THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ENGINEERING SCIENCE AND MECHANICS
AND SCHOOL OF MUSIC

MODELING INTONATION IN NON-WESTERN MUSICAL CULTURES

SAMUEL LAPP
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Engineering Science
with interdisciplinary honors in Engineering Science and Music

Reviewed and approved* by the following:

Matthew B. Parkinson
Associate Professor of Engineering Design and Mechanical Engineering
Thesis Supervisor

Mark Ballora
Associate Professor of Music Technology
Honors Adviser

Gary Gray
Associate Professor of Engineering Science and Mechanics
Honors Adviser

Judith A. Todd
Department Head
P. B. Breneman Chair and Professor of Engineering Science and Mechanics

*Signatures are on file in the Schreyer Honors College and Engineering Science and Mechanics Office.

Abstract

Musical cultures around the world use tuning systems very different from the tuning system used in Western music. Comparing and contrasting intonation from various cultures could give insight into the question of learned versus innate preferences in the perception of music, and into the cultural development of musical traditions. However, methods for analyzing intonation in Western music have not been generalized to non-Western contexts. Flexible and non-discrete tuning used in cultures such as Arabic maqam music pose a challenge for analyzing intonation. This paper attempts to generalize Stolzenburg's "Periodicity" method and Gill and Purves' "Similarity" method for microtonal music, and finds that they are ill-suited for the analysis of Arabic music. An alternative approach is developed, which analyzes the relative intonation of a series of consecutive notes. This approach reveals that Arabic intonation is fundamentally different from Western and Chinese intonation. While Chinese and Western intonation are based on intervals from a fixed scale, Arabic intonation is based on a continuous spectrum of small intervals.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are many people to thank for their generous support throughout this research. First, my thesis Supervisor Matt Parkinson for endless guidance and support, and believing in my crazy ideas. Thank you to my honors advisers, Gary Gray and Mark Ballora; scholars in the field who gave their input and thoughts, including Frieder Stolzenburg, David Huron, and Keith Mashinter; the Penn State professors and faculty who supported me, including Amanda Maple, Curtis Craig, and Paul Barsom; members of the Arabic music community, including Simon Shaheen, Neil van der Linden, Sami abu Shumays and Laila Mokhiber; friends who gave new perspectives and insight, including Joe Cosgrove, Matt Pennock and Small-Hand Luke; Zena for revising my writing; and everyone else who helped guide, advise, and revise my work.

Introduction

Throughout history, music and mathematics have been fundamentally linked. In the sixth century BC, Pythagoras was fascinated by whole number ratios, which described many physical phenomena. Pythagoras also found that simple ratios of frequencies sound pleasant together, and built a musical scale from simple ratios of 2:3 [1]. A fascination with mathematical relationships for musical tuning developed in many parts of the world. In Ancient China, standardized sets of brass bells were used to normalize tuning. The bells were made with such precision that they were also used for standard measurements of volume and weight [2].

Throughout modern history, the relationship between mathematics and music continued to fascinate scientists and musicians alike. Advances in the understanding of acoustical phenomena, the human auditory system, and psychoacoustic processes gave new insights into the connections between math and music. As a result, the histories of mathematics and music are closely linked and interdependent. For example, the adoption of equal temperament (a tuning system discussed in Chapter 1) was intertwined with the development of the logarithm [3].

In the last hundred and fifty years, scientists have delved deeper into the questions of why humans perceive certain combinations of frequencies in distinct ways. Pythagoras's ideal of simple ratio relationships between notes is insufficient for explaining the complex sets of tones used to create harmony and melody in Western music. Many studies have attempted to explain the mathematical relationships between the notes in Western musical traditions [4, 5, 6, 7, 8]. In some cases, these studies have produced successful models for describing the notes that are frequently used together. On the other hand, this research will demonstrate that these models generally fail to apply to non-Western musical cultures such as Arabic maqam music (introduced in Section 1.3).

Comparison of intonation across cultures provides a way to test the universality of theories on music perception. These theories have primarily been tested for Western music, and may or may not apply beyond Western music. In order to evaluate the importance of learned cultural norms versus innate preference for music perception, it is necessary to compare intonation systems across cultures.

Also, studying the differences between intonation systems can provide insight into the ways musical cultures develop, and are passed down. For example, Arabic music is traditionally taught aurally, in contrast to the Western practice of learning through notated music [9]. The vastly different styles of passing down music may lead to differences in the two cultures' intonation.

Finally, studying the intonation of other cultures gives insight into the potential implications of mapping a Western system of measurement onto artifacts of other cultures. In the case of Arabic music, Western notation and note names fail to capture the details of Arabic intonation.

This research begins with an investigation of contemporary computational models for predicting sets of tones that will be used together to make music. Two recent models, Stolzenburg's Periodicity model [4] and Gill and Purves' Similarity model [8], claim that their methods apply beyond Western music to cultures with other tuning systems. These methods are recreated and tested for Arabic music. Neither method adequately captures details in intonation that are essential to Arabic music performance.

To more accurately capture the features of intonation in Arabic music, this research develops a new method of analysis which retains continuous and contextual relationships between pitches. This method, described in Part Two, does not require intonation to adhere to a discrete set of pitch ratios. The method analyzes the frequency relationships of a series of consecutive notes. Instead of sorting frequencies into discrete note-bins, the method finds the decimal ratios between a series of frequencies extracted from the music. This ensures that resolution is high and context is not ignored. A sequence of two or more frequency ratios creates a musical "word". The method compiles the prevalence of all words used in a musical selection.

As a result, completely different patterns emerge for intonation in Western and Arabic music. Without this new approach for analyzing intonation, Arabic music performances would be forced into a measurement system that makes their intonation appear similar to Western music. By avoiding discretization of pitch relationships (such as major second, or minor third), the huge differences between intonation in Arabic and Western music become apparent. This form of analysis also reveals that even Western music is not tuned in a perfectly discretized way. Variation from "perfect" intonation adds to the emotional character of the music, and is an important tool for Western performers [10].

This research bring into question the validity of mapping foreign cultures' artifacts onto a familiar measurement system. While discrete measurement systems with limited resolution may be sufficient for one scenario, a higher resolution or continuous system of measurement may be necessary in other situations.

Part I of this document provides a background in music theory and the models for intonation that will be discussed, tested and reworked throughout the paper. Chapter 1 provides a brief introduction to musical theory and intonation for Western, Arabic, and Chinese cultures. Chapter 2 discusses the development of models for consonance and dissonance in Western music. Two preexisting methods for modeling intonation of scales are discussed in Chapter 3. Part II describes the development of a new model for intonation which is better suited for analysis of non-Western musical cultures. This includes an analysis of the two existing methods and a discussion of the issues arising from their approach to intonation. Finally, Part III details the results of the new method developed in this paper, and analyzes the implications for music analysis and modeling in general.

A set of audio clips is provided to accompany this paper. The Listening Tracks are referenced throughout text and listed in Appendix D.

Part I

Background

Chapter 1

Intonation in Western, Arabic, and Chinese Cultures

Sound is the brain's interpretation of periodic fluctuations of pressure in the world. Humans can hear acoustic vibrations with frequencies from around 20 to 20,000 Hertz (Hz, or cycles per second). The frequency of oscillations is perceived as pitch on a logarithmic scale. (Appendix A gives a more complete introduction to auditory perception.) The lowest frequency produced by a specific event is usually the strongest, and is called the "fundamental frequency. Musical notes are named by their fundamental frequency." For instance, the note 'A4' on a piano has a fundamental frequency of 440 Hz. Notes with different pitches (frequencies) are used simultaneously (harmony) and in sequence (melody) to create music.

"Intonation" refers to the relationships of the fundamental frequencies (referred to simply as "frequencies") of notes used in music. In general, the relationship between two notes is described by the ratio of their frequencies, which is called an "interval." Because humans perceive pitch on a logarithmic scale, the ratio (rather than the difference) of two frequencies describes the distance between pitches [11]. For example, the ratio 2:1 is the interval of one octave. The note 'A' with frequency 440 Hz and 'A' with frequency 220 Hz are an interval of one octave apart, because the ratio of the frequencies is 2:1. The notes 'A' at 880 Hz and 'A' at 440 Hz are also one octave apart with frequency ratio 2:1. Another example of a common interval is a fifth, which has a ratio of 3:2. The ratio from 'C' up to 'G' is a fifth (Listening Track 01). Table 1.1 shows the intervals used in Western music with their names and sizes.

1.1 Three Perspectives of Intonation

There are three perspectives for analyzing intonation in music. (1) Vertical analysis focuses on harmonies created by multiple notes sounding at the same time. (2) Horizontal analysis examines melodies, the lines created by the pitch changing in time. (3) Modal analysis describes the scale or mode, the set of frequencies that form the bases of an intonation system. All three forms of analysis are discussed in this paper. In general, the music of many non-Western musical cultures is primarily melodic (horizontal), while Western music relies heavily on harmony (vertical) as well as melody. Western studies often focus on vertical analysis [e.g. 12, 13, 5]. However, the lack of harmony in Arabic and Chinese cultures

Table 1.1: The intervals used in Western music are shown with their frequency ratios in twelve-tone equal temperament (12TET). The last column shows the approximate whole number ratio of frequencies with a maximum error of 1% from the exact decimal value.

