STOCHASTIC MODELLING OF MUSIC

THESIS

Submitted by

NAMRATHA S

in partial fulfilment of the requirements for the award of the degree of

MASTER OF TECHNOLOGY

In

ELECTRICAL ENGINEERING

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS,CHENNAI

MAY 2020

THESIS CERTIFICATE

This is to certify that the project titled “**Stochastic modelling of music** ” being submitted to the Indian Institute of Technology Madras by Namratha Sunkesula (EE15B115), in partial fulfilment of the requirements for the award of the degrees of Bachelor of Technology in Electrical Engineering and Master of Technology in Electrical Engineering is a bona fide record of work carried out by him/her under my supervision. The contents of this project report, in full or in parts, have not been submitted to any other institute or university for the award of any degree or diploma.

Dr. Bharath Bhikkaji
Project Guide Professor
Dept. of Electrical Engineering
Indian Institute of Technology, Madras

Dr.David Koilpillai
Head of the Department
Professor in Electrical Department
Indian Institute of Technology, Madras

PLACE:Chennai

May 1 2020

ACKNOWLEDGEMENTS

I take immense pleasure to place on this report to my heart full gratitude and thankfulness o the following people for helping me to complete the project successfully.

First and foremost, I would like to express my sincere and deepest gratitude and thankfulness to my project guide, **Prof Dr.Bharath Bhikkaji**, Department of Electrical Engineering, IIT Madras for providing me an opportunity, invaluable guidance throughout the course of the project.

I am highly indebted to him for devoting his valuable time. I sincerely thank him for the help and motivation he provided in order execute this work in good time. His moral support, unreserved cooperation and generosity and extreme patience, which enabled me to complete the work successfully, will be everlasting in my memory.

I would like to express my thankfulness to the College of IIT Madras for letting me being part of this great institute.

I would like to thank God for giving me wisdom ,strength, support and knowledge in exploring things and for giving me determination to make this study possible.

-NAMRATHA S

ABSTRACT

Musical sound patterns are created and performed by some humans in order to communicate with others. Perception details not just sensing the world but also making sense of it. When you listen to music from a live orchestra , you hear guitar, oboes, violins, piano,tabala and so on, each playing distinct notes. But the sound waves travelling towards your ears do not come packaged as distinct channels for the winds, the strings; the signal the ear recognizes is the air pressure changing as a function of time, $p(t)$. Effectively sound of the whole orchestra is compressed into a single audio line. Notes and chords, melodies and harmonies are all variations created by the brain interpretively; they are “hidden variables” to be evaluated from an analysis of the signal.

Along with speech, music is one of the most prominent among communicative sound patterns that our species generates. Music is complex and diversified in style, context, and specific function.

Stochastic modelling of music through pattern theory gives hope to decode the complexities in music and allows to perform various operations.

LIST OF FIGURES:

Table of Contents

CHAPTER 1
INTRODUCTION

1.1 MOTIVATION

Listening to music and finding its composition is a fairly easy task for humans, it is even for normal listeners without any particular musical training. However, building stochastic models to perform this process is a tough problem. Further, the amount of music available in digital platform has already become unfathomable.

Music generation is very important nowadays.Mostly it became commercial. It can be used in many applications. Musicians build and produce their work on what is generated by the computational machine.

The significance of stochastic modelling in present days is vigorous and end-reaching. Stochastic modelling is applied in various industries around the world.
Stochastic modelling is a major hope in the analysis of the real signal patterns generated in any modality by the world, with all their naturally occurring ambiguity and complexity with the goal of constructing the processes, objects and events that produced them.

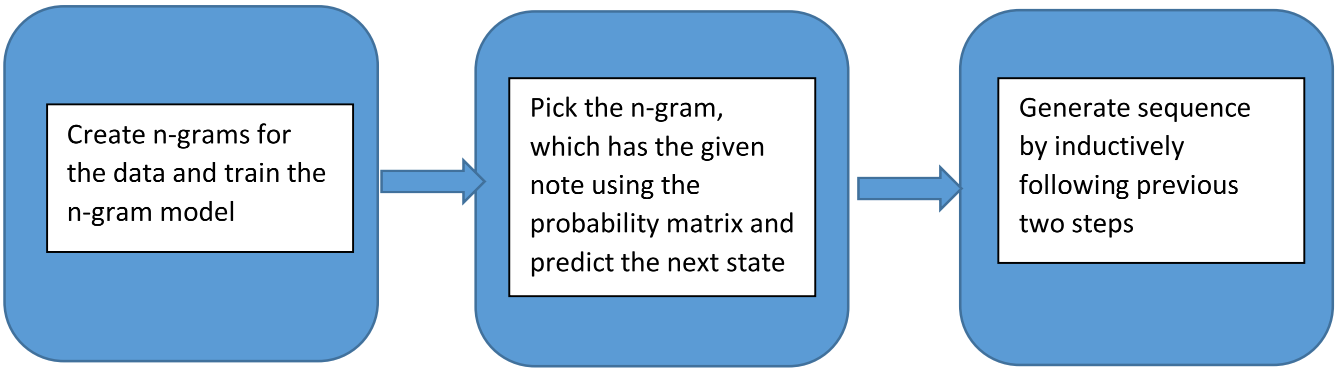
1.2 PROBLEM STATEMENT

The main aim of this project is to build stochastic model for music and this can be obatained by answering the following questions.

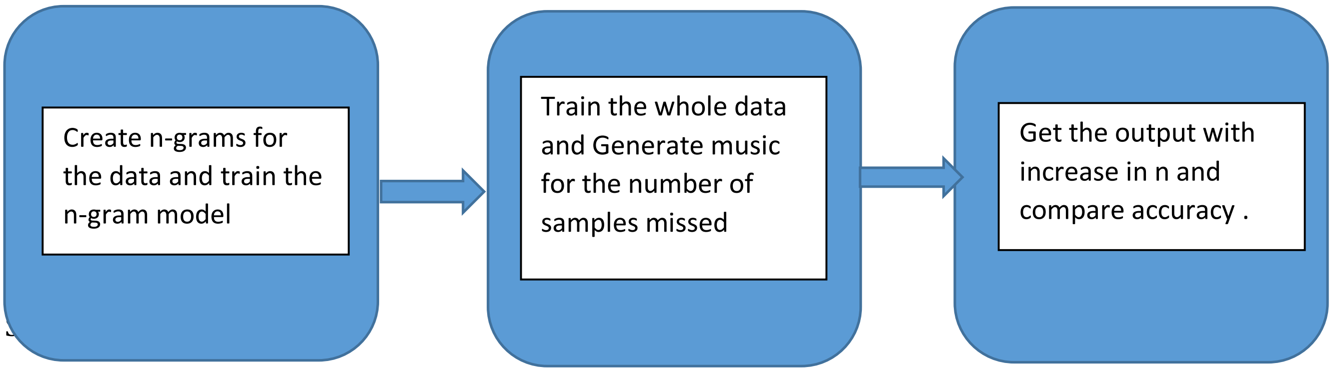
1. Generation of music using first order markov model and second order markov model.
2. Finding the broken music or missed notes using the trained markov model.
3. Comparing the outcomes of both first order markov model and second order markov model.
4. Building stochastic model for music using piecewise Gaussian distribution which is possible by building Gaussian model for each note of piano chords.On building model the main work is to segment the music.

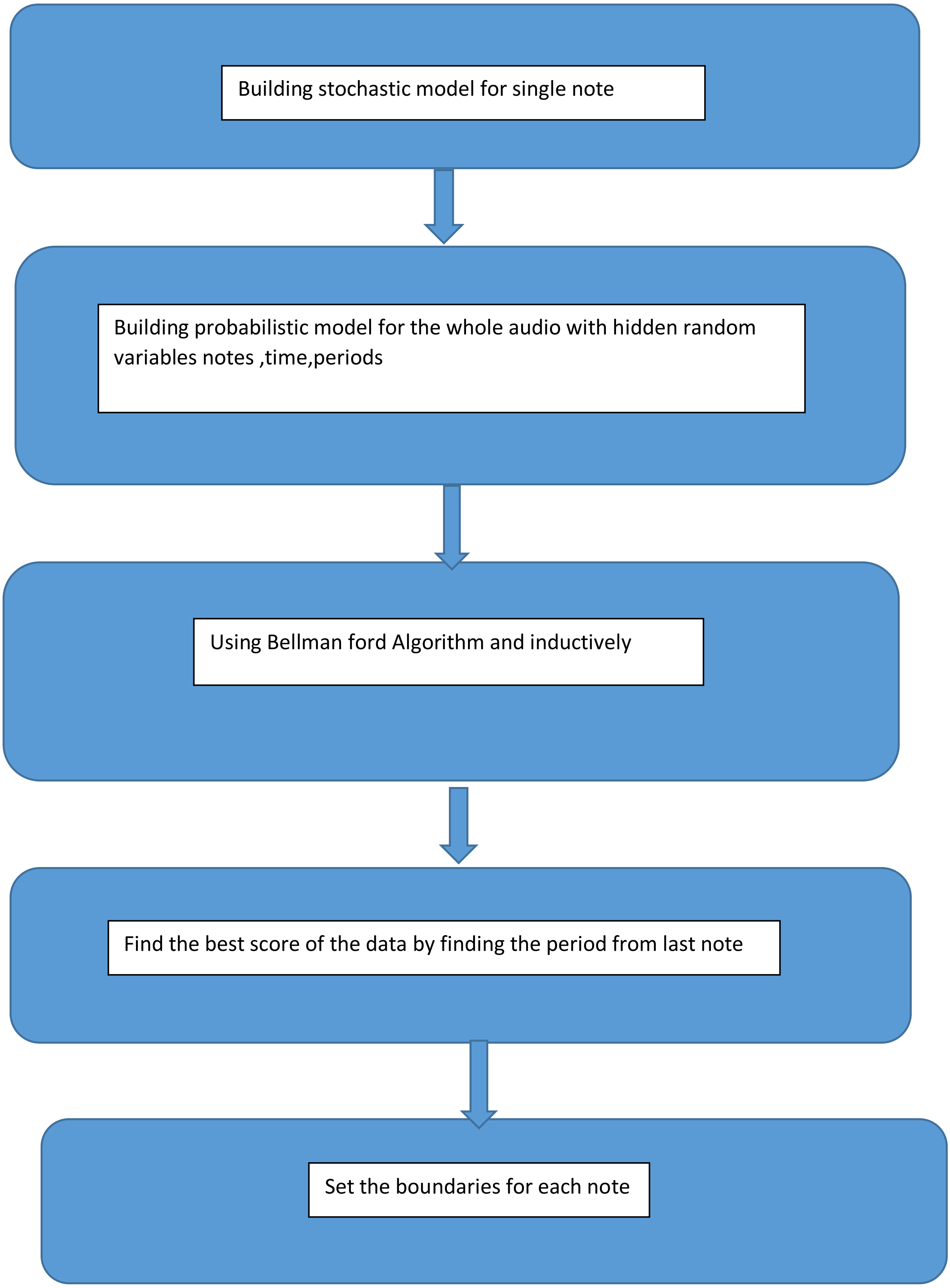
1.3 FLOWCHART

For Music Generation:



Comparing n- gram model with increasing n through finding music data missed :




Building stochastic model for single note
Building probabilistic model for the whole audio with hidden random variables notes ,time,periods
Using Bellman ford Algorithm and inductively
Find the best score of the data by finding the period from last note
Set the boundaries for each note

1.4 SOFTWARE USED

The software used to code the comaparison of n-gram markov model is Jupyter Notebook and programming language is python.

The software and the coding platform used for segmenatation of music is Matlab.

The music in .mp3 format is converted to .wav which is the input to Matlab code through Zamar Application.

1.5 SCOPE

By implementing python midi library package the music can be harmonized and can be used in real life.

And the stochastic modelling of music method can be applied to speech or image and the data can be reconstructed.

CHAPTER 2

BASICS

2.1 STOCHASTIC MODEL

A stochastic model represents a condition or situation where uncertainty exists.

In the real world, uncertainty is present every where , so a stochastic model could literally is applied to everything or *anything*.

Stochastic models contains some inherent randomness. The same set of initial conditions and parameter values will lead to an ensemble of different outputs.

Obviously, the real world is comprised by stochasticity But, stochastic models are considerably more hard and complicated.

Common Features of every stochastic model:

1. Reflecting all aspects of the question being studied,
2. Assiging probabilities to events within the model,
3. Those probabilities are used in making predictions or reveal otherinformation about the process.

2.2 MUSIC

Music is the systematical and chronological organisation of sounds; which makes particular sounds at particular times, which make sense in melodic, rhythmic and harmonic manner

Music is taken as an input which can be analysed through the frequency. Chords can be studied through the frequency is related.Here we are going to consider chords as inputs.

PITCH:

Pitch is defined as the highness or lowness of a sound.Some instruments may have high pitch while others might have a lo pitch.

For example, a flute has a high pitch, while a tuba has a low pitch.

A note or chord is a written representation of a particular pitch.

CHORDS:

There are totally 12 different *pitches*, or *notes*, in music.

Chords in music, are represented using the letters C, D, E, F, G, A and B.

These are names of notes, or pitches, as well as name of chords, or part of chord names. It is not the same thing though, a C note is just a note, whereas a C chord includes a couple of notes with C as the root note

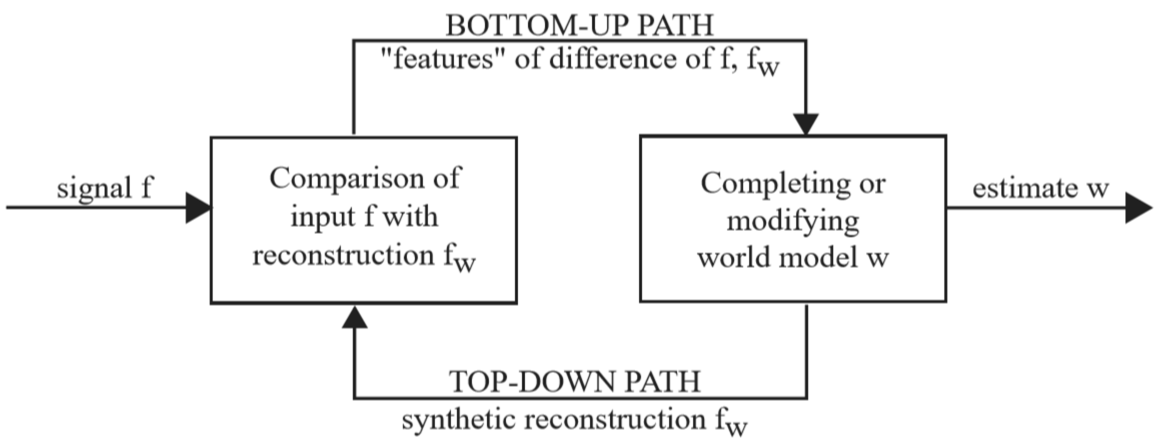
2.3 PATTERN THEORY

Real-world signals show two very distinct types of patterns.

We call these (1) value patterns and

(2)geometrical patterns.

The fundamental architecture of pattern theory follows as



CHAPTER 3

ST0CHASTIC MODELLING USING MARKOV CHAINS

3.1 BACKGROUND

Markov chains constitute the probability of transferring from one state to the next possible state in a sequence of events. Markov chains are generally used in learning algorithms when usually it is the abstraction of the data of probabilities which can be used to predict how the next forthcoming steps would be from the preceding steps that just have got completed or passed. Composition of music is an exciting content that can have Markov chains applied easily as a piece of music can be easily seen as a sequence of states, with each state as a chord or a note, for the specific length it is been played. As the notes available are finite, the options of lenth are finite either, even adding to the probability of many instruments, the categories of state will also be finite. Thus Markov chains which can be built with previous existing pieces of music of multiple genre and can be the basis for learning algorithm to make probabilistic decisions and create new musical pieces in the same genre.

3.2 FIRST ORDER MARKOV MODEL

In a first-order Markov model , the probability depends only on the present state, and not including any preceding transition state history. So, if we are currently present at State 1, and we want to know the probability of changing to State 2, we don't take any transitions prior to State 1

"Memorylessness" of this type is named as the **Markov property**: the probability of future actions is not dependent on the steps that led up to the present state.

TRANSITION MATRIX:

The transition matrix for a Markov chain is a stochastic matrix whose (i, j) entry gives the probability that an element moves from the jth state to the ith state during the next step of the process.

