Bach in Beta: Modeling Bach Chorales with Markov Chains

Michaela Suzanne Tracy

A Thesis
Presented to the Faculty
of Harvard University
in Candidacy for the Degree
of Bachelor of Arts

Recommended for Acceptance
by the Department of
Applied Mathematics
Advisers: Professor Mauricio Santillana and Professor Suzannah Clark

March 2013

Abstract

In this thesis I present a Hidden Markov Model capable of composing musical chorales inspired by the works of J.S Bach. The transition matrix for the model is constructed by sequentially incorporating musical information from approximately seven percent of the 371 chorales composed by J.S. Bach. The quality of the chorales produced by the computational model was tested experimentally by surveying 40 musical experts. Our results show that musical experts could not identify the differences between chorale harmonies composed by J.S. Bach and synthetic harmonic progressions produced with our algorithm. By performing statistical hypothesis testing on the proportion of correct responses, we found that the success rate of identifying an actual Bach chorale correctly, among experts, was similar to flipping a coin for a transition matrix trained with more than 10 chorales. This finding suggests that the computational methodology presented here is a successful way to create Bach-quality harmonic progressions.

Acknowledgements

I would like to thank my advisers, Professors Santillana and Clark, for giving me the attention and guidance in order to complete this project. Also, to Simon Lunagomez, for helping us design the experiment, and all his wonderful feedback throughout the process. I would like to thank my parents, whose laid-back parenting complemented with loving support motivates me to achieve all of my ridiculous goals. To Dr. Kevin C. Leong, whose sheer love of learning, teaching, and music have inspired me since my freshman year. His friendship can never be replaced, and I cannot imagine how I would have survived the last years without his substitute-parental concern. I also thank Nicandro, who was my first inspiration to take on this project, and our whirlwind friendship that never ceases to surprise me. Finally, to my brother, all the respect in the world for his perseverance and dedication to the things he loves.

To KL, NI, 'Olden and my family.

Contents

List of Tables

List of Figures

Chapter 1

Overview

The relationships between music and mathematics are natural to make, and have been made for centuries. From the determination of the mathematical ratio between audible pitches, to the numerical classification of pitches, and the regular summation of rhythmic events in musical measures, musicians use mathematics consistently in almost every realm of study, practice, and performance. In addition to the direct impact mathematical relationship and understanding have on music, mathematics can also be indirectly applied to music in order to gain better understanding. In the early 18th century, the ideas of quantifying music in terms of numbers became a sought-out practice. Groups of sounds began to be named in terms of numbers as opposed to musical letter names, and numerical relationships were described academically. Into the 20th century, composers began to use mathematics as the spark for musical ideas, such as replicating numerical patterns in their music. These less-consistent applications nonetheless inspire musician-mathematicians to find new ways to use math in music.

Before these secondary uses of mathematics were even formally addressed, J.S. Bach composed much of the work best known for inspiring formal music theory and the mathematics it implores. Bach (1685-1750), who lived his entire life in Germany, composed music of nearly every genre in great amounts. His contributions to solo repertoire, small and large-groups, and choirs are innumerable. One of the greatest contributions to Western music was Bach's chorale harmonizations. While he was director of music in Lutheran churches, he took the standard Lutheran hymns and reworked them for four voices, with one voice retaining the original melod. Bach used these four-part chorales in larger works, such as cantatas or oratorios. Today, 371 of these chorale harmonizations still survive.

What these chorales have provided for music theorists and historians is a large data-set of music from the Baroque era of Western music. By studying these works, we can tell how often Bach used certain musical devices in order to gain better understanding of the music. We can also study patterns, frequency, and dependency of events using mathematics. Applying the idea of Hidden Markov Models to the large data set of chorales, the possibilities of non-trivial results seem endless. However, certain results are more critical to the understanding of harmony, such as the regularity of chords and how replicable chorales are via statistical and probabilistic means.

Chapter 2

Musical Introduction

There is a great deal of vocabulary from the musical lexicon that will be used in this project. Hopefully, the exact musical significance of each term will not be necessary to understand the mathematical model and its construction. Despite the fact that there is no need to know all aspects of music theory to understand this project, it will be important to understand a few key concepts in order to understand the significance of the results. First off, a pitch (or note) has a specific frequency described in hertz. There are 12 pitch classes in Western musical notation. For example, the note A, which most modern orchestras tune their instruments to, has a frequency of 440Hz. When are person plays a note exactly twice this frequency, or 880Hz, that note is said to be "an octave" higher. Notes an octave apart are given the same pitch name, but are distinguished by their frequency, or the name of the octave.

The octave is then divided into 12 intervals, known as semitones. An octave is a musical distance between two pitces that is 12 semitones large. The pitch A440 is in the 4th octave, as where A880 is in the 5th octave. Besides the octave, there are several other important names to intervals. Two notes with identical frequencies are called unisons. The pitch that is in a 3:2 ratio to the original frequency is called the "perfect fifth." The octave, unison and perfect fifth are considered the most fundamental intervals in music. The following table describes the remaining relationships between the 12 pitches.

Number of semitones between	Name	Pitch ratio
0	Perfect unison	1:1
1	Minor second	16:15
2	Major second	9:8
3	Minor third	6:5
4	Major third	5:4
5	Perfect fourth	4:3
6	Augmented fourth Diminished fifth	45:32 64:45
7	Perfect fifth	3:2
8	Minor sixth	8:5
9	Major sixth	5:3
10	Minor seventh	9:5
11	Major seventh	15:8
12	Perfect octave	2:1

Figure 2.1: Ratio between pitches separated by the 12 semitones [11]

A key is a set of pitches (when played in order called a scale) that have predetermined spacing starting from the root of the key. There are two types of keys and scales, major and minor keys. Major scales start on the root pitch and then the other 6 notes are 2,4,5,7,9, and 11 semitones above the root pitch. A minor scale begins with the root pitch and has 6 other notes 2,3,5,7, 8 and 10 semitones above the root. Since there are twelves possible notes and two types of key, there are 24 possible keys that can be explored in this project. The important advent of the music of J.S. Bach is that it uses all 24 keys.

Chords are sets of three or more specific pitches, or notes, played simultaneously. In this project, only chords of three of four notes were studied, since they are the significant majority of all chords in Western Baroque and Classical music. Chords are given names based on which pitch the chord is "based on." This fundamental note of the chord is called the root. Chords of three notes can be if three positions:

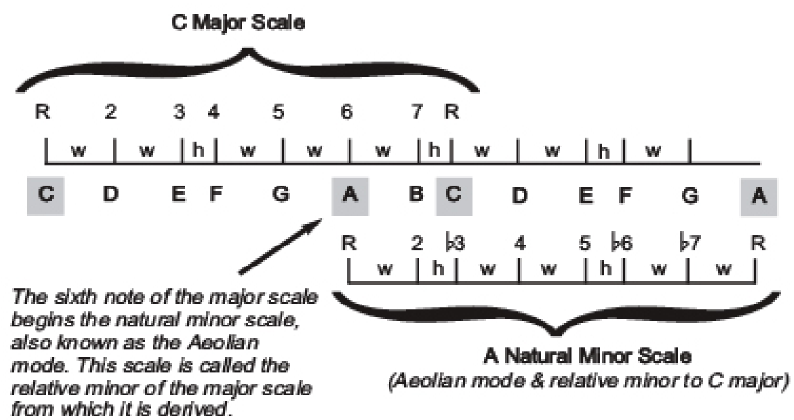


Figure 2.2: C major and A minor scale, who share the same pitches different order [5]

one where the root is the lowest frequency note in the chord (root position chord), another where the root is the middle frequency note (first inversion chord), and one where the root is the highest frequency note (second inversion chord). In this study, only a few types of chords have four notes, and they occur in four positions with the same naming system. Because of the twelve possibly notes and seven types of chords in inversion, there are 84 possible chords in Western harmony (to be studied in this project). There exists countless other types of chords, but these 84 types will be the bulk of the analysis of this project. Despite the great range of possibility, most pieces restrict themselves to 15-20 different chords per piece. Some chords never occur in Western music because of their perceived "dissonance" or "bad-sounding" nature compared to the tonic of a key, or the preceding chord.

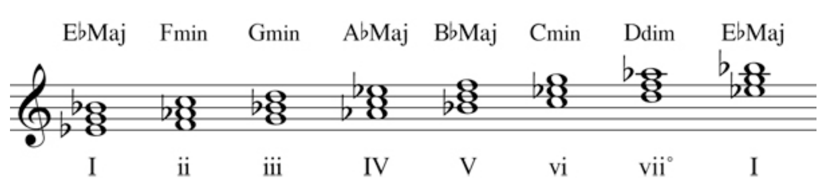


Figure 2.3: The seven chords based on E-Flat Major [3]

In music, we give chords Roman Numerals based on their relationship to the tonic. The tonic chord, based on the tonic and two notes above it in the major key, is labeled

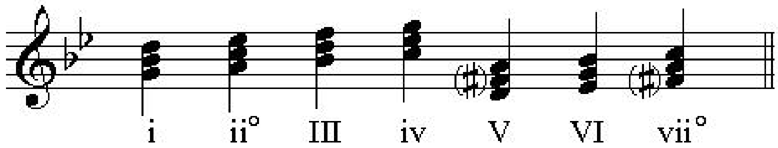


Figure 2.4: The seven chords based on G minor [3]

I. The second chord is called ii, and this pattern continues on. The Roman numeral is capitalized if the chord is a major mode, and is lower-case if the chord is a minor mode. In figures 2.3 and 2.4 we see the diffence in Roman Numeral naming in Major and Minor keys. These are the significant, three-note root position chords used in this project. Chords can occur in a variety of spacings, as well (since there are multiple octave-equivalent notes in music). The chords displayed in these figures are called closed-position chords. The definition of a closed-position chord is that no chord tones (notes belonging to a specific chord) are skipped or omitted from the starting pitch of the chord to the ending pitch. A closed-position chord may omit, for example, the second note of the chord and have it occur an octave higher or lower. The figure 2.5 below shows an F-major chord (the IV chord of C major, since it is based on the 4th pitch of the C major scale and its pitches classify this chords as being major) in an open position to contrast to all the closed position chords in Figures 1.3 and 1.4. The most important Roman Numerals to remember are the I (or i in minor), as this is the tonic chord of the key. The other most important chord is the V (almost always capitalized in major and minor keys) as this chord is known as the dominant, and as its name suggests it has a special function. The dominant usually functions as a musical signal that the tonic is arriving, and has a musical feeling associated with wanting to "go somewhere" or return home to the tonic.

Sometimes, there are special classifications of chords whose pitches do not come from the scale or key that the piece is in at that time. These chords are known as applied chords. For example, if a section of a chorale is in the key of C major, any

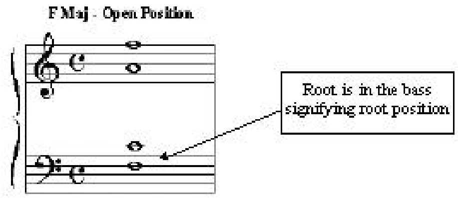


Figure 2.5: Example of an Open Root Position Chord [6]

pitches with sharps or flats are not part of that key. A common chord to see in these C major sections may be a D-major chord, which contains an F-sharp. This chord is known specifically as a secondary dominant, as D-major is the V or dominant of G major, (and G major is the dominant of C major). Secondary dominants can stack up onto each other and are noted as V/V, with the first Roman numeral describing the chord type, and the second Roman Numeral representing what the key or scale the chord belongs to.

Finally, the last idea integral to the overview of musical ideas in this project is the concept of musical harmonic progressions. A chord progression is the sequence of chords that occur in a piece or section of a work. A cadence is a musical pause or closing the ends an idea. In this era of music, most of the important cadences end with V-I in root position or inversion. Between an opening and a cadence, there is much more freedom for different harmonic events. The basic framework for a progression is tonic (I)-predominant (any type of chord, but frequently ii, IV and vi in major)-dominant (V)-tonic (I). This framework will be discussed as it relates to the results of the modeling experiment in subsequent chapters.

At this point, there is a baseline level of musical understanding in order to understand the model. The goal of this project is to catergorize the different chords that occur in the Bach Chorales, separating for major and minor, and normalizing for key.

Chapter 3

Mathematical Introduction

The mathematics behind modeling music is frequently used for many other topics. The main scope of this project deals with Markov processes, or the mathematics behind networks of interaction. A Markov chain is a mathematical system, generally in matrix form, that models transitions of a given phenomenon. The underlying concept of Markov chains is that the method shows the likelihood of moving to a state, or an event occuring, given we know our starting state, or previous event. The quintessiantial introductory example of a Markov chain is a theoretical example about predicting the weather given that the observes knows the weather of the current day. [1] A sample matrix of this situation may look like this:

Table 3.1: Theoretical Weather Markov Matrix

	Rain	Sunny	Snow
Rain	0.4	0.4	0.2
Sunny	0.3	0.6	0.1
Snow	0.4	0.3	0.3

In this example, there are three state: raining, sunny or snowy. The way the table is interpretted is that the current state (or weather is given by the row.) The column gives the probability that the next state is the given column name. For example, the

[1]I was introduced to the concept of Markov Chains with a similar example in Applied Math 115.

second row, first column number describes the probability that if we know today is sunny, the probability tomorrow will be rainy is 0.3.

