holds the rule that only one ground truth exists, a reasonable assumption for Rock and Pop but would not be valid for the varying expressivity styles of Jazz or perhaps Trinity R&P grades higher than 6 where more interpretation is allowed.

The XML files can be viewed in Musescore and by removing the chord information and exporting of a WAV file, a “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. It also indicates that there is information provided in the Song Description which is not reflected in the actual score. For “Bjean” it is required to pluck with short/jerky movement in the verse/chorus, but no dotted notation is used to reflect this.



*Figure 6: Billie Jean: Midi Rendering vs Human recording (shorter notes)*

The first function of the TCL dataset is to validate the algorithms short listed state of the art algorithms after testing with the IDMT dataset. The second function is to provide a basis to measure student performances.

The TCL professional or “ground truth” stem is taken to represent a grade of 100% in timing and rhythm.

The following table summarizes the individual musical features of each of the tracks a 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.

Table 5 summarizes the Technical Control parameters of the 6 chosen audio tracks. This table lays the basis for deciding on how to phrase the questions on the Grading Sheet given to the Bass teacher and also the description of the quality levels in each of the Multiple Choice answers.

*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) |
| bjean | intro | Accent just before chorus |  | -- | separate jerky quavers+ smooth, melodic material |  |
| just |  | Chorus: accented, syncopated motif.,  hard accent |  | Unexpected  subito p at bar 25. |  |  |
| brown |  |  |  |  |  | different note lengths and rests |
| road |  |  |  |  | Tenutu (underscore)  loud on beat 1 |  |
| wotm |  | syncopated  repeated notes |  |  |  | correct  separation |

In the experiment, truncation is applied to the songs, to facilitate grading a section of music once, for example “yellow” was truncated to remove the repeated verse, since no new musical features were introduced. The song “wotm” has been cut to only include the first 50%. The “wotm” song is symmetrical, the second half is a repeat of the first half. This opens the possibility to split the data and create more performances out of the student recpordings, however for this Tehsis this option was not performed. Technically speaking the second half of the song “wotm” has an ad-lib section on the bridge, but this wasn’t not observed by any student.

The WOTM verse has a long note duration with zero inter-note 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

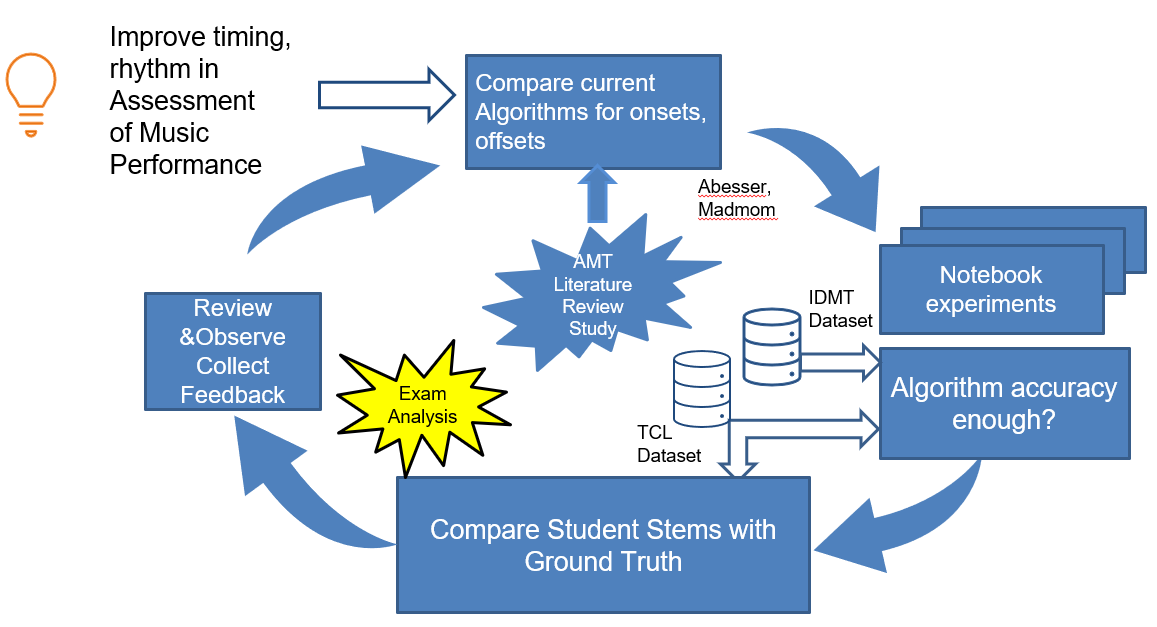
1. Methodology

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.[ 17] 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. Abesser used the standard MIR evaluation of 50ms for measuring onset accuracy in Guitar onsets and increased this to 200 ms for offsets due to difficulty in handling smoothly decreasing note envelopes.[17]

* Benchmark algorithms for accuracy
* Testing SOTA algorithms to measure student performances
* Using Multi-variable Linear Regression to train new a model with teacher graded student performances
* Analysis of teacher comments in student performances and correlation with student grades

There was no “straight run through” to get the best algorithm to grade all the Trinity songs equally well. A few iterations were required that involved modifying the algorithms to consider plucking style, developing a hybrid algorithm that would apply different threshold parameters to different songs and different offset strategies to different sections of a given song.

Even getting good PRF metrics for the professionally recorded ground truth it was a difficult task to produce student recordings that would return a precision value higher than 60%. Collecting sound Quality grades taken by the teacher were useful in annotating these recording imperfections.



* 1. Algorithm Evaluation

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

The 4 methods introduced in table 3 chapter 2 were 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)
* EnergyChecker and IndexedEnergyChecker (Paired ASP driven) +   
  SOP (Non Paired) ASP Driven

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.

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 | Energy Checker | Indexed Energy Checker |
| P |  |  |  |  |  |  |
| R |  |  |  |  |  |  |
| F |  |  |  |  |  |  |

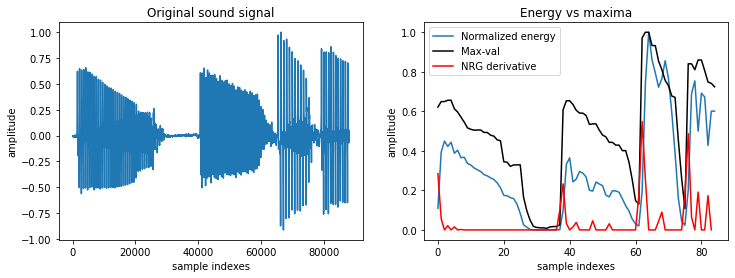
* 1. Indexed Energy Checker

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

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

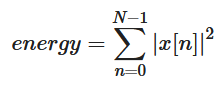
The main problem with this algorithm was that it did not calculate corresponding onsets within the same frame. The Onsets were calculated separately using standard HFC method, but these values were independent of the offset values. Too much effort would be needed to align them and fix this “un-paired”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

The IndexedEnergyChecker algorithm pairs the onsets and offsets, based on an Energy Threshold. It is based on the concept of a sound island. Depending on the energy threshold level set, it decides on how to split the wave boundaries. This improvement is based on code developed on this by Ramon Romeu []. His algorithm returns a set of start and stop indices representing onset and offset has a hard coded energy threshold based on experiments with the Saxophone. I parameterised this value for different bass stems songs.



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

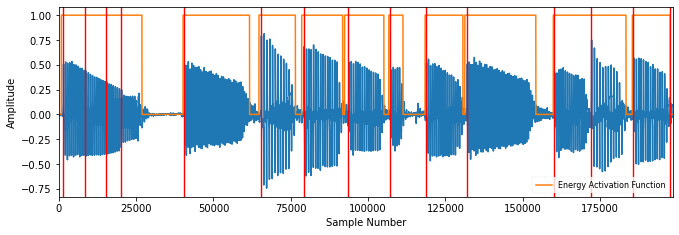
Energy is calculated using the normalise energy function



Equation 3 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

In addition to this function, an additional method based on the madmom SpectralOnsetProcessor [] can be used as a complimentary non-paired algorithm. This algorithm will be useful in scenarios where offset align with the next onset, hence avoiding the “unpaired” problem-.

