PAPER NAME

report_final.docx

AUTHOR

Abhimanyu Kumar

WORD COUNT

4248 Words

CHARACTER COUNT

20835 Characters

PAGE COUNT

16 Pages

FILE SIZE

1.2MB

SUBMISSION DATE

May 28, 2022 8:52 PM GMT+5:30

REPORT DATE

May 28, 2022 8:52 PM GMT+5:30

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

2. INTRODUCTION

Tempo and Beat are one of the most important perceptions in different kinds of music. It can help them to identify various characteristics of music which can help them to easily understand and replicate the music according to the need. The beat of a sic is somewhat similar to that of the beat of a human heart which is said to be a sequence of perceived pulse positions that are equally spaced in time and which dearer scribed mainly by two factors which are phase and the period. The second term which is also important in music is Tempo which can be defined as the rate of pulse and we can formulate it by using the formal equal to the reciprocal of the beat period. While listening to music we sometimes follow the music with a synchronize tapping by our foot or hand which is very often for may people and hence it is somewhat can be called beat and if we lost the track of it which may happen due to change in the tempo which leads to rhythmic displacement but at some point, of time, we are able to recover easily from it and start our taping again. But automated beat and tempo tracking are quite more difficult than that of our cognitive process by us humans. There are some techniques that can be used to beat and tempo from an audio file which helps in many applications related to the music.

Many applications require automatic beat monitoring, including musical analysis, automatic rhythm alignment of various musical instruments, audio editing cut and paste operations, and beat-driven special effects. The present time, there are various tools to extract beat and tempo from an audio file that uses various algorithms which differ according to the applications of the music. There are currently many algorithms for musical tempo identification and beat-tracking. Over the last decade identification and extraction of beat and tempo has received major popularity as it is used to suggest songs according to the mood with the help of Artificial Intelligence and beat and tempo identification. Beat tracking may sound like a simple and straightforward concept but actually, it is an unsolved problem till now. An effective algorithm can easily sort out the beat and tempo of the music for a simple tune or sound. But in real life, audio files are much more complex as they have different components such as file type, number of channels, channel mapping, etc. These audio files can also consist of noise that is unwanted and create a hindrance while identifying the beat and tempo. A beat is an interference pattern between two sounds of slightly different frequencies that are perceived as a periodic fluctuation in loudness with a pace equal to the difference between the two frequencies in acoustics. While learning music we cannot find the difference or inaccuracy in the original which is the source of the original music sample and the student’s music which is the testing sample. This leads the student to not analyze the music and cannot gain any benefit which learning. So, providing or developing the algorithm in which it can solve the problem where the student can identify its mistakes and find the time while visualizing it directly. This algorithm requires the use of beat and tempo which helps us to identify the music and hence helps to find the mistakes or miscalculations made by the student in his practice. The approach provided basically focuses on finding the beat and tempo which helps us to find different characteristics and al alas provides a visual form of the music played.

3. PROBLEM STATEMENT

- Music Learning has always been difficult for beginners as it is not easy to record the mistake made during practice and visualize it. Suppose a student wants to learn some musical instrument, in the process of learning the student will phase many problems. So how can the student identify the problem where it's gone wrong this is a tough process.
- It is Even difficult for the instructor to compare their audio through which learner can get oversight in visual form. In the same way the student phase the problem of understanding, the instructor also gets some difficulties during the comparison. Identifying the point where the student got wrong and played some wrong tune, it's a time-consuming as well as difficult process.
- There is a necessity to bridge this gap by developing an algorithm that helps to visualize this audio file. There is no algorithm or we can say an A.I based technique has been developed in this field to resolve the problem and give a proper visual output to understand the error completely.

3. LITERATURE REVIEW

1. **Ali Mottaghi proposed to obtain Real-Time Beat Tracking:**

   In simulations, Online Beat tracking OBTAIN outperforms the current state of the art findings according to prediction accuracy. The estimated tempos are used to create the Cumulative Beat Strength Signal CBSS. This study performs real-time beat tracking for an audio source. The onset strength signal is used to determine when the game begins and ends. Peak detection is performed on the periodic sequence of beats among all CBSS peaks. Real-time performance is tractable, according to simulations.

   Proposed Algorithm:

   1. *Generating Onset Strength Signal (OSS):* Because no onset can be determined while working with a single sample at a time, this technique must be run on a series of samples. It must process the frame of samples in order to gain access to an array of samples in order to understand the pattern of beats. The audio file in question was divided into many windows. They believe that each window contains samples. The overlap ratio is also taken into account." This is a novel approach to the problem.

   2. *Tempo Estimation:* The approach presented in this section, which employs all of the frames, differs from the one they use, which only uses the frames that have appeared thus far. The candidate tempos are scored using cross-correlation with optimal pulse trains. Pulse trains are scored using the greatest and lowest cross-correlation values. They apply the tempo estimate technique outlined in our baseline, despite the fact that it is an offline algorithm. In the following stage, all instance tempos calculated from OSS frames will be added together. The

maximum and minimum tempo restrictions are connected to autocorrelation time delays.