There are many important terms when it comes to Markov Chains and their analysis. One of the key ideas is reducibility. Reducibility is the idea that certain states may not be possible to ever transition to given a specific starting point. Accessibility is the concept that if you start in state "j" that through some amount of indeterminate time-steps it is possible to arrive in state "k" (therefore, state "k" is accesssible from state "j".) [4]

States are also known to be commutative if it is possible to go from state j to k and k to j with varying probabilities. A commutative class is a group of states where all other states are accessible and commutative, and no other states in the class can commuticate with any states outside the class. Relating back to reducibility, a Markov chain is said to be "irreducible" if its entire state-space is one commutative class[4].

After the transition matrix of the Markov chain is created, we can analyze it to learn about its important characteristics. A stationary distribution describes that the long-term probability of the states occuring. The stationary distribution can be solved in a few different ways. The first way is to set up a system of equations based on the matrix. In this case, there are three variables and three equations. For simplicity, let R=rain, S=sunny, and W=snow. We also have to add the constraint that the total probability is 1, R+S+W=1, (that is, there is a probability of 1 that there is some kind of weather on a given day).

Equation 1: $R = .4R + .3S + .4W$

Equation 2: $S = .4R + .6S + .3W$

Equation 3: $W = .2R + .1S + .3W$

Equation 4: $1 = R + S + W$

These equations can solved in a variety of ways, the easist method being using linear algebra to solve the Ax=b situation. One could also do elimination/substitution method for the three variables. Either way, the solution of these equations are: [Rain= 0.352112676248874, Sunny=0.478873238932554, Snow=0.169014084818572].

Solving a system of equations with multiple variables can be very time consuming. A quicker way is to use Markov theory to obtain the steady-state distribution. If we continue to apply the transition matrix to an arbitrary vector, then eventually the vector will converge to the stationary distribution. In this example, say we want to determine the distribution given we start in each state (separately). We can use the following MATLAB algorithm to continue to multiply the transition matrix until the difference of the norms between steps is within .001.

If a is the transition matrix above:

```
w(1)=1
bar(w)
ylim([0,1])
pause
for k=1:15
    t=w*a;
if norm(w-t) < .001, break, end
    w=t;
    bar(w)
    pause

end
```

After 7 steps, the matrix converges to the exact same probabilities from the analytic solution. The graph of transition probabilities is shown in figure 3.1.

What this diagram hopes to show is that no matter what state one starts in, over time (7 intevals of time, or days in this case) the probability of the specific next state occuring is the same.

Beyond the graphical, iterative analysis of the steady-state distribution, this analysis can also be done by using eigenvectors. Eigenvalues are normally defined by the equation $Ax = \lambda x$ or $(A - \lambda I)(x) = 0$, where λ is the eigenvalue, x is the eigenvector, A is the matrix one is trying to find the eigenvectors of and I is the identity matrix.

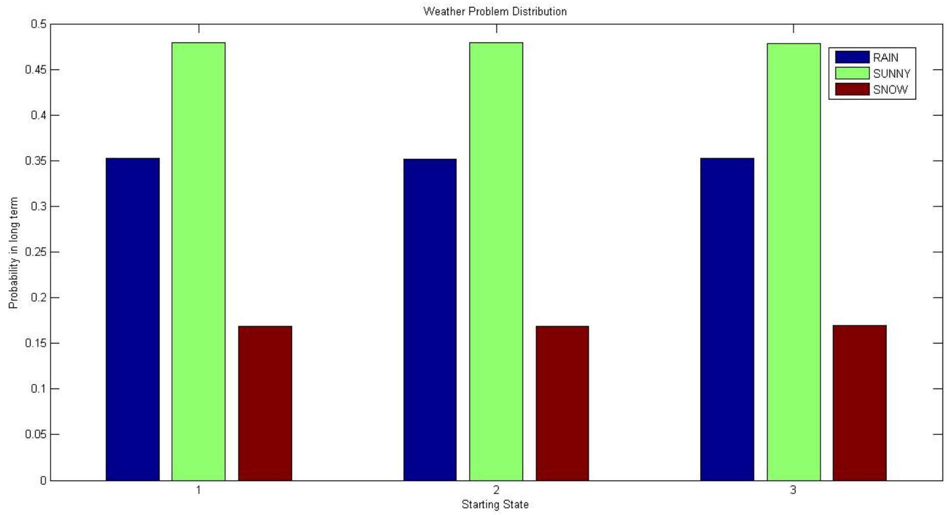


Figure 3.1:
Stationary Distribution of Weather Problem with the Aformentioned Algorithm

In the rainy, sunny, snowy example, there is an interesting concept that surfaces that is normally not discussed in basic Linear Algebra. Because Markov Chains, and this problem, deal with row vectors multiplied by the transition matrix (usually the matrix is multiplied first) the idea of eigenvectors has to be altered. The aforementioned eigenvector equation solves for column vectors, or the right-eigenvectors of the matrix. In the case of Markov Chain, since we are multiplying in a different order, we need to solve for left-eigenvectors, or rows. The equations for left-eigenvectors are $xA = \lambda x$ or $(A^T - \lambda I)(x^T) = 0$

The analytical way to solve for eigenvalues (the λ constants in the eigenvector equations) is to solve for the eigenvalues and solve for the null-space of the matrix corresponding to that particular eigenvalue. This process solves for the λ's and then solves the second form of the equation for the x term, which is simply the eigenvector. The simplest way to solve for the eigenvalues is to set up the characteristic equation for the matrix. The characteristic equation says that we want to find the values of λ that when subtracted from the original matrix, make the matrix singular, and give the matrix a determinant of 0. The equation for the characteristic equation is $det(A - \lambda I) = 0$. In the rainy, sunny, snowy example, we have a 3X3 square matrix where the characteristic polynomial would look like:

$$det \begin{pmatrix} .4-\lambda & .4 & .2 \\ .3 & .6-\lambda & .1 \\ .2 & .1 & .3-\lambda \end{pmatrix} = 0$$

By using determinant rules we can determine the eigenvalues of the matrix and then apply them to the left-eigenvalue formula. This method yields identical results to the other previously mentioned methods

Though the Markov model presented in this project is not often labeled this way, this process is more closely related to a Hidden Markov Model. A Hidden Markov Model (or HMM) is a process where the method involved is an underlying stochastic provess that can only be observed through another set of stochastic processes.[7]. Thus, in our model, we are trying to figure out Bach's personal harmonic model by the observable events of his chorale progressions. We learn more about the chorale harmonies in order to learn about Bach's probabilistic model of harmony. Embedded in the model is an overarching discovery of this HMM, and though we often refer to the model as a regular Markov chain, in reality we are also learning about this integrated hidden model.

With a basic understanding of how Markov Chains (hidden or not) work and how stationary distribution can be calculated, we can now begin to think about how Markov Chains can be used to describe chord progressions in Bach Chorales. These tools, such as solving for stationary distribution using iterative code and eigenvalues will be critical to the analysis of the method in this project.

Chapter 4

Model

In this chapter, we describe how we constructed the mathematical model for this musical phenomenon. Let us first look at a section of one Bach Chorale that was used to train the original model of this project. This chorale is a major-mode chorale from Bach's Christmas Oratorio (Part 2) entitled Brich an, O Schönes Morgenlicht (Break Forth, O Beauteous Heavenly Light). The chorale is in G-major, and thus would have the following seven chords available, with each root of the chord being one of the pitches from the G-major scale.

Figure 4.1: The seven chords based on G major [3]

With this set of available chords, we can begin to think about applying the concept of transition states to the chorale. Let us first look at the last section of the chorale, including the important, final cadence. In purple are the Roman Numerals for this section.

Figure 4.2: Closing Chords of Brich, an... Chorale in G major [2]

This closing section is very straightforward, and attempts to show how systematic cadences are in Baroque harmony. The final two chords of the piece are V and I (tonic and dominant) and they are preceded by a pre-dominant ii6 [1] . We see that in this example, the only state that does not have a certain state to follow is the V chord, which can transition to I or I6 with equal probability. When we determine the steady state of this situation using the left-eigenvector, and normalize so that the sum of all elements of the steady state is 1, we get a very regular fluctuation between .1 and .2 probabilities.

Table 4.1: Stationary Distribution of the Last Chords of G Major Chorale

Chord	Probability
I	0.2
I6	0.1
ii6	0.2
iii6	0.1
IV	0.1
V	0.2
vii0	0.1

[1] A chord in first inversion is called X6, second inversion is called X6/4.

However, when we try and solve for the stationary distribution using the iterative method, it takes exactly 196 times steps for the model to completely converge to the stationary distribution for all starting states of the chain. In the weather example in the introduction, the 3X3 matrix converged in 7 time-steps. One of the reasons why it takes much longer for this first Bach matrix to converge, as opposed to the weather matrix, is the size, length, and sparseness of the matrix. Note that many states are only attainable from one preceding state, and though the Markov chain is irreducible, it is still difficult to reach some states if you start in the state that is multiple degrees of accessibility away from the given state.

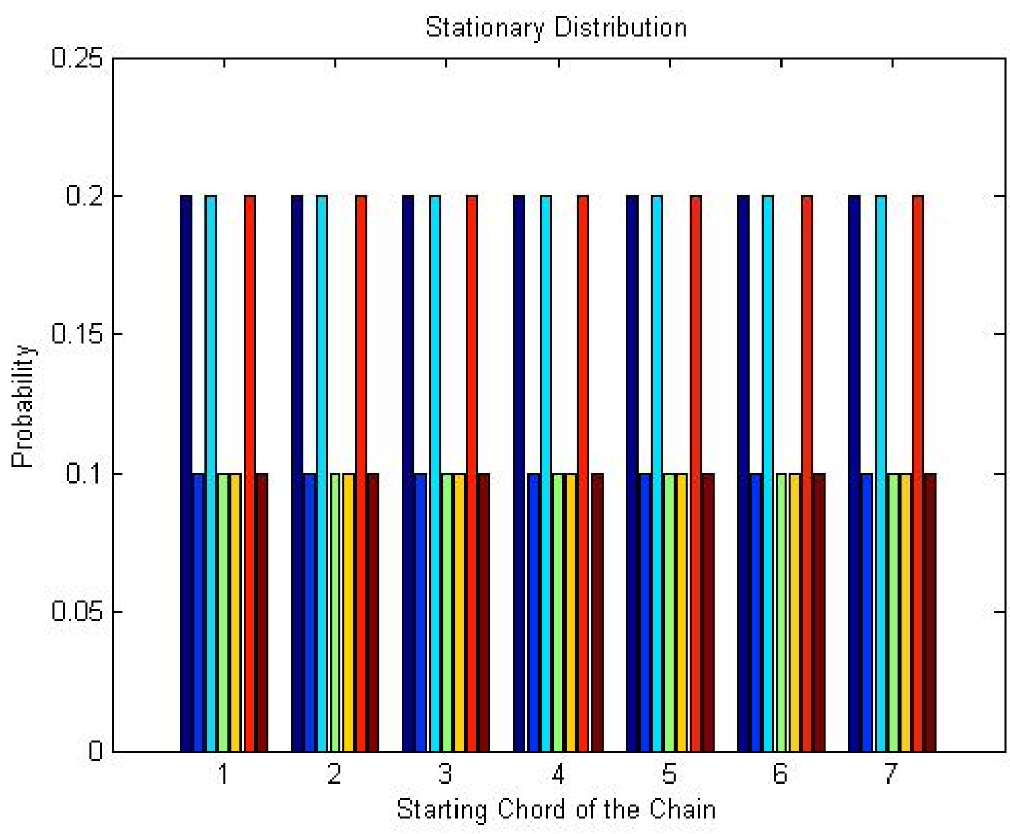


Figure 4.3: Stationary Distribution after 196 time-steps

What this initial set up also tells us is that this amount of data will not suffice in helping us create an idea for the Markov Chain. We need much more data in the hopes of creating a usable model. However, what this example shows is what we are trying to model. By taking the Roman Numerals, dividing up the occurances of each transition, and solving for the steady state, we are able to learn some important characteristics of this chain. First off, the three most likely states, with equal probabilities in the steady state analysis, are the I chord, the ii6 chord, and the V chord. Initially, it makes sense that I and V are most likely, since they are the only two chords in the passage that appear more than once. However, it is interesting to discuss why ii6 is also equally likely in this situation. In the steady-state analysis, contrary to intuition, the most popular or most frequent states in the training matrix do not always lead to the most likely state in the stationary distribution. The reason why ii6 is so probable is that is it well-connected to states such as I and V, which occur quite frequently.

The model, in principle, is supposed to represent all of Baroque harmony. Clearly, a set of 8 states (chords) in this case are not enough data to quantify all of Baroque harmony. Further analysis is needed to determine what types of patterns are common for Bach chorales. Also, this initial analysis highlights the important aspects of Markov chains in Bach chorales. The Markov chains in these chorales are irreducible, and therefore, every state is significant. Roman Numeral analysis is also useful in this setup because it describes relationships between chords and pitch classes, but is applicable to all keys. Since Bach's music uses almost every key, using chord symbols and relationships as opposed to notes is more useful. Overall, this method is defined to help solve the issues of harmony, and hopes to quantify a fully-accepted qualitative analytical method.