$\mathbf{M}^n\mathbf{p}$ is the probability vector after n steps of a Markov chain.

Here $\mathbf{p}$ is defined as the initial probability vector and $\mathbf{M}$ to be the matrix of transition.

A limit vector for a Markov chain is always a fixed point (a vector $\mathbf{x}$ such that $\mathbf{Mx} = \mathbf{x}$, if $\mathbf{M}$ is the transition matrix).

n-gram model:

Generally an ***n*-gram** is defined as a continuous sequence conataning *n* items from a given sample of speech or text .An ***n*-gram model** is a type of probability language model for being able to predict the next item in a sequence in the form of a $(n-1)$–order Markov model.

INPUT:

Same as natural languages we may think about music as a sequence of notes and consider chords as inputs.

3.3 MUSIC GENERATION:

3.3.1 BIGRAM

Here we will operate with chords.

If we take sequence of chords and learn its pattern we may notice the fact that certain chords might follow some particular chords more often, and other chords rarely follow that chords. We will construct our model to find and understand this pattern.

The steps need to be followed:

STEP1: A corpus of chords is taken as an input in csv file: Example of sequence.
[['B', 'C#m', 'B', 'A', 'E', 'B', 'A', 'E', 'B', 'A', 'E', 'B', 'A', 'B', 'A', 'E', 'B']

STEP 2: Calculate probability distribution for chords to follow a particular chord.
BIGRAM:

Bigram is a sequence consisting of two adjacent elements from a string of things, which are typically syllables, or letters, words. In this case it is chords.A bigram is an n-gram for n=2.

The bigram model evaluates the probability of a chord given all the chords existing previously by using only the conditional probability of one previous chord.

This assumption that the probability of a chord depends only on the preceding chord is also known as **Markov** assumption.

$$P(w_n|w_1^{n-1}) \approx P(w_n|w_{n-1})$$

Now let us assume chord F is taken as first chord in a sequence of chords,

18 bigrams are found that start with chord F:

STEP 3: PREDICT THE NEXT STATE

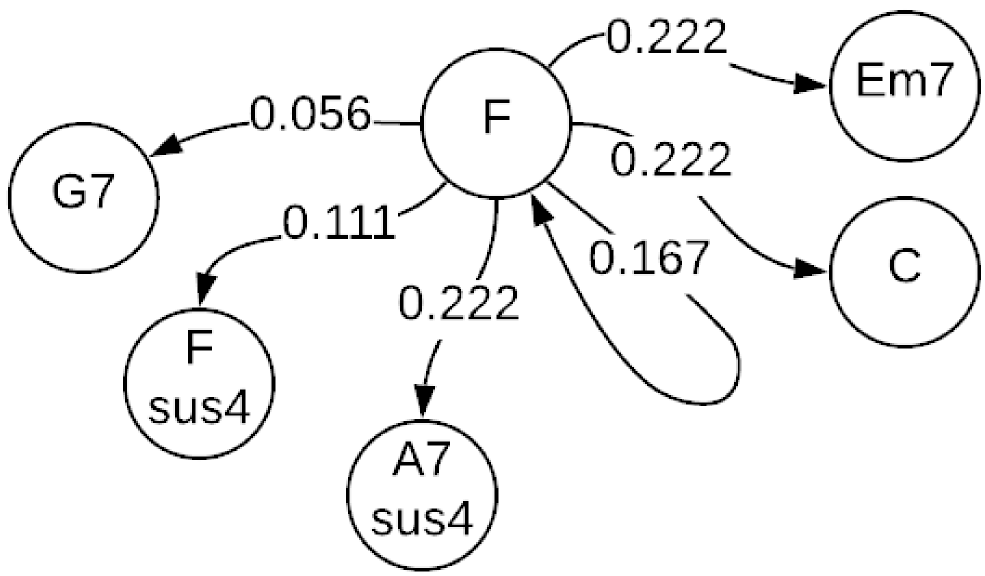
After creating the bigrams which start with the give state then we'll calculate the frequency of each unique bigram to appear in the sequence:

{'F Em7': 4, 'F C': 4, 'F F': 3, 'F A7sus4': 4, 'F Fsus4': 2, 'F G7': 1}

Then normalisation of the frequencies is done to get the proabablities.

{'F Em7': 0.222,
'F C': 0.222,
'F F': 0.167,
'F A7sus4': 0.222,
'F Fsus4': 0.111,
'F G7': 0.056}

This often can be interpreted in the form of a graph:



Weighted graph of possible next chord

Each node of this graph, except initial node F in the center, represents possible states that our sequence can achieve, in our case they are chords that may follow F. Some of the chords have higher probabilities than other, some chords can't follow F chord at all, for example, Am, because there was no bigram than combine this chord with F.

Now, Markov Chain is a stochastic process, or random process is you prefer. In order to move to the next state, we will be choosing chord randomly but according to the probability distribution, in our case, that means that we are more likely to choice chord C than G7.

Here we are predicting next chord in two different strategies

1.Random manner

The next chord is predicted randomly from the bigrams.

2.Probability wise

The next chord is predicted according to probability i,e preference is according to probability.

After calculating probabilities of each bigram containing the initial chord they are rounded off and multipled by 100.Now they are any natural numbers between 1 and 100.

Now using python listoflists we are generating lists mapping each bigram.

These list are numbers ranging according to proabblities.

Then a random number is picked from 1 to 100.

The list which conataing the random number is selected and the note mapped to it is the predicted as next state.

The bigram ['F','C'] is selected in a random way

The state predicted is 'C'.

STEP 4: GENERATING SEQUENCE

Create list.and by calling the predict next state sunction append all the next states into the list.There by new sequence of notes is formed.

The sequence is generate is

```
['Dm7', 'Bb', 'Dm', 'Gm6', 'C7', 'F', 'Em7', 'A7', 'Dm', 'Dm7']
```

3.3.2 TRIGRAM

Similarly to bigram,trigrams are created.

PROCEDURE:

STEP1: A corpus of chords is taken as an input in csv file: Example of sequence.

[['B', 'C#m', 'B', 'A', 'E', 'B', 'A', 'E', 'B', 'A', 'E', 'B', 'A', 'B', 'A', 'E', 'B']

STEP 2: Calculate probability distribution for chords to follow a particular chord.
TRGRAM:

Trigram is a sequence consisting of three adjacent elements from a string of things, which are typically syllables, or letters, words. In this case it is chords.A trigram is an n-gram for n=3

Trigrams are created like

```
['C#m B A', 'B A B', 'A B C#m', 'B C#m B', 'C#m B A']
```

STEP 3: PREDICITING THE NEXT STATE

Create list.and by calling the predict next state sunction append all the next states into the list.There by new sequence of notes is formed.

A second order Markov assumption on the state at time k would depend on state at time k-1 and time k.

$$P(w_1, \ldots, w_n) = \Pi_{i=1}^{n} P(w_i \mid w_{i-1})$$

In general representation the second order markov model is represented as above formula where n=2.

Now let us assume chord B and chord E are taken as first and second chord in a sequence of chords,

Trigrams with B followed by E are found.Trigram is picked according to most probability rule and the following note is selected as next state generated.

STEP 4: GENERATING SEQUENCE

The first state predicted is B.

Create list.and by calling the predict next state sunction append all the next states into the list.There by new sequence of notes is formed.

The sequence generated is

```
['B', 'A', 'B', 'C#m', 'B', 'A', 'C#m', 'B', 'A', 'E', 'B']
```

CHAPTER 4

RECONSTRUCT MUSIC FROM MISSING PARTS

4.1 BACKGROUND

We want to construct music in a song where some part is missing.

We will select a song and take the sequence of notes in text file.

And a part of song is missing.So we will use markov models to recreate it.

Here we will compare the original song with construted song and check the accuaracy.

In this way both first order markov model and second order markov model is used to construct back the song.And both models are compared interms of accuracy.

4.2 PROCEDURE

STEP 1:

Take sequence of notes of song.And remove some part of song.

Train the remaning notes of chords as input and generate music using the first order markov model as mentioned in above process.

STEP 2:

Similarly take the data with missing chords as input to second order markov model and generate music.

STEP 3:

Compare the output of both models with the missing part and check the accuracy of the chords matching in the generated sequences.

4.3 RESULTS OF INPUT 1

The input data is taken to be <u>data</u>.

And the data missed is
list1=['C#m','B','A','B','C#m','B','A','C#m','B','A','B','C#m','B','A','B','C#m','B','A']

Using the first order markov model the music generated is

List2= ['B', 'C#m','B','A','E','B', 'A','E','B', 'A', 'E', 'B', 'A', 'B', 'A', 'E', 'B']

The accuracy is calculated by measuring the number of elements common in the list which are: the element numbers 5, 6, 8, 9, 16.

So totally 5 chords are matched.

Similarly using the second order markov model the music generated is

List3= list2=['C#m','B','A','E', 'B','A','B','C#m', 'B','A', 'C#m', 'B', 'A','B', 'C#m', 'B','A', 'B']

The accuracy is calculated by measuring the number of elements common in the list which are: the element numbers   0 1 2 7 8 9.

So totally 6 chords are matched.

4.3.1 OBSERVATION OF OUTPUT 1

So by calculating the acccuracy of both first oder and second order markov model.

It is understood that second order model has good accuaracy compared to first order model.

By increasing n in n gram model and apply markov assumption accordingly the results will be better I,e with more accuracy.

4.4 RESULTS OF INPUT 2

The input data is taken to be data.

And the data missed is ['C','C','A','F','F','E','D','B♭','B♭','A','F','G','F']

Using the first order markov model the music generated is ['C', 'D', 'C', 'C', 'C', 'D', 'C', 'D', 'B♭', 'A', 'F', 'G','F']

The accuracy is calculated by measuring the number of elements common in the list which are: the element numbers   0 8 9 10 11 12

So totally 6 chords are matched.

Similarly using the second order markov model the music generated is

['C', 'C', 'D', 'C', 'F', 'E', 'D', 'B♭', 'B♭', 'A', 'F', 'G', 'F']

The accuracy is calculated by measuring the number of elements common in the list which are: the element numbers   0 1 4 5 6 7 8 9 10 11 12

So totally 11 chords are matched.

4.4.1 OBSERVATION OF OUTPUT 2

So by calculating the acccuracy of both first oder and second order markov model.

It is understood that second order model has good accuaracy compared to first order model.

By increasing n in n gram model and apply markov assumption accordingly the results will be better I,e with more accuracy

4.5 ACCUARCY COMPARISION BETWEEN FIRST ORDER AND SECOND ORDER

The accuracy between first order and second order model is compared by generating many output list through the 2 models and comparing them individually with the missed data list.

12 outputs generated from both first order and second order are taken to be comapared and the result obtained is:

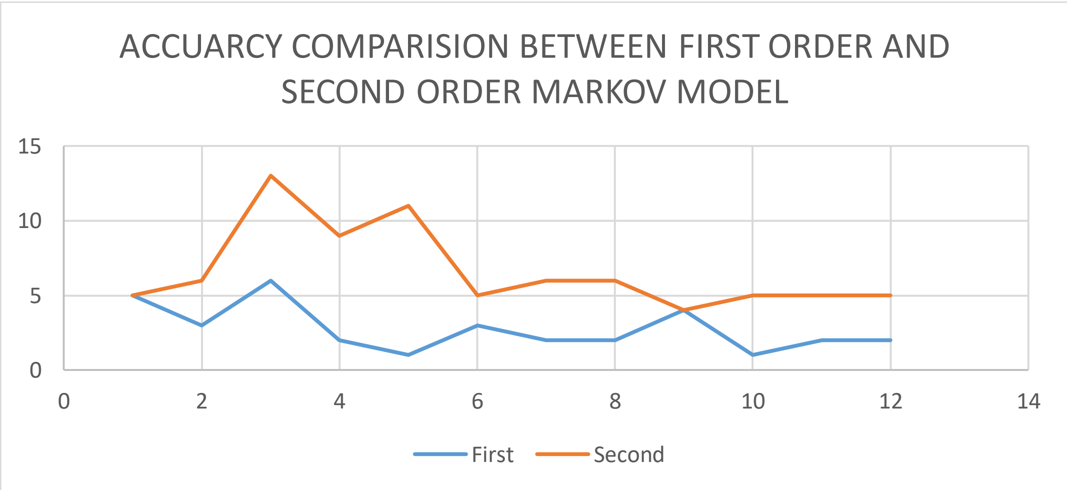


FIGURE 4. 1 ACCUARCY COMPARISION BETWEEN FIRST ORDER AND SECOND ORDER MARKOV MODEL

OBSERVATION FROM GRAPH

After comparing data from the hidden list with the generated lists after training the data to both first order and second order it is observed that out of 13 generated notes in a sequence first order has 5,3,6,2,1,3,2,2,4,1,2,2 number of notes exactly matched with given list.

Similarly out of 13 generated notes in a sequence second order has 5,6,13,9,11,5,6,6,4,5,5,5 number of notes exactly matched with given list.

We can say that second order results are matched more with given sequence compared to the first order markov model.

Similarly if we increase n in n-gram markov model the accuracy is increased.

CHAPTER 5

GAUSSIAN MODEL FOR MUSIC

5.1 BACKGROUND

Music is nothing but the air pressure varying with respect to time.We need to construct the stochastic model for the signal s(t) which represents music

A typical piece of musical data is taken with a sampling frequency of 8000 Hz so that if we consider 5 seconds of data ,we have a sequence s_k= s(kΔt),1≤k ≤40,000 of real numbers and we want a stochastic model for this finite-dimensional piece of data.

We need to put in the model extra hidden random variables which represent the patterns.

In this model, the main pattern i,e audio consists in what is usually called the "musical score". We need :

1. The number of notes m,
2. The times t_i = k_i(Δt) where new notes begin, 1 < k1 < k2 < .... < km < N
3. The frequency w_i of the note "i" in hertz (or its approximate integer period p_i ≈1/(Δt·w_i) ∈Z).

For developing this model, we will define a probability density p($\vec{s}$,m, $\vec{t}$, $\vec{p}$) in all the variables.Siginificantly we can use this distributon to rescore from a given signal,through recovering the hidden variables m, $\vec{t}$, $\vec{p}$ by maximizing the condtional proabability

$$p(m,\vec{t},\vec{p} \mid \vec{s}_{\mathrm{obs}}) = \frac{p(\vec{s}_{\mathrm{obs}}, m, \vec{t}, \vec{p})}{\sum_{m',\vec{t}',\vec{p}'} p(\vec{s}_{\mathrm{obs}}, m', \vec{t}', \vec{p}')}.$$

5.1.1 GAUSSIAN DISTRIBUTIONS

Let $\vec{x}$ =($x_1, \ldots\ldots, x_n$) denote a vector in R^n; we then define a Gaussian distribution onR^n; by its density

$$p(\vec{x}) = \frac{1}{Z} e^{-(\vec{x}-\vec{m})^t Q(\vec{x}-\vec{m})/2},$$

where $\vec{m} \in R^n$;, Q is a n × n symmetric positive definite matrix, and Z is a constant such that $\int p(\vec{x}\,)\mathrm{dx} = 1$.

CENTRAL LIMIT THEORM:

If $\vec{X}$ in Rn is any random variable with mean 0 and finite second moments, and if $(\overrightarrow{X^{(1)}}, \ldots\ldots\ldots \overrightarrow{X^{(N)}})$ are independent samples of $\vec{X}$, then the distribution of $(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \overrightarrow{X^{(K)}})$

tends, as N → +∞, to a Gaussian distribution with mean 0 and the same second moments

as $\vec{X}$

The definition of central limit theorm projects idea that the correct “default” probability models for continuous random variables are Gaussian.

The important properties used are:

1. $Z = (2\pi)^{\frac{n}{2}} (\det Q)^{-\frac{1}{2}}$.
2. $\int (\vec{x} - \vec{m}) p(\vec{x}) d\vec{x} = 0$, *which means that* $\vec{m}$ *is the mean of p, denoted by* $\mathbb{E}_p(\vec{x})$.
3. *If* $C_{ij} = \int (x_i - m_i)(x_j - m_j) p(\vec{x}) d\vec{x}$ *is the covariance matrix, then* $C = Q^{-1}$.

5.1.2 Fourier Analysis

When doing analysis of any pattern theory, we need to use only the discrete Fourier transforms (Fourier transforms on finite abelian groups), since the the transition from infinite to finite can cause much confusion.