3. *Cumulative Beat Strength Signal (CBSS):* The weighted sum of OSS in the corresponding frame and the value of CBSS of the last beat using various weights equals CBSS for a frame. The score assigned to each sample and the placement of the beats establish the beat power in the working frame. In order to construct CBSS for each frame, seek the preceding beat in the signal. CBSS is calculated by adding the words from the previous and current frames together. The frames can be graded based on their likelihood of being chosen as a beat. The overlap of the windows should decide whether the frames beat or not. The greatest value determines the optimum site.

2. Jean Laroche proposed Estimating Tempo, Swing, and Beat Locations:

This article describes approaches for calculating the swing and finding the beats in audio recordings under the premise that the tempo is constant. Some of the applications include automatic looping of audio tracks and synchronization of numerous audio recordings.

Proposed Algorithm:

1. Transient analysis: A percussive strike, note start, or quick shift in the signal might all create this. Despite the substantial reduction in the information given to the listener, a person can nevertheless appropriately tap his foot to this click track. Even though Transients aren't mentioned in the study, it explains a very conventional way of detecting them. We aim to detect when the signal's energy increases dramatically during the Transient analysis step. Replace the original audio with a succession of harsh clicks as an informal test of this idea. The signal is subjected to a short-time fourier transform.

2. Uses STFT (Short Time Fourier Transform)

3. **Miguel Alonso and Bertrand David research on Estimation of tempo and beat:**

This study describes an autonomous tempo tracking system that analyses recorded audio and calculates temporal beat position and beats per minute. To extract onsets, periodicity detection blocks, and temporal estimates, the frontend analysis is employed. The audio signal is considered to develop steadily from one to the other throughout the period of the analysis window. The result got out of the suggested technique is evaluated by the use of a huge library of snippets from various musical genres.

1. Onset Detection: A short-time Fourier transform, which converts the signal into a frequency domain, is used to examine the audio signal. The frequency-domain method outperforms methods related to the time-domain which is based on a direct analysis of the entire temporal waveform. The goal of onset detection is to

determine the location of the audio signal's most relevant elements. These occurrences are crucial in the perception of beats.

2. Periodicity Estimation: The detection function at the output of the onset detection stage is a quasi-periodic and noisy pulse train with big peaks during note assaults. The next step is to calculate the audio signal's duration. Traditional pitch determination procedures are employed in two ways. These methods have been utilized in the past.

3. Spectral Product: The principle behind the notion is that the signal is made up of strong harmonics. To find this frequency, the power spectrum is compressed by a factor.

4. Geoffroy Peeters proposed to obtain tempo detection and beat marking:

In this study, a unique technique to automated tempo estimation was described. The reallocated energy flow is the basis for the first suggested energy function. The Viterbi method is then used to find the most likely tempo and meter beat subdivision pathways. The suggested method's performance is evaluated using three databases. In this study, a unique technique to automated tempo estimation was described. This approach works for both percussion and non-percussion music.

1. Onset-Energy Function: To identify the tempo of a piece of music, it must be something relevant in terms of musical periodicity. The reallocated spectrogram can be used to increase temporal and frequency resolution, eliminate attack blurring and better distinguish extremely near pitches. A brief window is necessary for music without percussion. The frequencies are reassigned using the instantaneous frequency time derivative of the phase. The signal energy variation is employed in most approaches. The approach is based on the measurement of spectrogram fluctuations over time.

2. Tempo estimation: "Tempo states" Viterbi decoding The precise combination of tempo states is described as the set of considered tempo and the three considered meter/beat subdivision templates (mbst). This approach is similar to the onset energy function, but the observed periodicities, in this case, are dependent not only on the tempo frequency but also on the meter properties. The temporal course of tempo is the best technique to explain the observed periodicities. One of the three distinct meter beat subdivision templates that is examined is the triple simple. You might seek the most likely temporal succession of the condition based on your observation. At each point in time, the main periodicities are estimated. Viterbi decoding is the problem formulation.

5. Pavel Senin proposed obtain for Dynamic Time Warping Algorithm:

Handwriting and online signature matching, sign language recognition and gesture recognition, data mining and time series clustering, and computer vision and computer animation all employ dynamic temporal warping. The purpose of this project is to employ Dynamic Time Warping to help in software metrics analysis. By permitting the elastic modification of time series in order to detect comparable shapes with various phases, the time series similarity measure is an efficient technique to limit the impacts of shifting and distortion in time.

6. **Biswajit Das and Dhritikesh Chakrabarty proposed Lagrange's Interpolation:**

The idea method of interpolation entails representing the data that are numeric and have a suitable polynomial value and after that, we take the value of the dependent variable which we got from the polynomial and compute it corresponding to any given value of the independent variable, this requires the use of a formula to represent a given set of numerical data on a pair of variables. The value of the dependent variable corresponding to each value of the independent variable must be calculated again using the formula entered. The formula is required to express a collection of numerical data on two variables. The formula was created using the Lagrange formula. The derived formula was utilized to represent the whole Indian population. There has been a formula developed.

