European Journal of Computer Science and Information Technology
Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

ANOMALOUS NOTE CHANGE DETECTION OF NKNOWN MONOPHONIC MELODIES

K. L. Jayaratne
University of Colombo School of Computing

ABSTRACT: *Anomalous note changes in music melodies are a major concern in several automatic music content analysis tasks. When 'the musically scale un-related notes' occur in the note sequence, it will provide less pleasant melodies. These 'uncommon' notes are the situations which refer as the 'anomalous notes' in a particular melody. Identifying and eliminating the effects of these note changes will helpful to enhance the results in automatic melody evaluation, as well as in automatic melody transcription. In order to address the above issue, this paper proposes an approach to detect anomalous notes changes of unknown monophonic melodies. The proposed model is designed with two main phases with the applicable machine learning and signal processing techniques. Within the first phase, melodies are processed to have their pitch estimations. After the pitch estimation, a note event model has employed with the application of Long Short -Term Memory (LSTM) Neural Network for the detection of anomalous note changes. A set of recorded sample melodies are used as the dataset to evaluate the model. The dataset was collected only for the main seven major scales in music. The model was able to detect anomalous note changes with an overall accuracy of 68.2% for the used dataset.*

KEYWORDS: unknown monophonic melodies, anomaly detection, note detection, long short term memory (LSTM)

INTRODUCTION

Music is a creative way to express the emotions of human beings beautifully, as the combination of vocal and instrumental sounds. It is a type of language which is common for everyone in the world. People are tending to try new music creations all over the world. Today, it becomes popular with the usage of modern technologies. A typical approach to music creation has been to use music production software that has pre-programmed musical structures such as rhythm patterns, music clips [1, 2, 3] etc. The users would record, experiment and combine the musical structures in order to come up with a final music creation. However, non- musicians might find it difficult to use this approach due to lack of domain knowledge in music. A more suitable approach would be to provide users with an accompaniment music track to sing along with it. The accompaniment music track is usually be a pre-produced track based on a popular song.

Everyone has tried to hum a song even without the domain knowledge of music. Also, some people are trying to play musical instruments without prior knowledge of music. They may try their own new experimental creations as well. Though, for these situations, it is necessary to have some sort of evaluation process for their new melodies.

To address this, we can find several applications which provide the ability for users to have their own musical creations. Automatic melody evaluation is an important fact to be considered in such applications, especially for musically untrained users. While a process of performance evaluation of any melody is encountered, it will occur some patterns or any other rhythm of notes within that particular melody. But when non-musicians / musically untrained users are trying to have their own melodies, they may have some unnecessary note changes in their performances. However, in some occasions, even the musically trained users may have some ambiguous note events in a melody which lead to having anomalous note changes. Therefore, there may be some unclear/unpleasant states in that particular melody. However, identifying these kinds of effects is not a simple task.

When we try to have an automatic singing skill evaluation or any other kind of automatic music melody evaluation, we must perform various technical approaches to classify the input melodies according to their performance. However, 'the automatic melody evaluation' process depends on a set of decidable facts such as the features which we are using for the evaluation process, the noisiness of the input melodies, anomalous note changes in singing/playing melodies, fold responses of melodies, etc.
When we try to have an automatic singing skill evaluation or any other kind of automatic music melody evaluation, we must perform various technical approaches to classify the input melodies according to their performance. However, 'the automatic melody evaluation' process depends on a set of decidable facts such as the features which we are using for the evaluation process, the noisiness of the input melodies, anomalous note changes in singing/playing melodies, fold responses of melodies, etc.

Music melody is a collection of notes which are arranged in a musically satisfying manner. Therefore each melody should have a pattern of notes according to its pitch, chord and the scale. Thus, within this note arranging, there may be some situations which are uncommon and ambiguous regarding the pitch, chord and the scale of that particularly considered melody. If we consider melodies which include those type of anomalous notes, there is a probability of falsifiable results in melody evaluations. Also in melody transcription, these ambiguous situations lead to some false note key detections. The ambiguity occurs because of the harmonic and variations within the melodies. Note that all ambiguous situations in a note sequence cannot consider as anomalies because of the involvement of harmonies and variations. Therefore identifying anomalous notes changes and their effects, involves a valuable step in melody evaluations in the context of music content analysis. In here an anomalous note event is defined as a scale unrelated note occurrence of a melody.

BACKGROUND

Melody evaluation is a process of figuring out a set of feasible features and making decisions about them according to the structure of the music. According to the focused set of features, evaluation results can classify as 'good' or 'poor'.

For the process of feature extraction, we have a large collection of music features such as pitch, vibrato, tempo, etc. But it is challengeable to handle the interaction of these music information facets for such kind of evaluation of a melody. In the literature, they have already encountered this issue as ‘The Multifaceted Challenge’ [4].