If f is a periodic function of x with period 1, then the Fourier coefficients $\widehat{f_n}$ of f for n∈ Z and the inversion formula are:

$$\widehat{f}_n = \int_0^1 f(x)e^{-2\pi inx}dx; \qquad f(x) = \sum_{n=-\infty}^{+\infty} \widehat{f}_n e^{2\pi inx}.$$

5.2 The Gaussian Models for Single Musical Notes

we combine the ideas from n-dimensional Gaussian distributions and from discrete Fourier transforms. Let $\vec{s}$=(s_1,..., s_N) be a periodic signal (s_{N+1} = s_1). We first take the Fourier transform of $\hat{s}$ of $\vec{s}$.

The usual canonical basis of $\mathbb{C}^N$ is $\vec{e}^{(1)}, \ldots, \vec{e}^{(N)}$, where $\vec{e}^{(k)} = (0, \ldots, 1, \ldots, 0)$ (the 1 is at the k^{th} place). This basis is orthonormal. But instead we can choose another orthonormal basis: $\vec{f}^{(0)}, \ldots, \vec{f}^{(N-1)}$, where $\vec{f}^{(k)}$ is defined for $0 \leq k \leq N-1$ as

$$\vec{f}^{(k)} = \frac{1}{\sqrt{N}}(1, e^{2i\pi\frac{k}{N}}, \ldots, e^{2i\pi\frac{k(N-1)}{N}}).$$

This basis is the Fourier basis. If $\vec{s}$ is the signal, in the canonical basis we have $\vec{s} = \sum_{k=1}^{N} s_k \vec{e}^{(k)}$, and in the Fourier basis (using the inverse Fourier transform), we get

$$\vec{s} = \sum_{l=0}^{N-1} \widehat{s}_l \vec{f}^{(l)}.$$

Notice that if the signal $\vec{s}$ is real, then it has the property that $\widehat{s_{N-l}} = \overline{\widehat{s}_l}$. (This is analogous to the usual equivalence for the real Fourier transform: f is real iff $\widehat{f}$ satisfies $\widehat{f}(-\xi) = \overline{\widehat{f}(\xi)}$.)

5.2.1 CASE OF A MUSIC NOTE

Returning back to the question of finding a stochastic model for music.

We first construct a Gaussian model of a single note. Let ω be the fundamental frequency of the note being played and p = 1/ω be its period.

If the signal is s(t) then s(t + p) ≅ s(t), which means that the signal is close to be periodic, although in real music, there are always small residual variations.

By assuming that the expected value of $\int (s(t + p) - s(t))^2 dt$ is quite small

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

Take a discrete sample of the signal s and, for simplicity, we assume that s 'wraps around' at some large integer N, i.e $s_{N+k} = s_k$, and that p is an integer dividing N.

Let q = N/p, the number of cycles present in the whole sample. We'll analyze the simplest possible Gaussian model for s which gives samples which are periodic plus some small residual noise.

Its density is :

$$p_{a,b}(s) = \frac{1}{Z} e^{-a \sum_{k=0}^{N-1} (s(k) - s(k+p))^2/2 - b \sum_{k=0}^{N-1} s(k)^2/2} = \frac{1}{Z} e^{-\vec{s}^t Q \vec{s}/2}$$

where $a \gg b > 0$, $Q_{i,i} = b + 2a$, $Q_{i,i+p} = -a$, for $0 \leq i \leq N-1$ and otherwise 0.

On the one hand, we have

$$\sum_k (s(k) - s(k+p))^2 = \| s - T_{-p}(s) \|^2 = \| \widehat{s} - \widehat{T_{-p}(s)} \|^2.$$

Using the fact that $\widehat{s}(l) - \widehat{T_{-p}(s)}(l) = \widehat{s}(l)(1 - e^{2i\pi \frac{pl}{N}})$, we get

$$\sum_k (s(k) - s(k+p))^2 = \sum_l |\widehat{s}(l)|^2 |1 - e^{2i\pi \frac{pl}{N}}|^2 = 4 \sum_l |\widehat{s}(l)|^2 \sin^2(\frac{\pi pl}{N}).$$

On the other hand, we have

$$\sum_k s(k)^2 = \sum_l |\widehat{s}(l)|^2.$$

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi pl}{N}))|\widehat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi pl}{N}))|\widehat{s}(l)|^2/2}. \quad (1)$$

Then the expected power at frequency l is the mean of $|\widehat{s}(l)|^2$, which works out to be :

$$\mathbb{E}(|\widehat{s}(l)|^2) = \frac{1}{b + 4a\sin^2(\frac{\pi pl}{N})}.$$

Note that this has maxima $1/b$ if l is a multiple of N/p, that is all frequencies which repeat in each cycle ; and that all other powers are much smaller (because $a \gg b$).

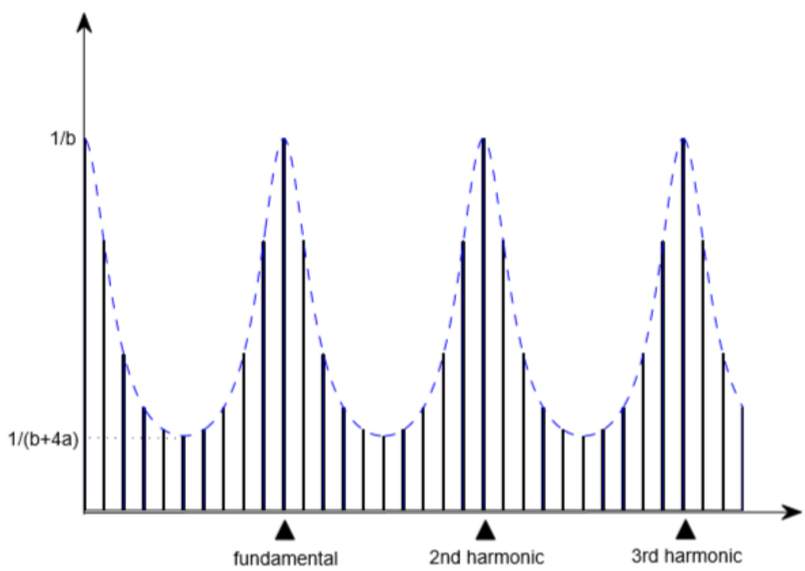


FIGURE: Expected power spectrum : $\mathbb{E}(|\widehat{s}(l)|^2) = 1/(b + 4a\sin^2(\frac{\pi pl}{N}))$.

This is not, however, an accurate model of real musical notes because the power in all harmonics (integer multiples of the fundamental frequency) is equally large. It is easy to change this and include extra parameters for the expected power of the various harmonics using the second expression in Equation (1).

5.3 The Geometric Model for Notes via Poisson Processes

The simplest model we can choose for the set of discontinuities is a Poisson process. These processes are precise mathematical description of what it means to throw down random points with a certain density.

5.3.1 Poisson Processes

A Poisson Process is a discrete model for a series of events where the information regarding *average time* between events is given, but the exact timing of events is random. The arrival of an event is independent of the event before (waiting time between events is memoryless).

Poisson disibution with mean λ(b−a) is given as

$$\mathbb{P}(\mathcal{D}_{a,b} = d) = e^{-\lambda(b-a)} \frac{(\lambda(b-a))^d}{d!}.$$

5.3.2 GEOMETRIC MODEL

The simplest model is gotten by taking the random variable $\vec{t}$ to be Poisson and each p_l to be independent of the other periods and uniformly sampled from the set of periods of all the notes the musical instrument is capable of producing (something like 'atonal' music). If per represents this set of periods, then this gives the form : p($\vec{p}$, $\vec{t}$m) = Ae^{-Cm} where

I{~ $\vec{p}$∈per^m}

Where a = log(|per|)+log((1−λ)/λ) and Z = $(1-\lambda)$.^{B}

5.4 MODEL FOR MUSIC

5.4.1 CONSTRUCTION OF MODEL

We construct the model for music in two stages

We recall that for this model, we need :

The sampled sound signal s, and hidden random variables :

The number of notes m,

The times $\vec{t}$ where new notes begin, and

The periods $\vec{p}$ of the notes.

The probability distribution $p(\vec{s}, m, \vec{t}, \vec{p})$ can be decomposed in the following way :

$$p(\vec{s}, m, \vec{t}, \vec{p}) = \prod_{l=1}^{m} p\left((\vec{s}|_{I_l}) \,\middle|\, p_l, t_l, t_{l+1}\right) \cdot p(\vec{p}, \vec{t}, m), \quad I_l = \{t \mid t_l \le t < t_{l+1}\}.$$

We have already constructed a Gaussian model for $p(s|_{[t_l,\ldots,t_{l+1}-1]} \mid p_l, t_l, t_{l+1})$:

$$p(s|_{[t_l,\ldots,t_{l+1}-1]} \mid p_l, t_l, t_{l+1}) = \frac{1}{Z} \exp\left(-a \sum_{n=t_l}^{t_{l+1}-p_l-1} (s_{n+p_l} - s_n)^2/2 - b \sum_{n=t_l}^{t_{l+1}-1} {s_t}^2/2\right),$$

where $a \gg b$. The simplest model is gotten by taking the random variable $\vec{t}$ to be Poisson and each p_l to be independent of the other periods and uniformly sampled from the set of periods of all the notes the musical instrument is capable of producing (something like 'atonal' music). If per represents this set of periods, then this gives the form :

$$p(\vec{p}, \vec{t}, m) = Ae^{-Cm}\mathbb{1}_{\{\vec{p} \in per^m\}}.$$

STAGE 2:

Segmentation.

5.5 FINDING BEST SCORE VIA DYNAMIC PROGRAMMING

Since music is a one-dimensional signal, we can compute by dynamic programming the best possible score or the mode of the posterior probability distribution in the hidden variables $\vec{m}$, p, $\vec{t}$.

The algorithm used for dynamic programming is Bellman Ford Algorithm.

5.5.1 BELLMAN FORD ALGORITHM

The dynamic programming algorithm of Bellman is a very efficient algorithm to compute the minimum of a function F of n variables (x_1,..., x_n), provided this function can be decomposed as the sum of functions fi(x_i,x_{i+1}).

The dynamic programming algorithm of Bellman is a very efficient algorithm to compute the minimum of a function F of n variables $x_1, \ldots, x_n$, provided this function can be decomposed as the sum of functions $f_i(x_i, x_{i+1})$.

Theorem

If $F(x_1, \ldots, x_n)$ is a real-valued function of n variables $x_i \in S_i$, S_i being a finite set, of the form

$$F(x_1, \ldots, x_n) = f_1(x_1, x_2) + f_2(x_2, x_3) + \ldots + f_{n-1}(x_{n-1}, x_n)$$

then one can compute the global minimum of F in time $O(s^2 n)$ and space $O(sn)$, where $s = \max_i |S_i|$.

The algorithm goes like this :

1. First initialize h_2 and Φ_2 by :

$$\forall x_2 \in S_2, \quad h_2(x_2) = \min_{x_1 \in S_1} f_1(x_1, x_2)$$
$$\forall x_2 \in S_2, \quad \Phi_2(x_2) = \operatorname{argmin}_{x_1 \in S_1} f_1(x_1, x_2)$$

2. We now loop over the variable k. At each stage, we will have computed :

$$\forall x_k \in S_k, \quad h_k(x_k) = \min_{x_1,..,x_{k-1}} [f_1(x_1, x_2) + ... + f_{k-1}(x_{k-1}, x_k)]$$
$$\forall x_k \in S_k, \quad \Phi_k(x_k) = \operatorname{argmin}_{x_{k-1}}(\min_{x_1,..,x_{k-2}} [f_1(x_1, x_2) + ... + f_{k-1}(x_{k-1}, x_k)]).$$

Then we define :

$$\forall x_{k+1} \in S_{k+1}, \quad h_{k+1}(x_{k+1}) = \min_{x_1,..,x_k} [f_1(x_1, x_2) + ... + f_{k-1}(x_{k-1}, x_k) + f_k(x_k, x_{k+1})]$$
$$= \min_{x_k}(h_k(x_k) + f_k(x_k, x_{k+1}))$$
$$\forall x_{k+1} \in S_{k+1}, \quad \Phi_{k+1}(x_{k+1}) = \operatorname{argmin}_{x_k}(h_k(x_k) + f_k(x_k, x_{k+1})).$$

3. At the end, we let $h = \min_{x_n}(h_n(x_n))$ and set :

$$\bar{x}_n = \operatorname{argmin}_{x_n}(h_n(x_n)), \quad \bar{x}_{n-1} = \Phi_n(\bar{x}_n), \cdots, \bar{x}_1 = \Phi_2(\bar{x}_2).$$

Then h is the minimum of F and $F(\bar{x}_1, ..., \bar{x}_n) = h$.

COMPLEXITY OF THE ALGORITHM:

If we look at the complexity of the algorithm, we see that at step k, for all x_{k+1} we have to search

$$\min_{x_k}(h_k(x_k) + f_k(x_k, x_{k+1})),$$

and since there are n steps, the complexity is in $O(ns^2)$. Moreover, we have to store all the $\Phi_k(x_k)$, which means that the complexity in space is $O(sn)$.

The key thing to realize about this algorithm is that at each intermediate stage, we don't have any idea what the best value of x_k is. But, for *each* value of this x_k, we will know the best values for all the previous variables. At first acquaintance, people often find the algorithm seems nearly trivial, but when they need to apply it, it is astonishingly strong. An example is the simple path problem.

5.5.2 FINDING BEST SCORE

The probability model for music is

$$p(\vec{s}, m, \vec{t}, \vec{p}) = Ae^{-Cm} \prod_{k=1}^{m} \frac{1}{Z_k} e^{-a(\sum_{t=t_k}^{t_{k+1}-p_k-1}(s(t+p_k)-s(t))^2/2)-b(\sum_{t=t_k}^{t_{k+1}-1} s(t)^2/2)}.$$

Then if we fix $\vec{s} = \vec{s}_o$ and define $E(m, \vec{t}, \vec{p}) = -\log p(\vec{s}_o, m, \vec{t}, \vec{p})$, we see that it is of the form $\sum_k f(t_k, t_{k+1}, p_k)$. We consider all possible scores on $[1, t]$ including a last note which ends at time t. The last note has a time of beginning $t' + 1 < t$ and a period p. Then for such scores we have :

$$E = E_1(\vec{s}_o|_{[0,t']},\ \text{notes up to } t') + E_2(\vec{s}_o|_{[t'+1,t]}, p) + E_3(\vec{s}_o \text{ from } t+1 \text{ on}).$$

Here E_1 assumes the last note ends at t', E_2 assumes there is one note extending from $t' + 1$ to t (so it has no other Poisson variables in it) and E_3 assumes a note begins at $t + 1$.
Using the algorithm of dynamic programming, we compute by induction on t the "best score" for the time interval $[0, t]$ *assuming* a note ends at t. Let $e(t') = \min E_1(\vec{s}_o|_{[0,t')},$ notes up to t') and assume by induction that we know $e(t')$ for all $t' < t$. We then find

$$e(t) = \min_{t'<t,p} [e(t') + E_2(\vec{s}_o|_{[t',t)}, p)],$$

and that continues the induction. Only at the end, however, do we go back and decide where the note boundaries are.

5.6 PROCEDURE

5.6.1 GENERATING INPUT SOUND THROUGH MATLAB

First we will take input data generated through MATLAB.

The notes of a piano are:

notes={'C' 'C#' 'D' 'Eb' 'E' 'F' 'F#' 'G' 'G#' 'A' 'Bb' 'B'}

And these above notes have their respective frequencies which are:

freq=[261.6 277.2 293.7 311.1 329.6 349.2...
370.0 392.0 415.3 440.0 466.2 493.9]

Now lets consider the input to be sequence of following notes:

song={'A' 'G' 'G' 'A' 'B' 'C' 'F' 'G'};

And the song is heard through inbuilt function sound.

THE INPUTS TO OTHER HIDDEN RANDOM VARIABLES

1.Number of notes m =8

2.Times at where the new notes begin :

A new note begins at every 0.5 second

the times ti = kiΔt where new notes begin, 1 < k1 < k2 < .... < km < N

And each note lasts for 4000 samples so a new note begins at every 4000 note

And as total no of samples are 8 then N=8*4000=32000 samples

K1 =4000,K2=8000……..