Chapter 5

"Original" Method

From the model setup in Part 4, we can see that this method has much to offer in terms of insight and analysis. Some limitations with our approach may arise if the constructed transition matrix is ill-conditioned. To overcome the ill-conditioned properties of the matrix, we must be sure to categorize the data appropriately. What this method is designed to achieve is information about the implicit Markov chains in Bach's chord progressions. By learning more about these progressions, we can then train a model with a set of raw data from the Bach chorales. With this data, we can run Monte Carlo trials (random trials using the probabilities in the Markov matrix) on the data set and see what kinds of progressions are selected by the transition matrix. We can then convert these progressions into sheet music and auditory examples, and analyze their quality.

The first step is to determine the transition probabilities in the Bach chorales. Unlike what was done in my previous work on this topic, I categorized the chords by related pitch (notated in Roman numerals) and what inversion the chord occurred. There were a few reasons why I chose this strategy as opposed to the one where I grouped chords into the 7 main pitch-class categories per scale type. First, when one notates a chord in the specific format, it actually describes the chord type while giving the exact bass note of the chord. The result is that less decisions have to be made for the other voices if the bass is already predetermined. Also, from multiyear studies of Western musical harmony and theory, there are very few paths to legitimately get from one chord to another, that is, in spacing and distance between notes in each voice part. My only disclaimer to this statement is that there are exceptions to every rule; however, factors like parallel fifth and octaves are almost never seen in Bach's 371 chorales.

To determine the transition matrices, I took four extremely famous Bach examples, three from Bach's St. Matthew Passion and one from the Christmas Oratorio and conducted a Roman Numeral analysis on each. By harmonic analysis, a person basically assigns each subinterval of the music with a Roman numeral value. Then, I counted the number of times the chorale had the same chord-transition and came up with transitions probabilities for each starting state. The rows were then normalized so that there is probability 1.0 that one ends up in a state. This method yielded separate matrices for each of the four chorales, and then I combined the data for the majors and the minors into two separate "combined transition matrices."

Even though I took into consideration pitch class and inversion, I did not take into consideration applied chords, or secondary dominants. Therefore, those chords were considered members of the pitch class of their root. Even though this omission seems like an oversight, it was made in order to keep the scope of the model manageable for this method. For example, all the matrices in this process were $13X13$ or $14X14$. If applied chords were used, it would almost double the size of the matrix and spread out the stationary distribution.

Below are the six transition matrices developed by this method of analysis. They are included to help categorize the differences in the types of chorale. Even though they are all nearly identical in size, there is a distinction between which sets of chords each mode of chorale is using. For example, minor chorales have III6 and vii6 chords, as where major chorales traditionally do not. It is also interesting to see how sparse the matrices are. However, once the data from two chorales is combined, it is already possible to see how less sparse the transition matrix is.

Table 5.1: Transition Matrix from Minor Chorale 1

	i	i6	ii	ii7	III	iv	iv6	V	V64	VI	VI6	vii	vii6
i	0.1	0	0.1	0.1	0	0.1	0.1	0.2	0.1	0.1	0	0.1	0
i6	0	0	0	0	0	0	0	1	0	0	0	0	0
ii	0	0	0	0	0	0	0	0	0	1	0	0	0
ii7	1	0	0	0	0	0	0	0	0	0	0	0	0
III	0.5	0	0	0	0.25	0	0	0	0	0.25	0	0	0
iv	0	0	0	0	0	0	0	0.5	0	0.5	0	0	0
iv6	0	0	0	0	0	0	0	0	0	0	0	1	0
V	0.63	0.13	0	0	0	0	0	0.25	0	0	0	0	0
V64	0	0	0	0	0	0	0	1	0	0	0	0	0
VI	0	0	0	0	0	0.25	0	0.25	0	0	0.25	0	0.25
VI6	0	0	0	0	0	0	0	0	0	0	0	0	1
vii	0	0	0	0	1	0	0	0	0	0	0	0	0
vii6	0.5	0	0	0	0.5	0	0	0	0	0	0	0	0

Table 5.2: Transition Matrix Based on Minor Chorale 2

	i	i6	ii	ii6	III	III6	iv	iv6	V	V6	V64	VI	vii	vii6
i	0.17	0	0	0.3	0	0	0	0.17	0.2	0	0	0.2	0	0
i6	0	0	0	0	0	0	1	0	0	0	0	0	0	0
ii	0	0	0	0	1	0	0	0	0	0	0	0	0	0
ii6	0	0	0	0	0	0.33	0	0	0.3	0	0.33	0	0	0
III	0	0	0.1	0	0.14	0	0	0.29	0	0.14	0	0	0	0.29
III6	0	0.25	0	0	0	0	0.3	0	0	0	0	0.3	0.25	0
iv	0.33	0	0	0	0.33	0	0	0	0	0	0	0	0.33	0
iv6	0	0	0	0	0	0.33	0	0	0	0	0	0	0.67	0
V	0.5	0	0	0	0	0	0	0	0	0	0	0.5	0	0
V6	1	0	0	0	0	0	0	0	0	0	0	0	0	0
V64	0	0	0	0	0	0	0	0	1	0	0	0	0	0
VI	0	0	0	0.3	0.33	0.33	0	0	0	0	0	0	0	0
vii	0	0.2	0	0	0.4	0	0	0	0	0.2	0	0	0	0.2
vii6	0	0	0	0	0	0.5	0	0	0	0	0	0	0.5	0

Interesting facets of these transition matrices point out the differences in the major and minor chorales is the use of III and vii in all inversions much more so in minor.

Table 5.3: Transition Matrix Based on Major Chorale 1

	I	I6	ii	ii6	iii	IV	IV6	V	V6	vi	vi6	vii	V64
I	0	0	0	0.2	0	0.2	0.2	0.2	0.2	0	0	0	0
I6	0	0	0	0	0	0	1	0	0	0	0	0	0
ii	0	0	0	0	0	0	0	1	0	0	0	0	0
ii6	0	0	0	0	0	0	0	1	0	0	0	0	0
iii	0	0	0	0	0.33	0.33	0	0	0	0.33	0	0	0
IV	0.33	0.33	0	0	0	0	0	0	0	0	0	0.3	0
IV6	0	0	0	0	0	0	0	0	0.5	0	0	0	0.5
V	0.43	0.14	0	0	0	0.14	0	0.14	0	0.14	0	0	0
V6	0.5	0	0	0	0	0	0	0	0	0.5	0	0	0
vi	0	0	0.3	0	0	0	0	0	0	0.25	0.25	0.3	0
vi6	0	0	0	0	1	0	0	0	0	0	0	0	0
vii	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5
V64	0	0	0	0	0	0	0	1	0	0	0	0	0

Table 5.4: Transition Matrix for Major Chorale 2

	I	I6	ii	ii6	iii	IV	IV6	V	V6	V64	vi	vi6	vii	vii6
I	0	0	0.1	0.1	0	0	0.1	0.4	0.1	0	0.2	0	0	0
I6	0.33	0	0.3	0	0	0.33	0	0	0	0	0	0	0	0
ii	0	0	0	0	0	0	0	0.2	0.2	0	0.2	0	0.2	0.2
ii6	0	0	0.5	0	0	0	0	0.5	0	0	0	0	0	0
iii	0	0	0	0	0	0	0	0	0	0	0	1	0	0
IV	0	0	0	0	0	0	0	1	0	0	0	0	0	0
IV6	0.33	0	0	0	0	0	0	0.33	0	0	0.33	0	0	0
V	0.5	0.3	0	0	0	0	0.1	0.1	0	0	0	0	0	0
V6	0	0	0	0	0	0	0	0	0	0.5	0	0.5	0	0
V64	0	0	0	0	0	0	0	1	0	0	0	0	0	0
vi	0	0	0.3	0.3	0.25	0	0.3	0	0	0	0	0	0	0
vi6	0.33	0	0.3	0	0	0	0	0	0	0	0	0.3	0	0
vii	1	0	0	0	0	0	0	0	0	0	0	0	0	0
vii6	0	0	0	0	0	0	0	0	0	1	0	0	0	0

In major, chords of all inversions based on IV and vi are much more predominant. Also, in the four matrices there is a trend that most transition probabilities are 0. At first glance, the lack of transition states seems to indicate that there are not many possibilities for chorales. This assumption is only somewhat true. Chords can only move to a limited number of other chords, but can happen in different orders or inversions, thus the possibilities are not limited.

Once I had obtained the transition matrices, I first wanted to take a look at the stationary distributions, to certify that the chains behaved as I would expect (tonic and dominant chords, I and V to be most popular). What I found was somewhat to be expected: either the I or V chord was most probable in any situation. Also, it took about 12-13 time steps for the stationary distribution to be reached within .001 accuracy. However, I seemed to be more probable in minor-mode chorales, and V seemed to be more influential in major-mode chorales. In figure 5.1, we can see how these transition probabilities lie graphically.

The first, highest purple column corresponds to the I chord, and the next highest purple column corresponds to the V chord. It is interesting to note that both the probabilities of I and V occuring at steady-state are above .2, which seemed to be an arbitrary baseline level determined by the model set-up in Chapter 4. It is hard to be certain, but the shortest columns are all actually equal in probability. In table 5.5, the steady state probabilities for this transition matrix are given.

It is also interesting to note that besides I and V, all of the other states share at least one other state with equal probability. This facet of the model seems to promote the idea of a state-hierarchy. With only one chorale, it is difficult to be certain that the hierarchy is data-independent, so it is helpful to look at other instances of this phenomenon to gain a better understanding of the mechanism that supports it.

Compared to Minor Chorale 1, the transition matrix based on Major Chorale 1 is somewhat more active. The row containing V has many more nontrivial values and

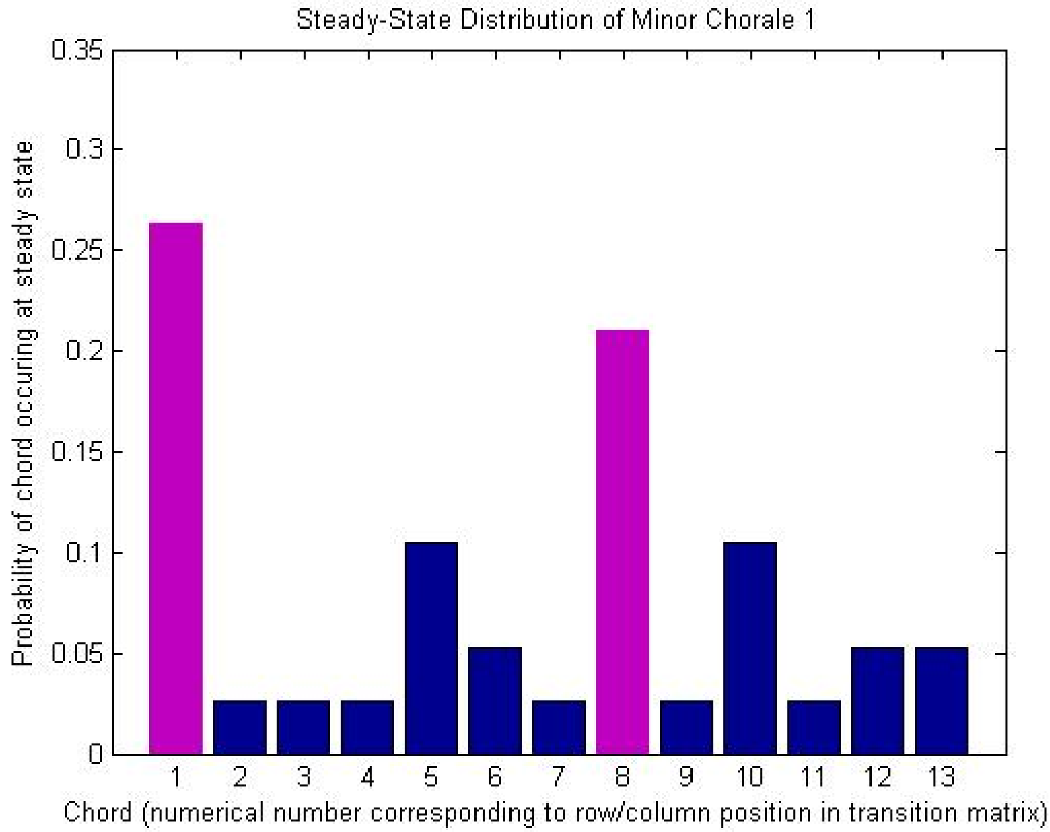


Figure 5.1: Stationary Distribution of Minor Chorale at 12 time-steps

interesting probabilities. However, many of the values are sparse just as it was in Minor Chorale 1. Also, the probability values of the I and V chord are much lower than their major counterparts. These two important chords are less distinguished in this model compared to the minor chorale. Also, this transition matrix took 18 time-steps to reach the stationary distribution. The increased amount of time seems to signal what happened in the model setup, that because the data was more spread out it takes much longer to settle. There are more rows and columns and this matrix with only one value, and therefore the accessibility of this matrix is matrix is challenging.