Here Lagrange's Interpolation formula is derived from using some variables let's say X and Y.

7. **Vijay Dahiya on Analysis of Lagrange Interpolation Formula:**

This paper explains how the interpolation formula was developed, how it was created, and what methodologies were employed. It demonstrates how several mathematicians contributed to projects such as The Imperial Standard Calendar was created using a second-order Gregory Newton interpolation method. An Indian astronomer and mathematician invented the method for second-order interpolation of the sine function in AD. The Lagrange interpolation formula is used to recreate the power series fitting. For picture scaling, the Lagrange Interpolation Formula is utilized. In this work, they describe the mathematical fundamentals of the Lagrange Interpolation technique. In medical image processing, the first comparison of interpolation algorithms was published. Many scholars are looking at the possibilities of employing the Hartley and Fourier transformers. They were unsuccessful.

4. METHODOLOGY

WORKFLOW

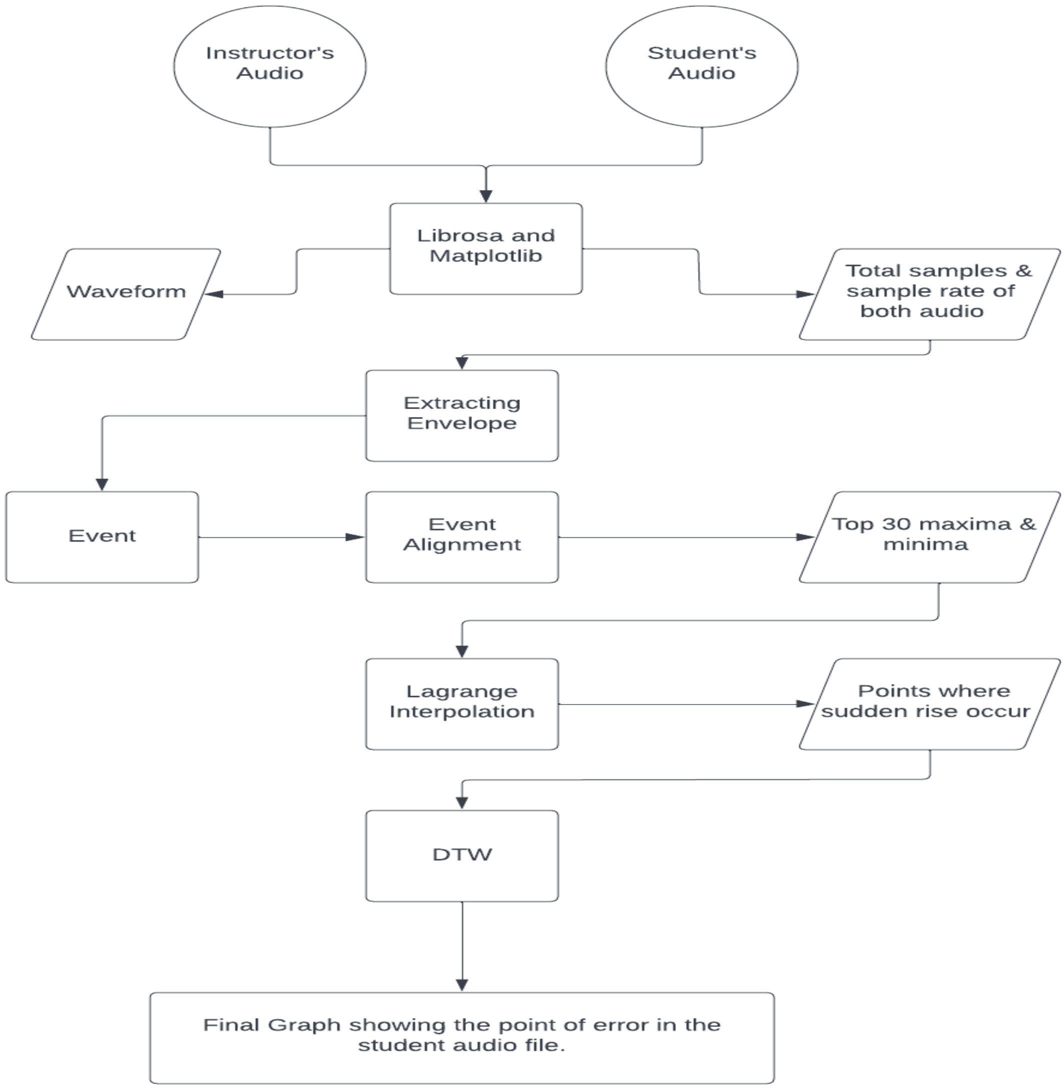


Fig 1

1. Audio File To Waveform

The audio files can be of different types like mp3, wav, flac, alac, and aac. Different audio files have different characteristics like mp3 file is a compressed file that will give low-quality audio and the size of the file is also small, whereas a wav file is a high-quality file with complete audio without any distortion, and the file size is also large.