3. Frequency of ith note:

song={'A' 'G' 'G' 'A' 'B' 'C' 'F' 'G'};

w_1 = 440 Hz
w_2 = 392 Hz
w_3 =392 Hz
w_4 =440 Hz
w_5 =493.9 Hz
w_6 =261.6 Hz
w_7 =349.2 Hz
w_8 =392 Hz

PERIODS:

'C' ' 'D' 'Eb' 'E' 'F' 'F#' 'G' 'G#' 'A' 'Bb' 'B'

The approximate integer period 8000/frequency of note .

C	261.6	period =30
C#	277.2	period =28
D	293.7	period =27
Eb	311.1	period =25
E	329.6	period =24
F	349.2	period =23
F#	370.0	period =21
G	392.	period =20
G#	415.3	period =19
A	440.0	period =18
Bb	466.2	period =17
B	493.9	period =16

But since input song has following notes:

song={'A' 'G' 'G' 'A' 'B' 'C' 'F' 'G'};

The periods are

p_1 =18

p_2 = 20

p_3 =20

p_4 =18

p_5 =16

p_6 =30

p_7 =23

p_8 =20

5.6.2 BUILDING GAUSSIAN MODEL FOR SINGLE NOTE

We will construct Gaussian model for every note

1.NOTE B :

Its frequency is 493.9 Hz.
Period is taken to be 16.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 16.19Hz

By assuming that the expected value of $\int(s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 4.13*10(-21)

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 161.080

Taking the discrete sample of signal s and it wraps around some large integer N

N is taken as 49 Such that $s_{N+k} = s_k$,

q=3; three cycles of data.

Lets take two values to be a=5,b=30.
Construct a matrix Q which is in positive definite quadratic form.

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z}\Pi_l e^{-(b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

With a=5,b=30 the proability is obtained 0.9978.

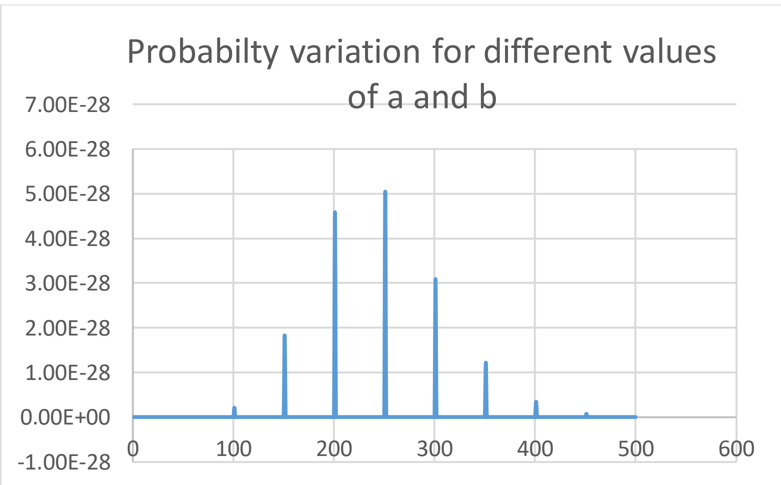


FIGURE 5.1 : Probabilty variation for different values of a and b for note B

EXPECTED POWER SPECTRUM:

Since we have taken values to be a=7;b=6

The limits are 1/b =0.033 1/b+4a=.002

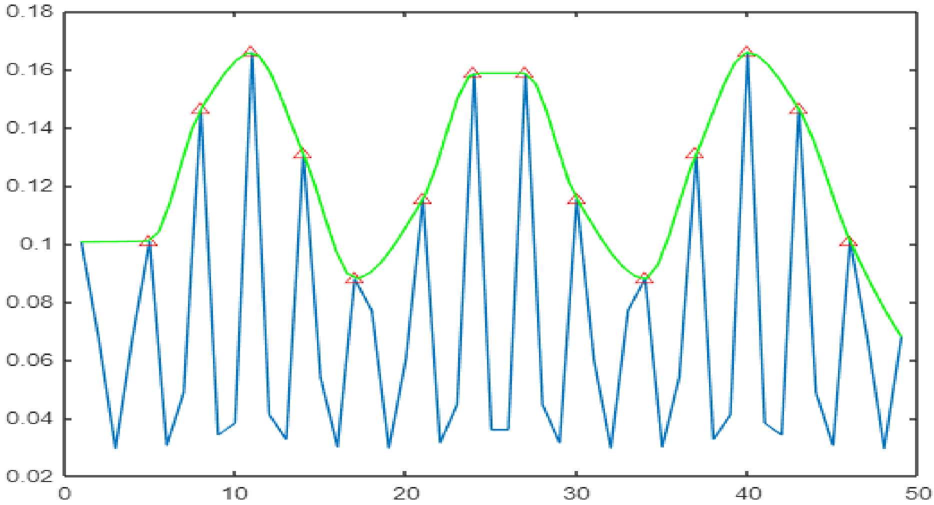


FIGURE 5.2 EXPECTED POWER SPECTRUM OF NOTE B AT a=7,b=6

A=5,b=1

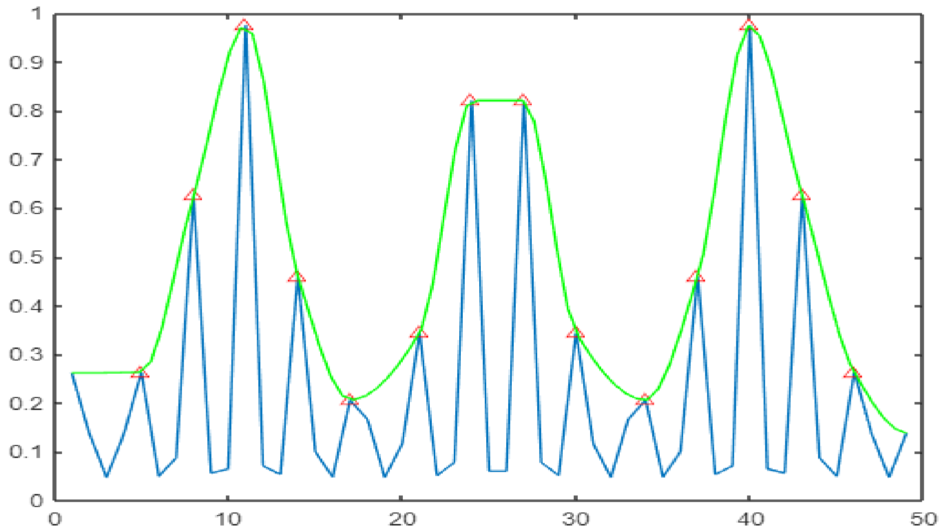


FIGURE 5.3 EXPECTED POWER SPECTRUM OF NOTE B AT a=5,b=1

2. NOTE Bb :

Its frequency is 466.2 Hz.

Period is taken to be 17.

Such that the property $s(t + p) \cong s(t)$ is valid.

By assuming that the expected value of $\int (s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 3.0774e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 171.4047

Taking the discrete sample of signal s and it wraps around some large integer N

N=52

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b + 4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

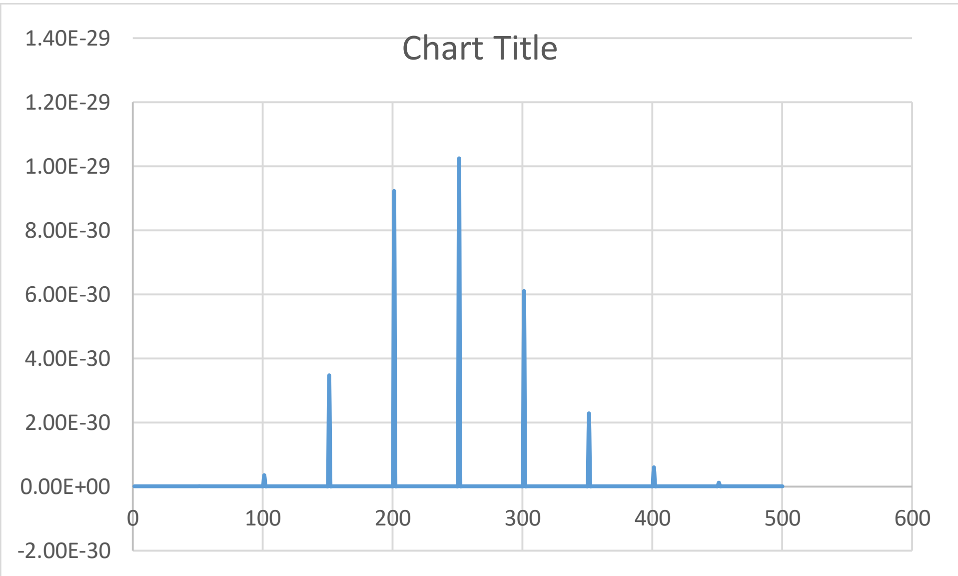


FIGURE 5.4 : Probabilty variation for different values of a and b for note B

A=5,b=1

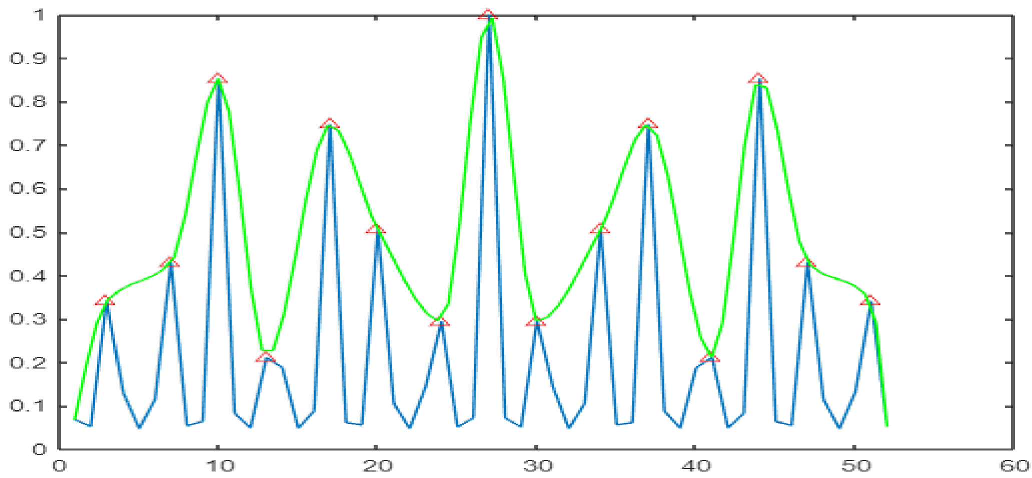


FIGURE 5.5: EXPECTED POWER SPECTRUM OF NOTE Bb AT a=5,b=1

A=5,b=10

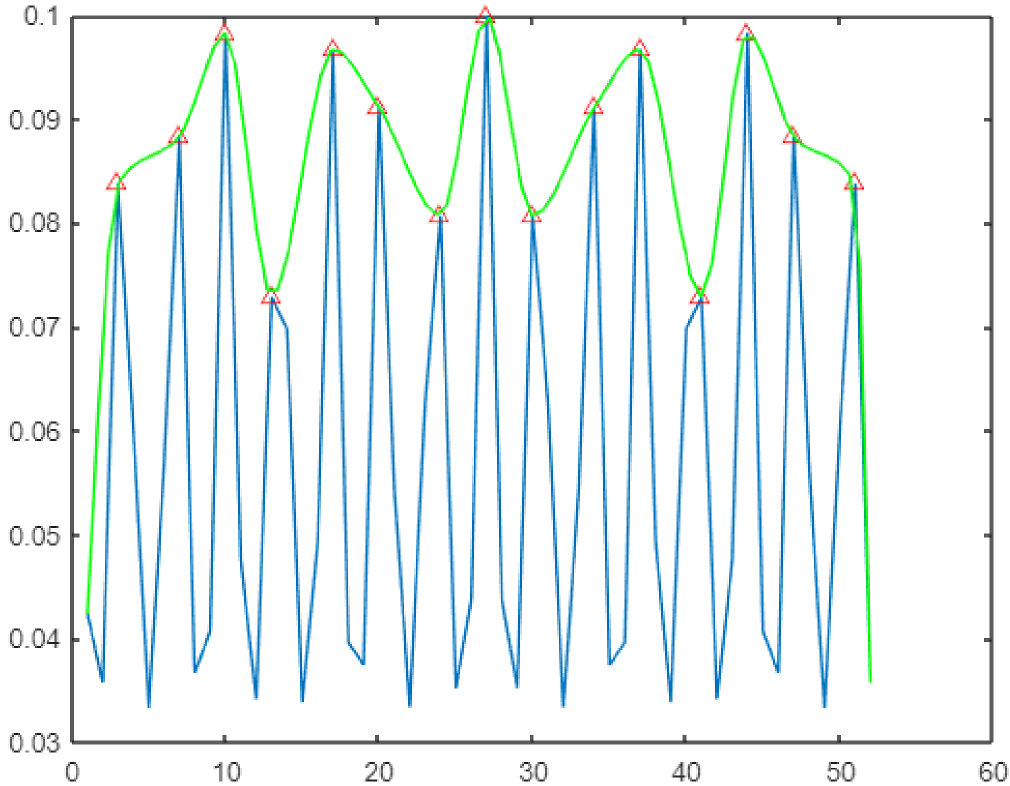


FIGURE 5.6: EXPECTED POWER SPECTRUM OF NOTE Bb AT a=5,b=10

3. NOTE A :

Its frequency is 440 Hz.
Period is taken to be 18.

Such that the property s(t + p) ≅ s(t) is valid.

By assuming that the expected value of $\int (s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 2.4923e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 181.2043

Taking the discrete sample of signal s and it wraps around some large integer N

N=55,. Such that $s_{N+k} = s_k$, q=3; three cycles of data.

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

a=5,b=30

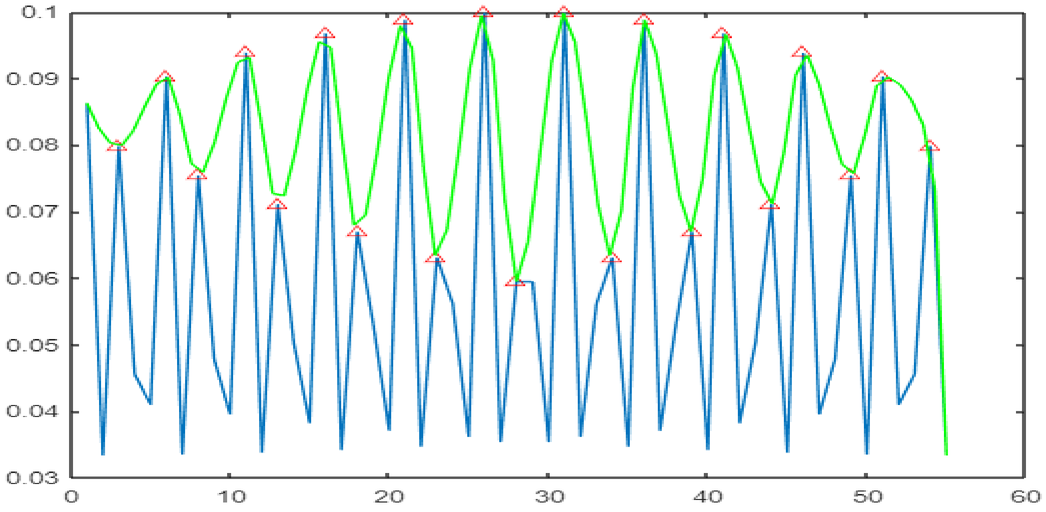


FIGURE 5.7 EXPECTED POWER SPECTRUM OF NOTE A AT a=5,b=30

4. NOTE G :

Its frequency is 392 Hz.
Period is taken to be 20.

Such that the property s(t + p) ≅ s(t) is valid.

By assuming that the expected value of $\int (s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 2.4923e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 181.2043

Taking the discrete sample of signal s and it wraps around some large integer N

N=62 Such that $s_{N+k} = s_k$, q=3; three cycles of data.