The highlighted purple columns of the bar graph again point to I and V. However, in this major-mode chorale, V has the highest probability value. If we take a look at the matrix, it makes sense that V is most dominant, because it is a more connected

Table 5.5: Steady State for Minor Chorale 1

Chord	Probability
i	0.263157895
i6	0.026315789
ii	0.026315789
ii7	0.026315789
III	0.105263158
iv	0.052631579
iv6	0.026315789
v	0.210526316
v64	0.026315789
VI	0.105263158
VI6	0.026315789
vii	0.052631579
vii6	0.052631579

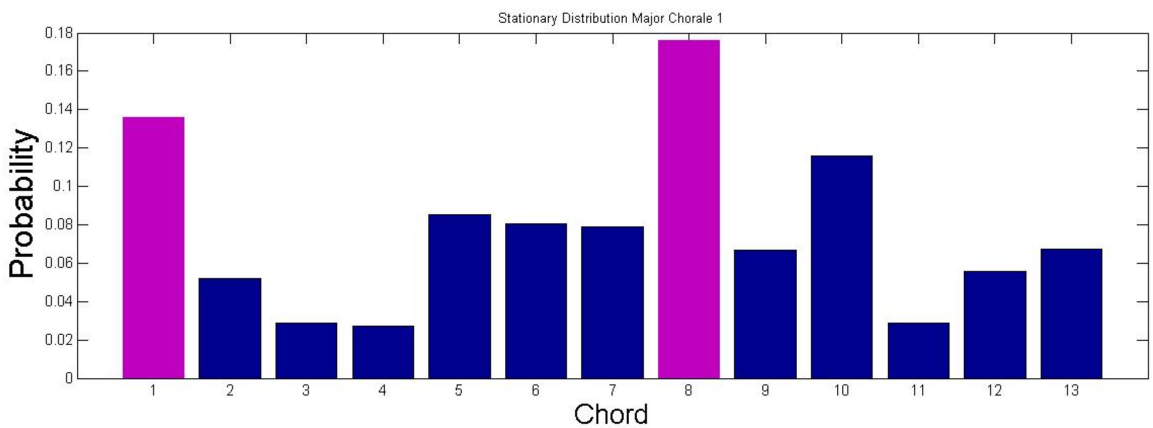


Figure 5.2: Stationary Distribution of Major Chorale at 18 time-steps

chord. Also, there is not paired-hierarchy as there was in Minor Chorale 1. Therefore, there is no pattern of matching values, as it depends on the matrix how the stationary distribution will appear. Another difference is that in the Minor Chorale stationary distribution, after I and V there was a large drop-off to the next highest probability. However, in the Major Chorale stationary distribution, the third highest probability is very close to the probability of I. The probability corresponds to the vi chord in root position. We can see that after I and V, vi has the most non-zero values in the row associated with vi in the transition matrix.

Table 5.6: Stationary Distribution of Major Chorale 1

Chords	Probability
I	0.135645665
I6	0.052034986
ii	0.028982315
ii6	0.02712913
iii	0.085389689
IV	0.080720498
IV6	0.079164111
V	0.175895475
V6	0.066711182
vi	0.11592927
vi6	0.028982315
vii	0.055888876
V64	0.067526487

Next, I decided to look at the difference between data from one chorale and data from two chorales of the same mode.

Table 5.7: Transition Matrix for Minor Chorales Combined

	i	i6	ii	ii6	III	III6	iv	iv6	V	V6	V64	VI	VI6	vii	vii6
i	0.14	0	0.1	0.2	0	0	0.1	0.14	0.2	0	0.05	0.1	0	0.05	0
i6	0	0	0	0	0	0	0.5	0	0.5	0	0	0	0	0	0
ii	0	0	0	0	0.5	0	0	0	0	0	0	0.5	0	0	0
ii6	0.5	0	0	0	0	0.17	0	0	0.2	0	0.17	0	0	0	0
III	0.25	0	0.1	0	0.2	0	0	0.15	0	0.07	0	0.1	0	0	0.15
III6	0	0.25	0	0	0	0	0.3	0	0	0	0	0.3	0	0.25	0
iv	0.17	0	0	0	0.17	0	0	0	0.3	0	0	0.3	0	0.17	0
iv6	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0.84	0
V	0.56	0.06	0	0	0	0	0	0	0.1	0	0	0.3	0	0	0
V6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V64	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
VI	0	0	0	0.2	0.17	0.17	0.1	0	0.1	0	0	0	0.13	0	0.13
VI6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
vii	0	0.1	0	0	0.7	0	0	0	0	0.1	0	0	0	0	0.1
vii6	0.25	0	0	0	0.25	0.25	0	0	0	0	0	0	0	0.25	0

Table 5.8: Transition Matrix for Major Chorales Combined

	I	I6	ii	ii6	iii	IV	IV6	V	V6	V64	vi	vi6	vii	vii6
I	0	0	0.1	0.2	0	0.1	0.2	0.3	0.2	0	0.1	0	0	0
I6	0.17	0	0.2	0	0	0.17	0.5	0	0	0	0	0	0	0
ii	0	0	0	0	0	0	0	0.6	0.1	0	0.1	0	0.1	0.1
ii6	0	0	0.3	0	0	0	0	0.75	0	0	0	0	0	0
iii	0	0	0	0	0.17	0.17	0	0	0	0	0.17	0.5	0	0
IV	0.17	0.17	0	0	0	0	0	0.5	0	0	0	0	0.17	0
IV6	0.17	0	0	0	0	0	0	0.17	0.3	0.25	0.17	0	0	0
V	0.46	0.22	0	0	0	0.07	0.1	0.12	0	0	0.07	0	0	0
V6	0.25	0	0	0	0	0	0	0	0	0.25	0.25	0.3	0	0
vi	0	0	0.1	0	0	0	0	0.5	0	0	0.13	0.1	0.13	0
vi6	0	0	0.1	0.1	0.63	0	0.1	0	0	0	0	0	0	0
vii	0.17	0	0.2	0	0.25	0	0	0	0	0.25	0	0.2	0	0
V64	0.5	0	0	0	0	0	0	0.5	0	0	0	0	0	0
vii6	0	0	0	0	0	0	0	0	0	1	0	0	0	0

The Minors combined data is pretty unique. It is the first time where V does not appear one of the two most popular chords, and in this case the popular chords and i, III, V, and VI, respectively. Curiously, the latter three chords are all major-mode, and they make up the majority of this minor-mode piece.

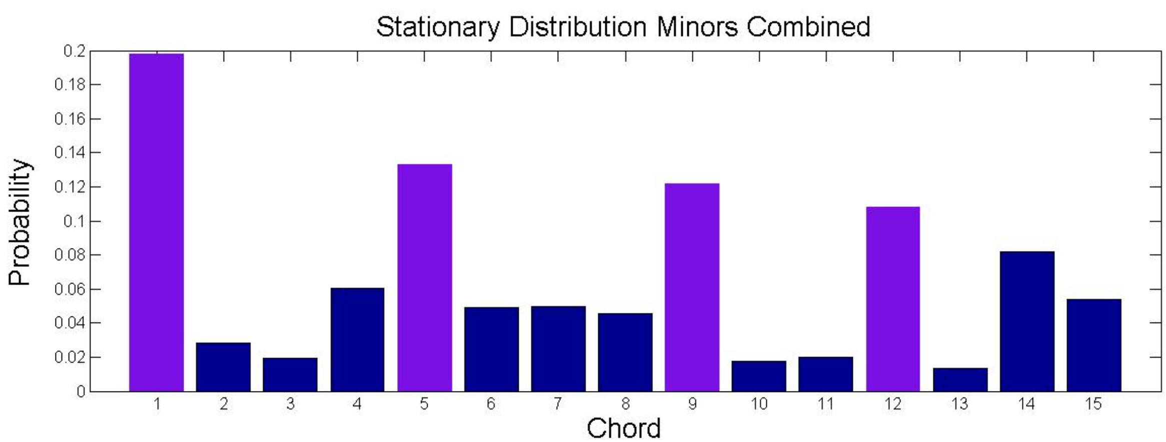


Figure 5.3: Stationary Distribution of Minor Combo at 11 time-steps

The majors combined data was very different than the minor data. First off, V was the most dominant probability, followed by I. However, there was a significant

Table 5.9: Stationary Distribution of Minor Chorale Combination

Chord	Probability
i	0.197954768
i6	0.028079424
ii	0.019392248
ii6	0.06053911
III	0.132923128
III6	0.049093316
iv	0.049694905
iv6	0.045382987
V	0.121881872
V6	0.017682987
cad	0.01998759
VI	0.107873007
VI6	0.013484126
vii	0.081884783
vii6	0.054145748

drop-off to the next most important probability. Therefore, it appears as though minor chorales have much more variance in the chords. Major chorales seem to stick to the tonic-dominant divide much more strictly.

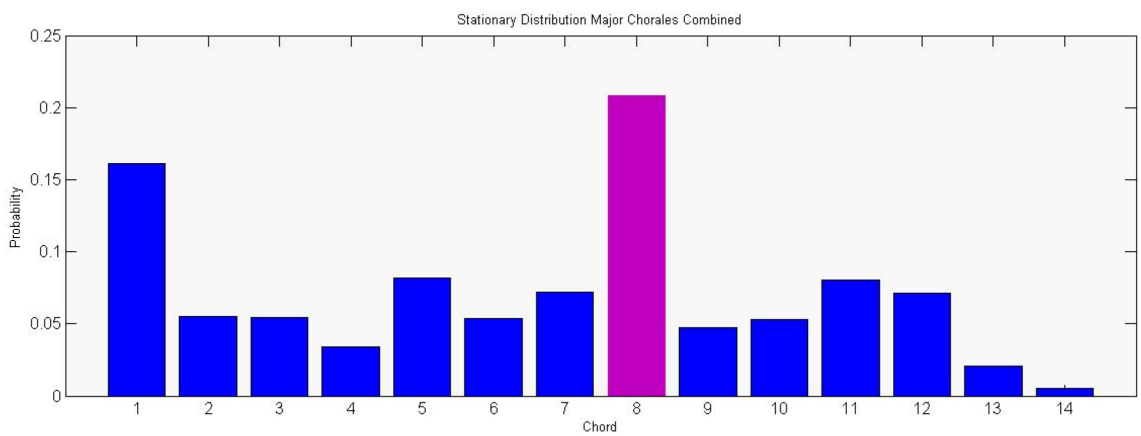


Figure 5.4: Stationary Distribution of Major Combo at 11 time-steps

Once I was certain of the validity of the matrices, it was now time to create Bach style chorales. To accomplish this task, I performed randomized trials with the aforementioned transition matrices. These trials involved generating a random

Table 5.10: Stationary Distribution Major Chorales Combo

Chord	Probability
I	0.161060512
I6	0.05501491
ii	0.05442212
ii6	0.034239453
iii	0.081910286
IV	0.053782533
IV6	0.072145311
V	0.207972152
V6	0.047637358
vi	0.053243492
vi6	0.080644499
vii	0.071423724
cad	0.021061469
vii6	0.005442182

vector with the same length as the number of chords desired for the chorale. The components of this vector are numbers between [0,1] with a uniform distribution. Then, the probabilities in the transition matrix acted as cutoffs. For example, if there were two possible transitions, each with probability .5, and the component of the vector was .4311, the chorale would then move to the state A. If the value was .6, then the chorale would instead move to state B as the next chord. In this manner, I was able to obtain chorales for each of the four transition matrices. Sample audio and visual examples can be found on the project website. [10]

What I noticed from the first four chorales is although they are not unrealistic, they are not very interesting. Part of the excitement of Bach's music is that he does somewhat unexpected things, and uses rhythm in an interesting way. However, the way these chorales are set up, there is too much regularity. Despite the limitations of the model, the trials had many redeeming qualities. The chorales were all successful, with no crazy transitions or jarring moments. The chorales have the capacity to add ornamentation, and the resulting melody is much more interesting that in my previous work. (see website for Major Chorale 2 MIDI [10].

Some of the issues I noticed were that the system could get caught in traps. That is, if one were able to move from I-IV and back from IV-I, there would be times where the chords would move back and forth two to three times. The commutatability of the chorales is helpful, however, with limited data there is a higher probability of consitently commuting between states. This repetition is very unlikely in Bach's music. Another somewhat unrealistic aspect I noticed is that one of the trials used the cadential V6/4 chord four times. The maximum number of cadential 6/4 chords I have ever seen in a choral is about two, so four would be extremely unlikely. What these two aforementioned examples show are the limitations to the model in terms of ability to create realistic music. If we told the model not to allow more than two cadential 6/4 chords to happen in a piece, we could reduce this musical error. All of these issues would have to be recognized and addressed in advance.

Figure 5.5: Last Four Bars of Major Chorale 2 Trial

In the last four bars of this trial, there is a lot of the going back and forth; a very unsettling quality for music. The end of chorale would most likely not end in such a fluxuating manner.

The last experiment I conducted was to assess what quality chorales would be developed if the results of the minor trials were averaged with each and the results of the major trials were averaged with each other. I hypothesized that the results would be worse and more jumbled sounding. However, even better results were obtained with more data in the data set, despite its combination. Subjectively, the combined minor trial result is the best of all 6 trials. The major combined trials chorale is also much better than the singular trials and has little cases of repeating changes over again. I also did a transition matrix for all the data combined, but it does not make sense to use that model because it is major and minor together, and the different properties of each type negate in a transition matrix. This transition matrix is added to show how different major and minor are, and when combined create a much different matrix.

Another addition to this experiment was the addition of passing 8th notes. When a voice moved by a musical interval of a third in subsequent beats, the model would fill in the third with a passing 8th note. This addition to the model gives the piece a more fluid and lyrical melodic line. The addition of the eight notes also seems to fill in the jumps and gaps that were slightly jarring in the previous trials.