Fig 2

When we convert an audio file into a waveform, the difference in audio file types high impact on the output waveform and its sample. Suppose you take an mp3 file and take the same audio in wav format. Now if you convert both the audio files into waveforms, both waveforms will be different from each other and the sample rate of both the file will also be different. The wav file will have a huge number of sample rates in respect of the mp3 audio file's sample rate.

There are several tools to convert the audio file into a waveform. We can also plot the waveform of an audio file by the use of different languages like Matlab, Scala, R, and Python.

We have used python for our waveform visualization where we have to use the Librosa library to load the audio file and after that, we have used IPython.display and matplotlib for displaying and plotting the waveform.

2. Envelope

- The concept of an envelope is to convert the idea of a continuous amplitude into an instantaneous amplitude. The envelope function might be based on time, space, angle, or any other variable in which we used the time and amplitude variable.
- Each half-wavelength of the modulating cosine wave influences both positive and negative values of the modulated sine wave, the modulation wavelength is double that of the envelope. The beat frequency is the same as the envelope, which is twice the modulating wave frequency, or 2f. Through this, we can use the concept of the envelope to find the beat as well as the tempo of an audio file.
- To find the envelope from an audio file we need to develop an algorithm that provides the maximum and minimum of each frame which are the small portion of the audio signal and when all the frame is combined, we can clearly see the envelope of the audio signal.
- There can be maxima and minima which were missed due to the local minima and maxima interference due to which we cannot get the ideal envelope which was required, so to eliminate that we have put a condition that if the maxima or minima in the frame is first or the last in the frame then we can eliminate or

avoid it. This is due to an increase in the waveform in the next frame which will lead to false maxima and wrong determining of the input.

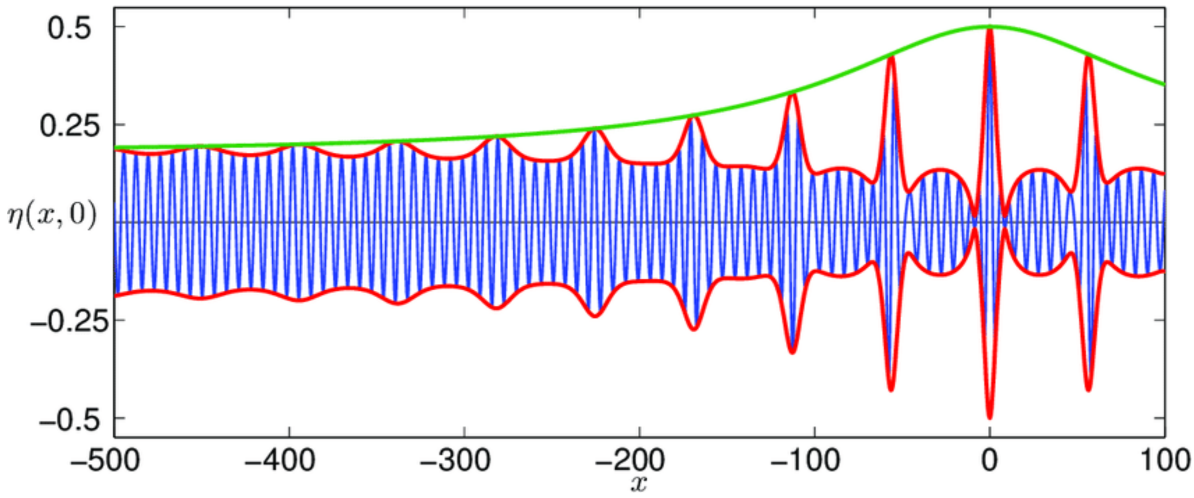


Fig 3

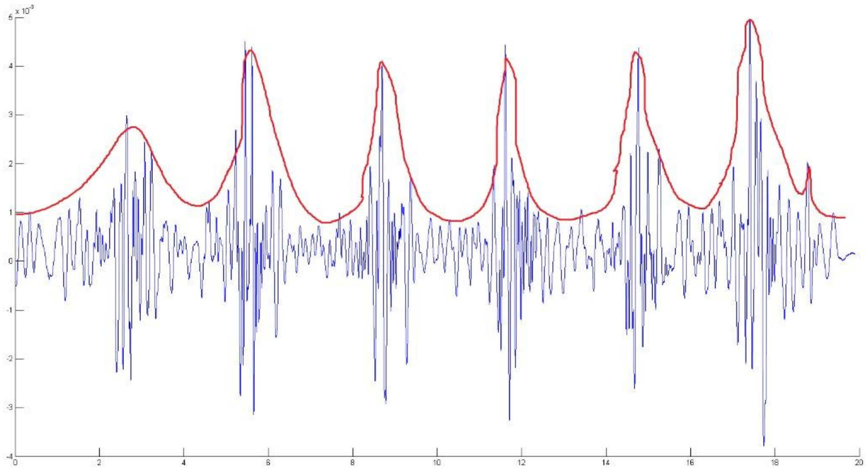


Fig 4

3. Lagrange Interpolation Polynomial (9)

- Interpolation is the method of estimating the value of a mathematical function for any intermediate value of the independent variable. Within the range of existing data points, there are new data points.