Semitone	Name	Symbol	Example	12TET Ratio	Fraction
0	Perfect Unison	P1	C — C	1.000	1/1
1	Minor Second	m2	C — Db	1.059	16/15
2	Major Second	M2	C — D	1.122	9/8
3	Minor Third	m3	C — Eb	1.189	6/5
4	Major Third	M3	C — E	1.260	5/4
5	Perfect Fourth	P4	C — F	1.335	4/3
6	Tritone	T	C — F#	1.414	17/12
7	Perfect Fifth	P5	C — G	1.498	3/2
8	Minor Sixth	m6	C — Ab	1.587	8/5
9	Major Sixth	M6	C — A	1.682	5/3
10	Minor Seventh	m7	C — Bb	1.782	16/9
11	Major Seventh	M7	C — B	1.888	15/8
12	Perfect Octave	P8	C — C	2.000	2/1

(discussed below) prohibit analysis of these cultures in the vertical dimension. In order to apply generally to cultures that may not use harmony, the method of analysis developed in this paper uses the horizontal perspective.

1.2 Western Music

This section briefly describes the construction of scales and chords in Western music theory. In this paper, Western music refers broadly to the classical musical traditions in Western culture from the Renaissance (thirteenth century) to the present. While this period contains many distinct musical periods and styles, some broad generalizations can characterize most (though not all) of this music.

1.2.1 scales

Western music theory is based on two diatonic scales, called “major” and “minor.” Both are seven note scales spanning one octave. The notes of all scales are taken from the twelve-note chromatic scale (all twelve notes in an octave on a piano). The seven note scale is built from two tetrachords of four notes each (Figure 1.1). The first note of the scale is the called the “tonic” and the first note of the second tetrachord is called the “dominant.”

A set of seven consecutive white keys on a piano can form seven distinct scales, called modes, depending on the starting note. Figure 1.2 shows all seven modes in the order of sharps and flats. Of these, the major and minor scale are the most commonly used. Figure 1.1 shows the major scale on the piano (Listening Track 02). The white keys from ‘C’ to ‘C’ an octave higher on a piano form a major scale. In the key of C major, ‘C’ is the tonic and ‘G’ is the dominant. The white notes from ‘A’ to ‘A’ form a minor scale. In A minor, A is

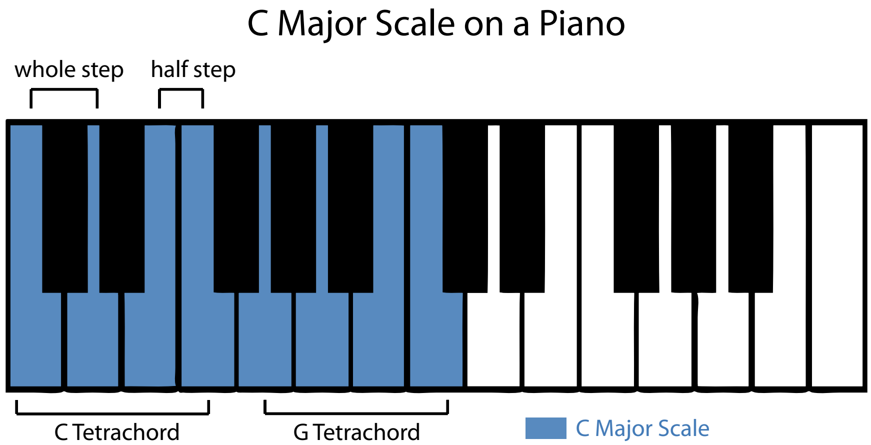


Figure 1.1: The C major scale uses the white notes on a piano, colored blue here. It is built from two tetrachords, {C D E F} and {G A B C}.

the tonic and E is the dominant. The major scale generally has a happy, euphonious, and bright sound, while the minor scale has a sad, dark and angry sound.

The most pleasant-sounding intervals have frequencies that are simple integer ratios, like 3:2 and 2:1 [14]. However, tuning strictly to perfect intervals (which is called “just intonation”) means an instrument will sound in tune in one key, but out out of tune in other keys. A hallmark of Western music is frequent modulation, or changes of key [15]. In order to be able to play keyboard instruments in all twelve keys, the twelve notes of the octave are tuned at equal intervals (called “equal temperament”). In equal temperament, none of the ratios between notes form perfect intervals like 3/2, but all keys are close enough to just intonation to sound in tune. Twelve tone equal temperament (12TET) has been common practice in Western music since the Baroque period [16]. Twelve tone equal temperament means that the pitches are tuned at equal distances one twelfth the size of an octave, on a logarithmic scale. Thus, the ratio of each note to its neighbor is always $2^{1/12}$. This means that the distance from C to D (skips over C♯), called a whole step, is twice the distance from E to F, called a half step (See Figure 1.1). Table 1.1 shows the intervals used in Western music, with their frequency ratios in twelve-tone equal temperament.

1.2.2 chords and harmony

Harmony refers to the sonority created by multiple notes sounded together. Any set of notes can be considered a harmony. The lowest note is called the root, or tonic. Chords are sets of notes separated by specific intervals that form a specific harmony when played together. For example, the major chord is formed by starting with the lowest frequency called the root, adding a tone an interval of a major third above the root, then adding a tone an interval of a fifth above the root.

Western music from the Baroque period (17th century) through at least the Romantic period (19th century) is based on triadic harmony. Triads are chords with three notes separated by intervals of a third. The major triad (spelled, for instance, C E G) is the

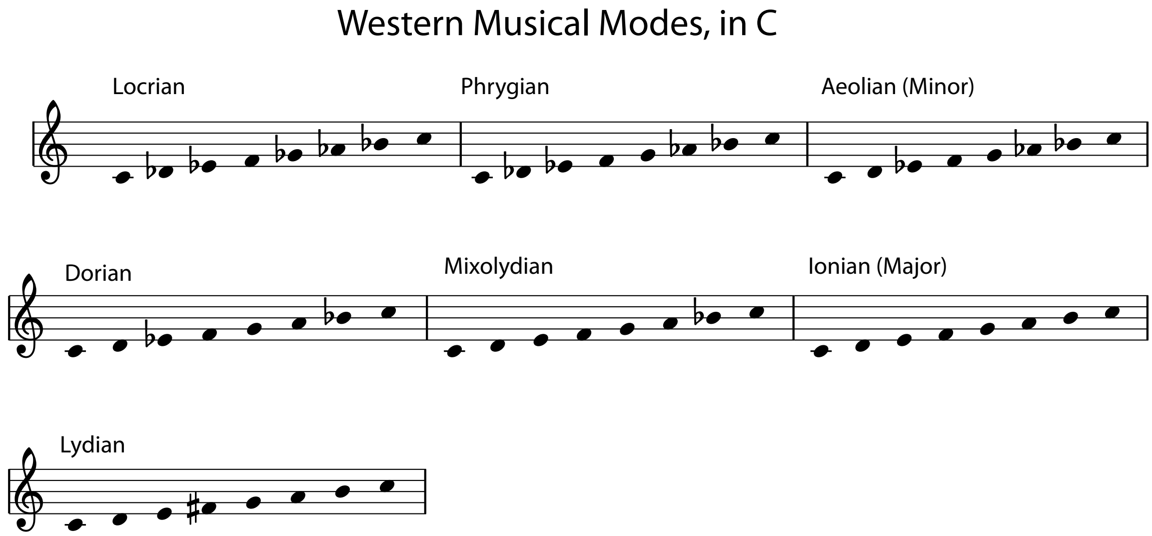


Figure 1.2: Western music uses seven diatonic modes, shown here in C, in order of increasing sharps.

most consonant, pleasant and euphonious. The minor triad (C E♭ G) is the second most consonant triad [15, p.64]. The concepts of consonance and dissonance are discussed in Chapter 2. Harmony is the foundational basis of Western music theory, and the major and minor chords account for a remarkably high portion of the harmony used in Western music.

1.3 Arabic Music

Arabic music is based on the melody form called the maqam (plural: maqamat, pronounced: mah-KAHM) and the the rhythmic meter called iqa. Listening Tracks 11-14 provide selections of Arabic music recordings.

A maqam, like a scale, describes the pitches used in a piece of music. However, the maqam contains much more information than just a scale. Maqamworld.com defines a maqam as “a set of notes with traditions that define relationships between them, habitual patterns, and their melodic development.” [17] There are dozens of maqamat, each with its own character, rules for melodic development, and tonic and dominant notes. The three most common maqamat are Rast, Sikah and Bayati. (Listening Tracks 06, 07 and 08 provide recordings of these three maqamat.) Their approximate pitch values in Western notation are shown in Figure 1.3.

Unlike Western music, Arabic music rarely uses harmony [18]. Shifts in character occur through the melodic development of the maqam, and modulation between maqamat [18]. The lack of harmony makes Arabic music fundamentally different from Western music, and means that the vertical perspective of intonation is rarely useful in Arabic music.