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

For a =5,b=30;

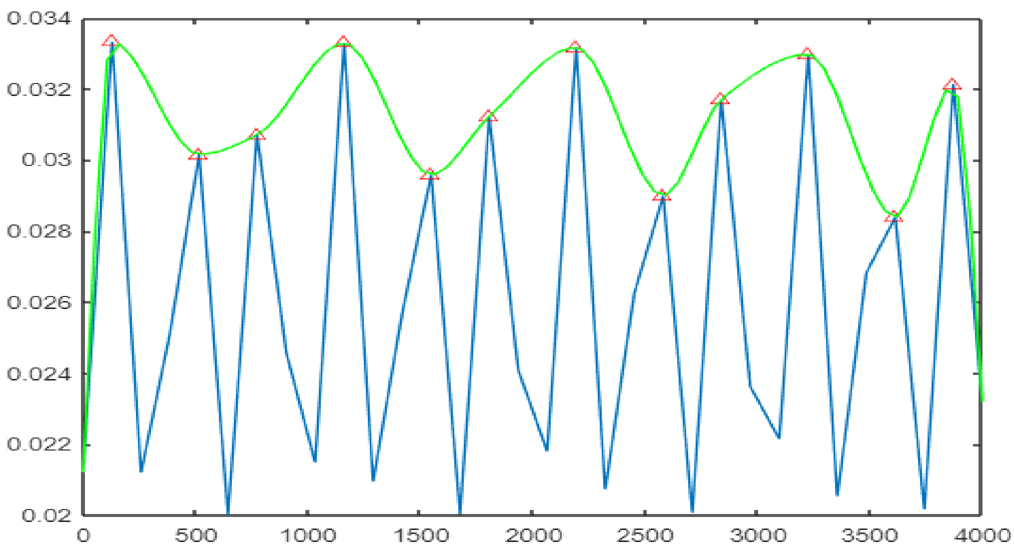


FIGURE 5.8 EXPECTED POWER SPECTRUM OF NOTE G AT a=5,b=30

5. NOTE G# :

Its frequency is 415.3 Hz.
Period is taken to be 19.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 19.26Hz

By assuming that the expected value of $\int(s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 3.9665e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 192.3735

Taking the discrete sample of signal s and it wraps around some large integer N

N=58 Such that $s_{N+k} = s_k$, q=3

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

For a=5,b=30

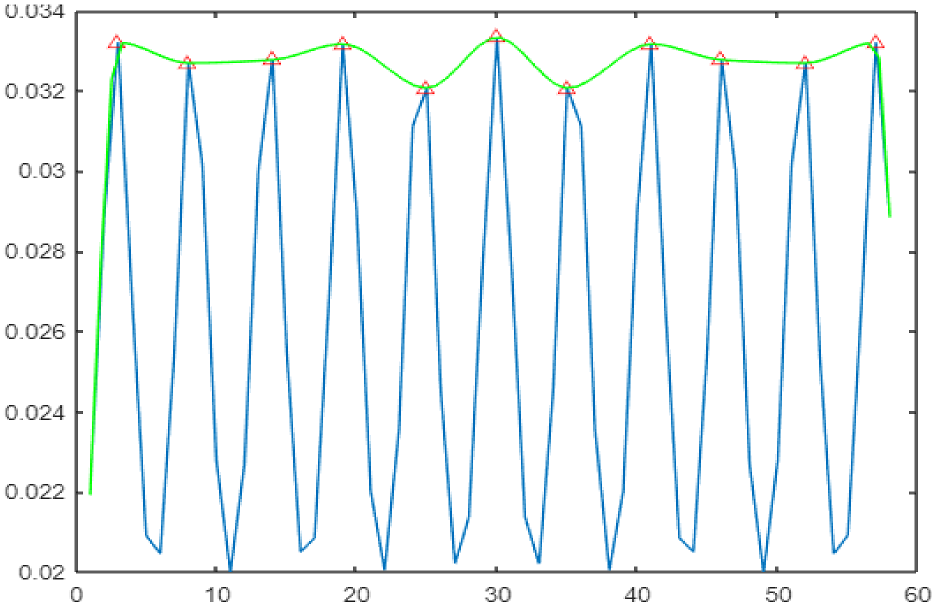


FIGURE 5.9 EXPECTED POWER SPECTRUM OF NOTE G# AT a=5,b=30

6.NOTE F:

Its frequency is 349.2 Hz.
Period is taken to be 23.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 22.909 Hz

By assuming that the expected value of $\int(s(t+p)-s(t))^2 dt$ is quite small

The value of obtained is 3.0172e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 228.9075

Taking the discrete sample of signal s and it wraps around some large integer N

N=69 Such that $s_{N+k} = s_k$, q=3

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z}\Pi_l e^{-(b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

For a=5,b=30

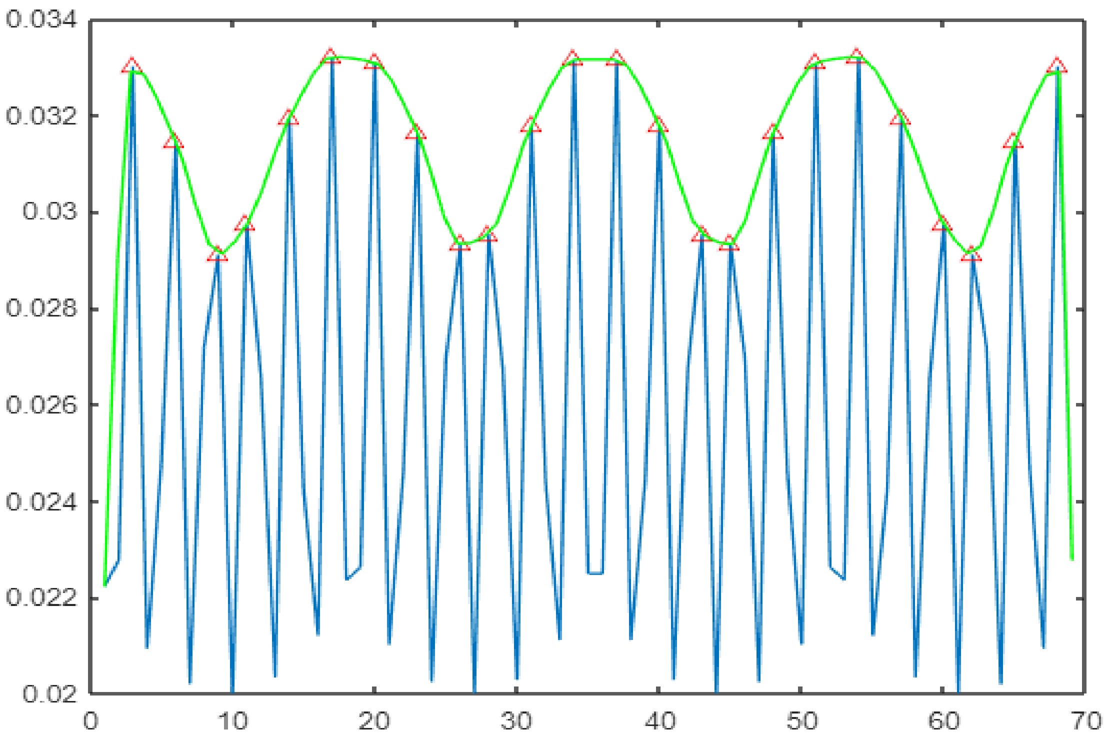


FIGURE 5.10 EXPECTED POWER SPECTRUM OF NOTE F AT a=5,b=30

7.NOTE F#:

Its frequency is 370 Hz.
Period is taken to be 21.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 21.621 Hz

By assuming that the expected value of $\int(s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 8.0879e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 215.7921

Taking the discrete sample of signal s and it wraps around some large integer N

N=63,. Such that $s_{N+k} = s_k$,

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

With a=5,b=30

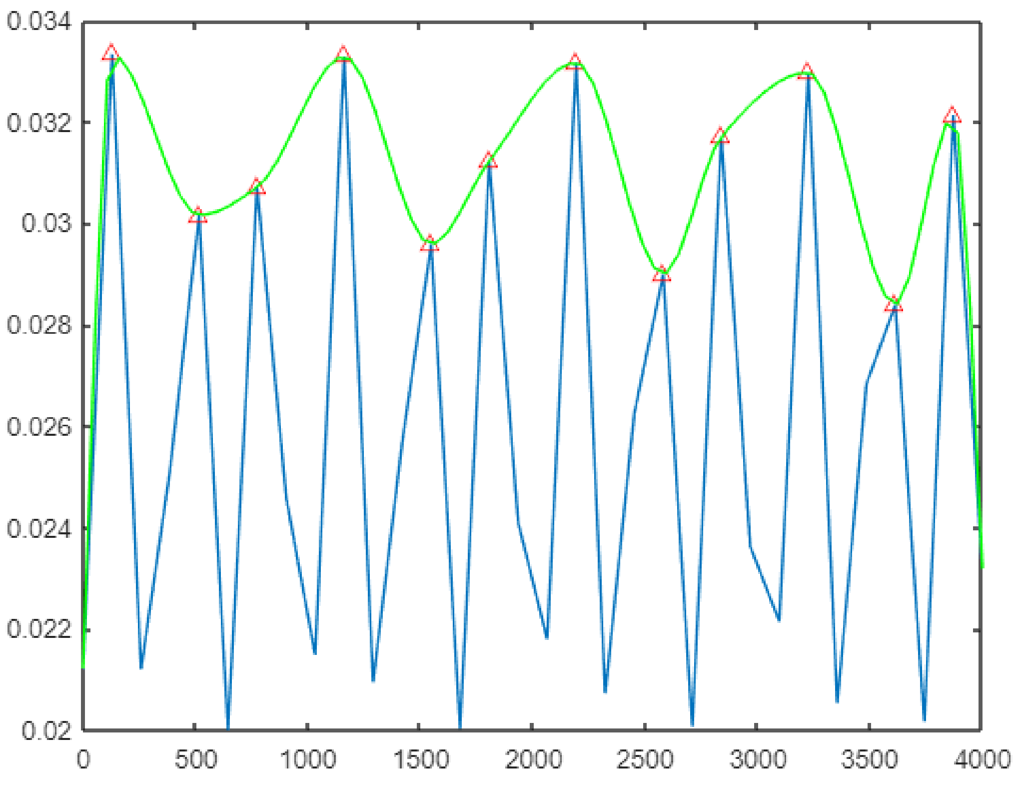


FIGURE 5.11 EXPECTED POWER SPECTRUM OF NOTE F# AT a=5,b=30

8. NOTE E:

Its frequency is 329.6 Hz.
Period is taken to be 24.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 24,271 Hz

By assuming that the expected value of $\int (s(t + p) - s(t))^2 dt$ is quite small

The value of obtained is 1.2355e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 242.2883

Taking the discrete sample of signal s and it wraps around some large integer N

N=73 ,Such that $s_{N+k} = s_k$

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

For a=5,b=30

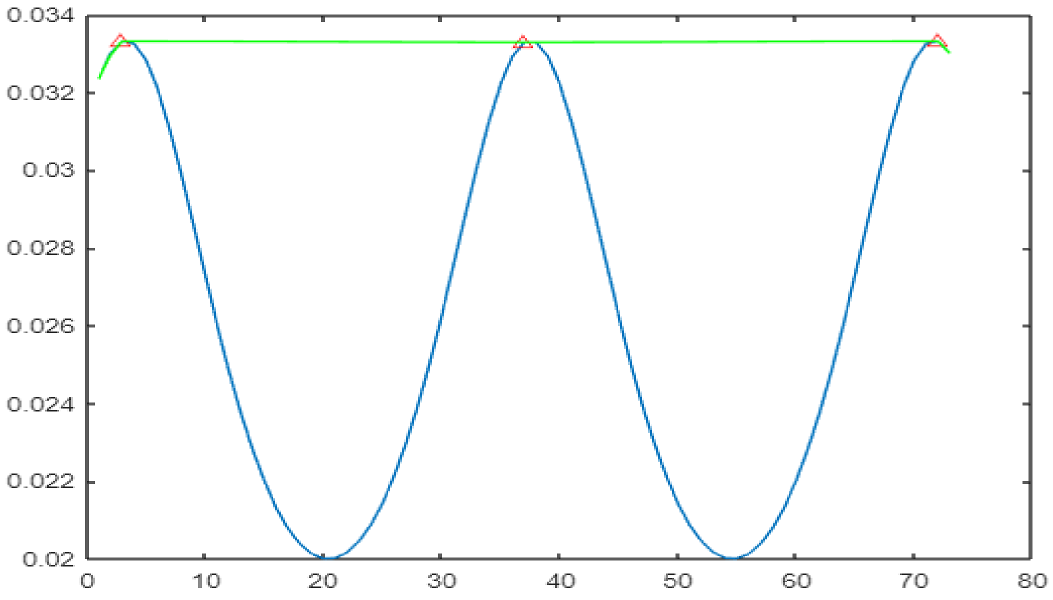


FIGURE 5.12 EXPECTED POWER SPECTRUM OF NOTE E AT a=5,b=30

9. NOTE Eb:

Its frequency is 311.1 Hz.
Period is taken to be 25.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 25.715 Hz

By assuming that the expected value of $\int (\mathrm{s}(\mathrm{t} + \mathrm{p}) - \mathrm{s}(\mathrm{t}))^2 \mathrm{dt}$ is quite small

The value of obtained is 8.0879e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 215.7921

Taking the discrete sample of signal s and it wraps around some large integer N

N=75 Such that $s_{N+k} = s_k$,

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z} e^{-\sum_l (b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z} \Pi_l e^{-(b+4a\sin^2(\frac{\pi p l}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

For a=5,b=30

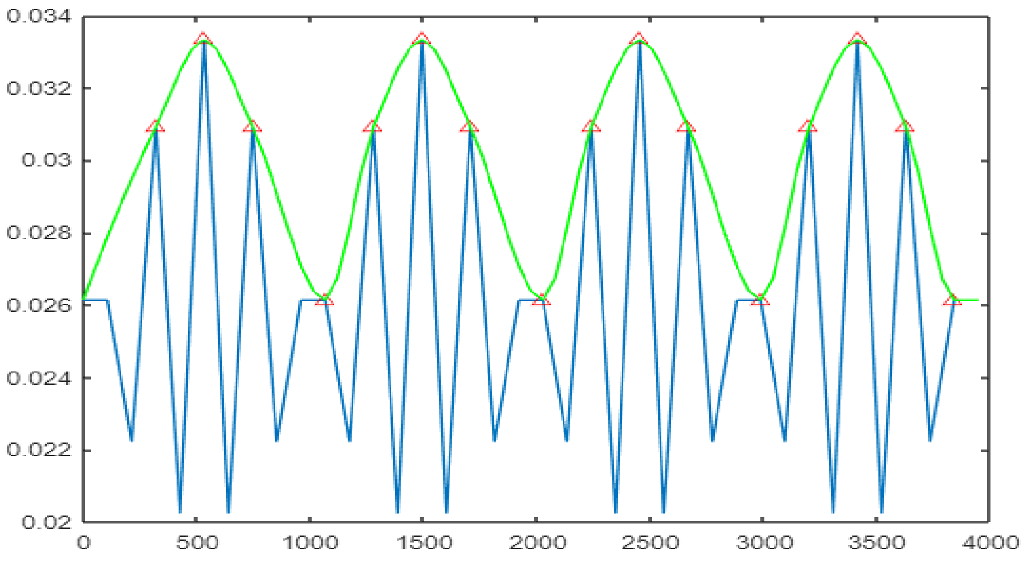


FIGURE 5.13 EXPECTED POWER SPECTRUM OF NOTE Eb AT a=5,b=30

10. NOTE C:

Its frequency is 261.6Hz.
Period is taken to be 30.

Such that the property s(t + p) ≅ s(t) is valid. But actual p= 30.58Hz

By assuming that the expected value of $\int(\mathrm{s(t + p) - s(t)})^2\mathrm{dt}$ is quite small

The value of obtained is 7.1568e-21

We then constrain the expected total power of the signal by bounding $\int s(t)^2 dt$.

The value obatained is 305.1973

Taking the discrete sample of signal s and it wraps around some large integer N

N=90,. Such that $s_{N+k} = s_k,$

Then later afer finding fft of the data the probability distribution is set

So

$$p_{a,b}(s) = \frac{1}{Z}e^{-\sum_l (b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2} = \frac{1}{Z}\Pi_l e^{-(b+4a\sin^2(\frac{\pi pl}{N}))|\hat{s}(l)|^2/2}.$$

Based on a,b the probability distribution is set for a single note.