In order to come up with a final music creation as a complete song, it should be a combination of at least a single vocal melody and some instrumental melodies [5]. The sound which consists of a particular pitch and a duration is referred to as a musical ‘note’ whereas the sequence of notes arranged in a musically satisfying manner is referred to as the ‘melody’. When a particular melody is consisting of only one singing voice (single vocal melody) or only one instrumental melody (single instrumental melody), it is called a ‘Monophonic Melody’. When a melody does not have its beat arrangement or previously estimated facts, it is called an ‘Unknown Melody’.

Notes can be identified as the fundamental units in a music melody. Each note has its own duration and a pitch, which is an identifiable fact that refers to a single sound in the complete range of sound [6, 7, 8]. Simply, these set of notes consists of different keys which can be named as C, D, E, F, G, A and B. Apart from these major keynotes there are five other note keys (C♯/D♭, D♯/E♭, F♯/G♭, G♯/A♭, A♯/B♭), which called as sharp (♯) notes or flat (♭) notes.

The pitch difference between two adjacent notes is referred to as an interval. These intervals with pitch frequency ratio 2:1 are called as octaves. Each octave is divided into twelve notes in western music and it is referred to as the ‘Chromatic Scale’. This twelve-note pattern defines the frequency of each note pitch using the following equation.

$$f = 2^{x/12} f_{base} \qquad (1)$$

The intervals between the two adjacent note pitches are measured in semitones [9]. This chromatic scale can be noted with sharp signs when ascending and flat signs when descending. Sharp (♯) raises the sound by one semitone and flat (♯) lowers the sound by one semitone. Figure 1 shows the twelve notes in the chromatic scale.

There are twelve major Scales correspond with one of the twelve notes in the chromatic scale and each scale consist with seven notes. The difference of the third and fourth notes for each of these scales is a semitone and the difference between all the other two adjacent notes is a tone. Each and every melody should include a set of these distinct notes in different octaves, according to the scale of the melody. Table 1 shows the main seven major scales with the relevant notes.

Vol.7, No.6, pp.23-41, December 2019

Published by ECRTD- UK

Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

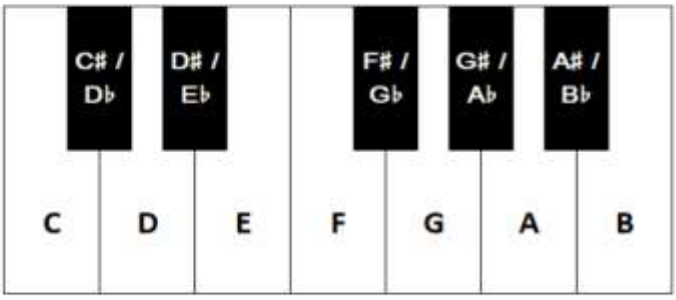


Figure 1: Chromatic Scale in Music

If we consider a simple three-note sequence in C major scale as C, D, and E♭, the ‘E♭ note’ is not a common related note within the C major scale (as shown in the Table 1). Thus, this situation can be considered as an anomalous note event in the sense of unknown monophonic melodies.

Table 1 : Seven Major Scales in Music with Notes

Scale	Set of Notes						
C Major	C	D	E	F	G	A	B
D Major	D	E	F♯	G	A	B	C♯
E Major	E	F♯	G♯	A	B	C♯	D♯
F Major	F	G	A	B♭	C	D	E
G Major	G	A	B	C	D	E	F♯
A Major	A	B	C♯	D	E	F♯	G♯
B Major	B	C♯	D♯	E	F♯	G♯	B♭

LITERATURE REVIEW

Anomaly detection of unknown monophonic melody is actually a combination of several computational music content analysis such as singing skill evaluation, melody transcription, music note detection and computational anomaly detection methodologies. In the context of unknown melodies, it also involves a way to identify the anomalies of a melody note sequences.

A. Singing Skill Evaluation Approaches

Singing skill evaluation approaches are the most common type of melody evaluation approaches that can be found in literature. According to the used features, we can classify the melody evaluation approaches into two main categories as pitch based approaches [10, 11, 12] and tempo-based approaches [13, 14]. In both approaches they have used several feature sets with different types of models and neural networks such as HMMs and ANNs. They have mentioned the limitations that they have observed in their methodologies to have less reliable results. However, in the context of singing skill evaluation, they were able to have applicable methodologies and techniques to obtain good results with a limited set of features.

B. Note Detection and Melody Transcription Approaches

Melody transcription is the process of converting the audio melodies into the sequence of notes. Simply, this is an automatic way of having a more precise score information for a vocal melody. Score information includes all the temporal estimations and timing estimations between each and every note pair. It tries to convert the melody signal into note keys by finding pitch estimations to the relevant MIDI standard note numbers. In order to do that, there should be a technique to identify and detect the notes in a melody.

Despite the difficulty, vocal melody transcription is still an important topic in music information retrieval field, as it enables computers to extract the information carried by singing and facilitates the direct musical interaction between humans and computers [15]. Though as in literature, generating a more accurate and precise note sequence is still a hard task. But there are some well-defined methods that they have tried to achieve this goal with the use of signal processing and machine learning theories.