Like Western scales, maqamat usually contain seven notes spanning an octave, built from two tetrachords called “ajnas” (singular: jins). However, ajnas sometimes have three or five notes, instead of four. Figure 1.4 shows how maqam Bayati is built from jins Bayati

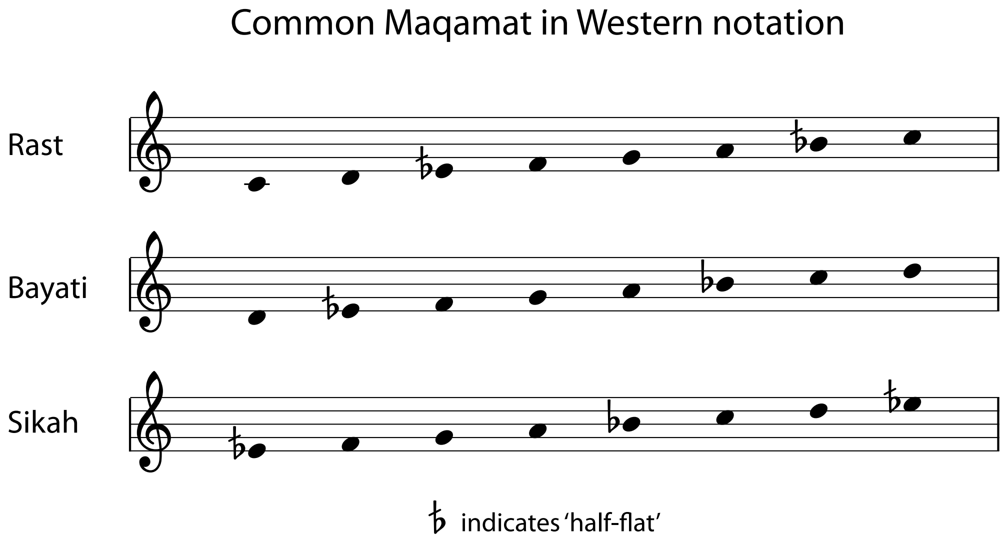


Figure 1.3: The three most common maqamat are Rast, Sikah and Bayati. Their approximate pitches in Western notation are shown here. A "half flat" lies between the natural and flat notes.

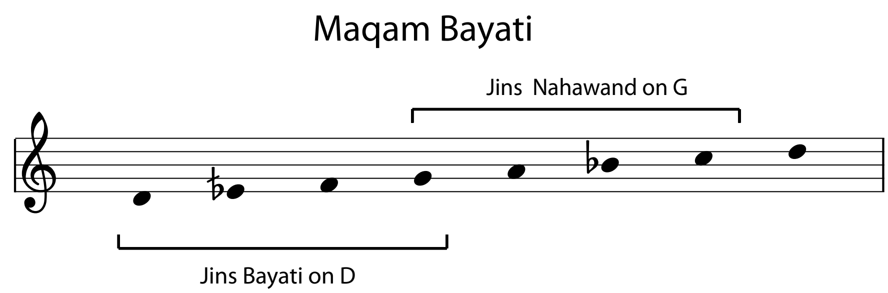


Figure 1.4: Like Western scales, Arabic maqamat are built from two smaller sets of (usually four) tones called jins. Maqam Bayati is built from two tetrachords, called jins Bayati and jins Nahawand.

tetrachord and jins Nahawand tetrachord.

The tuning of Arabic maqamat is not based on an equal temperament system. Each maqam has a unique interval structure, with intervals approximately the size of Western half steps, whole steps, and minor thirds. However, intervals in between the notes in the chromatic scale can be used. In order to approximate these tones with Western notation and names, these notes are often called "quarter tones" and are named with "half-flats" denoted with the symbol 𝄳 and "half-sharps" denoted with 𝄲 . For example, the second note of Bayati is "E half flat" (Figure 1.4). However, the quarter tone is not simply half way between its neighboring tones. The tuning of quarter tones and of maqamat in general is not equal tempered, and is learned by ear [17]. For instance, the E half flat occurring in Rast is tuned higher than the E half flat in Bayati (Rast and Bayati are shown in Figure 1.3) [9]. The tuning and variance of maqamat are discussed in detail in Chapter 6.

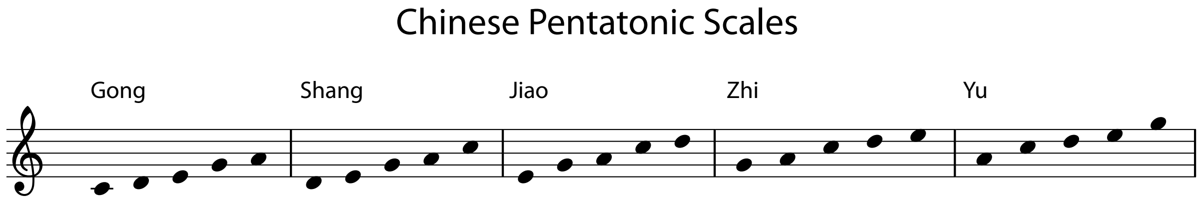


Figure 1.5: The five inversions of the pentatonic scale are shown on a Western staff, with their Chinese names.

1.3.1 instruments used in Arabic music

The most traditional Middle Eastern music ensemble is called the takht. The takht includes four melodic instruments and one percussive instrument. The melodic instruments are oud (a stringed instrument resembling a lute), nay (end-blown flute), qanun (plucked string instrument resembling a zither) and violin. Melodic instruments are classified as sahb (pulling, like violin and nay) or naqr (plucked, like oud and qanun). The percussion instrument can be riq (tambourine) or tabla (hand drum) [17].

1.4 Chinese Music

This paper primarily analyses Western and Arabic music, but also investigates the intonation of Chinese music in order to explore another non-Western system of intonation. Chinese music is based primarily on pentatonic (five note) scales similar to the Western major pentatonic scale (C D E G A). As with the western modes, variations of the pentatonic scale can be created by permuting it, or starting on a different note. Figure 1.5 shows the five permutations of the pentatonic scale used in Chinese music. Over time, Chinese music eventually also developed seven notes scales with "passing tones" filling in the gaps of the pentatonic scale (eg, C D E F$\sharp$ G A B) [19, p.165-167]. On rare occasions, a scale could even contain nine notes with an additional passing tone between two notes a half step apart (C D E E_2 F G A B$\flat$) [19, p.165]. In general, however, the pentatonic scale is the basis of Chinese intonation.

As in Western music, Chinese music uses twelve distinct, approximately equally spaced tones in the octave. However, Chinese music does not use equal temperament. Instead, the scale is built on perfect fifth intervals [19, p.175]. For example, the fifth is $3/2$, the major second is $(3/2)^2/2 = 9/8$, the major sixth is $(3/2)^3/2 = 27/16$ and so on. Because the scale is not equal-tempered, transposition (starting the same scale on a different note) changes the sound of the scale. Figure 1.6 compares the twelve tone tuning systems of Western equal temperament and Chinese temperament.

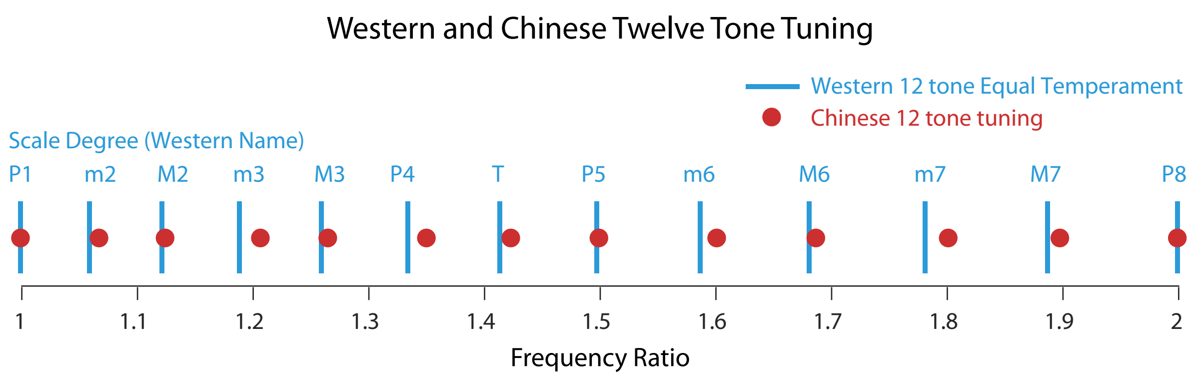


Figure 1.6: The twelve tone tuning system in Chinese music is based on perfect fifth intervals (3:2), and differs from Western twelve tone equal temperament.

Chapter 2

Harmony: Studies of Consonance and Dissonance

The concepts of consonance and dissonance are central to Western classical music practice and theory. While their distinct sounds are difficult to pin down in words, untrained listeners can easily distinguish consonance from dissonance. The music theory text *Music: An Appreciation* gives the following definitions:

> An unstable tone combination is a dissonance; its tension demands an onward motion to a stable chord. Thus dissonant chords are "active"; traditionally they have been considered harsh and have expressed pain, grief, and conflict.
>
> A stable tone combination is a consonance; consonances are points of arrival, rest, and resolution. [20, p.41]

Consonance is sometimes described as "pleasantness" or "euphony." However, pleasantness may be a separate quality because dissonant harmonies can also be perceived as pleasant [21].