* 1. Accuracy Metrics

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

*Table* *7: Accuracy Metrics for 6 TCL ground truths (combined vs non combined)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Song | P | | R | | F | |
| IEC | SOP | IEC | SOP | IEC | SOP |
| Yellow | 0.939 | 0.98 | 0.942 | 1.0 | 0.94 | 0.99 |
| **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 8: Accuracy Metrics for Billie Jean Hybrid Algorithm

|  |  |  |  |
| --- | --- | --- | --- |
| Song | P | R | F |
| Hybrid | Hybrid | Hybrid |
| **Billie Jean** | 0.9893 | 0.97555 | 0.9824 |

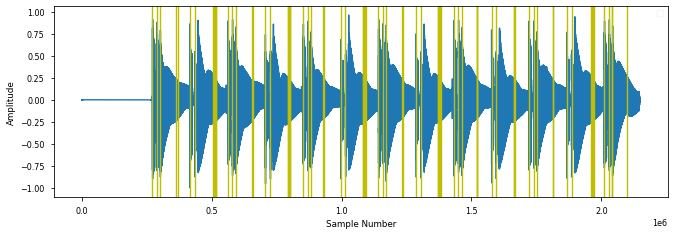
The worst performing EnergyChecker results were then benchmarked against the Salamon /Gomez results

*Table 9: 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

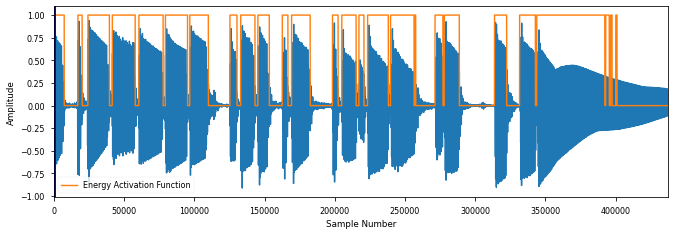


*Figure 9: Calculated IEC onsets for WOTM verse*

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

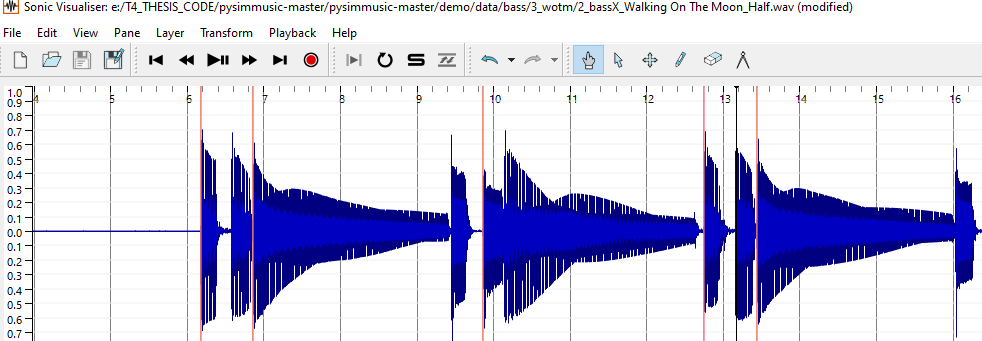
0.937 0.949 0.943 (with last note)

0.961 0.948 0.954 (Without last note)

****

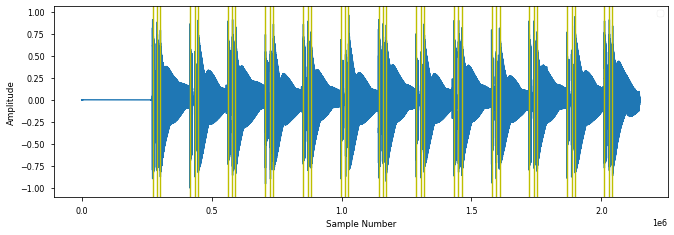
*Figure 10: Calculated IEC Sound Islands for WOTM bridge*

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 Figure 10. The IEC algorithm does not work well in the verse. A lot of onsets are missed



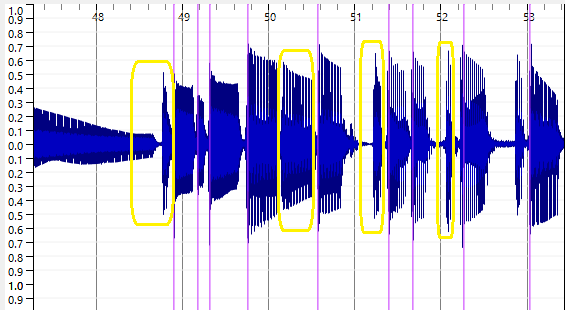
*Figure 11: Figure 12: Calculated IEC Onset for WOTM verse*

In contrast the Salamon/Gomez algorithm performed a lot better in the “wotm” verse with PRF measures of 0.974 1.0 0.987



*Figure 13: PitchMelodia derived Onsets for WOTM verse*

But did not work very well int eh bridge section of “wortm”



*Figure 14: 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 12 marked in yellow

It is important to clarify some local definition of “mute”, that it does not relate to the concept of the left hand muting technique described by Abesser [17]. 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.

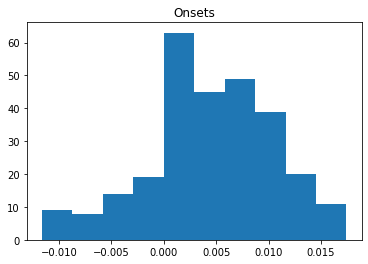
* 1. Deviation Metrics

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.

*Figure 15: IDMT Dataset GT Onset deviations ( SOP ) algorithm*

*Figure 16: IDMT Dataset GT Onset deviations ( IEC ) algorithm*

*Figure 17: IDMT Dataset GT Offset deviations ( IEC ) algorithm*

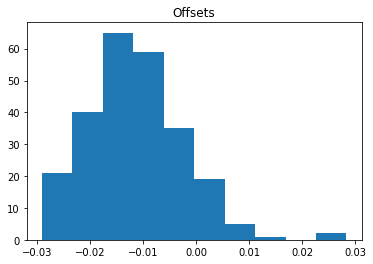


*Figure 18: Billie Jean GT onsets deviation (? Algorithm)*

Window 20ms

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

misses 9 Percentage Miss 3.147



*Figure 19: Billie Jean GT offsets deviation (? Algorithm)*

misses 39 Percentage Miss 13.636363636363637

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

The same methods are used to assess the student performances.

* 1. Correlation with Student Grades Metrics

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

1. Experiment

In this chapter, two 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.

Regarding the second experiment, it is worth recalling that in Chapter 2 the TCL was presented as incomplete at the start of the Thesis. Apart from the missing onset/offset annotations, it was necessary gather and process the student recordings for the six chosen TCL songs. Section 5.2 describes how the Stems were prepared for two purposes: (i) as synchronized stems ready for performance assessment using algorithms (ii) as inputs to mixing with the other stems so the Teacher could grade them.

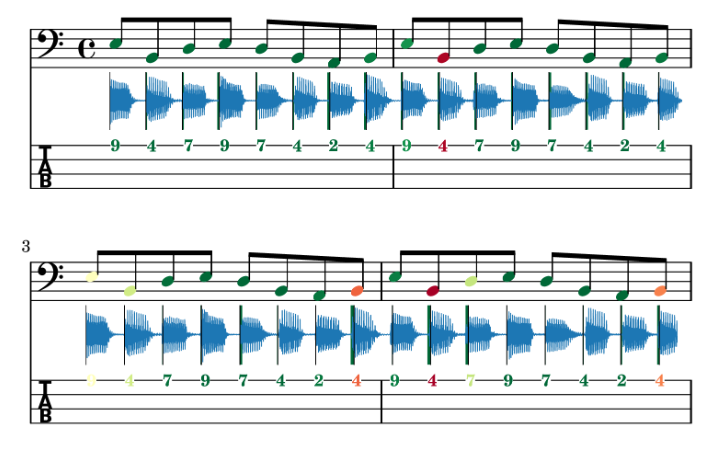
* 1. Pysimmusic End-to -End test

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 20: 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.