This model showed the power of Markov Chains in relation to Bach's music. With only a few predetermined conditions, and a knowlegde of the probabilities of moving from chord to chord, we can recreate Bach-like music. From the trials, it is already possible to see the extensions of this model, with more parameters, creating something even more spectacular. I think what this "Original" method showed is how important I and V are to major and minor chorales, but that their importance changes based on the mode and for each chorale.

What one can take away from this process is that it is indeed possible to create Bach chorales from Monte Carlo style simulations of the Markov chain. For this goal to be accomplished, however, many individual cases need to be considered (repeti-

Figure 5.6: Bars 6-9 of Minor Chorale Combo Trial

tions, number of rare chords in a section, number of cadence, etc.) and dealt with in the model. What we can definitively conclude is that Bach's musical prowess is certainly irreplaceable and even the most successful trials of chorales lack a sense of ingenuity behind them. Despite this lack of genius, mathematics and probability are an excellent substitute for creating interesting and surprising results.

Chapter 6

Improved Model

After creating some examples in the original model, we wanted to understand whether we would be able to produce Bach-quality chorales on a larger and more automated scale. We decided to pursue a more detailed and improved approach that took into consideration the oversights of the original model. In order to assess the quality of the synthetically-produced chorales we conceived an experiment (chapter 7) where 40 musical experts participated in a study that assessed the authenticity of chorales through audio examples. Afterwards, we determined that this model was indeed "improved", as the new methodology proved to trick musical experts into believing simulated examples were authentic Bach at the same probability of guessing. These results will be explained fully in the following chapters.

The first method of chorale creation did provide some interesting results. However, despite its positive attributes, there are many aspects of this method that needed be improved in hopes of demonstrating the success of the model. For example, the oversight of the applied chords was detrimental to the project. Even though they create larger matrices, they are crucial to the idea of Bach chorales. No Bach chorale stays in the same tonal area for the entire time. Bach tonicizes different harmonies (or makes different keys important) and often cadences to different keys throughout the shortest of chorales. This oversight was intentional, but for the model to be more successful, these chords needed to be taken into consideration.

Another problem with the original method is the inefficient use of time. Creating the Markov matrices from one's handwritten Roman Numeral analysis is extremely tedious. Then, getting the data and analyzing it for each trial is also time-consuming, as each chorale had to be dealt with separately. The audio trials and visual scores were all created manually, as the data had to be read off and transferred to the musical notation software Sibelius. Even though the scores are nice in the Sibelius format, there needed to be a better way to accomplish these tasks efficiently.

The reason for wanting to be more efficient stems from the goal to improve the model. In this model, we hope to take what was learned from the previous method and use more data to achieve better generalizations. We learned that two chorales created better trials than one chorale, but at what point does adding chorales not improve the model? Does more data always correlate to a "better" chorale trial?

In this next attempt, we tried to create transition matrices and chorale trials based on greater amount of data in a more efficient manner. Therefore, we had to find a way to get the Roman Numeral Analysis of more than four Bach chorales. This task was not necessarily easy. There is, however, a data resource started by a current MIT Professor during his time at the The Ohio State University. Professor David Huron and others on the project developed a programming system for music research called Humdrum, which can accomplish analytical tasks on large sets of musical data. [8] The music in this program has to be in a .KERN data format. In this format, each line of code represents the music that is occuring at that moment in time. Each line of code is the smallest metrical subdivision of the entire work.

While browsing the open source material for the .KERN data set, I discovered almost all of the Bach chorales have been converted to the .KERN format. However, the HUMDRUM program is not available for personal use outside of Stanford. I began to try and think of a code I could write but it became too difficult a task. The main problem is I had to think about how I could get a program to read the data and separate chords, because there are several different capitalizations which mean the same chord. It takes years of knowlegde of exceptions, rules and general trends, and it became difficult to execute given my limited programming skills. However, the last document in the page for each piece contains a Roman Numeral analysis. Apparently, someone on the HUMDRUM project did the analysis by hand and added it to the model. Therefore, I now have access to almost 100 Roman Numeral Analyses for Bach Chorales. [8]

```
**tshrm **tsroot          **kern  **kern  **kern  **kern  **root
!       !       !bass     !tenor  !alto   !soprno !       !
*       *       *Iorgan   *Iorgan *Iorgan *Iorgan *Iorgan *
*       *       *clefF4   *clefF4 *clefG2 *clefG2 *       *
*k[b-e-]        *k[b-e-]          *k[b-e-]        *k[b-e-]
*       *       *B-:      *B-:    *B-:    *B-:    *B-:    *B-:
*M3/4   *M3/4   *M3/4     *M3/4   *M3/4   *M3/4   *M3/4   *M3/4
*MM60   *MM60   *MM60     *MM60   *MM60   *MM60   *MM60   *MM60
vi      B-      8BB-      4d      4f      4b-     4BB-    I
.       B-      8AA       .       .       .       .       .
=1      =1      =1        =1      =1      =1      =1      =1
.       G       8GG       2d      2g      2b-     2GG     vi
.       G       8AA       .       .       .       .       .
.       G       8BB-      .       .       .       .       .
.       G       8C        .       .       .       .       .
.       D       8D        4d      4f      4a      4BB-    I7
.       B-      8BB-      .       .       .       .       .
=2      =2      =2        =2      =2      =2      =2      =2
IV      E-      8E-       2B-     2e-     2g      2EE-    IV
.       E-      8F        .       .       .       .       .
.       E-      8G        .       .       .       .       .
.       E-      8A        .       .       .       .       .
I       B-      8B-       4B-     4d      4f      4BB-    I
.       B-      8A        .       .       .       .       .
```

Figure 6.1: An example image of the KERN data format [8]

The problem then became what I should do with the data. Originally, I began to write out the transition states by hand and calculate thier occurance. However, I was able to upload the .KERN data into Excel and extract just the Roman Numerals, and calculate the occurance of each transition. The only slightly time-consuming aspect of this method is going from occurances to transition matrix, but in Excel one can write formulas to divide the number of occurance of one transition by all the possible transitions to normalize. Overall, this method is much faster and has made the process of data collection much simpler.

The first trial of this method was to compare the data from one chorale in this method compared to one chorale in the previous method. In this chorale, which I stopped categozing by name but by a number (to be blind in my selection of chorales), there were 18 chords used with the new system of classification. In the improved method, inversion chords are described with letters. That is, a ”b” after a chord represents the first inversion chord. ”c” and ”d” represent second and third inversion chords, respectively. Also, I considered applied chords as separate states. In this first trial, there was only one applied chord, the V7/IV, which is the dominant-seventh chord of the tonic in root position. In the simplistic method, this chord would have just been called a I or a I7.

Table 6.1: Transition Matrix Major Chorale 1

	I	Ib	ii	ii7b	iib	iiib	IV	IV7d	Ivb	V	V7	V7/IV	V7c	Vb	vi	vi7b	viio7d	viiob
I	0.2	0.067	0.067	0	0	0	0	0	0.067	0.33	0	0	0	0.2	0	0	0	0.067
Ib	0	0.2	0	0.2	0	0	0	0	0	0	0.2	0.2	0.2	0	0	0	0	0
ii	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ii7b	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
iib	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
iiib	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
IV	0.33	0	0	0	0	0	0	0.333	0	0	0	0	0	0	0	0	0.33	0
IV7d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ivb	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.5	0	0	0
V	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0.6	0	0	0
V7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V7/IV	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
V7c	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vb	0.5	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0.25	0	0
vi	0	0	0	0.25	0	0.25	0.5	0	0	0	0	0	0	0	0	0	0	0
vi7b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
viio7d	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
viiob	0.67	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

What I noticed right away about this method was that the matrix is even more sparse than before. There is also much more regularity in occurances of transition, as all probabilities in this case were 1/5, 1/4, 1/3, 1/2 and 1. Though I know when more data is added, these probabilities will become more complex, it is still interesting to see the regularity in most states. With the new matrix created, I thought I should do a steady state analysis of the matrix.

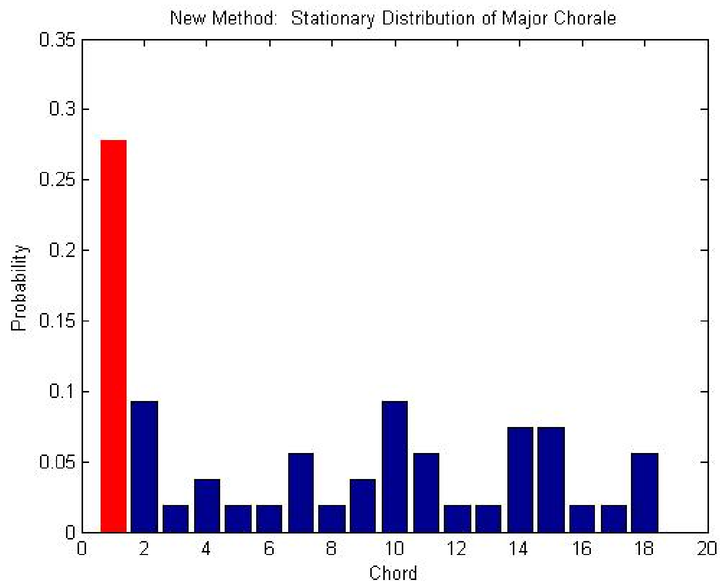


Figure 6.2: Improved Method Major Chorale 1 Stationary Distribution

However, it is interesting to note how significant the I chord is in this trial as demonstrated by figures 6.2 and 6.3. In the past, I and V are always close in range, but we can see that the I chord is almost three times more likely in this model than the V chord. However, it is hard to see from the diagram but V and Ib (I in first inversion) are tied for second-most likely chord at steady-state. Therefore, the principle that I and V are still most important has not been broken.

At this point in the analysis, it is time to create the data sets for additional analyses. What we learned from the simplistic method is that it is productive to keep minor chorale and major chorale data separate, and therefore we will continue that process. However, we continued the analysis in a more experimental way. Instead of creating more trials and giving my opinion, we developed an experiment to see how real musicians view the success of this model. We first compiled transition matrices based on 1, 2, 4, 8, 12, 16, 22 and 25 chorales for the major mode.

Table 6.2: Analytic Stationary Distribution of Major Chorale 1

Chord	Probability
I	0.277777778
Ib	0.092592593
ii	0.018518519
ii7b	0.037037037
iib	0.018518518
iiib	0.018518518
IV	0.055555556
IV7d	0.018518518
Ivb	0.037037037
V	0.092592592
V7	0.055555556
V7/IV	0.018518519
V7c	0.018518519
Vb	0.074074074
vi	0.074074074
vi7b	0.018518519
viio7d	0.018518518
viiob	0.055555556

The problem is that it means that there will need to be 8 new matrices and multiple new chorale trials per new transition matrix. I then needed to figure a more productive way to create chorale examples. WolfraAlpha Mathematica has the capability to take data and turn it into MIDI's of pitch. Therefore, while the analysis will still occur in the MATLAB format, the data can now be produced by Mathematica. This change will mean that the sheet music for each trial will not be readily available, despite this downside, we hope that the easily created audio examples will be more than adequate for musicans to analyze.

Chapter 7

Experiment

In this chapter, we present the results of our experiment conducted in the winter of 2013. Our findings show that the improved methodology proposed in chapter 6 was able to yield realistic sounding Bach harmonic progressions. Statistically, the results of the experiment demonstrate that the responses to the survey most accurately model guessing. In the sections of this chapter, we will look at the experimental setup that led to these results, as well as statistical hypothesis testing that supports the claim that the model sufficiently created Bach-like musical examples.

7.1 Experimental Design

We now have great range of data and chorale examples created by the improved method. However, my arbitrary and subjective analysis of how successful different types of chorales are in terms of representing Bach's harmonic language is not sufficient enough to make large claims about the success of the model. Therefore, we turn to a sampling of musicians to ask them about what they think sounds like real Bach or not.

Consulting with Professor Santillana and Dr. Simon Lunagomez, an experiment was devised to help quantify the qualitative difference between chorales created by different transition matrices. In this experiment, we looked at approximately forty musicians or “experts.” Each of them analyzed two sets of chorales examples: chorales by Bach, and chorales created by a random Markov process using one of the transition matrices created by the model. These examples were randomly distributed throughout the survey, however, there were equal amounts of Bach and model-created examples per track of the experiment. The participants were not told how many of each kind of example were in the track, or that they were indeed equal.

The data is assessed in a binary way: does this example sound like Bach or does it not sound like Bach? The decision to rate the quality of an expert in a binary manner was made in order to help reduce the complexity that quantification of the subjective quality. Also, having more gradations in the assessment means there is additional variance in the way different musicians perceived the sounds. Therefore, two choices seemed to be the best way to approach the model.

We then calculated the percentage of correct responses for each participant. We hoped that the number of correct responses would approach 1/2. A range of fractions of correct response around 1/2 would mean that it is equal probability that the person picked Bach or the model as the chorale that sounds like Bach.

After all of the data was collected from the experts (5 experts were chosen to analyze the sections trained by X number of chorales) statistical analysis was performed.

There were 8 levels of training, 1, 2, 4, 8, 12, 16, 22, and 25 chorales per track, and all examples were in the major mode. For the purposes of this experiment, only major mode chorales were used in the experiment to train the model. We expected 5-10 experts analyzing each of the levels of training, in hopes of getting a good average for the number of completed responses.

The chorales chosen to train each of the Markov chains were chosen at random, by using MATLAB’s random number generator. Therefore, no biases were made as to which of the chorales would be most effective at training the matrix to create Bach-like examples.