The perception of consonance and dissonance involves both learned and innate components [22]. The term "Sensory Dissonance" refers to dissonance arising from the physiological and psychoacoustical processes of hearing. (Appendix A explains the basic physiological process of hearing.) This dissonance, rather that culturally learned perceptions of dissonance, is most likely to be perceived universally.

Various treatises have set rules for composition of polyphonic music, called counterpoint. Unlike the fuzzy perceptual boundaries, these treatises strictly define which intervals should be considered consonances and dissonances. However, the definition of consonance and dissonance have evolved over time. Table 2.1 shows how intervals are classified in Johann Fux's sixteenth century treatise *The Study of Counterpoint*. This was the text which the Western composers Bach, Mozart, Hayden and Beethoven would refer to for the study of counterpoint [23]. Only octaves, fifths, thirds, and sixths were considered consonant intervals. Fourths, seconds, and sevenths were diatonic dissonances that could be used if resolved properly. Tritones and other chromatic intervals were prohibited [23]. Throughout history, composers have expanded their harmonic vocabulary and used increasingly more dissonance. This eventually led to an extremely free use of all intervals in the twentieth century, which Arnold Schoenberg called the "emancipation of dissonance" [24]. This research is primarily

Table 2.1: Fux's treatise on counterpoint [23] defines consonant and dissonant intervals in four categories. In strict sixteenth century counterpoint, diatonic dissonances can be used if treated properly, but chromatic dissonances are forbidden. Refer to Table 1.1 for definitions of the intervals.

Consonances		Dissonances	
Perfect	Imperfect	Diatonic	Chromatic
Perfect Unison	Minor Third	Perfect Fourth	Tritone
Perfect Fifth	Major Third	Minor Second	Augmented Intervals
Perfect Octave	Minor Sixth	Major Second	Diminished Intervals
	Major Sixth	Minor Seventh	
		Major Seventh	

concerned with the perception of intervals and the quality of their sound, rather than their definition according to music theory.

2.1 Computational Methods for Harmony Analysis

One key element for comparing the intonation and tuning systems of various musical cultures is the choice of notes used together to make chords and scales, harmonies and melodies. In Western music, the concepts of consonance and dissonance are central to understanding what notes are used together in different contexts. Scientists and academics throughout history have attempted to describe subjective perceptual phenomena with objective mathematical models and psychoacoustic theories.

2.2 History of Consonance and Dissonance Analysis in Western Music

In Western music, the study of how musical tones sound together apparently originated with Pythagoras in the fifth century AD [25]. Pythagoras stated that if an interval (a ratio of the frequencies of two notes) can be expressed as a whole-number ratio, the combined sound is more appealing to the ear. Ratios with the smallest whole numbers (for instance, 2:1 or 3:2), have the most appealing sound [22]. Pythagoras's theory generally aligns with perceived pleasantness of intervals, but does not explain the reason for this phenomenon.

In the nineteenth century, the German physicist Hermann von Helmholtz proposed that sensory dissonance is a result of the beating that occurs when two tones of slightly different frequencies are superimposed - a phenomenon he called roughness. Figure 2.1 illustrates how beating occurs when two sine waves are summed. The frequency of the beating is the difference of the frequencies of the summed tones. In this case, the frequencies of 200 and 210 Hz create beating at 10 Hz frequency. If beating is very slow (on the order of Hertz), it is perceived as oscillating loudness. When beating is rapid (about 20 to 100 Hz), it is perceived as roughness.

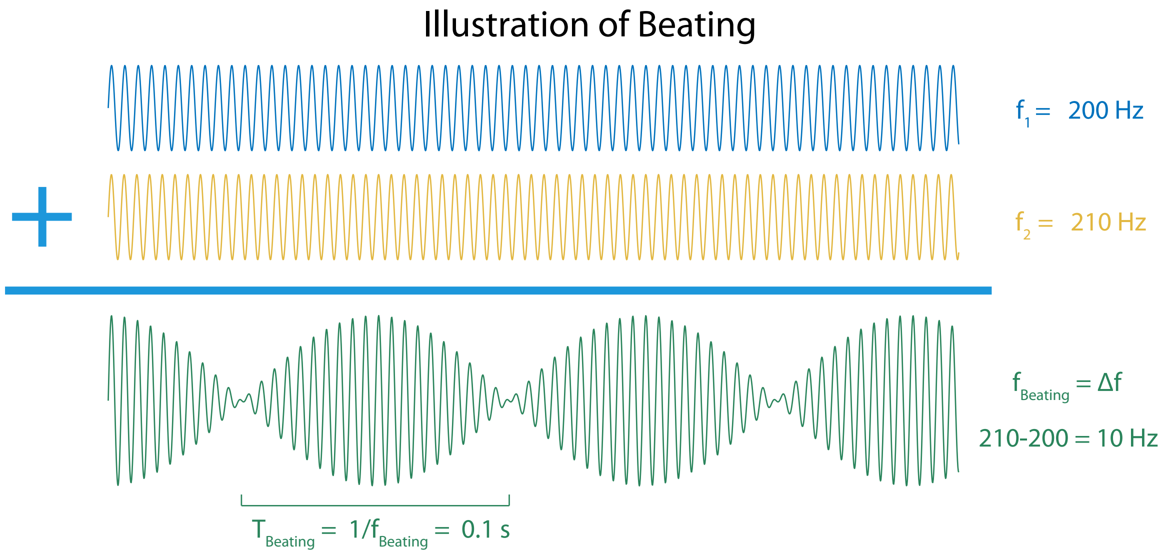


Figure 2.1: When two sinusoids with slightly different frequencies are superimposed, beating occurs at a frequency the difference between the two original frequencies.

Roughness arises from not only the fundamental frequencies, but also the upper partials of complex tones [12]. Helmholtz connected the idea of roughness with the physiological mechanism of hearing in the cochlea. Two pitches that stimulate the cochlea within a small region, called the critical band, cause an overlapping stimulation that leads to perceived dissonance [26].

Many more recent models are based on the fundamental idea of Helmholtz's roughness [26, 7, 6]. Later in the nineteenth century, Carl Stumpf claimed that tonal fusion (the tendency for two notes to sound like one) leads to perceived consonance [14]. However, later research showed that tonal fusion is a separate phenomenon from sensory dissonance [14, 27]. Instead, more recent research by Plomp and Levelt [26], Hutchinson and Knopoff [6] and others builds off of Helmholtz's idea that stimulation of the cochlea at frequencies separated by less than the critical bandwidth creates sensory dissonance.

2.3 Perception of Intervals of Simple and Complex Tones

2.3.1 simple and complex tones

A single sinusoidal pressure wave creates a "simple tone" with a pitch determined by the frequency of the wave (Listening Track 04). In the real world sounds are nearly always "complex tones." Complex tones are comprised of a fundamental frequency (the lowest and usually strongest frequency), and other frequencies at integer multiples of the fundamental frequency (Listening Track 05). These upper frequencies are called harmonics. The set of frequencies including the fundamental frequency and all integer multiples (harmonics) is called the harmonic series. The terms "partials" or "overtones" refer to frequencies present in the tone that may or may not be integer multiples of the fundamental frequency [28]. Figure

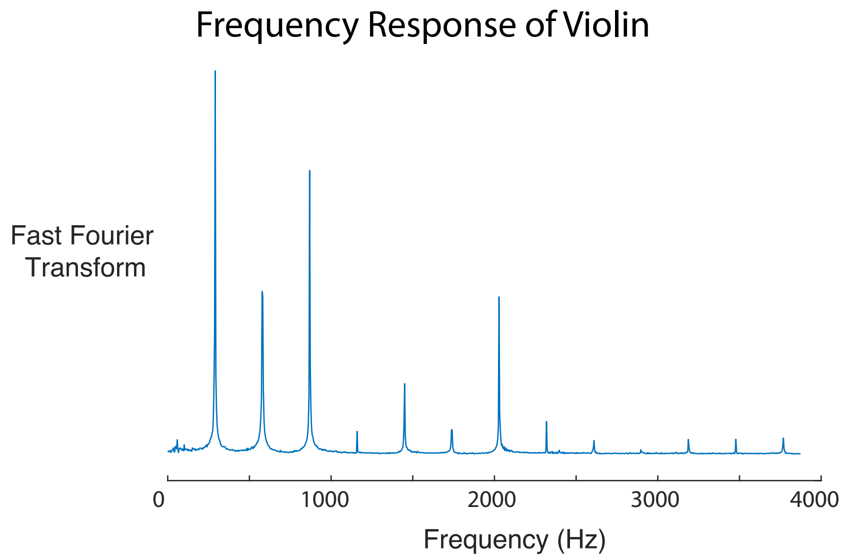


Figure 2.2: Peaks in the frequency response occur at integer multiples of the fundamental frequency. The relative strengths of these frequencies, called harmonics, help to define the sound of the violin.