* 1. Student Recording Portal

The second part of building the MTG Bass TCL 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. The Student presses the play button, listens and plays along. At the end they have the option to listen back on the stem and press “Submit”. Details of this Portal can be found in the appendix. Good headphones, that are capable of isolating the backing track from the microphone are required. A direct audio interface for the Bass is preferred instead of relying on positioning the microphone close to the bass amp.

* + 1. Latency Test

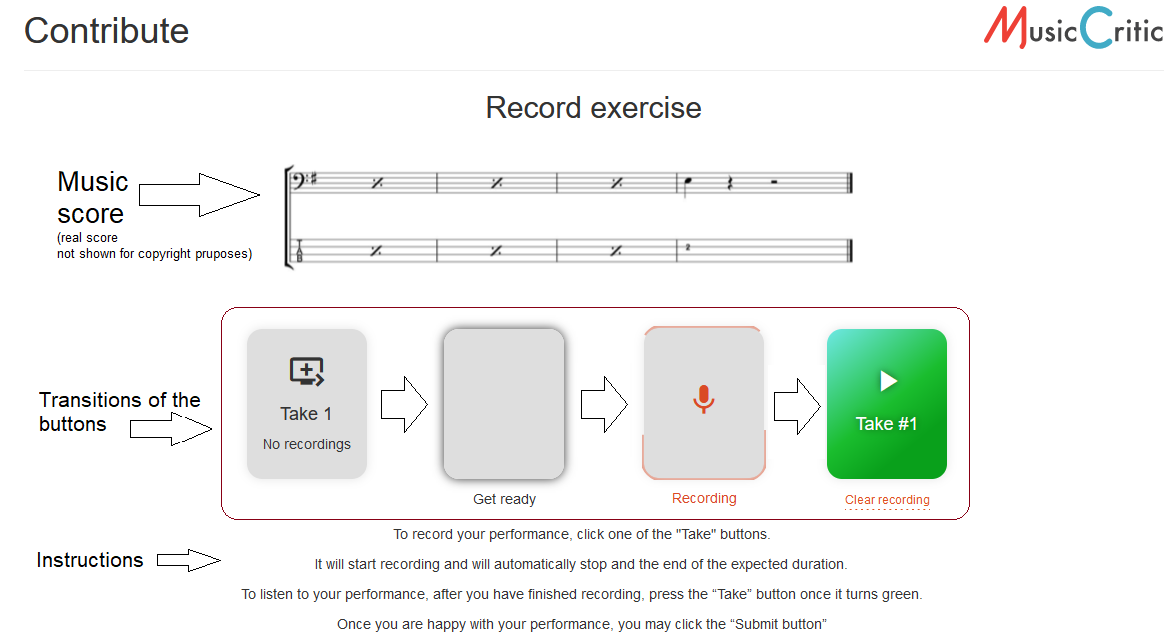
A pre-requisite before recording an to a backing track is to do the Latency Test. This involves placing the microphone close to the headphone speaker while recording the Click track. This allows to calculate a latency value which is then stored in the submissions.json file.

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 for microphone input for latency test.

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 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 21: Workflow on Student Portal*

After removing the latency from the recordings, it was noticed that the json calculated 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. The steps involved in going from having a bass stem recording on a server, to a clean bass stem ready for analysis and a clean mix ready for grading are quite elaborate.

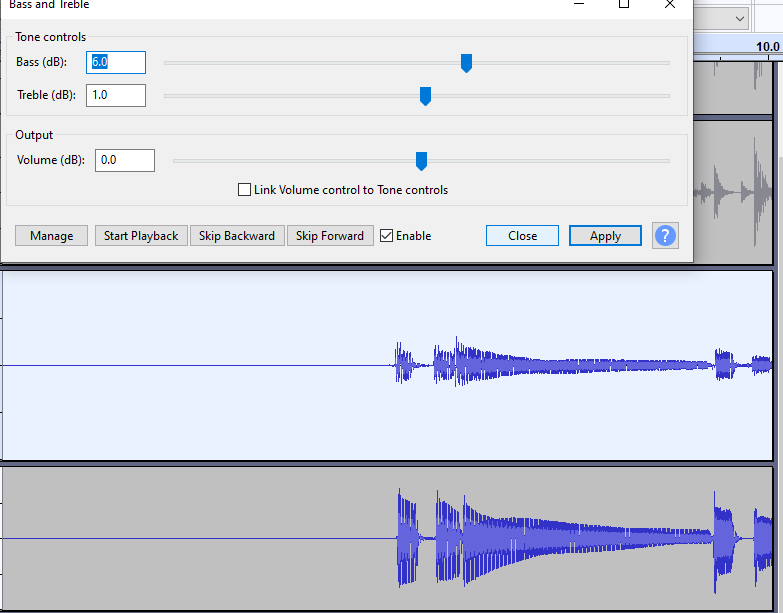
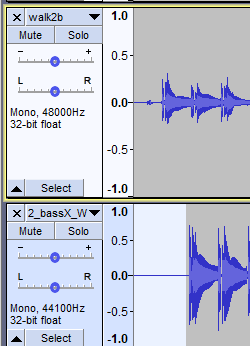
Steps:

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)

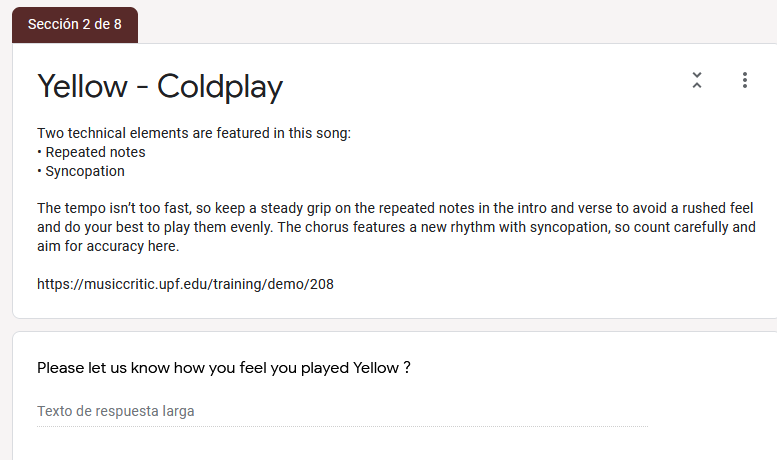
1. Don’t wait too long before you having the mixes graded

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



*Figure 23: Processing a stem in Audacity*

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



*Figure 24: Technical instruction and link to recording portal*

5.2.2 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)

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

The Onset and Offset Grades are common to all songs. There are variations in the content and number of Technical Control related questions. The table below summarizes the topics that the different questions cover

*Table 10: Technical Control Topics (TODO refer to other table related)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Q1 | Q2 | Q3 |
| Yellow | Repeated Notes | Syncopation | Sound Quality |
| Bjean | Articulation | Dynamics | Sound Quality |
| Just | Accented syncopation | Dynamics | Sound Quality |
| Brown | Groove | Sound Quality | ----- |
| Road | Syncopation | Articulation | Sound Quality |
| WOTM | Syncopation | Sound Quality | ----- |

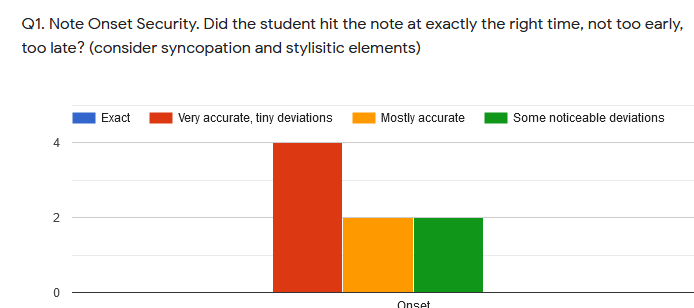
The original BJean Google Form has Q2 covering “Sound Quality” and Q3 covering Dynamics. In order for the parsing programs to work these columns had to be switched to read Q2 -Dynamics and Q3 Sound Quality as shown in table. Brown and WOTM