With a=5,b=30

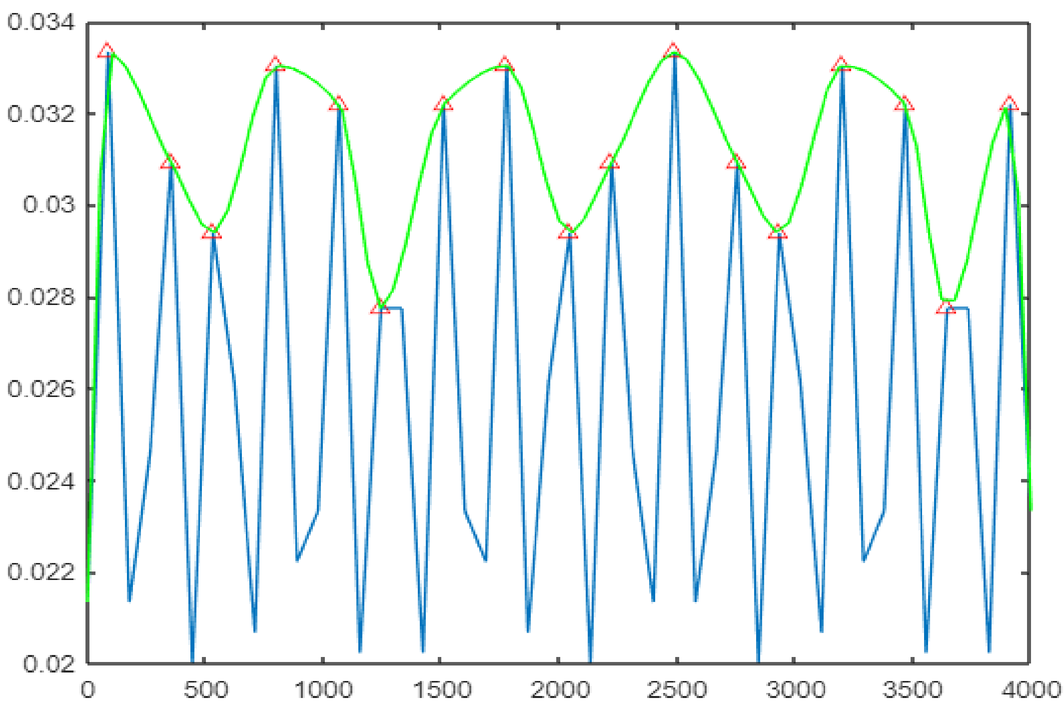


FIGURE 5.14 EXPECTED POWER SPECTRUM OF NOTE C AT a=5,b=3

5.6.3 BUILDING POISSON MODEL FOR TIME INTERVALS

To get a discrete model of a Poisson process, we just take $\Delta x = 1$, assume $\lambda \ll 1$, and describe each point of the sample $\mathcal{S}$ by the integral index $\mathcal{K}_l$ of the bin to which it belongs. Thus $\mathcal{S} = \{0 < \mathcal{K}_1 < \mathcal{K}_2 < \cdots < \mathcal{K}_m \leq B\}$ and

$$P(\mathcal{S}) = \lambda^m (1-\lambda)^{B-m} = \frac{1}{Z} e^{-a|\mathcal{S}|},$$

where $m = |\mathcal{S}|$, $a = \log((1-\lambda)/\lambda)$, and $Z = (1-\lambda)^{-B}$, similar to the Poisson density worked out above.

Assume no of notes in the total model to be n

λ=n/total no of samples

substitute in the above formula to get the poisson model

5.6.4 DYNAMIC PROGRAMMING

The goal is to find the note boundaries and the frequencies of the notes themselves by implementing the piecewise Gaussian model

Best possible music score from a given audio signal.

Since music is a one-dimensional signal, we can compute by dynamic programming the best possible score or the mode of the posterior probability distribution in the hidden variables .

Posterior Probability:

In Bayesian statistics a posterior probability, is the probability revised or updated of an event occurring after taking into consideration new information. Using Bayes' theorem the posterior probability is numbered after updating the prior probability Statistically the posterior probability is the probability of event 1 occurring given that event 2 has occurred.

We might make guesses about the note boundaries and periods based on local evidence, but this is often misleading.

Our probability model for music is

$$p(\vec{s}, m, \vec{t}, \vec{p}) = Ae^{-Cm} \prod_{k=1}^{m} \frac{1}{Z_k} e^{-a(\sum_{t=t_k}^{t_{k+1}-p_k-1} (s(t+p_k)-s(t))^2/2) - b(\sum_{t=t_k}^{t_{k+1}-1} s(t)^2/2)}.$$

Apply – logarithm to above formula and it results in the form which is applicable to bellman ford algorithm.

And the e function can be approximated to

$$\sum_{k=1}^{K} \sum_{n_{k-1}}^{n_k - p_k} (s(n+p_k) - s(n))^2 \approx 2 \sum_{n=1}^{40000} s(n)^2 - 2 \sum_{k=1}^{K} (C(n_k, p_k) - C(n_{k-1}, p_k)).$$

Where

$$C(n, p) = \sum_{k=1}^{n} s(k) s(k+p).$$

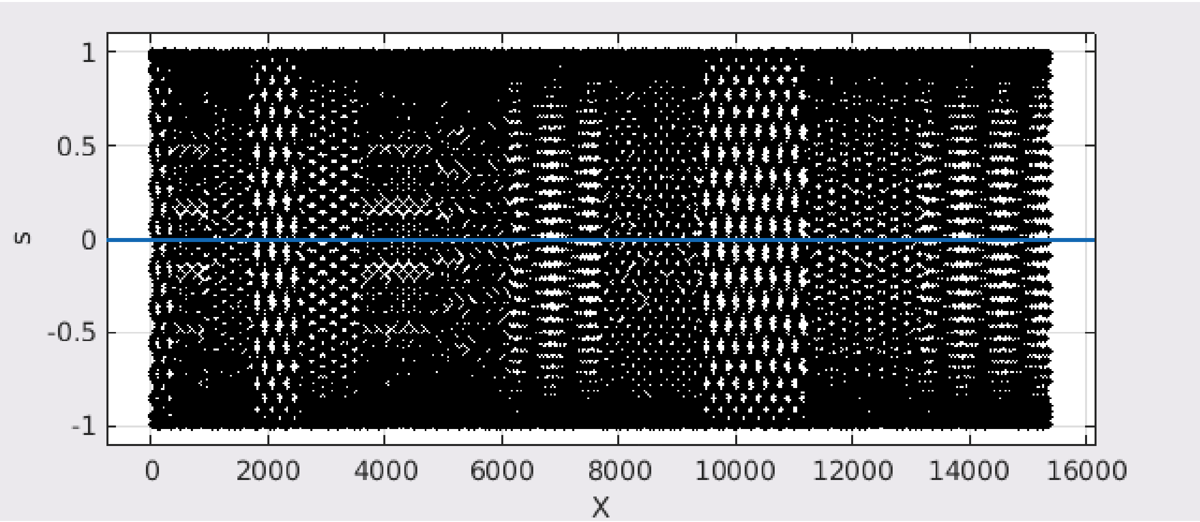
For the next note to be predicted ,its probability should be maximum which indirectly lead to max of above approximation.

Inductively predict all the possible periods of current boundary(k) and by solvation in next boundary(k+1) .The best score in k boundary is known.

5.7 SEGMENTATION

5.7.1 INPUT 1

Let the input data be



song={'C#m' 'A' 'E' 'F' 'G#' 'A' 'Eb' 'B' 'C' 'F' 'Bb' 'B'}

And the pitch of above audio signal plotted is:

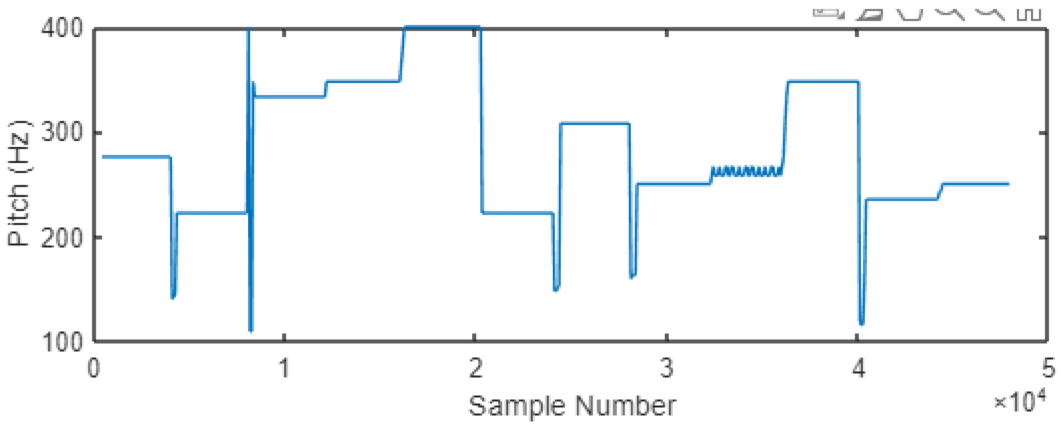


FIGURE 5.15 INPUT SIGNAL AND PITCH OF THE INPUT 1

SEGMENTATION AND BEST SCORE OBTAINED:

In the above sample it is noted that the total no of samples are: 48000

BOUNDARIES:

FIRST NOTE

We will first estimate the end point of first note.

Assume that the endpoint of first note to be n=1000 and obtain the periods

Possible periods for first note = 24 25 26 27 28 29

So increment the 'n' till the periods list remain same.

Here for n=4000
Possible periods for first note = 24 25 26 27 28 29
The period list reamin same.

Here for n=5000
Possible periods The period list changes so here we confirm that the first note ends at 4000.

SECOND NOTE:
Since we have obtained that the first note ends at 4000 sample point.
Second note is assumed to be started at sample number 4001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for second note =14 15 16 17 18
So increment the 'n' till the periods list remain same.

Here for n=8000
Possible periods for first note = 14 15 16 17 18
The period list reamin same.
On incrementing n the list changes.
The period list changes so here we confirm that the second note ends at 8000.
The best possible period for the previous boundary is 29

THIRD NOTE:
Since we have obtained that the second note ends at 8000 sample point.
Second note is assumed to be started at sample number 8001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for third note =

19 20 21 22 23 24

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the third note ends at 12000.
The best possible period for the previous boundary is 18.

FOURTH NOTE:
Since we have obtained that the third note ends at 12000 sample point.
Fourth note is assumed to be started at sample number 12001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for fourth note =

18 19 20 21 22 23

So increment the 'n' till the periods list remain same.

The period list changes so here we confirm that the fourth note ends at 16000.
The best possible period of previous boundary for all the possible periods of current boundary is 23.

FIFTH NOTE:
Since we have obtained that the fourth note ends at 16000 sample point.
Fourth note is assumed to be started at sample number 16001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for fifth note =

15 16 17 18 19

So increment the 'n' till the periods list remain same.

The period list changes so here we confirm that the fifth note ends at 20000.

The best possible period of previous boundary for all the possible periods of current boundary is 23.

SIXTH NOTE:
Since we have obtained that the fifth note ends at 20000 sample point.
Sixth note is assumed to be started at sample number 20001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for sixth note = 14 15 16 17 18

The period list changes so here we confirm that the sixth note ends at 24000.

The best possible period of previous boundary for all the possible periods of current boundary is 19.

SEVENTH NOTE:
Since we have obtained that the sixth note ends at 24000 sample point.
Seventh note is assumed to be started at sample number 24001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for seventh note = 21 22 23 24 25 26

The period list changes so here we confirm that the seventh note ends at 28000.

The best possible period of previous boundary for all the possible periods of current boundary is 18.

EIGTH NOTE:
Since we have obtained that the seventh note ends at 28000 sample point.
Eighth note is assumed to be started at sample number 28001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for eighth note = 13 14 15 16
The period list changes so here we confirm that the eighth note ends at 32000.

The best possible period of previous boundary for all the possible periods of current boundary is 18.

NINTH NOTE:
Since we have obtained that the eighth note ends at 32000 sample point.
Ninth note is assumed to be started at sample number 32001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for ninth note = 26 27 28 29 30 31
The period list changes so here we confirm that the ninth note ends at 36000.

The best possible period of previous boundary for all the possible periods of current boundary is 16.

TENTH NOTE:

Since we have obtained that the ninth note ends at 36000 sample point.
Tenth note is assumed to be started at sample number 36001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for tenth note = 18 19 20 21 22 23

The period list changes so here we confirm that the tenth note ends at 40000.

The best possible period of previous boundary for all the possible periods of current boundary is 31.

ELEVENTH NOTE:

Since we have obtained that the tenth note ends at 40000 sample point.
Eleventh note is assumed to be started at sample number 40001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for eleventh note = 13 14 15 16 17
The period list changes so here we confirm that the eleventh note ends at 44000.

The best possible period of previous boundary for all the possible periods of current boundary is 23.

TWELFTH NOTE:

Since we have obtained that the eleventh note ends at 44000 sample point.
Twelfth note is assumed to be started at sample number 44001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for twelfth note = 13 14 15 16
The period list changes so here we confirm that the twelfth note ends at 48000.

The best possible period of previous boundary for all the possible periods of current boundary is 17.

FINAL BOUNDARIES:

NOTE 1: [1,4000]

NOTE 2: [4001,8000]

NOTE 3: [8001,12000]

NOTE 4: [12001,16000]

NOTE 5: [16001,20000]

NOTE 6: [20001,24000]

NOTE 7: [24001,28000]

NOTE 8: [28001,32000]

NOTE 9: [32001,36000]

NOTE 10: [36001,40000]

NOTE 11: [40001,44000]

NOTE 12: [44001,48000]