7.2 Experiment Results

We analyzed the results of the experiment, and our main goal was to do various forms of hypothesis testing on the data. We wanted to see if we could determine any trends or cutoffs as to the number of chorales necessary to make the transition matrix able to produce Bach-like examples. In the experiment, the main statement we tried to prove is stated below.

Statement: The computational algorithm devised to create realistic sounding Bach harmonies for Bach chorales is successful when the transition matrix is trained by ten or more real Bach chorales.

7.3 Testing the Null Hypothesis

7.3.1 Analytical Approach 1

Null Hypothesis 1: The algorithm will have the same affect on the probability that a pariticipant of the experiment gets a questions right as if they were guessing. This affect is equivalent to saying that the probability of getting a correct response is $p = 0.5$, the approximate probabilty of correct response if a person simply guessing.

Figure 7.1 represents the data from all trials of the study. Judging from the box plot, there seems to be shift in the data trend after using eight chorale examples to train the transition matrix. After initially increasing towards almost a fraction of 0.7 correct responses (in the noise section), the results seem to decrease and converge towards a fraction of 0.5 correct responses.

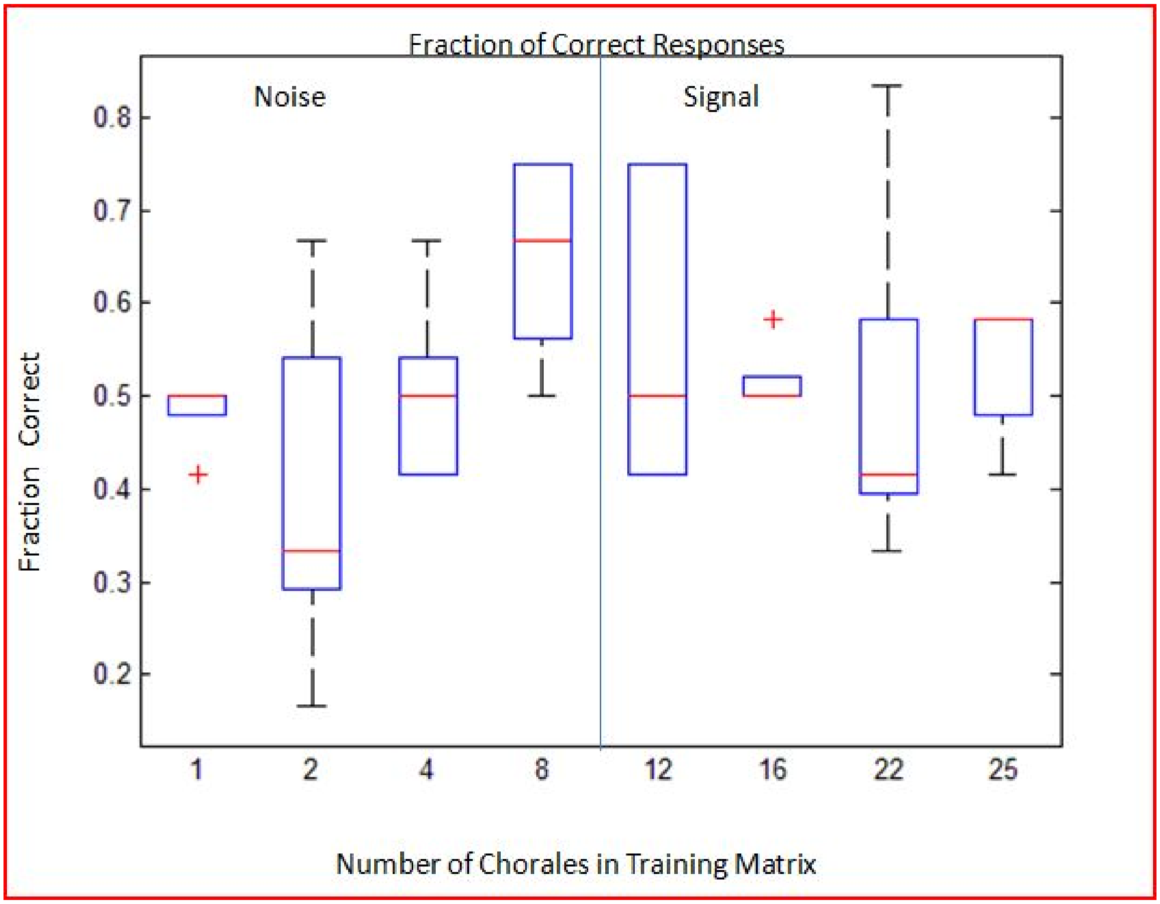


Figure 7.1: Box Plot of the Correct Responses Fraction vs. Number of Chorales in the Training Matrix

However, for chorales trained with more than 10 sample chorales, we cannot reject the null hypothesis. These values and boxes seem to converge towards $p = 0.5$. We can attempt to reject the null hypothesis by comparing the fraction of correct responses across all levels of training to the null hypothesis in terms of a binomial distribution. We assume that with enough data, this binomial distribution will model a normal distribution. Having 240 data points from the experts, we have enough data to approximate a normal distribution with a Bernoulli experiment that will lead to a binomial distribution for p ~ 0.5. With this data, we can apply the priniciples of the normal distribution, and determine the statistical likelihood of that fraction of correct responses occuring in this type of test. [1]

The null hypothesis states that the fraction of responses should be about 0.5, since the results of the experiment should match the model for guessing. For the 4 levels of training studied that used more 10 or more chorales to train the transition matrix, there were 20 participants, five per each level of training, and 12 questions per participant. The total correct responses was 125 of 240 questions, with a fraction of correct responses of 0.521.

After using statistical hypothesis testing and reducing the fraction of correct responses to its most simplified form 25/48, the p-value of this two-tailed hypothesis test was determined to be 2*0.11 = 0.22. This p-value is well over the significance level of 0.05. Therefore, we cannot reject the idea that this data behaves like a Gaussian distribution with average probability of 0.5.

The average fraction and its p-value do not reject the null hypothesis, which in turn, supports our original statement that the algorithm is capable of producing realistic sounding Bach harmonies when at least 10 chorales are used in the training set. In other words, participants believed that synthetic chorales were really Bach, and vice-versa.

7.3.2 Analytical Approch 2

Null Hypothesis 2: The algorithm is inherently bad, so people should be able to distinguish between model-created examples and real Bach. The probability of correct response should be $p \geq 0.75$. This high probability would mean that the model would fail. If we can reject this null hypothesis, we can say that the model is not inherently bad.

This null hypothesis can be rejected by performing an analysis similar to part 1. Instead, we use a one-sided hypothesis test. The fraction of correct response does not change for the region of [12,25] chorales used in the training set, and is 0.521. Because the total fraction of correct responses in this "tail" region of the graph, the one-tailed analysis of this null hypothesis concludes that the null hypothesis is rejected, as the p-value is $3.310 * 10^-4$, which is not outside the 0.05 significance level.

On a superficial level, we have statistically shown the statement we are trying to prove is true; that is, the algorithm is tricking participants. When there are ten or more chorales in the training set, there is statistical evidence to support the claim that the algorithm is at least successful in this range.

7.3.3 ANOVA

Another way to reject the null hypothesis is through the use of one-way analysis of variance (ANOVA). This analysis is included as to prove yet another method for analyzing the results of the experiment, hoping to show that the statement cannot be proven false. The ANOVA tests whether the subdivision of the experiments by the number of chorales in the training set has any effect on the fraction of correct responses. This test is performed to justify that this method of chorale creation is reasonable, and the reasonably justify that after 10 chorales, there is a less noticeable difference in average correct responses per level of training.

The analysis of variance (ANOVA) of the data in this experiment shows almost an identical finding to the box plot. Similar to the box plot, the general trend matches the upwards swing around 8 chorales, and dips lower afterwards. Again, this test supports the idea that the region of greatest importance is the region when there are 10 or more chorales in the training set.

The p-value for this one-way ANOVA was $p = 0.1610$. This p-value suggests that there is no significant difference between the number of chorales in the training set and the number of correct responses. Similar to the analysis of the box plot (medians and quartile relationships), this ANOVA also shows the trend towards a 0.5 fraction of correct responses. Therefore, we cannot reject the original null hypothesis (once again) that states that the probability of a correct response is 0.5 for 10 or more chorales in the training set. The reason that 0.5 is critical to proving the success of the model is that if correct responses model a Gaussian (like un-biased coin-flipping), we have proven that experts have had difficulty distinguishing between Bach and model-created harmonies.

7.4 Exploratory Data Analysis

We thought is was important to provide additional data and analysis that further supports the claimes of the quality of the improved method. This additional analysis involves different factors that were not considered in the hypothesis testing.

In the next section of data analysis, we are looking at the fraction of Correct and Incorrect Responses from the experiment. All of the figures represent the average fraction of responses in that category. Fraction "tricked" means that the example was created by the model, but was identified by the participant as Bach. Fraction Model Correct represents model-created examples being identified as fakes by the participant.

Table 7.1: Mean Fractions of Correct or Incorrect Responses from the Experiment, Training Sets 1-8

Number Chorales in Training	1	2	4	8
Fraction Correct	0.483 (.042)	0.4 (.212)	0.514 (.107)	0.65 (.121)
Fraction Bach Wrong	0.567 (.091)	0.633 (.139)	0.472 (.167)	0.367 (.139)
Fraction Bach Correct	0.433 (.091)	0.367 (.139)	0.528 (.164)	0.633 (.139)
Fraction "Tricked"	0.467 (.075)	0.567 (.253)	0.5 (.149)	0.333 (.204)
Fraction Model Correct	0.533 (.075)	0.433 (.253)	0.5 (.149)	0.666 (.204)

7.5 Additional Analysis

In this section, we provide further mathematical thoughts regarding the results of the experiment. Though we have already proved that the improved method is likely

Table 7.2: Continuation of Table 1 for Data from Training Sets 12-25 (Standard Deviation in Parenthesis)

Number Chorales in Training	12	16	22	25
Fraction Correct	0.567 (.191)	0.472 (.125)	0.5 (.219)	0.548 (.088)
Fraction Bach Wrong	0.4 (.253)	0.5 (.118)	0.633 (.274)	0.524 (.167)
Fraction Bach Correct	0.6 (.253)	0.5 (.105)	0.367 (.274)	0.476 (.150)
Fraction"Tricked"	0.467 (.247)	0.556 (.228)	0.333 (.204)	0.381 (.185)
Fraction Model Correct	0.533 (.247)	0.444 (.228)	0.633 (.183)	0.619 (.185)

to be successful, it is interesting to look at other factors besides fraction of correct response and their relationship to the findings.

More evidence is available to support the claim that additional chorale data in the training set creates more Bach-like, or indistingushiable-from-Bach examples. Below, we see that the relative percent error between L2 norms of the transition matrices are steadily decreasing and converging towards zero. Also, to support our claim that the noise-signal break occurs between 8 and 12, the dropoff between 8 and 12 on this plot of norm differences is most steep in this interval.

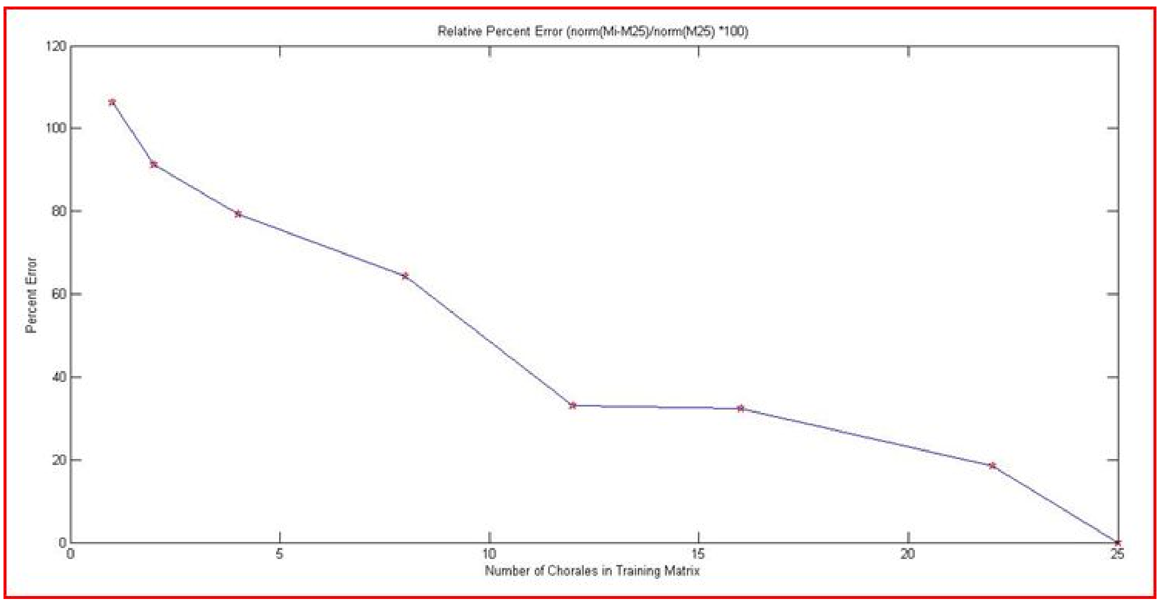


Figure 7.2: Relative Percent Error of Transition Matrix in Relation to the M_{25} Transition Matrix

Also included in the exploratory data is the box plot of the musical experience of participants in this experiment. What the box plot hopes to demonstrate is that there is no significant trends in the years of musical experience and what training set was analyzed in the experiment. The p-value of the ANOVA of experience was 0.75, and suggests that the means and variances are all closely related. This high p-value does not reject the original null hypothesis, but rather adds that each track had fairly similar musical experience among its experts analyzing it. The lack of a trend also helps quantify the qualitative fairness of the experiment.