- In our Algorithm, it is used to find a sudden rise in the slope by calculating the equal distance between two base points with unequal distance

- The formula is represented as:

$$P_j(x) = y_j \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}.$$

Fig 5

- This formula can also be used for finding the value of the function even when the arguments are different.

- It can also be used to find the value of the independent variable.

Let's understand with different sets of graphs and equations

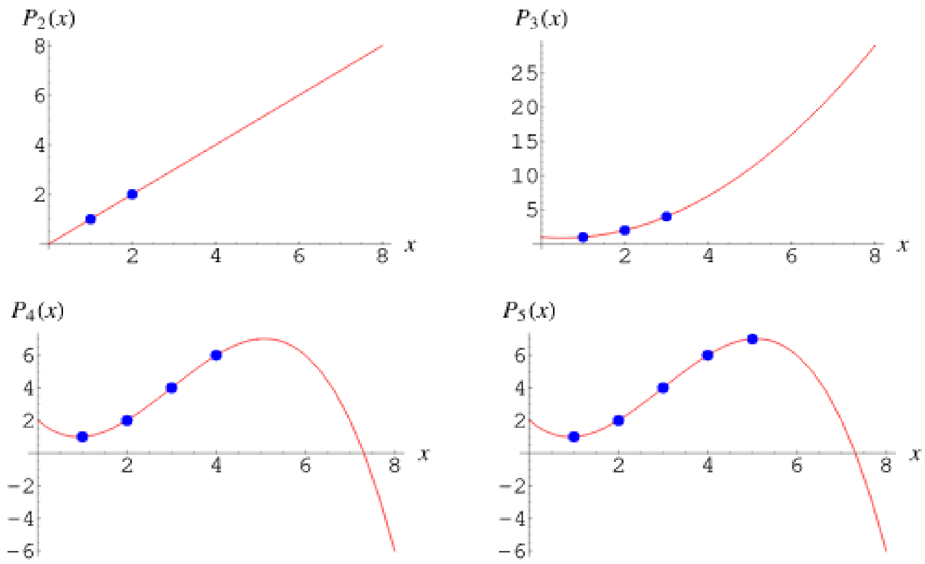


Fig 6

Here P(x) is the polynomial of degree <=(n-1) that go through n points like (x1, y1 = f(x1)), (x2, y2 = f(x2)), ……………, (xn, yn = f(xn)) and is given by

$$P(x) = \sum_{j=1}^{n} Pj(x)$$ where,

Some basic graphs

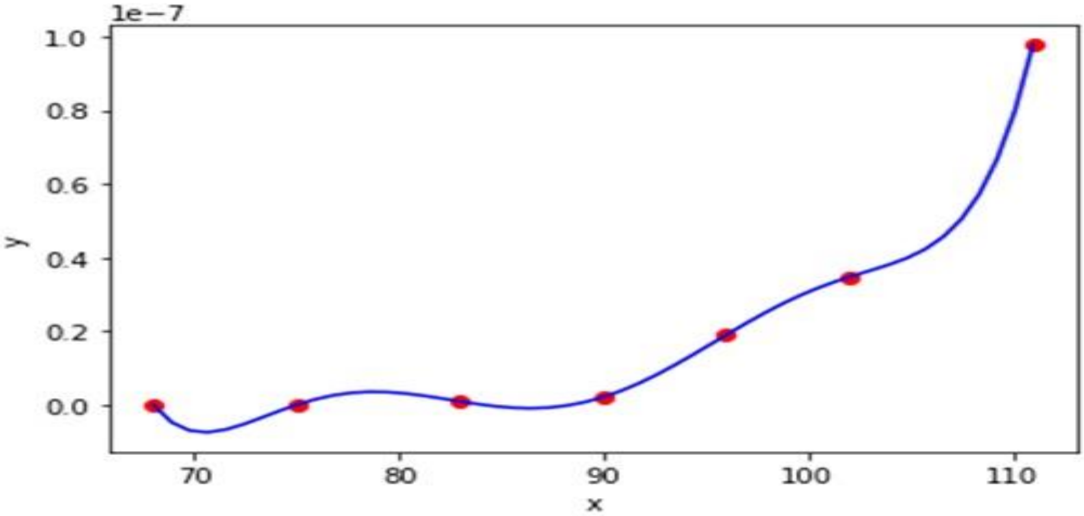


Fig 7

4. Dynamic Time Warping (DTW)

When the time index between comparison data points is not correct, dynamic time warping is employed to compare time series data. Once you grasp the notion of dynamic time warping, it's easy to find instances of its uses in everyday life as well as its fascinating future applications. Many diverse fields can benefit from dynamic time warping. You may imagine how beneficial it is to detect the wake words used in your gadget if your speech is sluggish and you haven't had a cup of coffee.