2.2 shows the frequency-domain plot of a solo violin playing a ‘D’ at 291 Hz. The lowest frequency peak is the fundamental frequency. The other peaks are harmonics, and occur at integer multiples of the fundamental frequency. The relative strengths of the harmonics help to define the violin’s tone quality, called “timbre.”

2.3.2 dissonance of dyads

The study of dissonance and consonance starts with the most basic harmony possible, two notes played together. A harmony of two notes is called a dyad, and the distance between them is an interval. In the twentieth century, various studies explored empirical rankings of the consonance and dissonance of dyads. The distinction between simple and complex tones is critical for these experiments. An investigation by Plomp and Levelt [26] showed that perceived dissonance of dyads of simple tones follows a smooth curve, showing that people do not prefer simple ratios of pure tones. For complex tones, simple ratios are preferred because they minimize the beating (called roughness by Helmholtz) of interacting harmonics. Building off of this, Hutchinson and Knopoff [6] empirically derived a curve for critical bandwidth. The critical bandwidth describes the frequency band in the cochlea where roughness will occur between two tones. According to the study, the maximum dissonance occurs when two tones are separated by a distance 25% of the size of the critical bandwidth.

Building off this empirical research, Kameoka and Kuriyagawa [7, 29] showed that the perceived dissonance of two pure sine tones follows a v-shaped curve as the distance between the tones increases logarithmically (see Figure 2.3). This means that the perceived dissonance of an interval of two sine tones begins at zero when they are the same frequency, increases to a maximum dissonance, then decreases back to zero at the interval of an octave. When the experimental rankings for dissonance are plotted with the frequency ratio on a logarithmic

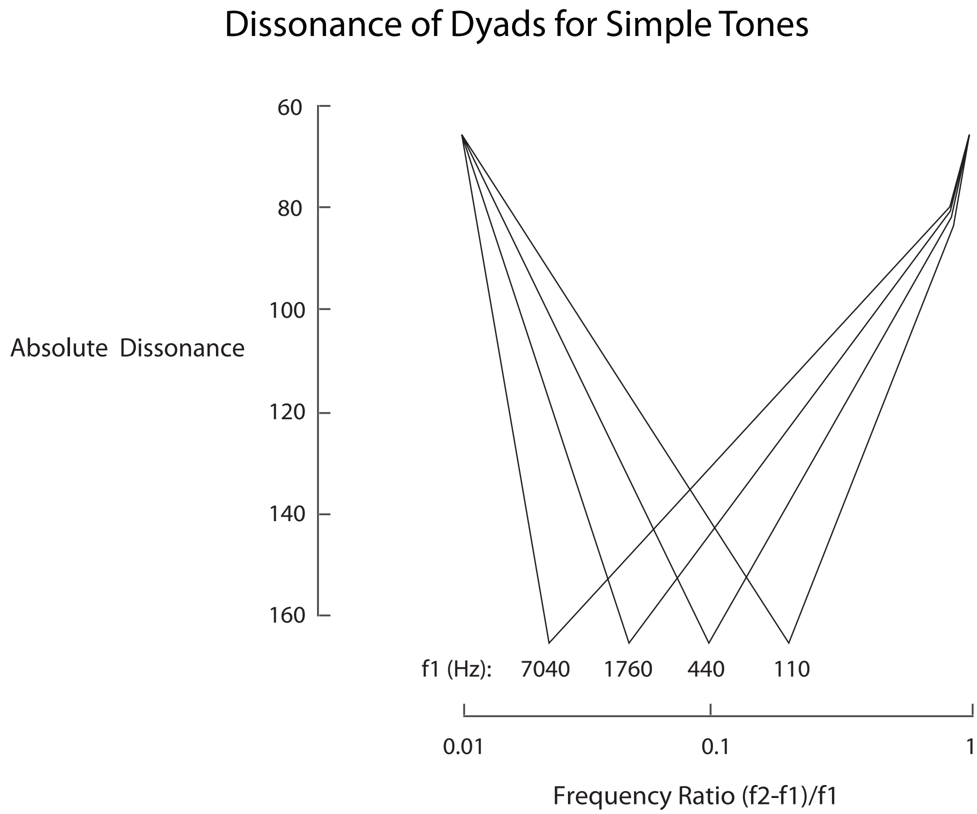


Figure 2.3: The dissonance of simple-tone dyads follows a v-shaped curve on a logarithmic ratio scale. Dissonance increases in the downward direction. The highest dissonance occurs around a frequency ratio of 0.1. Adapted from Mashinter [22], Figure 1. Original figure from Kameoka and Kuriyagawa [7], Figure 4.

horizontal axis and increasing dissonance in the downward direction, the V-curves in Figure 2.3 emerge.

For a dyad of two complex tones (with several harmonics as well as the root), sharp dips perceived dissonance arise at the simple ratios used in just intonation, such as 3:2 and 4:3. The paper demonstrates a model that closely agrees with the empirical ranking of dissonance for complex dyads (Figure 2.4). The premise of the model is that the interaction of all pairs of frequencies present in the two complex tones adds to the total dissonance. The dissonance of each frequency pair is determined primarily by the v-shaped curve of dyad dissonances, but also accounts for the effects of pitch register and sound pressure level.

2.4 Factors Influencing Consonance and Dissonance

Harmony perception involves learned and innate components. In reality, this complex perceived quality of harmony cannot be fully characterized by a model based only on frequency ratios, because there are many factors that influence the perception of dissonance. Even for sensory dissonance, perceived dissonance depends not only on frequency ratio but

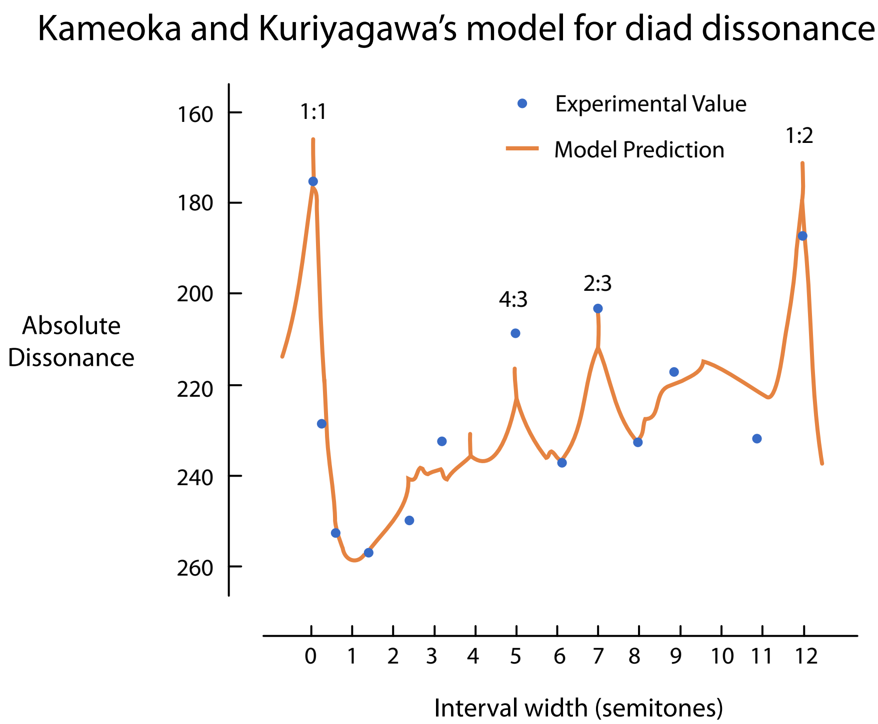


Figure 2.4: When the tones contain six harmonics in addition to the fundamental frequency, intervals at simple integer ratios are ranked as less dissonant. The peaks from the model of Kameoka and Kuriyagawa [29] (solid line) agree well with empirical rankings (plotted points). Adapted from [29], Figure 7.

on the pitch register (high or low) and sound pressure level (perceived as volume) of the harmony [7]. While these variables could be accounted for in a model (as in [29]), other factors in a holistic picture of dissonance are much more complex. Sensory dissonance does not account for culturally learned preferences, which play an important part in perception. As Plomp states, "our consonance perception is indeed profoundly influenced by the development of Western music and musical training." [26] Also, the perception of harmony depends on context. In two different chord sequences or musical settings, the perception of a harmony can change completely [13].

2.4.1 pleasantness versus consonance

Pleasantness is often used interchangeably for consonance in describing an appealing aspect of harmony. A study by Plomp and Levelt [26] showed that Western listeners rank the pleasantness and consonance of harmonies similarly. Studies have also shown that all people perceive the fast beating of close tones called "roughness" by Helmholtz [13]. However, the pleasantness of consonant harmonies may not be universal across all listeners [30, 31].