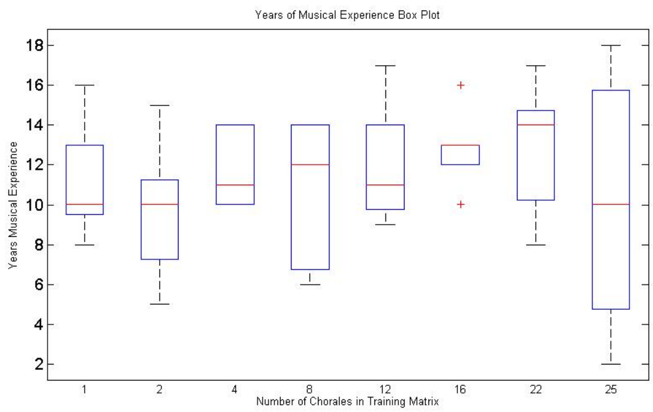


Figure 7.3: Box Plot of Experience, per level of training

7.6 Musical Analysis

The success as described by the statistics of the experiment is only one piece of the interdisciplinary puzzle of this project. Musically, it is difficult to analyze the Bach-qualities of the examples, since they lack a traditional melody and do not follow the typical rules of voice leading. However, we can look at the examples that tricked participants and see if they successfully adhered to the tonic-predominant-dominant-tonic model.

The tonic-predominant-dominant-tonic model describes the musical framework for music of Bach's and later generations. Typically, sections vary greatly in length, but usually the first tonic or the dominant sections are the longer sections. Tonic sections can have a series of different chords based on the first scale degree (and most of the others in some form) and functional harmonies around it. Predominant chords are mostly based off scale degrees 2, 4, and 6, with variations of inversion, key and function. The dominant group has the least amount of options for harmonies, with only chords based around scale degrees 5 and 7 being available.

Ideally, a choral would wander through many tonic-grouped chords, have an important predominant moment, followed by one of the two types of dominant chord before returning home to the tonic. Below is an example of the first 16 chords of a trial that was trained by 12 chorales.

I, I, $IV^{\binom{6}{5}}$, vii^{o}, I, V^{6}, I^{6}, I, I^{6}, V, I^{6}, ii, ii^{6}, V, I...

In this example, we can see the consistent fluxuation between I and V chords, followed by a predominant ii^{6} chord, a dominant V and back to I. This exerpt adheres to the TPDT models. However, the fluxuation is not something one would see in a Bach chorale. Bach would use different chords instead of going back and forth between the two most important chords in the key. In the Markov Model, I and V are commutative states with high probabilities. The stochastic system has the ability to fluxuate between these two chords because of their relatively easily ability to commute. This fluxation happens with other higher probability commutative pairs as well in other chorale trials.

There are over 30 examples of music created by the model, and looking at them all, most adhere to the TPDT model.[1] [10] In the examples used to create the model, the conditions of the matrix are that the exerpt has to end with a dominant-tonic cadence. The only examples that do not adhere to the TPDT framework are those

[1]See the experiment website to listen to examples, take the survey, or browse the raw data.

whose antepenultimate chord is a I chord (or another related tonic chord). This additional tonic chord sometimes makes the last three chords some form of I-V-I, which is not entirely unlikely in Bach. However, these three chords do not fully align to the TPDT structure if there is no predominant chord.

Despite the commutative concerns and the occassional departure from the TPDT framework, the majority of the examples closely model real Bach harmonic progressions that trained the model. This success, musically speaking, concludes that the model has the inherent property to follow Baroque harmonic guidelines with out pre-selecting all of the exceptions and restrictions.

Chapter 8

Application to Education

Throughout my studies at the Harvard Graduate School of Education[1] I have been exposed to progressive views of math education in the United States. These progressive facets of math education involve evolving the curriculum so it supports the mathematical knowledge students will need in their futures, not their present.

One of our classes in T-214, (the math methods of teaching course) discussed technology and its impact on math curriculum. The idea that once students learned how to compute square roots by hand, and that now we rely solely on calculators for this computation led me to think about what other concepts will be discontinued in mathematics curricula in the coming years and decades. Rudimentary concepts that seasoned teachers have viewed as critical to understanding may become obsolete. Personally, I think that it necessary to determine what is and is not important in the math education of our future students. I also think that reducing time spent on learning concepts that are not necessary to our future students success, we may have more time to focus on really interesting topics in mathematics that get brushed over.

A topic very important to me that I feel is not taught adequately in schools is matrices and introductory linear algebra. Based on my experiences as a high school

[1] This chapter and the unit plan it references were completed as part of final exam requirements for T-214: Teaching Mathematics at the HGSE. However, I believe that it is important to see a different application of Markov Chains to another field, especially my Applied Math field of application.

student and of many Harvard undergraduates, matrices are touched upon in school, but never delved into. I remember my first day in linear algebra my sophomore year of college the professor assumed we knew how to multiply matrices. I remember calling my mother, a high school math teacher, that night to ask when I was supposed to have learned matrix operations. Apparently, the Algebra 2 curriculum includes matrices, but it is one of the last chapters of the textbook and often times get skipped. I was pretty upset to hear that I had not had the opportunity to learn about matrices in high school, since it put me at a disadvantage in my linear algebra class (and frankly, I took Algebra 2 twice in separate years and different schools, so it seems to be more of a problem than just bad luck on my part).

What I hope my story introduces it the idea that end of the book topics are often not given the time they deserve in the classroom. My hope with this unit is to delve into two topics that are often overlooked: probability and matrices, and their intersection with a concept called Markov chains. Markov chains are generally taught to linear algebra students, and it may be surprising at first for people familiar with the Algebra 2 curriculum to see something so challenging. However, looking through sample curricula, I believe that if matrix algebra is taught well, and all the skills in the curricula are covered, that Markov chains are not a difficult stretch to make in terms of learning. I was a teaching fellow this semester for a more advanced linear algebra class at the School of Engineering and Applied Sciences, and we actually did not get enough time to really learn all there is to about Markov Chains. This unit will also not give them the proper study (as they could be a course on their own) but hopes to address many of the interesting facts about these Markov matrices.

The study of Markov Chains is becoming its own subfield of applied math. Think about this project, which applies Markov Chains to music. Markov Chains can be applied to almost any application area, and they are used heavily in economics, finance, sports, music, games, and even how Google ranks pages in a search. I think if students realize how applicable this concept is, they will get past the initial fear of a "challenging topic. The students will eventually do a problem based on my work, and the work of two of my students this semester (Amol Pai and Naor Brown) on Taylor Swift songs and Markov Chains.

The other unique aspect of this unit is I am not exactly sure in what type of class the unit could be taught in. During my UTEP pre-practicum, I observed Honors Geometry, and a college preparatory class called Functions, Statistics and Trigonometry. I do not believe this unit fits either curriculum. I then thought that this unit would be a great extension for AP Statistics. However, after talking with former AP Stats students and reading the curriculum online, I did not feel as though the unit could be in the already packed curriculum. Instead, the target class for this unit is an Algebra 2 class, preferably honors/accelerated/college-preparatory. I believe that the review of probability will be helpful as students prepare for the PSATs and SATs. Also, matrices should be introduced in algebra, and not swept under the rug as they sometimes do.

My goals for the unit are for students to really feel like they have learned something they can apply to their lives. After the tenth lesson, students will have seen just how Markov Chains work in a real problem that I have been studying for two years now. I also hope that students understand the basics of matrices. I will not be discussing the ideas of rotation and how matrices transform vectors, as I believe basic vector calculus is necessary to really understand that concept. However, I do believe that some students will enjoy using matrices to solve systems of linear equations, and even if they do not appreciate Markov Chains, they may like the formulaic approach to solving systems of equations.

Other goals I have in this unit are exploration and discovery. I hope that some of my exercises, like what transpose is, or what is the identity matrix, help the students discover the definitions and applications of these topics without me giving them notes.

Also, after the unit introduction, students will be responsible for taking their own notes. I did not have many opportunities to take my own notes in high school, and in my observations I see many teachers using guided notes. If this theoretical classroom is truly college-prep, note taking will become an important skill later in life. My content goals for the unit are tri-fold: that students understand basic probability and how to represent it, that students understand basic matrix operations, and finally that students can grasp the context of Markov chains and use it to solve a real-life problem.

The assumed prior knowledge for this unit is very straightforward. The students should know how to solve systems of equations with two or three equations. They should also have a baseline understanding of probability. They should be good at multiplying without a calculator, and with a calculator. My instructional emphasis will be on acquiring the new skills, while focusing on how applicable these skills are to many contexts. I also plan to use a computer lab three times during the unit.

Students will have one quiz during the unit and one test. Homework will be graded periodically to make sure that students are keeping up with the material. Matrices may not be important for every students future, but I still think they are very applicable and very useful. Some students prefer using the matrix methods to solve systems of equations. Students will have some difficulty with the matrices at the beginning. Multiplying matrices is very confusing at first, which is why I developed the cutting/lining up exercise so students can start to visualize how that process is completed. I also think students may have trouble really understanding what Markov Chains are, and how to solve for the steady state. I really hope that the way I have tied the unit together with the unit problem will combat this difficulty.

This project was my first exposure to unit planning, however, I have employed it many times since writing this plan. Though I may never teach this unit, I think this project has been a really useful exercise. I have been able to connect college linear algebra to high school algebra 2. Also, this unit was made for my schools schedule, which is daily 85-minute blocks for a semester. Though 10 days worth of lessons may not seem like a lot, it would have equated to four weeks worth of lessons in my high school, which has every-other-day block scheduling. I really hope that this gets people thinking about what is possible in classrooms, and that technology like WolframAlpha can be used in interesting ways in the classroom. Included in the appendix, available online, is this unit plan that would introduce students to linear algebra and Markov models. Hopefully, students someday will have the opportunity to study this type of mathematics in high school.[2]

[2]See appendix for lesson plans, assessments, homeworks, and Power Point slides for this theoretical unit.

Chapter 9

Conclusion

We presented a successful methodology in terms of creating examples that tricked musical experts. The Hidden Markov Model was studied in order to help gain understanding of the probabilitistic unknown that is Bach's harmonic language. Throughout the process of this project, there have been several different strategies employed to create the model. After successful baseline approaches, a method was created to help increase the amount of data involved in the project. Data from twenty five Bach chorales was used in this model, and this larger data set allowed for more gradations in the number of chorales training individual transition matrices. Then, an experiment was conducted to see if the model was successful. It was determined through statistical hypothesis testing that there is evidence to support the claim that the model represents a binomial distribution with a probability of correct reponse of 0.5 when there are greater than 10 chorales in the training set. This finding suggests that with 10 or more chorales, a person using this model can create Bach-like chorale progressions that can trick musical experts into believing the examples are real Bach.

9.1 Future Work

Due to the limits of any project, there are some intentional oversights that should be considered in future studies. In terms of the model, there should be an easier way to compile data. The transition matrices were all created by manually inputting the frequency of transitions. By automating this process, adding the rest of the chorale data would be much easier. Ending the study with 25 chorales in the sample set means that only about 7 percent of all available chorales were used. There is no way to determine if the omission of chorales has an effect on the overall outcomes.

Major revisions should be made to subsequent experiments of this nature. First off, it is extremely difficult to control for all the subjective variables. People analyzing the chorale trials should have experience with Bach specifically, not just general musical ability. Also, many participants complained about the organ sound for the musical output. In testing, the fake voice output was equally troublesome, so a synthetic piano sound might be a nice alternative. Beyond the logistics, there should be more data involved in the experiment. It was mutually decided that 5 participants per level of training was adequate, but there is much more certainty in the finding when there are five to ten times that amount of participants per track. There are countless difficulties in the quantification of a qualitative field like music, but more data would hopefully settle the large ranges in the box and ANOVA plots.

Also, a huge omission in the experiment was the lack of minor-chorale examples. Minor chorales behave much differently harmonically than major-mode chorales. Due to restraints of the project, it seemed more beneficial to go into more depth about one type of chorale as opposed to limited analysis of each type. However, it would be interesting to compare the results of minor mode chorales to major chorales using an almost identical process. Hopefully, the model would be as successful, but the variety of possible progressions in minor-mode chorales might make this outcome less likely to occur.

9.2 Take-Aways

A project like this one, with numerous components spanning multiple disciplines, can become unwielding in terms of determining its significance to any field. However, I believe that there are a few things that can be extracted as the important results of this thesis, and of this methodology. First off, there are multiple studies of hidden Markov chains in Bach chorales. Some examples, such as an undergraduate thesis by Christopher Thorpe '98, tend to look at chorale harmonization.[9] Most other papers and theses about Markov chains and Bach focus on harmonization as a means of utilizing the hidden Markov Chain. The approach in this paper hopes to do something different: to focus only on harmony and the creation of authentic-sounding harmonic progressions.