Use Cases

Sound Pattern Recognition: The sound pattern of a similar type can be detected. If we analyze the person's voice, we might be able to collect the Hello soundtrack from one situation. We need to match up the soundtracks of varying durations with the same individual in order to distinguish them from one other.

Stock Market: People usually want to anticipate the future of various stocks. However, employing a standard machine learning technique is time-consuming since it takes a large number of tests and training sets to have an equivalent dimension of features.

How We are Applying DTW in our project:

This method may be applied to pattern matching as well as anomaly detection. Examine the graph's red and blue lines and attempt to recall the classic time series matching. We all know that matching is a pretty limited procedure. Because the axes are different, dynamic time warping allows the two curves to meet up. This may be thought of as a strong dissimilarity score, with a lower number indicating that the series is more similar.

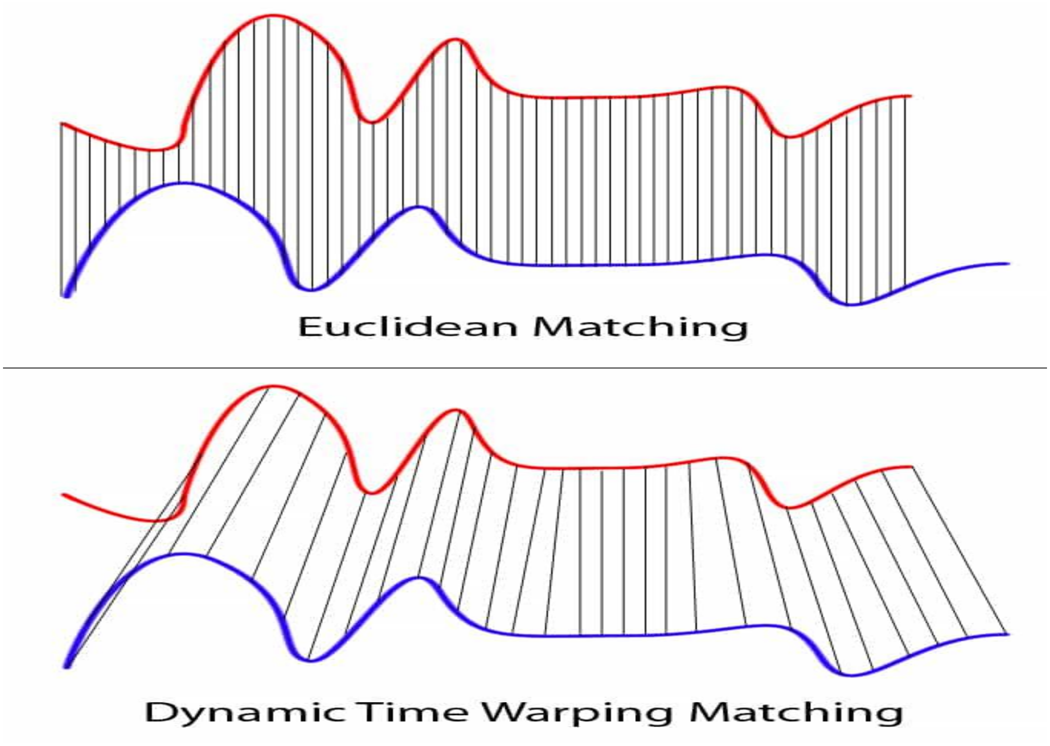


Fig 8

As can be seen in the accompanying graphic, the blue curve in these two series is significantly longer than the red. If we use the one-to-one match, the mapping is not flawless, and the tail of the blue curve is left out. So, in our method, we employed DTW to solve these concerns by one-many matching, which is why the troughs and peaks with the same pattern are precisely matched, and the two curves at the bottom top have no left out. This allows us to clearly distinguish between the audio of the teacher and the audio of the students.

In our situation, we used two audio files: one is the audio file of an expert, most likely a music teacher, and the other is the audio file of a student attempting to learn by referencing the instructor's audio. Our objective now is to apply dynamic temporal warping across audio samples in order to detect the similarity of those clips and assist both the instructor and the student in identifying and correcting errors.

Pseudo Algorithm

- Inputs: $x_{1:N}$ and $y_{1:M}$
- Cost matrix : $D \in \mathbb{R}^{(N+1)*(M+1)}$
- Initialization:
    - For a= 1 to I: $D_{a,0} = \infty$
    - For b= 1 to J: $D_{0,b} = \infty$
- Calculate cost matrix:
    - For a=1 to I:
        - For b=1 to J:
            - $D_{a,b} = d(x_a, y_b) + \min \{ D_{a-1,b-1}$ (match) / $D_{a-1,b}$ (insertion) / $D_{a,b-1}$ (deletion)
- Get alignment: Trace back from $D_{I,J}$ to $D_{0,0}$