Ryynnen has done several approaches for melody transcription for different kind of melodies such as polyphonic melodies, homophonic melodies, and monophonic melodies. Among those, he has proposed a model-based method for automatic note transcription of monophonic melodies based on two probabilistic models [16]. First one is a note event model which is based on Hidden Markov Models (HMM). It has used to represent note candidates. As the second model, a musicological model has used for the key estimation and likelihoods of different note sequences to examine the transitions between note candidates.

He has used a HMM to limit the dynamics of their pitch estimates. Though, this approach will be helpful to identify the process of generating note sequences. In this study, there is an attempt to eliminate some kind of predefined ambiguous notes during their transcription model.

With the idea based on Ryynnen’s method of using HMM, Sebastian Bock and Markus Schedl have proposed another way of piano note transcription with Recurrent Neural Network (RNN) [17]. They have targeted the polyphonic melodies in their approach. But it is based on a RNN to simultaneously detect the onsets and the pitch values of the notes from spectral features. The memory units in RNN are used in a bi-directional neural network to model the context of the notes.

Mauch et al. has proposed a software tool (called as Tony Software) for the interactive annotation of melodies from monophonic audio recordings [18]. In their approach, they have done a melody transcription based on their own pitch detection and note detection algorithms. They have employed two valuable post-processing steps to increase the accuracy of the system. In the first step, an amplitude-based onset segmentation has used to separate consecutive notes of similar pitches. In the second step, they have used

Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

a threshold value to discard the unnecessary note events. This is a special kind of technique to eliminate exceptional notes which can be considered as anomalous notes.

Fundamental frequency estimation and pitch detection are the most fundamental steps in note detection. This is the main step that followed by several other steps to achieve the final detection of notes. This kind of pitch detection is involved in each and every computational music-content analysis technique. Therefore, each melody transcription approach has conducted a special considerable analysis on pitch detection.

Among them, YIN algorithm has proposed a more precise way to detect the corresponding pitch for melody notes [19]. YIN includes a number of mathematical formulas for the estimation of the correct pitch. In above mentioned Ryynnen's proposed method [16], he has analyzed each step of this algorithm and stated a conclusion of the algorithm for the better performances. Also in Tony software [18], they have used their own defined algorithm called as pYIN [21], which has defined based on the YIN algorithm.

C. Anomaly Detection Approaches

As automatic ways of detecting anomalies, different anomaly detection approaches have been analyzed to identify their applicability for the music melodies. As in [17], there are some combinations of common anomaly detection methods with the music note events handling approaches. Thus, this study was conducted to analyze the features and their applicability of common anomaly detection methods which can be used in the domain of melody note estimation.

In [21], a time series anomaly detection approach has been done to utilize machine learning and statistical approaches to classify anomalous drops in periodic, but noisy, traffic patterns without a large body of labelled examples. As the first step of their approach, they have used TensorFlow to train their various models including DNNs, RNNs, and LSTMs to perform regression and predict the expected value in the time series. As the second step, they have created anomaly detection rules that compared the actual values to predicted values. At the later part, they have provided the confusion matrices for each model that they were investigated. In above mentioned automatic piano transcription [17], they have already used these kinds of LSTM application to detect the notes in a more accurate way.

In the study of Collective Anomaly in [22], a Long Short Term Memory Neural Network has employed to identify the anomalies in collective data set. They have provided descriptive details in each step to analyze the process of LSTM in the sense of collective anomalies.

Another study on anomaly detection in time series has done based on the long short term memory networks in [23]. In their approach, the network has trained on non-anomalous data and used as a predictor over a number of time steps. They have proposed a stacked LSTM based prediction model for the experiments and within that

Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

proposed model they were able to detect anomalies using the prediction error distribution.

DESIGN

Analysis of the limitations of melody evaluation was able to figure out the important facts of music notes anomalies such as what are the possible note patterns to have ambiguities, what type of uncommon changes are there, how the note distances are going to have uncommon changes etc. As an example, we can find that there is a low probability to have flat notes in the scale of C major. Therefore, if a melody has that kind of note patterns, it is a considerable point in the design phase. Thus, that kind of patterns and estimations are encountered in parallel to other analyzing stages. Though, the anomalies within note events are defined as any type of scale unrelated note changes.

However, in music melodies, it is possible to have scale related/unrelated variations and harmonic parts within melodies. Especially, harmonization can be found easily in singing. But when it comes to the context of monophonic melodies, the probability of having that kind of situations is very lower than the other textures of the music. Though, this anomalous note change detection has designed under the assumption that the unknown monophonic melodies do not contain any harmonic/variation point within a single type of sound. The overall design is proposed to handle only the monophonic melodies under the above assumption.

The overall design of the proposed approach is composed to follow several steps. These noted steps have addressed throughout the literature review as a combination of different areas. This approach should follow a flow of steps which is more like the melody transcription process. But in advance, required steps should follow a set of well-defined techniques to have a more accurate way in anomaly detection. It is easy to observe that the required process does not need the entire melody transcription here.