2.4.2 multiple dimensions of harmony

The perception of reality is very complex, yet models attempt to project the entire perception of harmony onto a single dimension of consonance to dissonance. In reality, there are many features of harmony which are distinct and contribute to the holistic perception of its sound. For instance, the quality of roughness can be distinguished from, or be seen as one attribute of the general phenomenon of dissonance [32]. Similarly, Stumpf's tonal fusion—the degree to which different tones sound like one tone—can be perceived separately from consonance and dissonance [14, 27]. In general, there are many different qualities of harmony beyond their consonance or dissonance, and attempting to characterize harmony with a single dimension can only capture a piece of true human perception. This view of harmony perception may explain the non-additive phenomenon of dissonance. Some chords, for example, a C major seventh chord (spelled C, E, G, B), are perceived as more consonant than intervals they contain (C, B) [25]. None of the models above allow for decreased dissonance when tones are added to a chord. A multidimensional perspective of dissonance could allow the interaction of multiple dimensions of harmony, creating a more flexible and holistic picture of perception. However, the concept is beyond the scope of this paper.

Chapter 3

Two Models for Predicting Common Scales

Though a single dimensional view of consonance and dissonance struggles to describe harmonic preference in Western Music, modeling innate preferences for sets of frequencies is still possible. Recent studies by Stolzenburg [33] and Gill and Purves [8] provide computational methods for predicting the harmonies and modes commonly used in Western music. (Other models exist [e.g. 34], but are beyond the scope of this paper.) These approaches predict the most prominent harmonic and modal construct s of Western music (the major and minor triads and diatonic scales) without involving music theory or concepts of harmony.

The fact that harmony is not present Arabic music [17] suggests that a universal or widely applicable model of intonation should focus on melodic or scalar ("horizontal") constructs rather than harmonic ("vertical") analysis. The following methods predict not only commonly used, "consonant" harmonies, but also commonly used scales in Western music.

3.1 Periodicity Model for Dissonance

Stolzenburg's model for ranking the "harmoniousness" of an arbitrary number of simple tones is remarkably straight forward. The input is a set of frequencies representing simple sine tones. The output is a scalar "periodicity" value which ranks the relative harmoniousness of the set of tones.

The algorithm is as follows. First, superimpose the sinusoidal waveforms of all frequencies in the set. Call the period of this aggregate waveform the "total period." The periodicity is the total period divided by the longest individual sinusoid's period. Figure 3.1 illustrates the process visually for the C major triad, containing the notes C, E and G. The ratio of the frequencies of E to C (a major third) is 5/4, and the ratio of G to C (a perfect fifth) is 3/2. The individual sinusoids of C (blue), E (middle) and G (bottom) are shown, as well as their superimposed waveform. The time axis is normalized so that the period of C is one. The periodicity is defined as the period of the aggregate waveform on this normalized scale. The C major triad has periodicity of four, because the aggregate waveform has a period four times the period of the lowest frequency. The periodicity can also be found by computing the least common denominator of the frequency ratios, which is also four.

An auto-correlation function can find the period of the aggregate waveform. However,

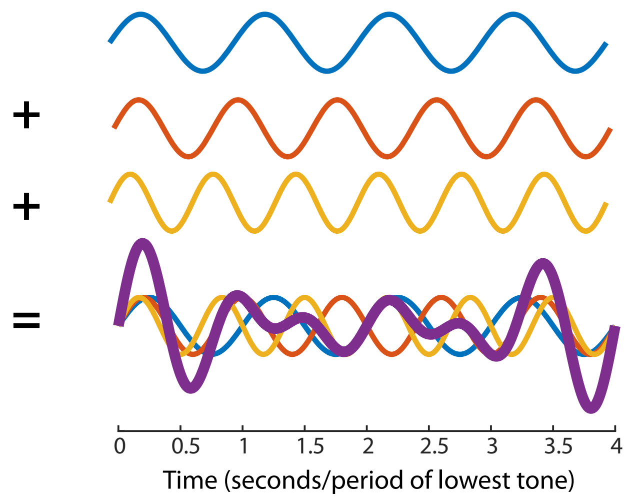


Figure 3.1: The Periodicity method finds the period of the aggregate waveform of all frequencies. This figure illustrates the periodicity of a C major triad. The aggregate waveform has a period of 4 times the lowest frequency's period.

this is unnecessary if the algorithm can represent the set of frequencies as ratios of the lowest frequency. Note that if two frequencies in the set are only slightly different from whole number ratios of each other, their waveforms will be barely misaligned and will take a long time to coincide. This is problematic because perfect ratios won't appear in empirical experiments, but small enough deviation from perfect coincidence will be unnoticeable to the human ear. This suggests that the ratios should be rounded to a whole-number fraction with some "tolerance." The appropriate size of this tolerance is unclear, but Stolzenburg claims 1% is a reasonable value [33]. A simple algorithm (see Appendix B) finds the simplest whole-number ratio that represents the original frequency ratio within the error tolerance. Then, the periodicity is simply the least common denominator of the whole number ratios, divided by the smallest ratio (usually 1/1). For example, consider input frequencies of 200 and 301 Hertz, and a tolerance of one percent. The ratios of each frequency to the lowest frequency are 1:1 and 1:1.505. The first fraction is 1/1, and the second is 3/2 because the error in expressing 1.505 as 3/2 is less than the tolerance of 1%. The set of harmonies as whole ratios is $\{1/1, 3/2\}$. The least common denominator is 2, and the smallest ratio is 1/1, so the periodicity score is $2/1 = 2$.

Interestingly, modeling complex tones instead of simple tones (by including harmonics) will not change the periodicity. Since the frequencies of harmonics are whole-number multiples of the fundamental frequency, their ratios will have the same denominator. Recall that in the Kameoka and Kuriyagawa [7] model for dissonance of dyads, using complex rather than than simple tones created the spikes at simple ratios. The fact that Stolzenburg's model does not depend on complex tones suggests that the Periodicity model is describing a preference for a quality of harmony wholly different from Helmholtz's roughness of upper partials.

The paper also presents a modification of the Periodicity method called "smoothing."

Smoothing recalculates the periodicity of the set of frequencies using each frequency as the root of the tuning system. The Smoothed Periodicity is the arithmetic average of these shifted periodicities. Finally, the Smoothed Logarithmic Periodicity is the average value of $\log_2$(periodicity). Stolzenburg justifies the use of logarithmically scaled periodicity based on Fechner's law, the principle that human perception scales logarithmically with the intensity of the stimulus [28].

3.1.1 method, claims and results

Using these ranking methods, Stolzenburg [33] analyzes combinations of two, three, four and more tones taken from a twelve-tone equal temperament scale (the Western chromatic scale). These rankings can be compared to experimental data on the perceived consonance and dissonance of various three and four tone harmonies. The experimental data is from Johnson-Laird et al. [13], which reports perceived rankings of consonance and dissonance for harmonies.

For this research, the Periodicity method was recreated in Matlab. A script analyzed the Smoothed Periodicity and Smoothed Logarithmic Periodicity of all 220 possible combinations of three tones from the chromatic scale. The top twenty ranked triads are shown in Table 3.1. In the Periodicity model, the best-ranked sets of three tones coincide very well with empirical rankings. The three inversions of the major triad, the most important three-note chord in Western music, are ranked as the top three triads. The minor triad and its inversions also appear in the top twenty of 220 total combinations analyzed.

The predictive power of the Periodicity model is not limited to harmonies. Table 3.2 shows the twenty top ranked heptatonic scales out of 792 possible combinations of seven notes from the chromatic scale. The best-ranking sets of seven tones coincide perfectly with the most used heptatonic scales in Western music. The major scale, the most commonly used scale in Western practice, is ranked highest. The top seven scales are the seven diatonic modes. Several scales are variations of the eight-note bebop scale (spelled C D E F G A Bb B C, common in jazz) with one note missing. The remarkable success of this method in ranking the most used heptatonic scales suggests that Stolzenburg's Periodicity method is a useful method for predicting likable sets of tones. Whether the success applies outside the context of Western music will be examined in Chapter 5.

The success of this method in predicting the most commonly used scales and chords in Western music is remarkable because it uses no music theory to aid in the ranking process, unlike other methods such as the dual process theory [13]. This suggests that something in the method emulates an innate biological preference for certain combinations of tones. Indeed, Stolzenburg points out that the auditory cortex has the ability to detect the period of a signal [4, 35]. Periodicity may be a natural way of describing human perception of harmony. If this measure of harmoniousness does reflect innate preference, it should perform well in non-Western musical cultures. However, Chapter refchapter:applyToModels will show that the method does not generalize to Arabic music.

Table 3.1: The Periodicity method ranks the three inversions of the major triad as the top combinations of three tones chosen from the chromatic scale, out of 220 possible combinations.