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

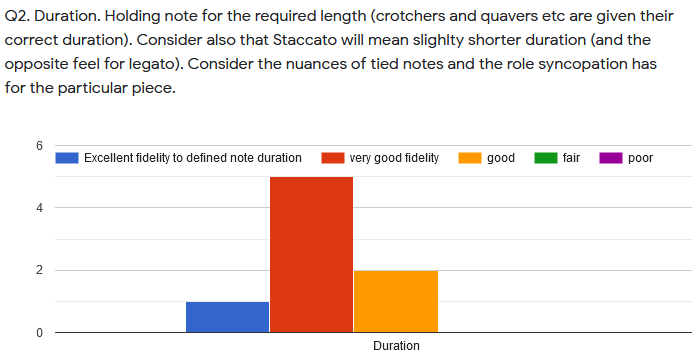
*Table 11: Selected comparison: IEC vs SG*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 |

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.



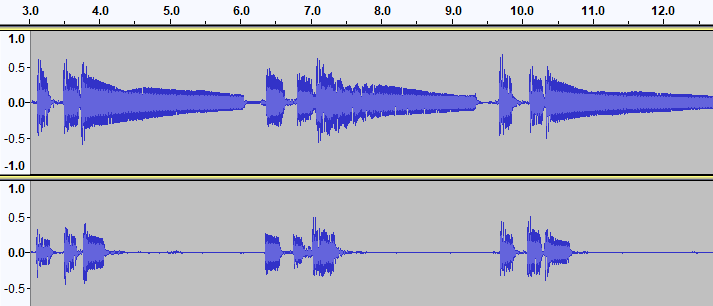
*Figure 25: Grading Histogram Onsets: Billie Jean*



*Figure 26: 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:



*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**

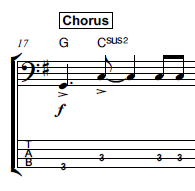
*En esta cancion las notas deben ser mas largas al prinicipio y màs cortas en el "bridge".*

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

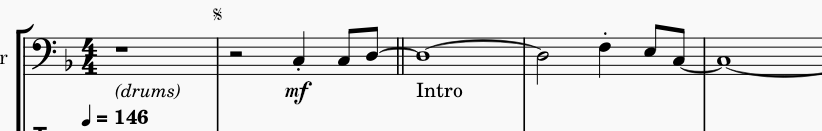
*Figure 27: Contrasting comments of WOTM performances*

In TCL exams[4], comments alongside grades are obligatory, and applying this principle to the experiment, meant we have more rationale behind why a particular recording was graded a certain way. The tagging is also another means of securing the integrity of the recordings by appending comments to the tags field.

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. on the “and” beat after 2, when counting 1+ 2**+** 3+4.

*Figure 28: 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 the Rhythm CSV file that indicates Syncopation and this could flag the onset detection algorithm to check for greater energy. The same musical property occurs in the chorus of “Yellow”. For WOTM the syncopation manifests itself in the audio, with the first bass note occurring half a beat after where it is indicated in the score.  


*Figure 29: Syncopation example: WOTM*

The dot is used to indicate syncopation 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 Musescore. 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. Therefore, the Syncopation Technical Focus grade should correlate with the combination of onset and offset 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.

* 1. 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.

* 1. Outcomes 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.

Additional recordings were made to bring the total recordings of “bjean” up to 15, “just” up to 12, “yellow” up to 12 and for “wotm” up to 12.

1. Results

The student recordings were analyzed for their onsets and offset measurements and compared against the ground truth stems. Linear Regression Machine Learning algorithms were used to train the models with multiple variables. The Cross Validation strategy was that 30% of the total cases are used as Test Cases. The X inputs were the Precision, Recall and F-Measure (PRF for short) for both predicting both Onset and Duration Grades. The additional inputs are the Mean or Absolute Mean and Standard Deviation of the Onset/Offset Deviations.

*Table 12: Grades Actual vs Predicted*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ONSET | | DURATION | |
| **Song**. | Actual | Predicted | Actual | Predicted |
| yellow | 76.5  76.5  49.5 | 60.924392  69.760548 47.442291 | 90 63 76.5 | 79.76 78.3 66.13 |
| bjean | 76.500000  68.849998  49.500000  49.500000  76.500000 | 68.763970  55.128501  42.868460  63.165241  78.933811 | 76.5  81.0  63.0  36.0  76.5 | 64.393139  72.570609  67.022827  64.313710  74.541020 |
| just | 79.199997  72.000000  90.000000  56.700001 | 72.854928  73.866432  90.703671  79.140241 | 90.000000  79.199997  90.000000  79.199997 | 78.190844  78.908711  92.321572  83.463884 |
| Brown | 79.2  79.  72? | 77.53  81.1  77.53? | 72  90  90 | 79.56  86.55  79.56 |
| Road | 79.2  79.2  72 | 76.87 79.54  92.96 | 79.2  90  79.2 | 94.5  86  74.6 |
| Wotm | 63.0  76.5  49.5  63.0 | 83.797603 60.819976 64.732879 65.781555 | 76.5  76.5  49.5  63.0 | 99.302107 68.546260 74.259509 74.963361 |

*Table 13: Grade Prediction Errors*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Song.** | **#** | **Algo.** | **MAE Onset** | **RMS Error Onset** | **Extra Inputs** | **MAE**  **Duration** | **RMS Error Duration** | **Extra Inputs** |
| yellow | 8 | SOP | 8.12 | 9.87 | Mean | 11.97 | 12.2 | ABS Mean |
| bjean | 15 | IEC/ SOP | 8.84 | 9.46 | - | 8.84 | 9.85 | ABS Mean Standard Deviation |
| just | 11 | SOP | 7.84 | 11.7 |  | 4.67 | 6.38 | Mean |
| brown | 8 | IEC | 3 | 3.5 |  | 7.14 | 7.7 | - |
| road | 8 | IEC | 7.87 | 12.17 | Mean | 7.95 | 9.5 | Mean |
| wotm | 11 | SOP/IEC | 13.62 | 15.15 |  | 16.87 | 18.3 | Mean |

Table 17 gives an overview of how the different songs responded to predicting the main grades: Onset and Duration.

Four songs used a single algorithm, while two of the songs, “bjean” and “wotm” used a blended algorithm, i.e., different algorithms were applied to different song segments. The criteria for algorithm, selection was the “muteness” of the string after plucking. The muted property is also annotated in the Three-Column “rhythm” csv-format annotation file: [Onset, Muted, Offset] using the IEC algorithm.

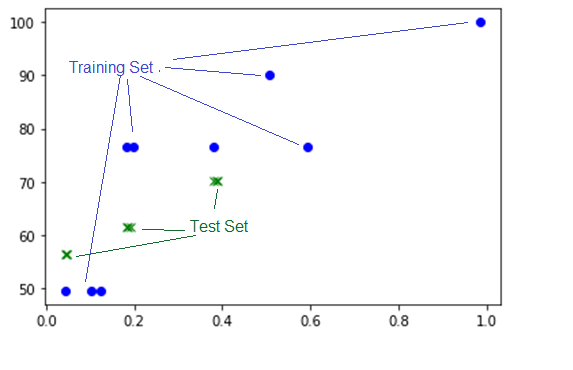
As table 17 shows, Sometimes the addition of these additional inputs did note reduce the Mean Absolute Error. One explanation for this might be the following:  
a student can have a low PRF but it also has a lower Duration Mean and Duration Absolute Mean. The reason this happens is that less onset/offset pairs are considered (through higher missed notes) for calculating the deviations, thus the low number of deviations values for low PRF scoring. The remainder of this chapter discusses the relationship between the predicted grades and the nuances and musical properties of the audio tracks. The final section discusses the Technical Control grade predictions followed by a review of the role played by the Teachers Comments in the grade prediction.