NOTE FREQUENCIES

Note 12: period =16 ,frequency =500 Hz

Note 11: period =17 ,frequency =470.5 Hz

Note 10: period =23 ,frequency =347.82 Hz

Note 9: period =31 ,frequency =258.06 Hz

Note 8: period =16 ,frequency =500 Hz

Note 7: period =18 ,frequency =444.44 Hz

Note 6: period =18 ,frequency =444.44 Hz

Note 5: period =19 ,frequency =421.05 Hz

Note 4: period =23 ,frequency =347.82 Hz

Note 3: period =23 ,frequency =347.82 Hz

Note 2: period =18 ,frequency =444.44 Hz

Note 1: period =27 ,frequency =296.3 Hz

THE SEGMENTATION PLOT

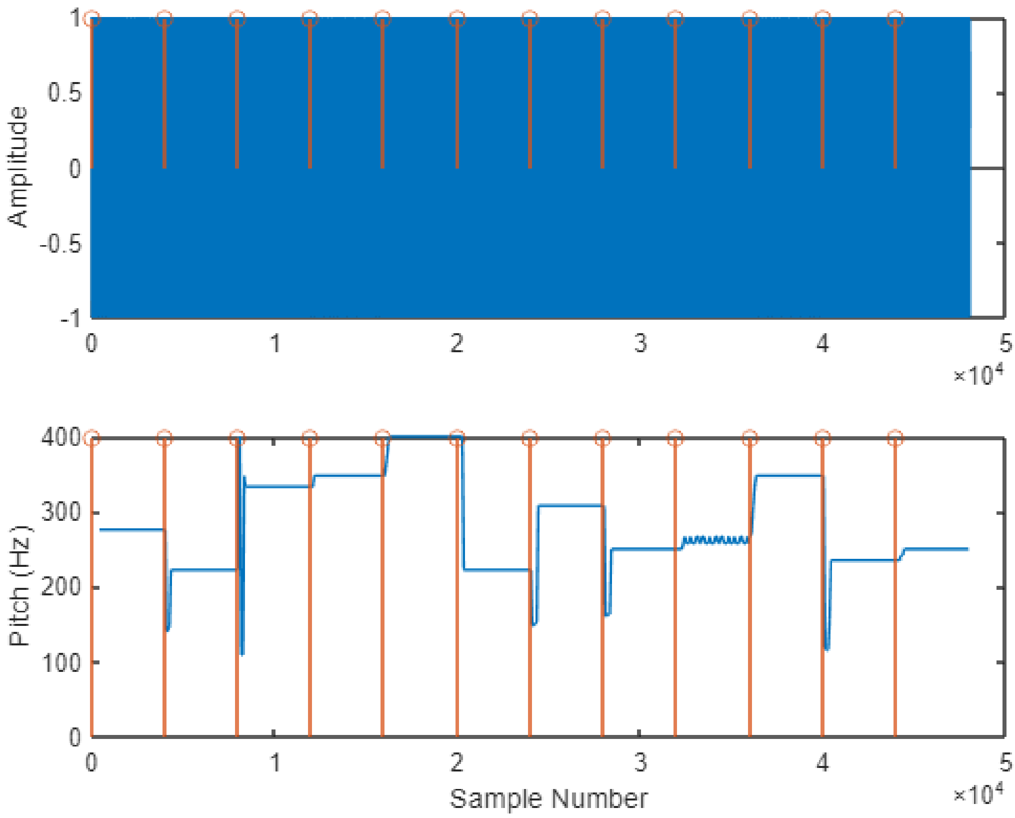


FIGURE 5.16 REPRESENATION OF BOUNDARIES OF NOTES FOR INPUT 1

THE PITCH GRAPH OF SEGEMENTATION RESULT

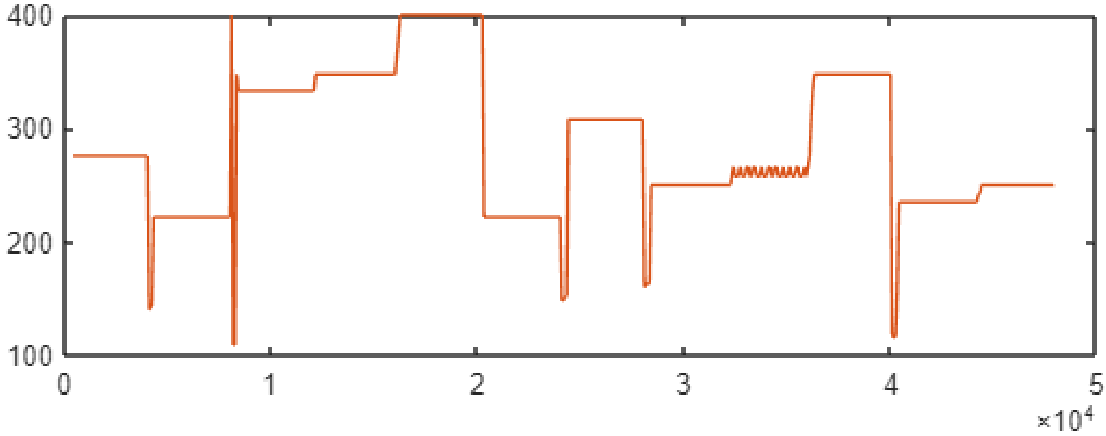


FIGURE 5.17 PITCH OF THE SIGNAL OBTAINED VIA SEGMENTATION OF INPUT 1

COMPARING PITCHES:

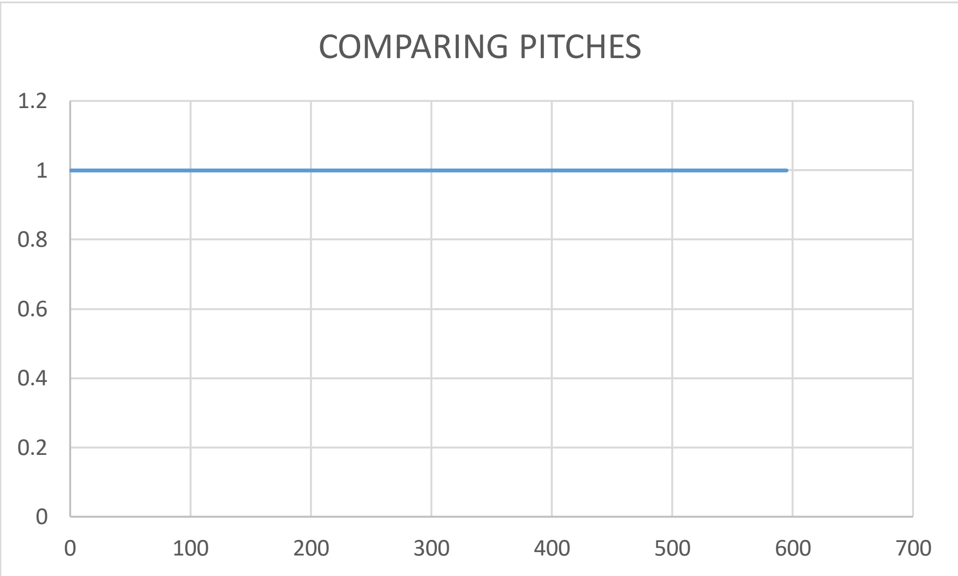


FIGURE 5.18 COMPARING FREQUENCIES OF ORGINAL INPUT 1 AND THE SEGMENTED OUTPUT 1

OBSERVATION:

All pitches are matched.

This represent perfect score.

All the notes present in the input signal is present in output signal too.

5.7.2 INPUT 2

INPUT 2:

The input and pitch graph is taken to be

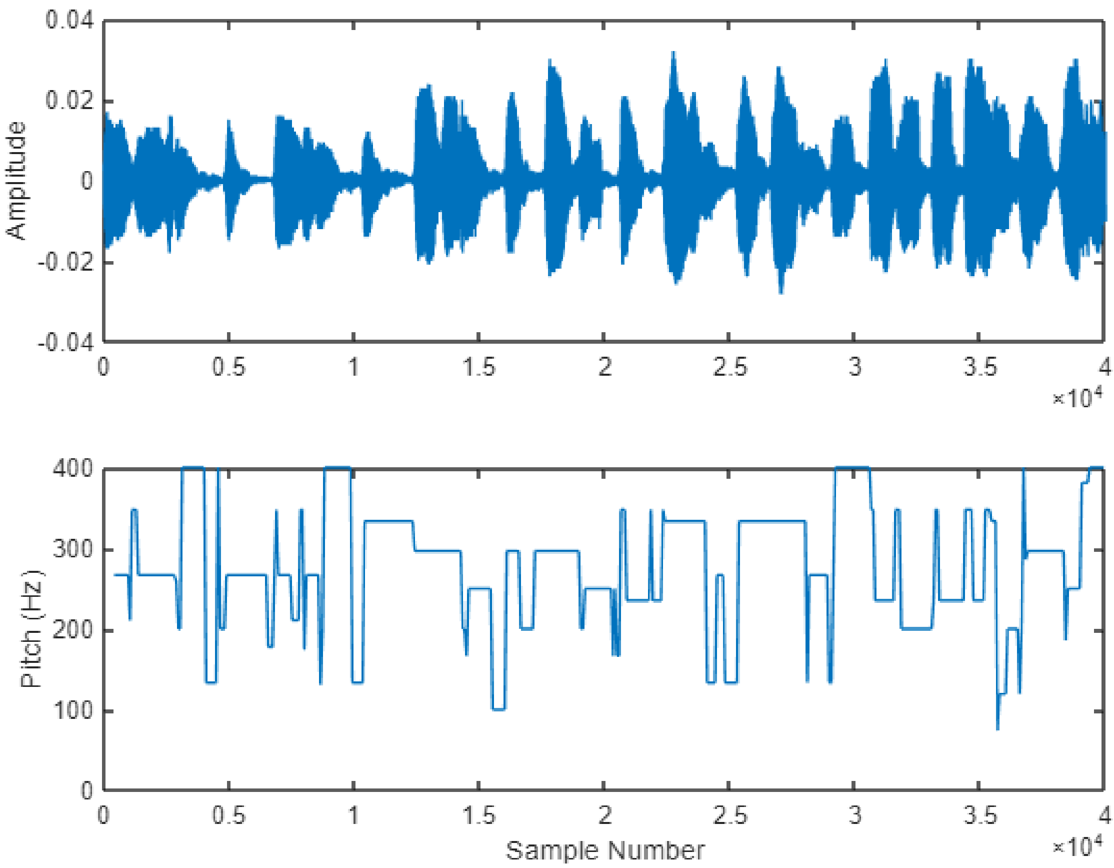


FIGURE 5.19 INPUT SIGNAL AND PITCH OF THE INPUT 2

The information of audio is

```
  NumChannels: 1
   SampleRate: 8000
 TotalSamples: 40000
     Duration: 5
        Title: []
      Comment: []
       Artist: []
BitsPerSample: 8
```

SEGMENTATION AND BEST SCORE OBTAINED:

In the above sample it is noted that the total no of samples are: 40,000

BOUNDARIES:

FIRST NOTE

We will first estimate the end point of first note.

Assume that the endpoint of first note to be n=30 and obtain the periods

Possible periods for first note = 7 8 15

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the first note ends at 1500.

SECOND NOTE:

Since we have obtained that the first note ends at 1500 sample point.
Second note is assumed to be started at sample number 1501.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for second note =13 14 15
So increment the 'n' till the periods list remain same.
Here for n=3000
The period list changes so here we confirm that the second note ends at 3000.

THIRD NOTE:

Since we have obtained that the second note ends at 3000 sample point.
Third note is assumed to be started at sample number 3001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for third note = 8 9 10 19 20
So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the third note ends at 4800.

FOURTH NOTE:

Since we have obtained that the third note ends at 4800 sample point.
Fourth note is assumed to be started at sample number 4801.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for fourth note = 7 8 15
So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the fourth note ends at 6600

FIFTH NOTE:

Since we have obtained that the fourth note ends at 6600 sample point.
Fifth note is assumed to be started at sample number 6601.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for fifth note = 8 15 30
So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the fifth note ends at 6900

SIXTH NOTE:

Since we have obtained that the fifth note ends at 6900 sample point.
Sixth note is assumed to be started at sample number 6901.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for sixth note = 8 15
So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the sixth note ends at 7200

SEVENTH NOTE:

Since we have obtained that the sixth note ends at 7200 sample point.
Seventh note is assumed to be started at sample number 7201.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for seventh note = 8 15
So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the seventh note ends at 8700

EIGHTH NOTE:

Since we have obtained that the seventh note ends at 8700 sample point.
Eighth note is assumed to be started at sample number 8701.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for eighth note = 9 10 19 20

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the eighth note ends at 10400

NINTH NOTE:

Since we have obtained that the eighth note ends at 10400 sample point.
Ninth note is assumed to be started at sample number 10401.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for ninth note = 12

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the ninth note ends at 12500

TENTH NOTE:

Since we have obtained that the ninth note ends at 12500 sample point.
Tenth note is assumed to be started at sample number 12501.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for tenth note = 27

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the tenth note ends at 14700

NOTE 11:

Since we have obtained that the tenth note ends at 14700 sample point.
Eleventh note is assumed to be started at sample number 14701.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for eleventh note = 16

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the eleventh note ends at 16100

NOTE 12:

Since we have obtained that the eleventh note ends at 16100 sample point.
Twelfth note is assumed to be started at sample number 16101.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for twelfth note = 27

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the twelfth note ends at 19200

NOTE 13:

Since we have obtained that the twelfth note ends at 19200 sample point.
Thirteenth note is assumed to be started at sample number 19201.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for thirteenth note = 16

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the thirteenth note ends at 20700

NOTE 14:

Since we have obtained that the thirteenth note ends at 20000 sample point.
Fourteenth note is assumed to be started at sample number 20001.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for fourteenth note =32

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the fourteenth note ends at 20600

NOTE 15:

Since we have obtained that the fourteenth note ends at 20600 sample point.
Fifteenth note is assumed to be started at sample number 20601.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for fifteenth note = 23, 32

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the fifteenth note ends at 21500

NOTE 16:

Since we have obtained that the fifteenth note ends at 21500 sample point.
Sixteenth note is assumed to be started at sample number 21501.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for sixteenth note = 11 34

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the sixteenth note ends at 22400

NOTE 17:

Since we have obtained that the sixteenth note ends at 22400 sample point.
Seventeenth note is assumed to be started at sample number 21501.
Similarly with incrementing n for every thirty points the possible periods list is obtained.

Possible periods for seventeenth note = 6 12 24 36

So increment the 'n' till the periods list remain same.
The period list changes so here we confirm that the seventeenth note ends at 23200

NOTE BOUNDARIES AND POSSIBLE FREQUENCIES IN IT:

18. 23200-24000: 6 12 24

19.24000-25000: 6 7 15 30

20.25000-25700: 6 12 24 36

21: 25700-28200: 6 12

22.28200 -29100 : 7 15 30

23.29101-30650: 9 10

24. 30650-31400: 6 11 34

25. 31400-31700:6 23

26. 31700-33400:6 13 40

27.33400-35200: 6 11 34

28. 35200-35800: 6 12

29.35800-36700: 6 8 13 40

30.36700-38300:8 9

31.38300-38400: 10 11 23 34

32.38400-39000: 10 11 21 32

33. 39000-39700 : 10

34.39700-39800: 10 20

35:39800-40000:10

Segmentation:

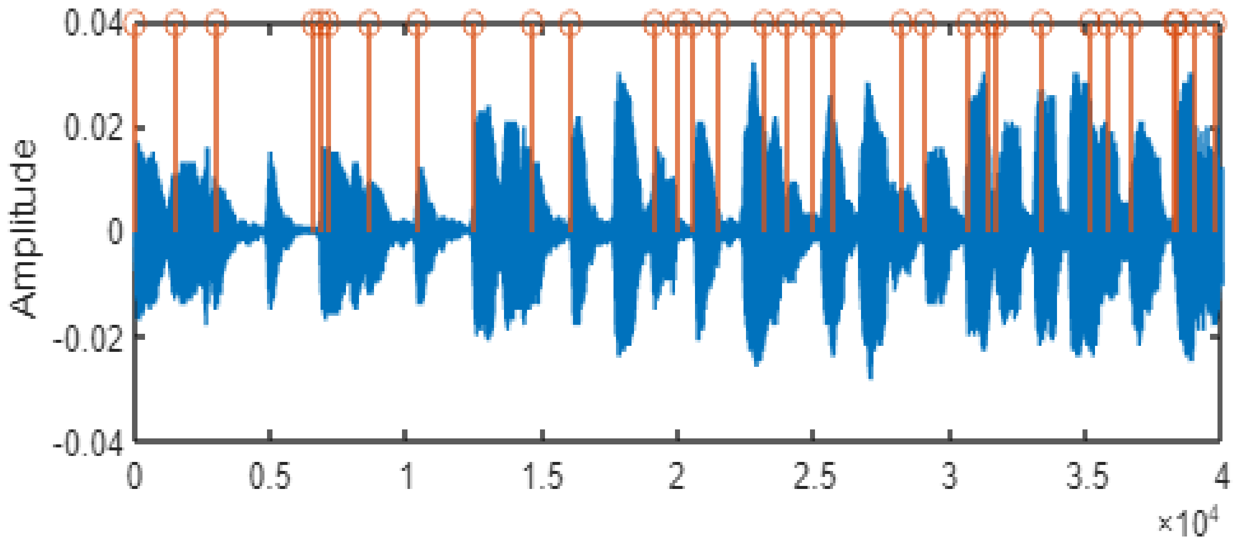


FIGURE 5.20 BOUNDARIES AT WHICH SEGMENATATION IS BEING DONE INPUT 2

So this is segmentation of 40,000 samples

These are the possible notes at particular note boundaries:

Now we are segmenting with period in each segment taken to be

pr=[15 15 20 15 30 15 15 20 12 27 16 27 16 32 23 34 36 24 30 36 12 30 10 34 23 40 34 12 8 8 23 32 10 20 10]

The new signal is found and the pitch of the signal is

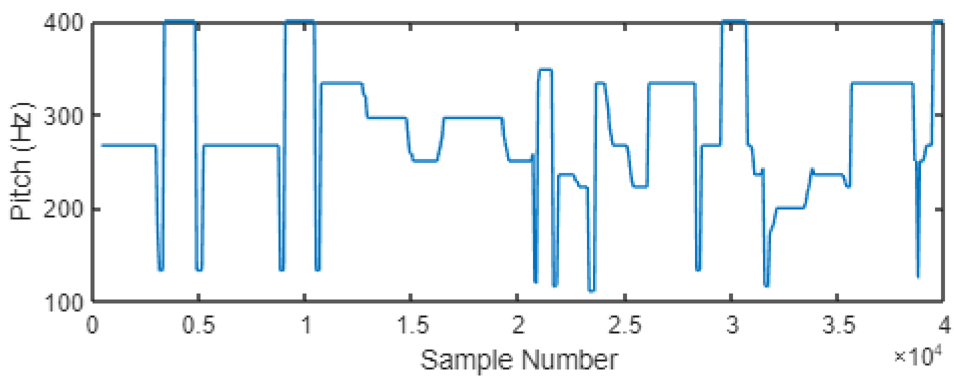


FIGURE 5.21 PITCH OF THE SEGMENTED AUDIO 2

Comparing pitches of signal taken and new signal created.