Pseudo Code:

```
int distance (s: arr [1..i], t: arr [1..j]) {
        dynamic:= arr [0..i, 0..j]
        for a:=1 to i
                for b:=1 to j
                        dynamic[a, b] := infinity
        dynamic[0, 0] := 0
        for a:=1 to i
                for b:=1 to j
                        cost:= d(s[a], t[b])
                        dynamic[a, b] := cost + minimum (dynamic[a-1, b 1,//deletion
                                                        dynamic[a , b-1], // insertion
                                                        dynamic[a-1, b-1]) // match

        return dynamic[i, j]
}
```

(Ctrl)

5. RESULT

1. The first visual form of an audio file is formed by taking the audio file from the user and as can be seen the wavform depends on two parameters time and amplitude on x and y axis respectively.

```
plt.figure(figsize=(500,100))
plt.xlim(0,0.3)
librosa.display.waveplot(audio_array,sr=sr);
```

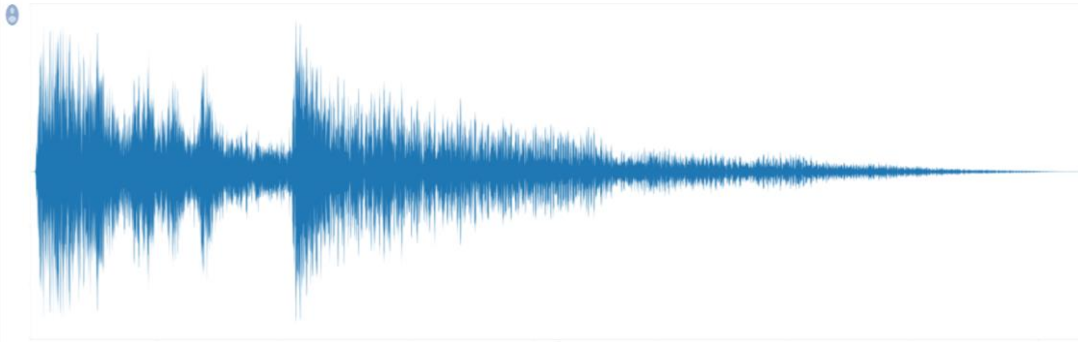
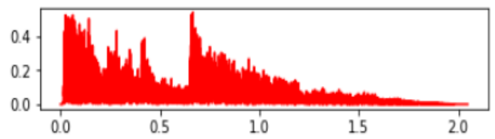


Fig 9

2. After that we applied an envelope to the audio to locate maxima and minima for each frame which is equal to 256 samples. Therefore, we set the hope size ranging from 7 to 21to reduce the frequency of the error in missing the maxima or minima of a frame.

```
[<matplotlib.lines.Line2D at 0x7f26d7b2f650>]
```




```
frames = range(len(ae_min))
t = librosa.frames_to_time(frames, hop_length=HOP_LENGTH)
# amplitude envelope is graphed in red

ax = plt.subplot(3, 1, 1)
plt.plot(t, ae_min, color="r")
```

```
[<matplotlib.lines.Line2D at 0x7f26d674c110>]
```

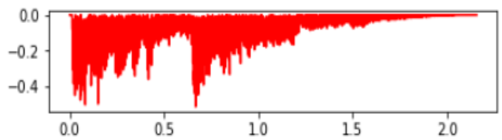


Fig 10

3. Then we applied the Lagrange interpolation technique to calculate the equal distance between two base points with unequal distance and final get slope which suddenly rise at some points. We have taken a small interval frame from the audio file to calculate LaGrange interpolation due to the large size of the audio file and sample due to which the sample was not clearly visible and observation cannot be made easily.

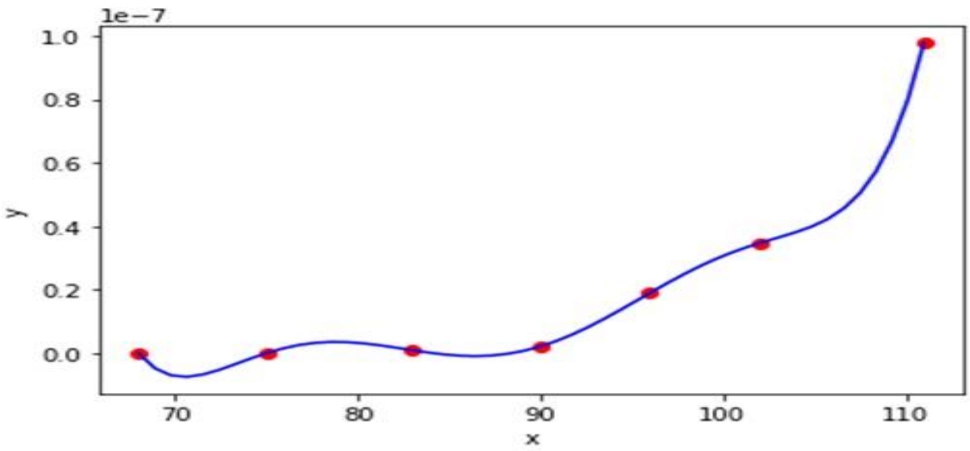


Fig 11

4. The final outcome will be a graph that will depict the difference between the original audio file to that of the testing file. So each point describes the time at which errors or mistakes were made by the testing file as a point can have one too many or many to one relationship with the other file sample.
   In this figure, we can see that there are two graphs parallel to each and showing some connections. The red one is the graph of an expert who is trying to teach the student who is graph is the blue one. Here when we see from red to blue if one point in the red match with multiple points in blue it's depicted that the student has played a tune or sung a Sargam multiple times by mistake and from blue-red if multiple points in red depicted to a single point blue, its represent student has missed some tune or any Sargam.
   As of now, we are applying this algorithm only to Sa Re Ga Ma Sargam song by an expert and a student. Then compare the audio of the two to get the difference.