* 1. Yellow

Yellow, being a Grade 0 song is the most basic: there are no rest notes, just repeated notes in the verse with some syncopation and tied notes in the bridge, thus with very short inter-note gaps it was more suitable for the SOP algorithm. While the PRF score of the Ground Truth was close to 100% the highest graded Student 5 only scored a 50% precision while 2nd highest graded Student 4 for a P.R.F score with a 59% precision.

*Table 17:Best two “yellow” students*

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



*Figure 32: Plot of Onset Grades vs Precision*

Green plot: Test Set  
Blue Dots: Training Set

One explanation for the higher grade given by the teacher for Student 5 was because it was a different student with different instrument, with an overall superior sound quality.

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

Student 1 was an outlier. Apart from a slight lack of consistency in hitting the A notes, when you listen to the Audio you can hear some clicks which cannot be located on the audio waveform. These noise sources were probably due to the settings or the environment of the Sound Card. For Student1, it was noticed that the IEC algorithm returned a 12 % rather than a 4% precision from SOP. It may be that the sound island approach is less sensitive to noise. Usually when doubts like these occur, the best solution is to take more recordings. Three additional ungraded recordings Student 9,10,11 have been made to allow for future validation checks.

* 1. Billie Jean

The song was divided into three sections: 1st verse (Muted), Bridge (Non-Muted), Chorus (Muted).

The M.A.E increased to 11% when including 'Onset Mean', 21%, when including ‘Absolute Mean’ and 11% and when including Onset Standard Deviation. Overfitting can be a problem with five or more input variables. Low PRF results in a lower deviation count because “bad onsets” are filtered out. If you have two student recordings with very similar Precision, Recall , F Measure Values, then you can do a fair comparison of the Statistics

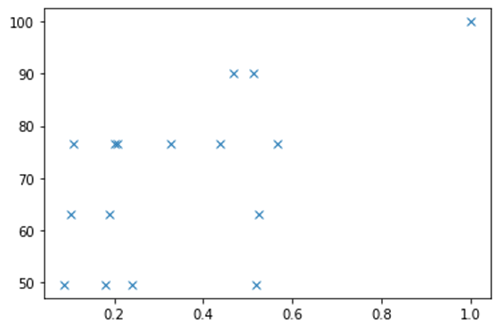
The interesting thing about the Table 17 is that it’s the only result that produces the minimum error when the Standard Deviation of the Duration is included Although relatively speaking, the Mean Absolute Errors, compare well against the other songs, but there are huge prediction errors for the low grades. For Test 3 in table 17, the predicted grade is the same as Test 0, i.e. 64.3% but the actual grades are widely different: 36% and 76.5%. The failed grades of 36 are from Students 9 and 10 have very low PRF scores. However, if these two rows are removed and the Linear Prediction of the Duration Grade is applied on the reduced set , the MAE jumps to 63%, but removing the Duration Standard Deviation would then bring it back down to 14%. Billie Jean is the song with most Student recordings; however they are more divergent in quality than other songs.

*Table 18:Billie Jean: Actual vs Predicted Grades Overall*

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

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

The low MAE for the overall grade (which is give out of 5) does not fit well with the prediction for test 3. Figure 33 shows 3 over estimated outliers and two underestimated outliers.



*Figure 33: Plot of Precision (X) vs Onset Grades (Y) Billie Jean*

Lets examine the 50% grades with precision value above 0.5. Listening back on the stem and looking at the waveform there are no major differences with the ground truth, but the problem was the synchronization with the bass drum. The teachers comment was

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

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

You can see that the best (Student 14) and worst (Student 10) student performance show similar statistics, partly because of the “filter effect” of the minimum window size.

*Table 16:Best vs Worst wotm students*

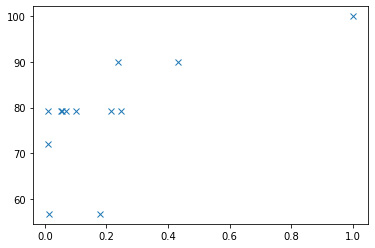
|  |  |
| --- | --- |
| High Grade: Student 2 155 Onset devistion | Worst Grade: Student 10  62 = deviations |
|  |  |

* 1. 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, so the SOP algorithm was be applied

Student 4 had best accuracy with a precision of around 0.43, suggesting high sensitivity of algorithm or an overall distinct performance characteristic from ground truth. This student was also the highest graded student summing all the grades together. Student 6 had a very poor PRF 0.1). The best predictions for the Onset and Offset grades came from only considering the PRF inputs.

The modified SOP algorithm is used to calculate this for these Non-Muted notes.  
TDO : Point out Student 6 and 4., look at comments

*Figure 36: Just Looking: Grades (Y) Precision (X)*

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

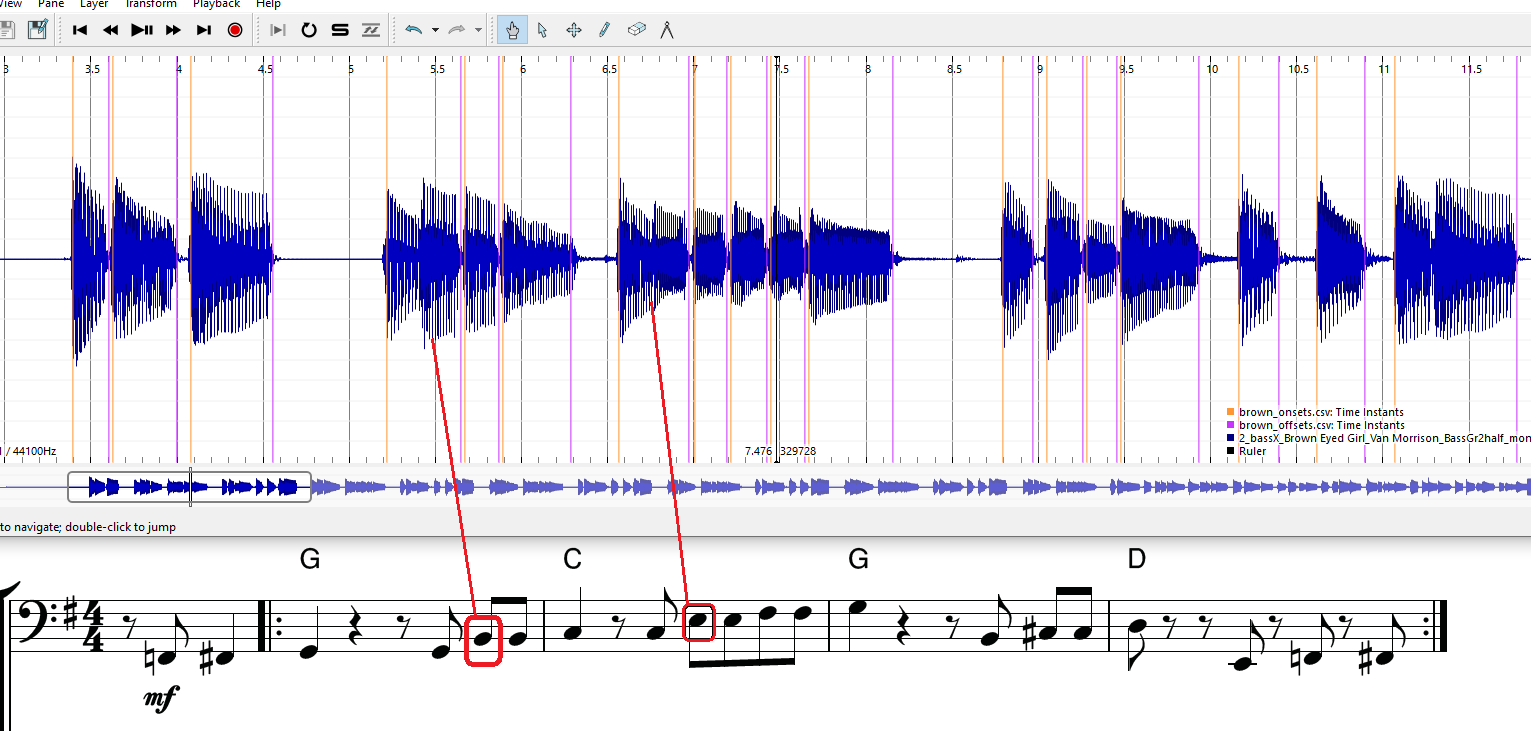
* 1. Brown Eyed Girl

“brown” from grade 2 is a good example in highlighting the limits of all the state of the two main algorithms shortlisted. The Recall metric for the IEC methods was only 82 % and the SOP not much better at 85%. Although IEC had higher precision the closeness of the notes generated “Sound Archipelagos” instead of sound islands.