Therefore, this research is designed to have a combined process of different steps. Each step has identified as a special part of a model. Figure 2 shows the interconnection of these identified phases as a high-level research design. Melody processing consists of several steps to follow and the audio melodies are sampled into several frames. The note event model has employed with a LSTM, which is a special kind of RNN. The note event model is the most important part of this anomaly detection approach.

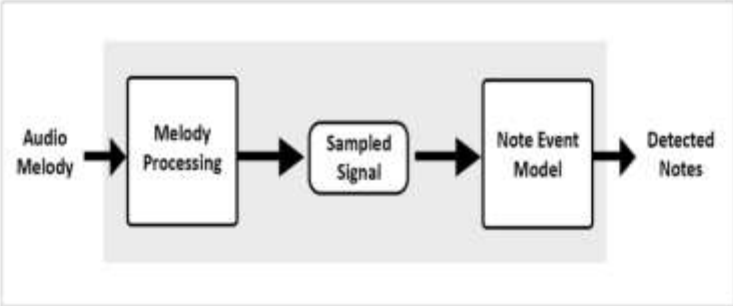


Figure 2: High-Level Research Design

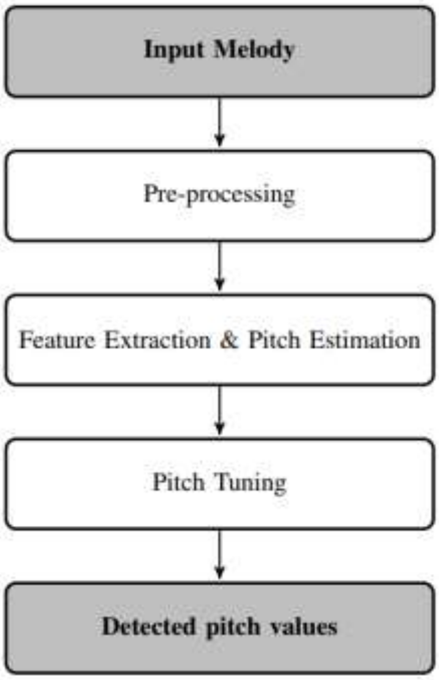
A. Melody Processing Steps

The main aim of this study is to have a way to determine the note sequence anomalies of melodies. Typically, each note is directly mapped to have a pitch value. Though, as the first phase, music melodies should process to identify the fundamental estimations for their pitch values. This involves several audio processing techniques and methodologies to follow. Figure 3 shows the basic steps that need to follow, as the melody processing steps.

As the unknown melodies are involved, recorded melodies already contain noisiness and some points in silence. Those silent points can be considered as the time gaps between some notes in the melody. As this study only considers the note changing events, these silence points make no sense in the context. Therefore, a pre-processing technique is used here to minimize the captured unwanted points in recorded vocal melodies.

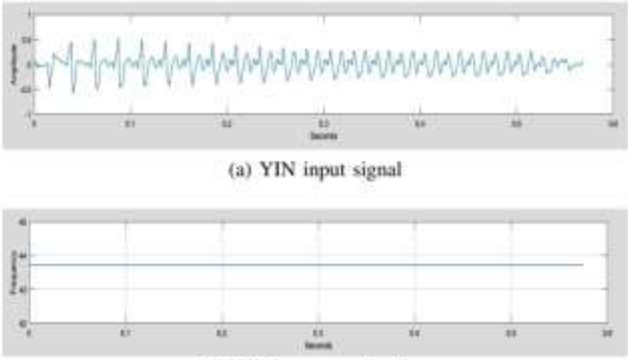
As an important fact of the needed pre-processing, it is not going to cut off any part which has at least a single point of sound. Each sound point may provide valuable information regarding a minor note change in the melody. There can be unnecessary out-pitch estimations for this kind of minor estimations as well. This pointed issue has already handled under the tuning phase after the pitch estimation.

Figure 3: Melody Processing Steps



Pitch detection is the valuable steps of melody processing. This step is going to use the facts of vocal melodies to have some meaningful estimations. Therefore, pitch' is the main feature that is going to be extracted from audio recordings in order to estimate the pitch intervals. There are two types of pitch detections as Time Domain Pitch Detection' and Frequency Domain Pitch Detection', whereas this study follows the frequency domain pitch detection technique. Pitch value is denoted as a frequency value for each point of melody (measured by Hertz (Hz)). YIN algorithm is chosen to use in this study for pitch detection, as originally proposed by Cheveigne and Kawahara in [19]. Figure 4 shows the YIN pitch estimation for a single note melody.

Typically, in the context of music, pitch values are defined based on the key distribution in a piano. A standard modern piano has 52 white keys and 36 black keys which distribute across seven octaves and three extra keys. As mentioned in the literature review, an octave consists of twelve keys as seven white keys for natural notes and five black keys for sharp/flat notes. Piano keys are named to represent their music notations with their location of the octave. For an instance, the A' note key in the 4 th octave is named as the key A4' in a piano. Thus, a standard modern piano has named its keys from A0 to C8 and the middle C' is denoted as C4' in this format. When playing the piano, this 4 th octave is the most used set of keys. In here, this representation is referred to as the 'Pitch Note Representation'.