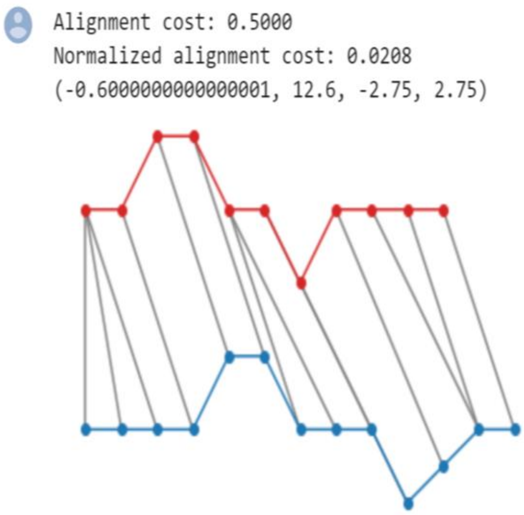


Fig 12

6. Comparison with other existing work and Analysis

1. Our project is to analyze the two Audio files where one of the audios is of an expert who is trying to teach their students. Presently the students take the reference of the audio of the instructor to become familiar with the music, in the process, the student makes a few mix-ups that will be detected by our program which will help to improve them.

2. Now we know that there are lots of manual ways the teacher can help the student in their learning. But during our research, we didn't find any A. I based on technology which is doing the same work in a visual way which helps the student to identify their mistake by seeing the visual representation with the point of an error made or we can time where they are playing the wrong tune.

3. So, our work is completely different and unique as we are trying to bridge this gap by providing this unique solution. We are trying a different approach to solve this problem.

7. Concluding Remarks

- Accuracy in the algorithm- One of our major shortcomings was that their algorithm was not fully accurate to provide the result in each stage of the process which leads us to lose some of the information which can be used for better accuracy for our application. One of the steps in which we have to find the envelope of the audio file was not accurate which lead us to miss some of the points where we require the maxima or the minima. It can be improved by modifying the algorithm with more sets of conditions which helps us to find only vital information which is required by us

- Limitation of Resources- our work is completely different and unique as we are trying to bridge this gap by providing this unique solution. We are trying a different approach to solve this problem which has not been used earlier by anyone. This leads us to the limitation of the approach as we require self-analysis to provide the next suitable approach to follow so that the previous work is unaffected and provides the result that we require in optimized time.

- Audio Types/Format – An audio file can be of various file formats such as .wav, .mp3, FLAC, .MPEG-4, .AAC, etc. have their own unique characteristics which differ from each other. These files can be compressed, or uncompressed to reduce the file size which is often using the lossy compressions. This makes us to limit our audio files to certain file formats. This can be overcome or improved by using an application or developing an algorithm that will provide the conversion from one audio type to another audio type.

Future Scope

This algorithm does not provide accurate output at certain points such as while constructing the envelope which leaves out some of the major maxima and minima from the envelope which leads to less accuracy.

This algorithm can further be integrated into the user interface where the user can input the audio of an expert as well as of the student and can analyze the mistakes in the students' audio file.

This interface can also be used as a real-time application in which, while playing the music the user will visualize errors simultaneously to instantaneous output.

6% Overall Similarity

Top sources found in the following databases:

- 3% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	en.wikipedia.org	Internet	1%
2	Michigan Technological University on 2019-10-21	Submitted works	<1%
3	Universiti Tenaga Nasional on 2016-01-21	Submitted works	<1%
4	Higher Education Commission Pakistan on 2021-08-31	Submitted works	<1%
5	eprints.hud.ac.uk	Internet	<1%
6	British University in Egypt on 2020-06-12	Submitted works	<1%
7	Yeditepe University on 2018-05-31	Submitted works	<1%
8	selby college on 2018-01-17	Submitted works	<1%

9	**UT, Dallas on 2011-06-16** Submitted works	<1%
10	**CSU, Los Angeles on 2013-08-17** Submitted works	<1%
11	**etheses.whiterose.ac.uk** Internet	<1%
12	**Miguel Alonso, Gaël Richard, Bertrand David. "Tempo Estimation for Au...** Crossref	<1%
13	**University College London on 2022-02-01** Submitted works	<1%
14	**Shah Muhammed Abid Hussain, A. B. M. Harun-ur Rashid. "User indep...** Crossref	<1%