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

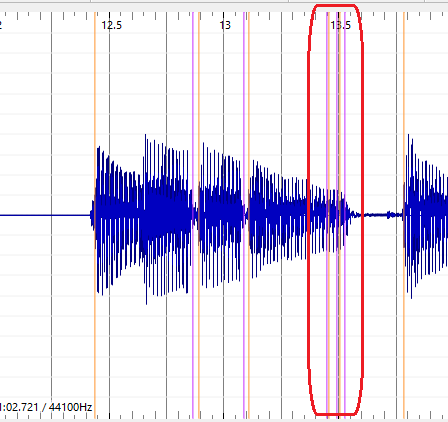
*Table 16:First 8 onsets of Brown*

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



*Figure 37: Brown: Missed Onset pattern*

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



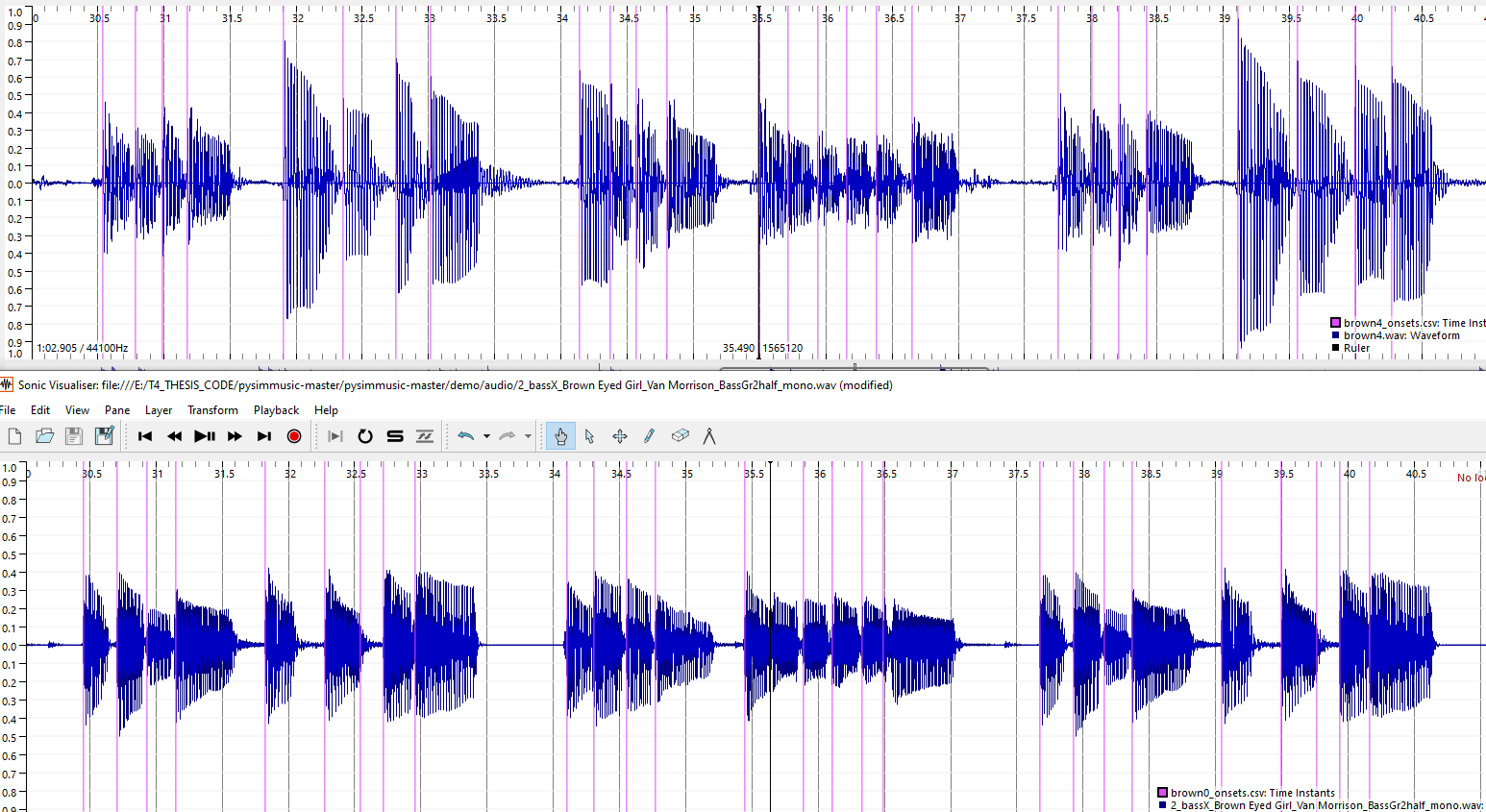
*Figure 38: Brown: False Onsets*

Notwithstanding these limitations, the same procedure of evaluating students was applied like the other songs. However, it would be useful to have “confidence formula” that depended on the best ground truth measurements. However, the student performances were well distributed. The best graded student had PRF results in the 10%, so this raises question about the confidence and robustness of the IEC algorithm to short notes and note intervals.

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

*Figure 39: Brown: Grade Plots*

The big outlier is the Student 4 who scored 90% grade , but a precision around 10%. The investigate this the wave from of the Ground Truth and Student are compared in the figure below. At 36.5 secs Student 4 correctly hits 6 offsets but the GT only hits 5 onsets and also as the song progresses. There is almost a 10ms drift at 39.5 second timestamp.



* 1. 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 cutoff point was made at 34.5 seconds to remove them all. This resulted in getting a better PRF score using the IEC algorithm. The Grade Prediction for Onset and Duration worked optimally when considering the Mean Onset and Mean Duration respectively alongside the PRF

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

*Figure 8: Roadrunner: Onset & Duration Grades*

* 1. Walking on the Moon

The PRF before applying this muting classification for WOTM was 0.701 0.932 0.8.

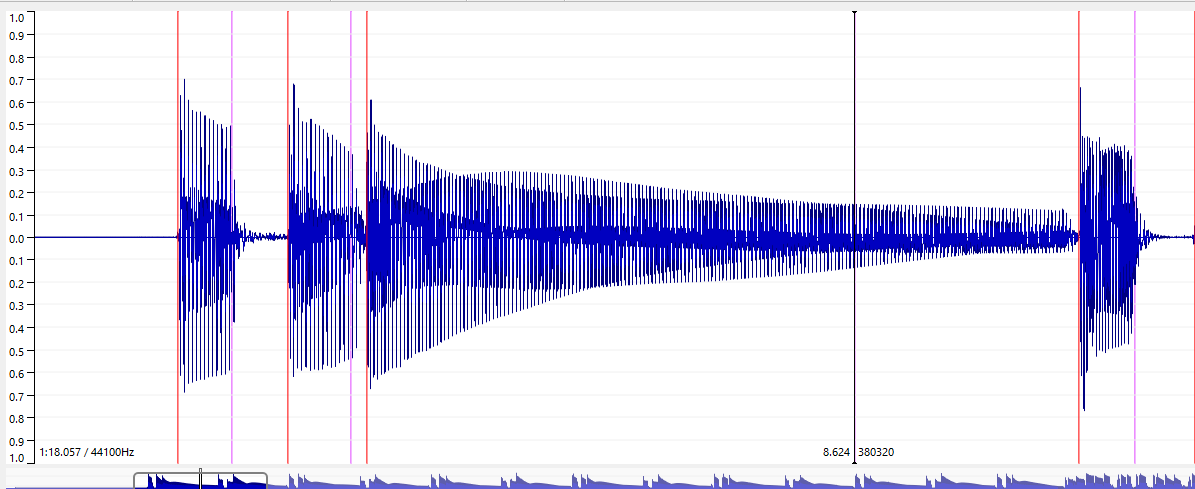
The precision was low for the stem and the range of student precisions was also very low in comparison.