(a) YIN input signal

(b) YIN output estimations

Figure 4: YIN pitch estimation for single note melody ('E Note' in first octave)

Each and every pitch notation in the above representation has its own unique frequency value. These pitch values are used to estimate the corresponding note for each time lag in a melody. In literature, all melody transcription techniques have followed this standard pitch notation values to obtain the notations even from a simple melody. These values are very useful in MIDI representation of melody transcription. Each of these notes has its own MIDI number in a standard way. But for this study, MIDI numbers are not needed, and hence only the estimated pitch values are used for further enhancement. Standard fixed pitch values for 4th octave notes are presented in Table 2. However, the pitch values for human singing can differ from these values within small range [24]. But they can be mapped to the standard representation.

Table 2 : Frequency values for Notes in 4th Octave

Note	Frequency (Hz)
C4	261.63
C♯4 / D♭4	277.18
D4	293.66
D♯4 / E♭4	311.13
E4	329.63
F4	349.23
F♯4 / G♭4	369.99
G4	392.00
G♯4 / A♭4	415.30
A4	440.00
A♯4 / B♭4	466.16
B4	493.88

European Journal of Computer Science and Information Technology
Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

Pitch tuning is the process of determining more accurate values for the pitch estimations by applying technical improvements. In this approach, a tuning process has applied as the final step of the YIN implementation. It uses a simple methodology as a part of the parabolic interpolation to obtain better estimation. Figure 5 shows a sample pitch estimations before and after the tuning process

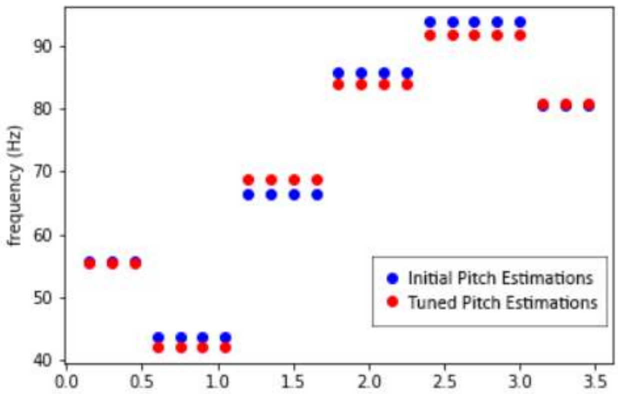


Figure 5: Pith Tuning Example

But even after the tuning process, it is very hard to find the correct note pitch for each and every point in time. For an example, a melody can have a value of 268 Hz as the estimated pitch value at one point. But that value is not the exact value for any note in any octave. 268 Hz is in between C4 and C#4 in the pitch notation representation. As a solution for this, a min-difference technique has applied to figure out the relevant pitch note. Following is the basic difference function which is used here.

$$\Delta x = | x_{est} - x_{ref} | \tag{2}$$

Here, xest is the estimated tuned pitch value and x ref is the reference pitch value which is taken for the closest two pitch note values in pitch note representation, one by one at a time. That means the estimated tuned pitch value is taken to pair with every note value in the pitch note representation. So, the absolute difference is taken as Δx for each pair and, the minimum value is considered as the correct one. Though, the xref value is assigned as the pitch note estimation when the absolute difference between that xref value and the estimated tuned pitch value (xest) is the minimum.

B. Note Event Model

As this proposed approach is mainly focusing on the context of note changes in a melody, the note changes should be identified through the reference pitch values. After identifying the relevant note changes, a machine learning model has employed to determine the anomalous note changes in a particular melody. Thus, this note event model is composed of two major components as shown in Figure 6.

European Journal of Computer Science and Information Technology
Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

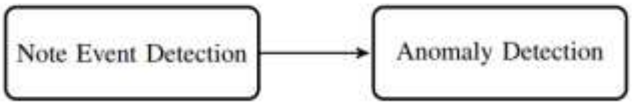


Figure 6: Note Event Model Components

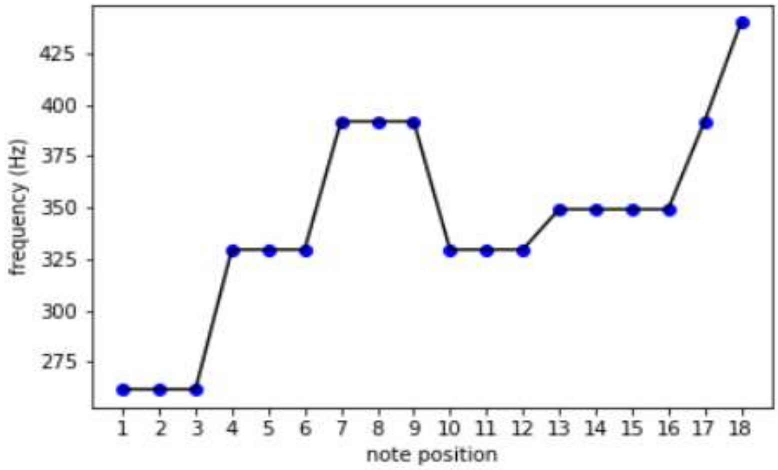
The term ‘Note Event’ is a widely used concept in different domains of music content analysis. In the context of anomalous note change detection, a note event has identified as a simple change of a note estimation which only based on its pitch value. Time intervals or time gaps between two estimation points are not an important fact in order to identify a note change. Therefore, only the pitch values are considered for this note event estimation. But in advance, note events are defined within a single octave range. All detected reference pitch values are given a score regarding its note by ignoring its octave. This provides two notes in two octaves which have the same key note, as a single score in note event range. For an example, there can be two notes in key A in two different octaves as A3 and A4. But when the note event is encountered, both of these two notes are getting a single score which represents the note A in note event range.