Subjectively, the quality of harmonizations seem to be more dependent on the rigor of the model as opposed to the complexity of the Markov analysis. In sections 4 and 5, we are able to see that Markov analysis of steady-state in Bach chorales, which is incredibly interesting and important to empirical musicologists. Well-known facts, such as the I and V chord being most important in major-mode pieces are reaffirmed with quantitative data and thorough analysis.

The intriguing part of this analysis was the creation of realistic-sounding progressions. Though not all participants were tricked equally, there is strong statistical evidence that supports the randomly generated Markov chains were representative of Bach idioms and harmonic progressions. Comments from participants ranged from "did I win" to "I had no clue what was Bach and wasn't." Though the process could be improved and there are numerous other analysis to be made before the method is deemed unfallable, it seems as though people were tricked, and that hidden Markov chains can describe Bach's harmony.

Appendix A

Code

A.1 MATLAB code

I wrote routines to create the progression using the transition matrices and a random number generator.

```
function x =coolio (A,b)


[m,n]=size (A);

s=size (b);

c=1;

d=1;

x=zeros (s,1);

for i=1:s

    for j=n:-1:1

        f=d-A (c,j);

        if f>b (i,1);

            d=f;

        else
```

```
            x (i,1)=j;
            c=j;
            d=1;
            f=1;
            break
        end
    end
end
```

A.2 Mathematica code

The following code was used to take the chords created by the previous routine (they were named by number) and turn them into chorale examples.

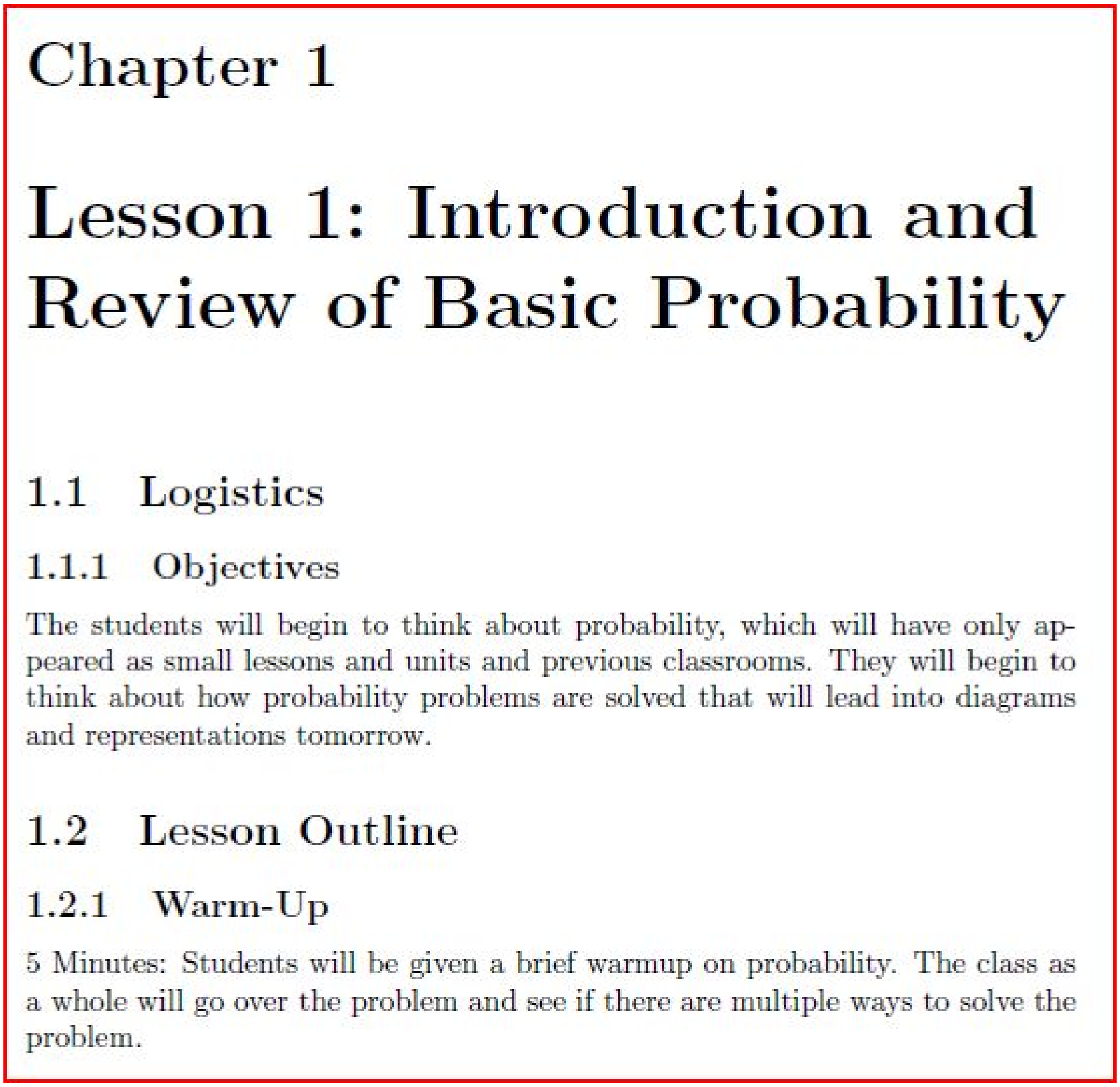
```
one = {-12, -5, 4, 12}; oneseven =
 one + {0, 0, 0, -2}; onesevenb = {-8, -2, 7, 12}; oneb = {-8, 0, 7,
  12}; onec = {-5, 0, 4, 7}; two =
 one + {2, 2, 1, 2}; twosevenb = {-7, 0, 9, 14}; twob =
 oneb + {1, 2, 2, 2}; twoseven = two + {0, 0, 0, -2}; three =
 one + {4, 4, 3, 4}; threeseven =
 three + {0, 0, 0, -2}; threeb = {-5, -1, 4, 11}; four =
 three + {1, 1, 2, 1}; fourseven =
 four + {0, 0, 0, -1}; foursevenb = {-3, 3, 5, 12}; fourb =
 foursevenb + {0, -3, 0, 0}; five = {-5, -1, 7, 14}; fiveoftwo = {-3,
  1, 4, 9}; fiveoffive = two + {0, 0, 1, 0}; fiveofsix =
 three + {0, 0, 1, 0}; fiveseven =
 five + {0, 0, -2, 0}; fivesevenoffive =
 fiveoffive + {0, 0, 0, -2}; fivesevenoffour = {-12, -2, 7,
```

```
16}; fivesevenb = {-13, -5, 5, 14}; fivesevenboftwo = {-11, -3, 7,
16}; fivesevenboffive = {-6, 0, 9, 14}; fivesevenc = {-10, -5, 5,
11}; fivesevencoffour = {-5, 0, 10, 16}; fivesevencoffive = {-3, 2,
6, 12};  fivesevend = {-7, -1, 7, 14}; fivesevendoffive = {0, 6, 9,
14}; fiveb = {-13, -5, 2, 7}; fiveboffive = {-6, -3, 2,
9}; fivebofsix = {-4, 4, 11, 16}; fivec = {-10, -1, 7,
14};  six = {-15, 0, 4, 12};  sixseven =
six + {0, 0, -2, 0}; sixsevenb = {-12, -3, 4, 9}; sixsevend = {-5, 0,
 4, 9};  vib = {-12, -3, 4, 9}; viib = {-10, -1, 5,
14}; viiboffive = {-3, 0, 6, 12}; viio = {-13, -7, 5,
11}; viiooffive = {-6, 0, 9, 12}; viioofsix = {-4, -1, 8,
14}; viioseven = {-13, -7, 5, 9}; viiosevenb = {-10, -3, 5,
11}; viiob = viib; viioboffive = viiboffive; viioD7ofsix = {-4, -1,
5, 14};
```

```
twenfive1 =
 Sound[SoundNote[#, 1, "Organ"] & /@ {one, fiveb, sixsevenb, twob,
    six, fivesevenboffive, fourseven, five, fourb, viioseven,
    sixseven, twoseven, fivesevend, six, five, fivebofsix, fourb,
    onec, fiveseven, one}]
```

A.3 Education

On the thesis website listed in the references, is the entire unit plan created. This unit plan involves probability and an introduction to Markov chains. It was not included because it is well over 70 pages, and is more useful to see as a PDF. However, I have included several images to get an idea of what a unit plan looks like.



Chapter 1

Lesson 1: Introduction and Review of Basic Probability

1.1 Logistics

1.1.1 Objectives

The students will begin to think about probability, which will have only appeared as small lessons and units and previous classrooms. They will begin to think about how probability problems are solved that will lead into diagrams and representations tomorrow.

1.2 Lesson Outline

1.2.1 Warm-Up

5 Minutes: Students will be given a brief warmup on probability. The class as a whole will go over the problem and see if there are multiple ways to solve the problem.

Figure A.1: Opening of a lesson plan [10]

Lesson 1: Intro to Unit, Probability Review

- Problem of the Unit. Initially, 10% of all people drink Pepsi and 90% of all people drink Coke. Pepsi creates a new ad-campaign which dramatically changes who drinks which beverage. Every week, 70% of the people who used to drink Coke switch to Pepsi, and only 20% of all people who were already drinking Pepsi switch to Coke. After one week, what percentage of people are drinking Pepsi?

Figure A.2: Unit Problem [10]

Cornell Notes	Topic/Objective: Probability Re-Fresher	Name:
		Class/Period:
		Date:
Essential Question: What are the basic concepts employed by probability?		
Questions:	Notes:	
What is probability?	Probability is a measure of the expectation that an event will occur or a statement is true. Probabilities are given a value between 0 (will not occur) and 1 (will occur). The higher the probability of an event, the more certain we are that the event will occur.	
What is a sample space?	A sample space of an experiment or random trial is the set of all possible outcomes. For example, if the experiment is tossing a coin, the sample space is the set {head, tail}.	

Figure A.3: Sample Notes [10]

Name:________________

Worksheet on Unit Problem

- Initially, 10% of all people drink Pepsi and 90% of all people drink Coke. Pepsi creates a new ad-campaign which dramatically changes who drinks which beverage. Every week, 70% of the people who used to drink Coke switch to Pepsi, and only 20% of all people who were already drinking Pepsi switch to Coke. After one week, what percentage of people are drinking Pepsi?

Guiding Questions

1. What is the initial composition of Pepsi and Coke drinkers?

2. What is the probability that a person switches from Pepsi to Coke in week 1?

Figure A.4: Sample Worksheet [10]

Name: ____________________

Homework Lesson 2: Representations and Tree Diagrams

1. Tina's favorite meal is pasta, followed by ice cream for dessert.
   Tina's Mom cooks pasta once a week.
   If she cooks pasta, then the probability Tina gets ice cream for dessert is $\frac{2}{3}$
   If she doesn't cook pasta, then the probability Tina gets ice cream for dessert is $\frac{1}{4}$
   What is the probability that Tina gets ice cream for dessert?
2. Teddy has a two pairs of black shoes and three pairs of brown shoes. He also has three pairs of red socks, four pairs of brown socks and six pairs of black socks.
   If Teddy chooses a pair of shoes at random and a pair of socks at random, what is the probability that he chooses shoes and socks of the same color?
   The probability of a fine day is $\frac{3}{7}$ and the probability of a wet day is $\frac{4}{7}$

Figure A.5: Sample Homework Question [10]

Name:________________________________ 100 points

Unit Test : Probability, Matrices, and Markov Chains

Directions: You will have 80 minutes to complete this test. You may use calculators. The test will not be curved. If you get below a 75, you may re-take the test for credit up to a 75.

Question 1: 15 Points (a is worth 5 and b is worth 10).

a. A jar contains 10 red, 3 green, 2 purple and 5 yellow marbles. A marble is chosen at random from the jar. After replacing it, a second marble is chosen. What is the probability of choosing a purple and then a red marble?

Figure A.6: Sample Test Question [10]

Bibliography

[1] A. Agresti and C. Franklin. *Statistics: The Art and Science of Learning from Data*. Pearson, 2011.

[2] Bach, J.S. Brich an... Chorale. http://conquest.imslp.info/files/imglnks/usimg/9/9e/IMSLP139858-WIMA.50fb-2202.pdf, March 2013.

[3] Couglin, K. Skyleap Music: Fundamentals of Music. http://www.skyleapmusic.com/MusicFundamentals/Harmony/diatonic-chords-major-harmonic-system.html, March 2013.

[4] M. DeGroot and M. Schervish. *Probability and Statistics, Fourth Edition*. Pearson, 2011.

[5] Lotus Music. Parallel Minor Scale. http://www.lotusmusic.com/lm_parallelminor.html, 2009.

[6] MusicDish. Open Position and Voicing of Chords. http://www.musicdish.com/mag/?id=4100, March 2013.

[7] L. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceeding of the IEEE*, pages 257–286, February 1989.

[8] Sapp, C. KernScores. http://kern.ccarh.org/cgi-bin/ksdata?location=ccarh/class/2008/craig/bachchorale&file=chor001.krn&format=info, March 1999.

[9] Christopher Thorpe. CPU Bach. Master's thesis, Harvard, March 1998. http://www.eecs.harvard.edu/~cat/papers/Thesis-CPU-Bach.pdf.

[10] Tracy, M. Bach in Beta. http://people.seas.harvard.edu/~mtracy/, March 2013.

[11] Wikipedia. Interval (music). http://en.wikipedia.org/wiki/Interval_(music), March 2013.