After segmenting the verse into the SOP algorithm and the bridge into the IEC algorithm, the new PRF were as follows;

|  |  |  |
| --- | --- | --- |
| 0.983193 | 0.991525 | 0.987342 |

It is important to point out a trade off made in using SOP for the first part. The very first note of WOTM has a subtle gap, which is ignored if you use the SOP, “offset=next onset” algorithm. How would you determine if this subtle gap is respected in a Student recording? Checking the Energy would be one way, but that requires a lot of effort in determining what is the perceptual drop off point. A better approach is to use an existing library function from Essentia that has already researched the topic [x]

If you take a measurement of the “Effective Duration” using a Threshold setting of 0.05 of the first three notes WOTM you will obtain offset points which approximately align with the human hearing threshold



*Figure 43: WOTM: Offsets using Effective Duration*

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

* 1. Technical Control Grades

Bjean having the most recordings discussed here under the two technical control2% was achieved elements and sound quality. A Mean Absolute Error of 10.6 % in predicting the Articulation Grades using the PRF and Mean Duration Deviation as inputs. The smallest MAE error on predicting the Volume Control Technical Control grades was 7.9% and this was achieved with considering PRF only. For the sound quality grade, no grade prediction was attempted the audio features that the algorithms extract don’t relate directly to sound quality. This requires other audio features to be extracted, such as background noise level, stability of the dynamics which are not the subject of this thesis. Apart from collecting valuable data for future research, having the recordings labelled with Sound Quality in trying understanding how robust the osnet detection is in the presence of noise

*Table 2: Tech Control Grades Actual vs Predicted*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Technical Focus 1 | | Technical Focus 2 | |
| **Song**. | Actual | Predicted | Actual | Predicted |
| yellow |  |  |  |  |
| bjean | 63.0 67.723573  1 81.0 61.580665  2 76.5 70.486409  3 36.0 55.916098  4 76.5 73.553451 |  |  |  |
| just |  |  |  |  |
| Brown | 79.199997  79.199997  72.000000 | 76.876749  77.229587  76.876749 | - | - |
| Road |  |  |  |  |
| Wotm |  |  |  |  |

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

*Table 3: Technical Grade Prediction Errors*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Song. | # | Algo. | MAE Onset | RMS Error Onset | Extra Inputs | MAE  Duration | RMS Error Duration | Extra Inputs |
| yellow |  |  |  |  |  |  |  | ABS Mean |
| bjean |  | 10.6 | 12.9 | MeanOnset, Mean Duration |  | 7.910 | 11. |  |
| just | 11 | SOP |  |  |  |  |  |  |
| brown | 8 | IEC | 3 | 3.3 |  |  |  | - |
| road | 8 | IEC |  |  | - |  |  | - |
| wotm | 11 | SOP/IEC |  |  |  |  |  |  |

* 1. Teacher Comments

This part introduces the topic of text analysis without going into real depth. The intention is to uncover patterns in the text that can help towards predicting scores.  
After Translating the comments using deepl online translator.

A simple correlation of word count with the grades showed that poorer grades resulted in more words in the case of “brown”.

https://deepai.org/machine-learning-model/sentiment-analysis

Some real example of teacher text was used for testing purposes.

https://deepai.org/machine-learning-model/sentiment-analysis

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.

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

Classification for handling the text processing of the examiners report.

1. Conclusion

Regarding the Tenuto scenario, a starting point for looking into it would be to examine the two extreme cases: Max Case: All notes equal , Min Case the third note is absent.  
Between these two extremes: an ideal energy statistic should be chosen for tenuto, considering the limit of human perception.

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.

Possible Improvements to Duration Accuracy

The main challenge was to do a score fidelity check on long notes, typically compose of tied quarter, eight notes. If a particular song yielded better results for the full extract using SOP, then this meant that there was effectively no strict validity check on whether a long note was properly held. The scope of the project was limited to blending two algorithms. There are other library functions that could be incorporated into the mix to identify the duration of an audible segment in a frame. EffectiveDuration() in Essentia returns the audible segment within a parameterizable threshold. Let us propose that a Student recording could obtain 100% Onset accuracy and that the only imperfection to be measured was the offsets. Let us also consider that this is the non muted scenario where all offsets are the next onset. In this scenario we have a student recording perfectly aligned with the stem onset and offset. In these conditions you could measure the EffectiveDuration of each segment and compare for similarity.

If we measure the opening riff of WOTM the first Effective Duration of note 3 is is xx sec. The Table below captures the manually calculated effective Durations of note 3 of all eight recordings an manually compares them to the grade.

Optimising Mean Absolute Error

We found that in some cases, adding the deviation statistics information did not help in improving the predicted grade. In some cases there were improvements with adding the absolute mean rather than the mean. In all cases adding the Standard Deviation statistic did not improve the prediction.

Logic would dictate, that if the PRF of the students was generally high, then adding the deviations statistics would improve the predictions. This observation was tested in the following table summary:

Choosing Mean over ABS Mean in the statistics, suggest that the grade is more sensitive to whether a note is late or early or too long or too short. Ti make a more in depth analysis to this it would be required to have significantly large number of recordings from diverse students so that separate statistics on early and late onset deviations could be studied.+ The Same could be said of separating the duration deviations into the shorter durations and the longer durations. Musically speaking , the particular song style would dictate the teacher tolerance is to lateness/vs earliness and note longitude.

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 the previous grades allocated by the Bass Teacher were downgraded 90% and the additional grades to bring the Students up to 12 were maintained. The downgrading is built into the code for the specific students 1-8 for each song.

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

Adding a Tenuto marking in the Rhythm Files, could be an extension to the existing methodology to customize algorithms for certain technical focus elements,

1. List of figures

(Optional)

[*Figure 1: 4 crotchets, 4 quavers and 4 semi-quavers.* 5](#_Toc78887595)

[*Figure 2: dynamic threshold setting* 10](#_Toc78887596)

[*Figure 3: Extract from IDMT Dataset. File 002.xml* 14](#_Toc78887597)

[*Figure 4: Note duration of Staccato notes* 15](#_Toc78887598)

[*Figure 5: Audio 002.wav (IDMT DATASET) with onsets* 16](#_Toc78887599)

[*Figure 6: Billie Jean: Midi Rendering vs Human recording (shorter notes)* 19](#_Toc78887600)

[*Figure 7: Normalized Energy of Audio 002.wav sample (IDMT DATASET)* 25](#_Toc78887601)

[*Figure 8: Sound Islands of Audio 002.wav sample (IDMT DATASET)* 26](#_Toc78887602)

[*Figure 9: Calculated IEC onsets for WOTM verse* 28](#_Toc78887603)

[*Figure 10: Calculated IEC Sound Islands for WOTM bridge* 28](#_Toc78887604)

[*Figure 11: Figure 12: Calculated IEC Onset for WOTM verse* 29](#_Toc78887605)

[*Figure 13: PitchMelodia derived Onsets for WOTM verse* 29](#_Toc78887606)

[*Figure 14: PitchMelodia derived Onsets for WOTM bridge* 30](#_Toc78887607)