A melody is also consisting with repetitions of the same note in nearby estimations. But in the context of identifying note changes, only the first occurrence of a note estimation is considered, and all the immediate repetitions after that position are ignored. However, if the same note occurs once after some other notes (when it is not a consecutive occurrence), it is considered as another note event.

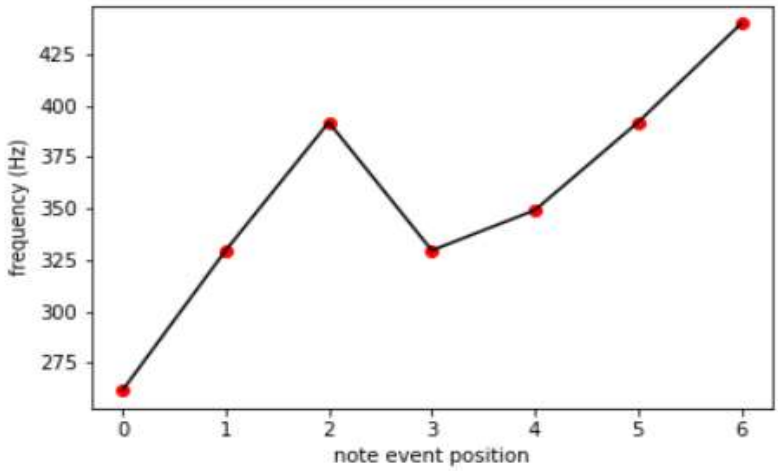
For an instance, Figure 7 shows a simple note sequence of a melody. The corresponding note sequence for that western music notation is ‘C, C, C, E, E, E, G, G, G, E, E, E, F, F, F, F, G, A’. It is easy to see that, there are some repetitions of the same note in nearby positions in the sequence. ‘Note C’ is repeated three times in first three positions. ‘Note E’ is repeated three times in second, third and fourth positions and again three times from ninth position. ‘Note G’ and ‘Note F’ also have same kind of repetitions within the sequence as well. Therefore, the corresponding note events are identified as ‘C, E, G, E, F, G, A’ for the above note sequence. These are the note changes that can be identified in that particular note sequence which denoted in the Figure 7 and Figure 8 shows the pitch value distribution and corresponding note event estimations for the above note sequence.

Figure 7: Example note sequence for note event detection

Detected note events are the inputs for the most important phase of the note event model. These note events are also represented by a pitch value. Therefore, it is considered as a note sequence which does not have the same note consecutively. Each collection of these note events should have a set of unique note patterns according to the scale of the considered melody. However, in the domain of harmonic melodies, this can differ as well. But for monophonic melodies, the set of notes associated with each scale in music can be easily determined. Although, these set of notes for each scale performs a musically satisfying note pattern within a melody. In here, once a scale unrelated note occurs within a pattern, it is considered as an anomaly. These patterns only represent notes, which occurs after a change of pitch value. These note event changes are considered as anomalous note changes.



(a) Pitch value distribution



(b) Note event distribution

Figure 8: Note event estimation example

After estimating the note events within the note event model, it uses a LSTM for the anomaly detection of those note event patterns. In here, a LSTM model has used as a sequence prediction-based model for the scale unrelated note detection as the anomalies. The model first converts the detected note event sequences as sub sequences

Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

based on the predefined note sequence length. Each predicted note event is then compared with the actual note event in order to determine the scale relationship between those two. Final anomaly prediction is then decided based on that procedure based on the trained LSTM.

EVALUATION AND RESULTS

As the required monophonic melody dataset, a set of recorded audio melodies are considered as the unknown monophonic vocal and instrumental melodies. All of these are recorded in WAV format, which is a lossless audio format. Each sample audio contains 12 to 16 seconds long melodies. However, the proposed design is analyzing a set of notes as a sequence one at a time. Therefore, it is necessary to have a set of melodies which includes a large set of patterns in all the scales.

The model has designed to use only major scale patterns, and therefore the dataset includes melodies with different types of patterns of all major scale as much as possible. This is a very important fact to train the LSTM model with more possible patterns. But in the real world, some music scales occur in very few songs or melodies. As an example, there are only a few Sinhala song melodies in B major scale. Therefore, it is hard to have an equal no of melodies for each scale type. The dataset has performed with melodies in all seven major scales as much as possible.