Semitones			Spelling	SP	SLP	Chord
0	5	9	C F A	3.0	1.58	Major, 2nd Inversion
0	4	7	C E G	4.0	2.00	Major
0	3	8	C Eb Ab	5.0	2.32	Major, 1st Inversion
0	5	7	C F G	6.0	2.58	Sus4
0	2	7	C D G	8.0	3.00	Sus2
0	2	11	C D B	8.0	3.00	
0	4	11	C E B	8.0	3.00	Major 7
0	7	11	C G B	8.0	3.00	
0	5	10	C F Bb	9.0	3.17	Sus2, 2nd Inversion
0	9	10	C A Bb	9.0	3.17	
0	3	7	C Eb G	10.0	3.32	Minor
0	7	8	C G Ab	10.0	3.32	
0	4	5	C E F	12.0	3.58	Major 7, 2nd inversion
0	4	9	C E A	12.0	3.58	Minor, 1st inversion
0	7	9	C G A	16.3	3.78	
0	1	3	C Db Eb	15.0	3.91	
0	1	5	C Db F	15.0	3.91	
0	1	8	C Db Ab	15.0	3.91	
0	5	8	C F Ab	15.0	3.91	Minor, 2nd inversion
0	2	9	C D A	16.3	3.92	

Table 3.2: The twenty top-ranking heptatonic (seven-note) scales are shown, out of all 462 unique seven-note combinations possible from the chromatic scale. The last two columns show the smoothed periodicity and smoothed logarithmic periodicity. The major scale, the most commonly used scale in Western practice, is ranked highest.

Rank	Spelling	Name	Smoothed	Smoothed Log
1	C D E F G A B	Major (Ionian)	129.7142857	6.453079015
2	C D E F# G A B	Lydian	139.2857143	6.584155849
3	C D E F G A Bb	Mixolydian	135.2857143	6.606909814
4	C D Eb F G A Bb C	Dorian	145.2857143	6.615323198
5	C D Eb F G Ab Bb C	Minor (Aolian)	151.5714286	6.766593725
6	C Db Eb F G Ab Bb	Phrygian	162.8571429	6.777567382
7	C Db Eb F Gb Ab Bb	Locrian	163.4285714	6.790062074
8	C D F F# G A B		135.4285714	6.818992447
9	C D E F G Bb B	Bebop Dominant -6	136.4285714	6.830772756
10	C D Eb F Ab A Bb	Bebop Minor -5	142.1428571	6.850159797
11	C D E G Ab A B	Bebop Major -4	148.2857143	6.852487484
12	C D F G A Bb B	Bebop Dominant -3	163.5714286	6.855047756
13	C D Eb F G Bb B	Bebop Harm. Minor -6	148.1428571	6.885175868
14	C D F F# A Bb B		148.2857143	6.889257175
15	C E F G A Bb B	Bebop Dominant -2	161.5714286	6.8926241
16	C Eb E G Ab Bb B		175	6.906890596
17	C Db Eb F G# A Bb		155.5714286	6.907123181
18	C D E F A Bb B	Bebop Dominant -5	166.1428571	6.951915171
19	C Eb E F# G Ab B		166.1428571	6.954401358
20	C Eb E F G A Bb	Bebop Dorian -2	151.8571429	6.984603556

3.1.2 generalizing to microtonal

Some cultures' musical modes are not derived from an equal-temperament division of the octave. For example, in Arabic music, the maqam is constructed using cumulative frequency ratios built on the previous tone, rather than by selecting tones from a some equal division of the octave [17]. The smoothed periodicity method must be modified in order to analyze the periodicity of these tuning systems. The original model assumes each frequency can be shifted to the first pitch of the tuning system, which doesn't work without an equal temperament system. Stolzenburg suggested the an alteration to the smoothed periodicity algorithm to generalize the method: instead of shifting each frequency to the root of a fixed tuning system, recalculate the frequency ratios for each frequency as the denominator.

For example, the frequencies of 100 Hz and 150 Hz will have two ratio sets, {1/1 3/2} and {2/3 1/1}. The periodicities of the sets are 2 and 3, respectively. Taking the average, the new smoothed periodicity is 2.5.

For western triads and heptatonic scales, the new version of the algorithm should produce similar rankings to the original. Table 3.3 shows that the rankings of this method do not correlate as well with empirical studies as the original methods. The modified method still predicts the commonly used triads and scales in Western music, but does not predicit common scales.

3.1.3 analysis

In general, the Periodicity model is very successful in ranking Western chords and scales taken from the twelve-tone equal temperament scale. However, there are limitations with the model, especially concerning its generalization to microtonal tunings. First, the ranking of heptatonic scales only finds three of the seven modes and does not rank the major or minor scales well (see Table 3.3). Second, the paper claims that the method can be generalized to Turkish and Central African music. In 2010, Stolzenburg [11] claims that average values for maqam periodicities were similar to the scores of Western modes, but provides little evidence. The application of the method to maqam melody types will be investigated further in Chapter 5.

3.2 Similarity Model for Dissonance

Gill and Purves [8] propose another method, called "Similarity," for numerically ranking scales or modes. This method is designed for arbitrary microtonal tunings. The premise is similar to Stumpf's tonal fusion. The score, called "percent similarity" or simply "similarity," measures the overlap of the harmonics of each tone.

3.2.1 methods, claims and results

The premise of the Similarity scoring algorithm is that the scales used to make music use tones with similar harmonic series. The harmonic series of a tone is the set of frequencies at all integers of the fundamental. Taking a dyad of any two frequencies, their "missing fundamental" or virtual pitch is the highest pitch of which both pitches could be a har-

Table 3.3: When the microtonal implementation of the Periodicity method is used to rank seven-note scales (SLP Micro), only three of the diatonic modes occur in the first twenty scales. The rank (SLP Rank) and score (SLP Score) from the original implementation of Smoothed Logarithmic Periodicity are shows for reference.

Rank	Spelling	Name	SLP Micro	SLP Rank	SLP Score
1	C Eb E F G A Bb		6.238350443	28	6.984603556
2	C D E F G A Bb	Mixolydian	7.32017517	4	6.606909814
3	C Db Eb F Gb Ab B		7.565372494	40	7.027226359
4	C D Eb E G A Bb		7.807069309	274	7.494180783
5	C D E F# G A B	Lydian	7.823191745	2	6.584155849
6	C Db F Gb Ab Bb B		7.883239144	45	7.089077703
7	C Db E F G A Bb		7.936724218	282	7.504347788
8	C D Eb F G A Bb C	Dorian	7.943017448	6	6.615323198
9	C Eb F G Ab A Bb		7.963227795	37	7.01729194
10	C D Eb E F Ab Bb		7.973733576	604	7.896149525
11	C Db Eb F G# A Bb		7.984806342	23	6.907123181
12	C Db Eb E Ab Bb B		8.002357306	410	7.656424967
13	C Db Eb Gb A Bb B		8.013381286	389	7.622929931
14	C Db D E Bb G B		8.033591633	175	7.352844675
15	C Db Eb F Ab Bb B		8.082141633	311	7.537842824
16	C Eb E F Ab A Bb		8.102938323	84	7.219440154
17	C Db D E F Ab Bb		8.128717704	590	7.882848181
18	C D E Gb G Bb B		8.144111444	228	7.436410747
19	C Db D F Gb Ab Bb		8.146650687	302	7.526062516
20	C Db E F G A Bb		8.173604426	130	7.303006226

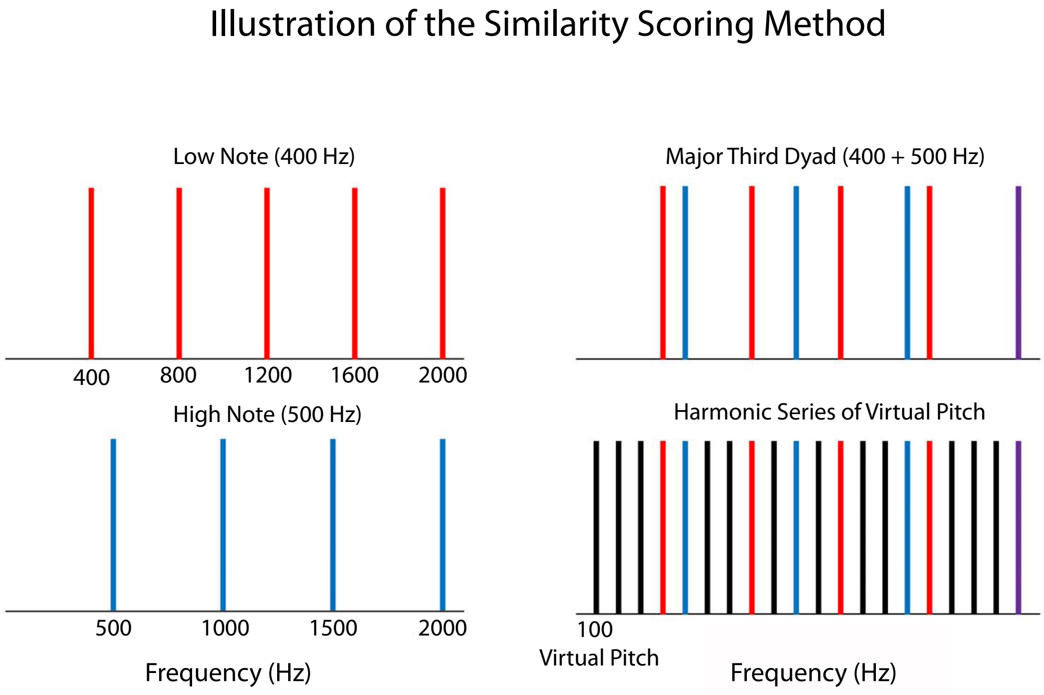


Figure 3.2: Each figure shows the harmonic series of a note or dyad. The Percent Similarity of a dyad is the percentage of the virtual pitch's harmonic series that is included in the combined harmonic series of the two dyad frequencies. The percent similarity score is 40. This figure was adapted from Gill and Purves [8], Figure 2.

monic frequency. Mathematically, the virtual pitch is the greatest common divisor of the frequencies, because this will make them integer multiples (harmonics) of the virtual pitch.