[*Figure 15: IDMT Dataset GT Onset deviations ( SOP ) algorithm* 31](#_Toc78887608)

[*Figure 16: IDMT Dataset GT Onset deviations ( IEC ) algorithm* 31](#_Toc78887609)

[*Figure 17: IDMT Dataset GT Offset deviations ( IEC ) algorithm* 31](#_Toc78887610)

[*Figure 18: Billie Jean GT onsets deviation (? Algorithm)* 31](#_Toc78887611)

[*Figure 19: Billie Jean GT offsets deviation (? Algorithm)* 32](#_Toc78887612)

[*Figure 20: End to End test of Music Critic with Billie Jean GT* 35](#_Toc78887613)

[*Figure 21: Workflow on Student Portal* 37](#_Toc78887614)

[*Figure 23: Processing a stem in Audacity* 38](#_Toc78887615)

[*Figure 24: Technical instruction and link to recording portal* 39](#_Toc78887616)

[*Figure 25: Grading Histogram Onsets: Billie Jean* 42](#_Toc78887617)

[*Figure 26: Grading Histogram Offsets: Billie Jean* 42](#_Toc78887618)

[*Figure 27: Contrasting comments of WOTM perfromances* 43](#_Toc78887619)

[*Figure 28: Syncopation example Just Looking* 44](#_Toc78887620)

[*Figure 29: Syncopation example: WOTM* 44](#_Toc78887621)

[*Figure 32: Plot of Onset Grades vs Precision* 50](#_Toc78887622)

[*Figure 33: Plot of Precision (X) vs Onset Grades (Y) Billie Jean* 52](#_Toc78887623)

[*Figure 36: Just Looking: Grades (Y) Precision (X)* 54](#_Toc78887624)

[*Figure 37: Brown: Missing Onsets* 55](#_Toc78887625)

[*Figure 38: Brown: False Onsets* 56](#_Toc78887626)

[*Figure 39: Brown: Grade Plots* 57](#_Toc78887627)

[*Figure 8: Roadrunner: Onset & Duration Grades* 58](#_Toc78887628)

[*Figure 43: WOTM: Offsets using Effective Duration* 59](#_Toc78887629)

1. List of tables

(Optional)

[*Table 1: Abessers Data Driven algorithms vs Salamon/Gomez* 6](#_Toc78887630)

[*Table 2: Abesser Signal Processing vs Salamon/Gomez* 7](#_Toc78887631)

[*Table 3: State of the Art algorithms for Onset/Offset detection* 8](#_Toc78887632)

[*Table 4: Six chosen songs from TCL song list* 16](#_Toc78887633)

[*Table 5: Tech. Control Parameters for 6 TCL songs* 20](#_Toc78887634)

[*Table 6: Benchmark results of Algorithms* 24](#_Toc78887635)

[*Table 7: Accuracy Metrics for 6 TCL ground truths (combined vs non combined)* 26](#_Toc78887636)

[Table 8: Accuracy Metrics for Billie Jean Hybrid Algorithm 27](#_Toc78887637)

[*Table 9: Selected comparison: IEC vs SG* 27](#_Toc78887638)

[*Table 10: Technical Control Topics (TODO refer to other table related)* 41](#_Toc78887639)

[*Table 11: Selected comparison: IEC vs SG* 41](#_Toc78887640)

[*Table 2: Grades Actual vs Predicted* 48](#_Toc78887641)

[*Table 3: Grade Prediction Errors* 49](#_Toc78887642)

[*Table 17:Best two “yellow” students* 50](#_Toc78887643)

[*Table 18:Billie Jean: Actual vs Predicted Grades Overall* 52](#_Toc78887644)

[*Table 17:Best vs Worst wotm students* 53](#_Toc78887645)

[*Table 17:Best two “yellow” students* 53](#_Toc78887646)

[*Table 2: Tech Control Grades Actual vs Predicted* 60](#_Toc78887647)

[*Table 3: Technical Grade Prediction Errors* 60](#_Toc78887648)

[14. *Table 14:* Inputs (yellow) vs outputs (green) for Billie Jean Onsets 73](#_Toc78887649)

[17. *Table 15:* Inputs (yellow) vs outputs (green) for Billie Jean Offset 74](#_Toc78887650)

1. List of symbols

(Optional)

1. Bibliography

For example a journal article is cited as1

**0. Stevenson, G. N., Collins, S. L., Ding, J., Impey, L. & Noble, J. A. 3-D Ultrasound Segmentation of the Placenta Using the Random Walker Algorithm: Reliability and Agreement. *Ultrasound Med. Biol.* 41, 3182–3193 (2015).**

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)

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.

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

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.

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

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)

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)

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

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

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)

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

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

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)

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

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-handguitar playing techniques detection, In Proc. Of the Int. Conf. on New Interfaces for Musical Expression (pp. 1–4).

16. G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," CUIDADO I.S.T. Project Report, 2004

17 Automatic Tablature Transcription of Electric Guitar Recordings by Estimation of Score-and Instrument-Related Parameters. Authors: Christian Kehling, Jakob Abeßer, Christian Dittmar, Gerald Schuller. Date de publication 2014/9/1, Conférence DAFx, Pages 219-226

18. Romain Hennequin, Anis Khlif, Felix Voituret, and 171 Manue Moussallam: “Spleeter: a fast and efficient music 172 source separation tool with pre-trained models”, The Jour-173 nal of Open Source Software, DOI:10.21105/joss.02154 174 (2020)

Github links

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

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

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

“Homelesss reference”

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).

1. Stevenson, G. N., Collins, S. L., Ding, J., Impey, L. & Noble, J. A. 3-D Ultrasound Segmentation of the Placenta Using the Random Walker Algorithm: Reliability and Agreement. *Ultrasound Med. Biol.* **41,** 3182–3193 (2015).

1. Stevenson, G. N., Collins, S. L., Ding, J., Impey, L. & Noble, J. A. 3-D Ultrasound Segmentation of the Placenta Using the Random Walker Algorithm: Reliability and Agreement. *Ultrasound Med. Biol.* **41,** 3182–3193 (2015).

1. Stevenson, G. N., Collins, S. L., Ding, J., Impey, L. & Noble, J. A. 3-D Ultrasound Segmentation of the Placenta Using the Random Walker Algorithm: Reliability and Agreement. *Ultrasound Med. Biol.* **41,** 3182–3193 (2015).

1. Appendices
   1. Appendix 1

*WOTM; This account opened with clear rhythmic drive and placement in the first lines, as the style was portrayed effectively and consistently. A few moments missed complete rounded tone or proficient legato but movement across the instrument with clear picking was achieved. Musical details were observed very well and some confident moments of execution rendered a pleasing musical flow overall.*

*Marks: 8, 7, 9 = 24*

[

"Positive",

"Negative",

"Positive"

]

Brown Eyed Girl The rhythms were accurate in this, and the rests were well counted. Most of the note-lengths were correct apart from the doted-crotchet C in the second bar of the Chorus, which was played short. The notes were mostly fine, and the few alterations in the Pre-Chorus were all fine. For some reason the dynamics were ignored; the crescendo in the 2nd time bar at 28 was absent, and the f chorus was no louder than the mf verse. It was all good otherwise—it just needed more shape. Marks: 7/10/8 = 25

[

"Negative",

"Negative",

"Negative",

"Negative",

"Positive"

]

Yellow Repeated notes were broadly steady, although underlying pules wavered at times, and the dynamic drop in bar 13 was effective. The odd placement error affected flow a little and attack in the chorus was not fully controlled, but syncopation was handled well on the whole. Marks: 7/9/8 = 24

[

"Positive",

"Negative"

]

* 1. Appendix 2

1. Technical Focus Grades are not included in tables 14 and 15.
2. *Table 14:* Inputs (yellow) vs outputs (green) for Billie Jean Onsets

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

1. *Table 15:* Inputs (yellow) vs outputs (green) for Billie Jean Offset

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

1. 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)