The model has trained with a set of melody note patterns in main seven major scales which includes more precise scale related notes. The model was designed to train on note event sequences of predefined length and to predict the next note event of that sequence. In order to have an unbiased training process, the training data has collected across all considered seven major scales as presented in the Table 3. The LSTM model was trained with for 150 epochs with the batch size of 64 samples.

Table 3 : Training Dataset

Training Data		
Scale	**Samples**	**Total Note Event Patterns**
C Major	29	1698
D Major	24	1260
E Major	23	991
F Major	39	1822
G Major	25	1263
A Major	27	1390
B Major	11	673
Total	**178**	**9097**

European Journal of Computer Science and Information Technology
Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

The trained model has tested with a set of unknown monophonic melodies, which already contains melody samples of considered seven major scales. Table 4 shows the testing dataset which has also collected across the major scales. This test data set was collected from a melody set which has less pleasant melodies. The reason for taking that kind of melodies for the testing process is to have much more scale unrelated notes. Note events are detected through the estimated pitch values of the melodies. As in the training process, note event sequences of the same predefined length are tested to predict the next note event of the pattern. The predicted note event is then tested with the actual note event of the sequence. Final prediction decision has taken upon these two note events based on the distance measure.

Table 4 : Testing Dataset

Testing Data		
Scale	**Samples**	**Total Note Event Patterns**
C Major	19	1423
D Major	14	980
E Major	12	891
F Major	19	1445
G Major	18	1087
A Major	13	995
B Major	4	203
Total	**99**	**7024**

Error Rate and Accuracy for the anomaly predictions are calculated using following two equations.

$$ErrorRate = \frac{FalsePredictions}{TruePredictions + FalsePredictions} \quad (3)$$

$$Accuracy = \frac{TruePredictions}{TruePredictions + FalsePredictions} \times 100\% \quad (4)$$

Table 5 represents the results obtained from the model for each pattern in each group of scale. The performance of the anomaly detection was measured with error rate for each major scale category. As mentioned there, the model was able to achieve 68.2% overall accuracy.

Vol.7, No.6, pp.23-41, December 2019
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

Table 5: Results for test data

Scale	True Predictions	False Predictions	Error Rate
C Major	1002	421	0.295
D Major	675	305	0.311
E Major	620	271	0.304
F Major	1090	355	0.246
G Major	659	428	0.393
A Major	642	353	0.355
B Major	102	101	0.497
			0.318

Apart from this noted testing dataset, an additional evaluation has done for the note event model with a monophonic piano melody data set. The main aim of this additional evaluation is to compare the performance of the model between two different data representations. The piano melodies that used here are collected with their MIDI pitch value representations. Therefore, those melodies do not need the initial pitch detection as their pitch values are already known. Results for this sample piano melody dataset are presented in Table 6.

Table 6 : Performance of the model for piano melody dataset

Scale	True Predictions	False Predictions	Error Rate
C Major	695	45	0.061
D Major	673	92	0.120
E Major	345	89	0.205
F Major	541	93	0.147
G Major	269	86	0.242
A Major	433	145	0.251
B Major	108	42	0.280
			0.161

The results and the analysis of the pitch detection phase indicate that the pitch estimations were obtained in more accurately with the use of a tuning algorithm. If this pitch detection was not capable of providing more reliable estimations, then there is a chance to have falsifiable predictions within the note event model.

The note event model was designed with a LSTM Neural Network as a sequence prediction model. The evaluation results for the unknown monophonic data indicates that the design of this prediction model is applicable to the detection of anomalous note changes.

With the comparison with piano melodies of MIDI note references, it is clear that the model can perform better when the pitch estimations and note event estimations have their best values. Though, a better pitch detection methodology can improve the performance of the entire model.

CONCLUSION

This approach was designed to have a more generalized way of anomalous note detection in the context of computational music content analysis. Melody evaluation approaches were analyzed to identify its limitations and barriers which lead to having weak performances. The effect of anomalous note changes was then highlighted among them and had an individual analysis based upon that effect.

Anomalous note changes were defined as the scale unrelated note changes of a melody. Effects of harmonies and other variations of melodies are ignored, and all the melodies were considered as just simple monophonic melodies. Among the identified uncommon note events, a rule set for its scale was defined to remark the anomalies in the note sequence in a melody. LSTM Neural Network has used for the final the anomaly detection which have already used in computer-based music content analysis approaches such as melody transcription, melody creation etc. The overall accuracy has obtained as 68% for the used dataset of unknown monophonic melodies.

FUTURE WORK

This study was mainly designed to focus on unknown monophonic melodies. A monophonic melody is the simplest texture of music. It is the biggest limitation as an approach for anomalous note change detection of any kind of melody. The monophonic melodies are considered to have no harmonic or any other variations within the melodies in order to detect the scale unrelated note changes as anomalous note changes. Therefore, this study of anomalous note change detection was designed to follow an accurate way of anomaly detection in a selected category of music melodies under certain conditions.