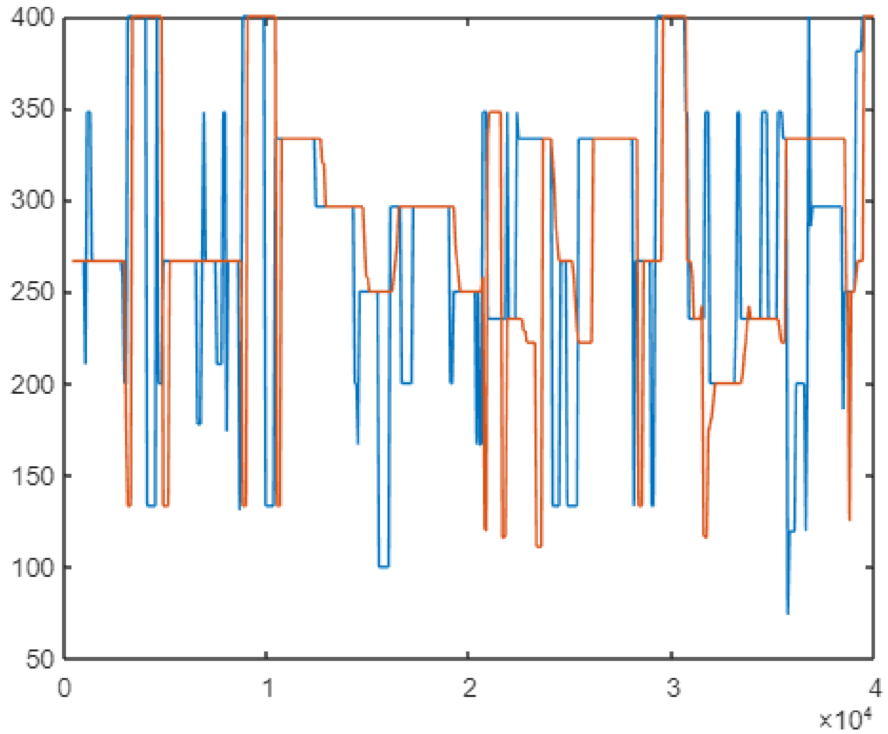


FIGURE 5.22 COMPARING PITCHES OF INPUT 2 AND SEGMENTED AUDIO 2

On observation it is known that out

295 matched out of 495.

5.7.3 INPUT 3

The input is taken to be

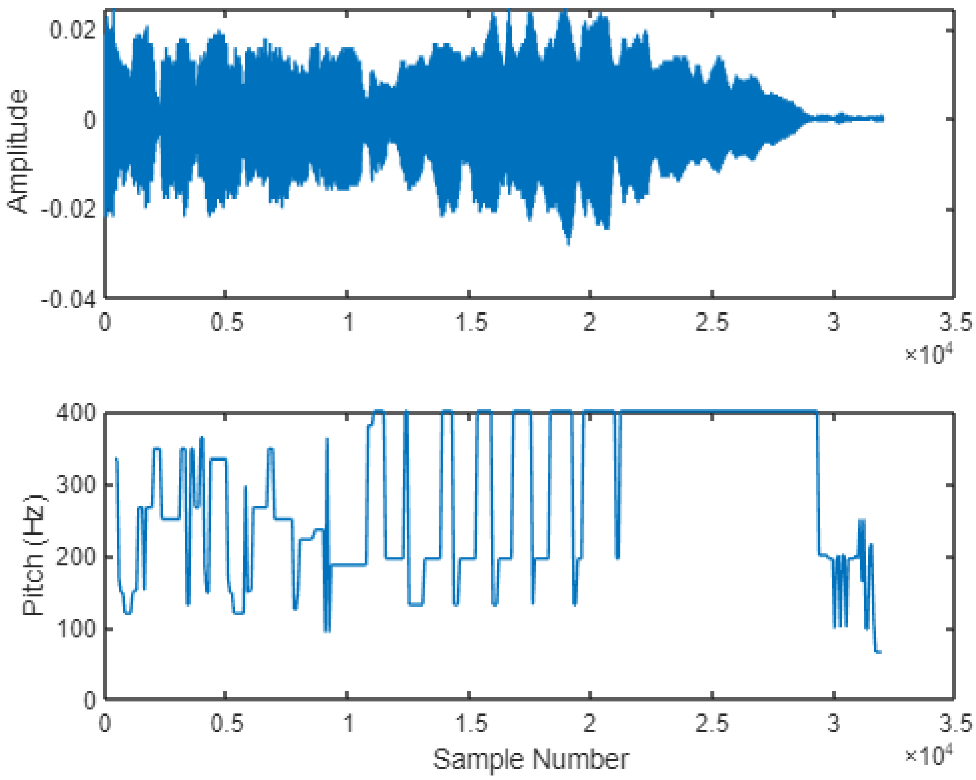


FIGURE 5.23 INPUT SIGNAL AND PITCH OF THE INPUT 3

The information of audio is taken to be:

Total number of samples 32000

Sampling frequency 8000 Hz

Total time is 3 seconds

SEGMENTATION:

SETTING BOUNDARIES:

Boundaries are set and periods which are the hidden random varaiables are known through considering the posterior probability.

FIRST NOTE:

Lets assume that the first boundary ends at 200.

So the interval of first note is [1,200] samples.

On following the bellman ford algorithm the periods should have min value of e function.

So the possible periods are: 12 24

SECOND NOTE:

Second note is started at 201 sample.

And by induction the periods which give minimum for e function is obtained.

So the possible periods are: 12

And the best possible periods for the previous boundary is 12,24 both.

boundary ends at 950.

THIRD NOTE:

Third note is started at 951 sample.

And by induction the periods which give minimum for e function is obtained.

The code is only looped over the immediately preceding note boundary and possible p's in the last time segment.

So the possible periods are: 12,13

And the best possible periods for the previous boundary is 12.

boundary ends at 1050.

FOURTH NOTE:

Fourth note is started at 1051 sample.

And by induction the periods which give minimum for e function is obtained.

The code is only looped over the immediately preceding note boundary and possible p's in the last time segment.

So the possible periods are: 7,15,23

And the best possible period for the previous boundary is 13.

boundary ends at 1050.

FIFTH NOTE:

Fifth note is started at 4301 sample.

So the possible periods are: 12 23 36

And the best period for the previous boundary is 23 for all possible periods

boundary ends at 4450.

SIXTH BOUNDARY:

Sixth boundary is started at 4451 sample.

So the possible periods are: 13 23

And the best period for the previous boundary is 36 for all possible periods

boundary ends at 5900.

SEVENTH BOUNDARY:

Seventh boundary is started at 5901 sample.

So the possible periods are: 15 16 23

And the best period for the previous boundary is 13 for all possible periods

boundary ends at 7500

EIGTH BOUNDARY:

Eighth boundary is started at 7501 sample.

So the possible periods are: 8 9 17 18

And the best period for the previous boundary is 16 for all possible periods

boundary ends at 8600

NINTH BOUNDARY:

Ninth boundary is started at 8601 sample.

So the possible periods are: 7 17 22 34

And the best period for the previous boundary is 18 for all possible periods

boundary ends at 9100

TENTH BOUNDARY:

Tenth boundary is started at 9101 sample.

So the possible periods are: 7 8 11 21

And the best period for the previous boundary is 34 for all possible periods//do once more///

boundary ends at 9400

ELVENTH BOUNDARY:

Eleventh boundary is started at 9401 sample.

So the possible periods are: 7 8 20 21

And the best period for the previous boundary is 21 for all possible periods

boundary ends at 13000

TWELETH BOUNDARY:

Twelfth boundary is started at 13001 sample.

So the possible periods are: 6 15 20 21

And the best period for the previous boundary is 21 for all possible periods

boundary ends at 15500

THIRTEENTH BOUNDARY:

Thirteenth boundary is started at 15501 sample.

So the possible periods are: 6 7 20 21

And the best period for the previous boundary is 20 for all possible periods

boundary ends at 23000

FOURTEENTH BOUNDARY:

Fourteenth boundary is started at 23001 sample.

So the possible periods are: 6 7 8 13 19 20 21

And the best period for the previous boundary is 20 for all possible periods

boundary ends at 30000

FIFTEENTH BOUNDARY:

Fifteenth boundary is started at 30001 sample.

So the possible periods are: 6 7 8 9 20 29 30 31 33 40

And the best period for the previous boundary is 20 for all possible periods

boundary ends at 31000

SIXTEENTH BOUNDARY:

Sixteenth boundary is started at 31001 sample.

So the possible periods are: 5 7 34 37 38

And the best period for the previous boundary is 40 for all possible periods

boundary ends at 31500

SEVENTEENTH BOUNDARY:

Seventeenth boundary is started at 31501 sample.

So the possible periods are: 5

And the best period for the previous boundary is 38 for all possible periods

boundary ends at 32000

So the best score is calculated from backwards

If we segment the total audio sample of 32000 samples it results in 17 notes

FINAL BOUNDARIES:

NOTE 1: [1,200]

NOTE 2: [201,900]

NOTE 3: [951,1050]

NOTE 4: [1051,4300]

NOTE 5: [4301,4450]

NOTE 6: [4451,5900]

NOTE 7: [5901,7500]

NOTE 8: [7501,8600]

NOTE 9: [8601,9100]

NOTE 10: [9101,9400]

NOTE 11: [9401,13000]

NOTE 12: [13001,15500]

NOTE 13: [15501,23000]

NOTE 14: [23001,30000]

NOTE 15: [30001,31000]

NOTE 16: [31001,31500]

NOTE 17: [315001,32000]

FINAL NOTES:

Note 17: period =5 ,frequency =1600 Hz

Note 16: period =38 ,frequency =210.5 Hz

Note 15: period =40 ,frequency =200 Hz

Note 14: period =20 ,frequency =400 Hz

Note 13: period =20 ,frequency =400 Hz

Note 12: period =20 ,frequency =400 Hz

Note 11: period =21 ,frequency =380.9 Hz

Note 10: period =21 ,frequency =380.9 Hz

Note 9: period =34 ,frequency =235.9 Hz

Note 8: period =18 ,frequency =444.44 Hz

Note 7: period =16 ,frequency =500 Hz

Note 6: period =13 ,frequency =615.3 Hz

Note 5: period =36 ,frequency =222.2 Hz

Note 4: period =23 ,frequency =347.82 Hz

Note 3: period =13 ,frequency =615.3 Hz

Note 2: period =12 ,frequency =666.67 Hz

Note 1: period =24 ,frequency =333.33 Hz (or) period =12 ,frequency =666.67 Hz

The pitch of the newly created signal

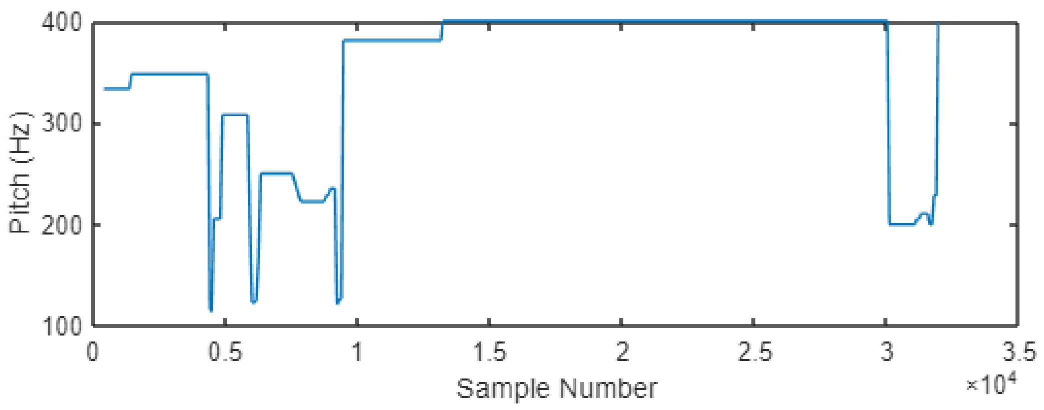


FIGURE 5.24 PITCH OF THE SEGMENTED OUTPUT 3

COMPARING THE PITCH OF INPUT AND OUTPUT

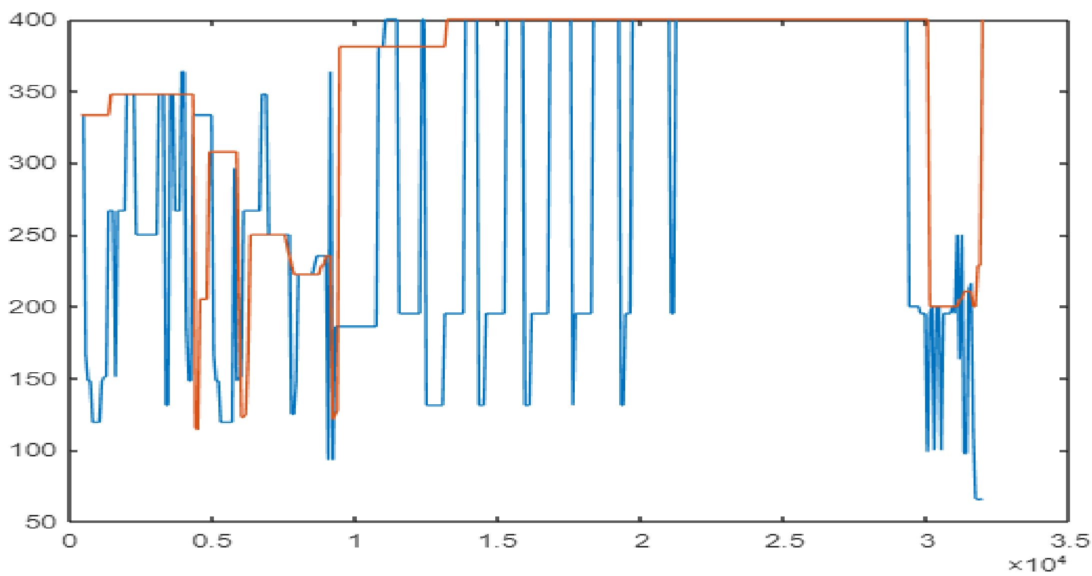


FIGURE 5.25 COMPARING PITCHES OF INPUT 3 AND SEGMENTED OUTPUT 3

NUMBER OF MATCHES OF PITCHES

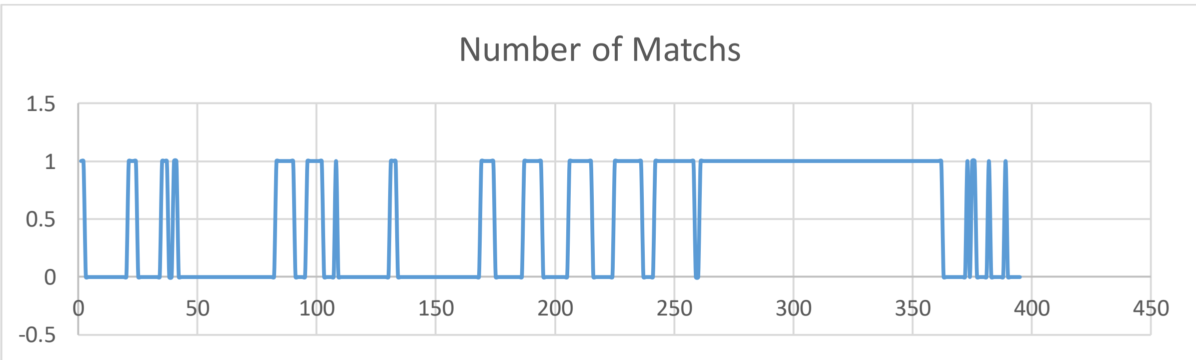


FIGURE 5.26 NUMBER OF MATCHES FOR PITCHES

Here if Y axis =1 then match.

Y axis = 0 then not a match.

References

1. http://desolneux.perso.math.cnrs.fr/MVA/Cours-Gaussian-Music.pdf
2. https://www.dam.brown.edu/people/mumford/vision/pattern.html
3. https://towardsdatascience.com/markov-chain-for-music-generation-932ea8a88305
4. http://eita-nakamura.github.io/articles/Nakamura_etal_ARHSMMForScoreFollowing_ISMIR2015.pdf