Identification of different types of variations of the pitch can provide more reliable results for anomalous note change detection regarding the scale unrelated note events. Thus, this study can be further implicated to have such a way to distinguish the variations of the pitch from anomalous note changes. However, for polyphonic melodies, it is difficult to identify anomalous note events among two or more simultaneous pitch sounds. But it is possible to conduct a more complicated approach based on this study, to perform under polyphonic melodies as well.

REFERENCES

[1] Dhanith Chaturanga and Lakshman Jayaratne, “Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches,” International Journal of Computing (JOC) by Global Science and Technology Forum (GSTF), 3(2):137-148, 2013.

[2] Rajitha Peiris and Lakshman Jayaratne, "Musical Genre Classification of Recorded Songs Based on Music Structure Similarity," European Journal of Computer Science and Information Technology (EJCSIT) by European Centre for Research Training and Development UK, 4(5):70-88, 2016.

[3] Dhanith Chaturanga and Lakshman Jayaratne, "Musical Genre Classification Using Ensemble of Classifiers," Proceedings of fourth International Conference on Computational Intelligence, Modeling and Simulation (CIMSim 2012), Kuantan, Malaysia.

[4] J. Downie, "Music information retrieval," Annual Review of Information Science and Technology, vol. 37, no. 1, pp 295-340.

[5] M. G. Viraj Lakshitha and K. L. Jayaratne, "Melody Analysis for Prediction of the Emotion Conveyed by Sinhala Songs" Proceedings of IEEE International Conference on Information and Automation for Sustainability (ICIAfS 2016), Sri Lanka.

[6] E D Nishan Senevirathna and Lakshman Jayaratne, "Audio Music Monitoring: Analyzing Current Techniques for Song Recognition and Identification", International Journal of Computing (JOC) by Global Science and Technology Forum (GSTF), 4(3):23-34, 2015.

[7] E D Nishan W Senevirathna and Lakshman Jayaratne, "A Highly Robust Audio Monitoring System for Radio Broadcasting". International Journal of Computing (JOC) by Global Science and Technology Forum (GSTF), 3(2):87-98, 2013.

[8] Rajitha Amarasinghe and Lakshman Jayaratne, "Supervised Learning Approach for Singer Identification in Sri Lankan Music," European Journal of Computer Science and Information Technology (EJCSIT) by European Centre for Research Training and Development UK, 4(6):1-14, 2016.

[9] "Understanding Music", www.musictheoryacademy.com/, [online] accessed on 2018-10-12.

[10] M. Mauch, K. Frieler, and S. Dixon, "Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory," The Journal of the Acoustical Society of America, vol. 136, no. 1, pp. 401-411, 2014.

[11] P. Zwan, "Automatic singing quality recognition employing artificial neural networks," Archives of Acoustics, vol. 33, no. 1, pp. 65-71, 2008.

[12] T. Nakano, M. Goto, and Y. Hiraga, "An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features," pp. 1706-1709, 2006.

[13] C. Toh, B. Zhang, and Y. Wang, "Multiple-feature fusion based onset detection for solo singing voice," Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR), pp. 515-520, 2008.

[14] M. Alonso, B. David, and G. Richard, "Tempo and beat estimation of musical signals," Proc. International Conference on Music Information Retrieval, vol. 4, pp. 158-163, 2004.

[15] A. Mesaros, "Singing voice identification and lyrics transcription for music information retrieval," Proceedings of International Conference on Speech Technology and Human-Computer Dialogue, Cluj-Napoca, Romania, pp. 1-10,2013.

[16] M. Ryynanen and A. P. Klapuri, " Modeling of note events for singing transcription," Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio, 2004.

[17] S. Bock and M. Schedl, "Polyphonic piano note transcription with recurrent neural networks," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 121-124, 2012.

[18] M. Mauch, C. Cannam, R. Bittner, G. Fazekas, J. Salamon, J. Dai, J. Bello, and S. Dixon, "Computer-aided melody note transcription using the tony software: Accuracy and efficiency," Proceedings of the First International Conference on Technologies for Music Notation and Representation, Research Institute in Musicology, pp. 23-30, 2015.

[19] A. de Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," The Journal of the Acoustical Society of America, vol. 111, no. 4, pp. 1917-1930, 2002.

[20] M. Mauch and S. Dixon, "pyin: A fundamental frequency estimator using probabilistic threshold distributions," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 659-663, 2014.

[21] D. Shipmon, J. Gurevitch, P. Pisellis, and S. Edward, "Time series anomaly detection; detection of anomalous drops with limited features and sparse examples in noisy highly periodic data," 2017.

[22] L. Bontemps, J. McDermott, N. Le-Khac, and V. Loi-Cao, "Collective anomaly detection based on long short-term memory recurrent neural networks," International Conference on Future Data and Security Engineering, Springer, pp. 141-152, 2016.

[23] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," European Symposium on Artificial Neural Networks, vol. 23, 2015.

[24] S. Mahendra, H. Patil, and N. Shukla, "Pitch estimation of notes in indian classical music," 2010.