Figure 3.2 illustrates the method with an example. The virtual pitch of frequencies 400 Hz and 500 Hz is 100 Hz, because 100 is the greatest common divisor of 400 and 500. The method defines the measure called Percent Similarity as the percentage of the virtual pitch's harmonics which are included in the combined harmonic series of the dyad. In this case, the frequencies 400 Hz (Figure 3.2, upper left) and 500 Hz (lower left) have virtual pitch (greatest common divisor) of 100 Hz. In this case, the combined harmonic series of 400 and 500 Hz (Figure 3.2, upper right) make up 40% of the harmonic series of the virtual pitch (lower right). The similarity score of a scale is the average percent similarity of all possible dyad pairings within the octave.

One advantage of this approach is that it is built for microtonal pitch sets, rather than equal tempered or justly tuned scales. It is possible to analyze the similarity of any set of set of pitches in relation to a tonic, or root frequency. The only limit to the resolution of microtonal scales that can be examined is computational expense. Instead of just the twelve intervals of the chromatic scale, the 60 rational intervals with the best percent similarity

Table 3.4: The top scoring scales in Similarity method include Western modes and non-Western scales.

Rank	Name	MPS Gill	MPS Stolzenburg	MPS Lapp
1	phrygian	40.39	28.6750	40.3484
2	dorian	39.99	34.9425	39.6587
3	major	39.61	34.8485	39.6131
4	husayi	39.39	33.1595	37.4357
5	minor	39.34	30.5486	38.3603
6	lydian	38.95	31.8755	38.5603
7		38.83	27.9981	38.6593
8	kardaniya	38.76	30.0143	38.7644
9		38.69	29.0528	38.6933
10	mixolydian	38.59	36.9664	38.5931

are selected as candidates for creating scales. Then, the paper analyzes the similarity scores of scales that can be made with these intervals. In particular, five note (pentatonic) and seven note (heptatonic) scales are analyzed. Table 3.4 shows the top scoring scales given by the method. The results are impressive, but not as convincing as those from Stolzenburg's Periodicity method. The top ten scales contain all of the diatonic modes except the Locrian mode, and two non-western scales are also identified. The authors name the fourth result Husayni (also Husseini). According to maqamworld.com [17] and other sources [36], this scale is called Bayati and is common maqam built from the jins Bayati and Nahawand. It could also be seen as the descending form of maqam Husayni, but the ascending form of Husayni is built from the jins Bayati and Rast. This results (approximately) in the sixth scale degree being half-flat, rather than flat [17, 36].

The paper also claims the eight ranked collection is an Arabic scale called called Kardaniya, though this is not a maqam mentioned in any of the resources found in this research.

3.2.2 recreating the similarity algorithm

Recreating the results of the Similarity method was more difficult than the Periodicity study. In the actual implementation, the authors allowed a tolerance of 22 cents in choosing the best ratio for each dyad (see [8] for a full explanation). Matlab scripts implemented the model as described in the paper, but achieved slightly different similarity scores for some scales. The authors of the similarity method paper did not respond to an email inquiring about these discrepancies. It is likely that one of the tolerance parameters was implemented slightly differently than reported, or that some aspect of the algorithm was not fully described. Table 3.4 shows the results of the similarity method for heptatonic scales from the original study, from Stolzenburg's implementation, and from the implementation in this study.

3.2.3 analysis

While the similarity method gave high rankings to diatonic Western modes and found one common maqam, other common maqamat (or scales from other cultures) do not appear in the top results. Also, unlike in Stolzenburg's Periodicity results, the order of the modes does not reflect how common they are. For instance, the major scale is ranked third and the much less used Phrygian scale is ranked first. Still, Stolzenburg's method shows that at least Western diatonic modes posses tones with high degrees of similarities in their harmonic spectrum.

Part II

Methods

Chapter 4

Developing a Model to Analyze Intonation

This chapter describes the development of a method to extract information about intonation in musical recordings. The method takes audio recordings of music and extracts the frequencies at small time steps. This yields a frequency versus time representation of the music. Testing the model reveals that it is accurate for some types of music but innaccurate for other types of music.

4.1 An Algorithm for Analyzing Music

The general approach consists of breaking an audio file into short segments, finding the fundamental frequency of each segment, and compiling the results into a frequency versus time representation of the music. For each short segment of audio, Fourier analysis transforms the signal from the time domain into the frequency domain. This means that given the original signal indicating the sound pressure level over time, the Fourier transform gives the prevalence of each frequency of pressure oscillations during the time interval. Figure 2.2 shows how Matlab's Fast Fourier Transform function extracts the relative prevalence of each frequency from the time-domain signal of the audio file.

4.1.1 finding the fundamental frequency

Once the signal is represented in the frequency domain, the program must decide what the fundamental frequency of the current note is. Notice that the frequency-domain signal (Figure 4.1) has many peaks (harmonics) at integer-multiples of the fundamental (lowest) frequency. It is possible for the amplitude of a harmonic to be higher than the amplitude of the fundamental. The program chooses the fundamental frequency to be the lowest frequency peak who's amplitude is at least half the amplitude of the highest peak. This method is generally accurate, but sometimes chooses the first or second harmonic instead of the true fundamental frequency. In the example shown in Figure 4.1, the true fundamental frequency has an amplitude of less than half of a harmonic frequency. This means that the algorithm will incorrectly select the harmonic frequency instead of the true fundamental. However, allowing the algorithm to choose peaks less than half of the maximum amplitude

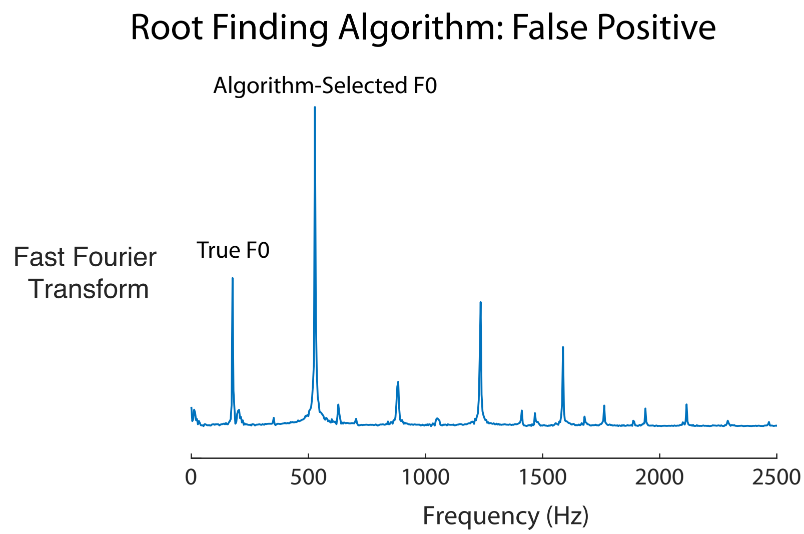


Figure 4.1: The Fourier transform of a pitch played by a clarinet reveals a harmonic stronger than the fundamental. The fundamental frequency algorithm selects the lowest frequency pitch that is at least half the amplitude of the highest peak (in this case, the first harmonic).

introduces noise and decreases overall accuracy. The accuracy of the algorithm used is about 97%, and incorrectly selects a harmonic for 3% of the samples analyzed.

4.1.2 subdivision of notes

When a note is played on any instrument, the sound changes from when the note is first heard to its sustain and decay. The initial burst of sound, called the transient, contains many inharmonic frequencies and incoherent noise. This quickly turns into regular, periodic frequencies giving the harmonic spectrum. From a mathematical perspective, this corresponds to a transient effect giving way to a steady state solution. The transient is very short (on the order of 50 milliseconds), but has a large effect on the perception of the instrument's timbre. Timbre, also called tone color, is the sound of the instrument apart from pitch and volume. The difference between a piano and a flute playing the same note is due to the different timbres of the instruments [28].

As a result of the transient, the frequency response of a note changes over time. For example, consider the response of a piano when the hammer hits the string. The transient is a messy aggregation of noise, the sustain is stable or oscillates slightly, and finally the note decays. Figure 4.2 shows the frequency-domain plots for four segments of a note played by a piano. Frequency response taken over 0.05 seconds show the piano's frequency response at the beginning of the note (top), a fortieth of a second later (middle), and a twentieth of a second later (bottom). The aperiodic transient characteristics quickly give way to a steady state solution with peaks at integer-multiples of the fundamental frequency.

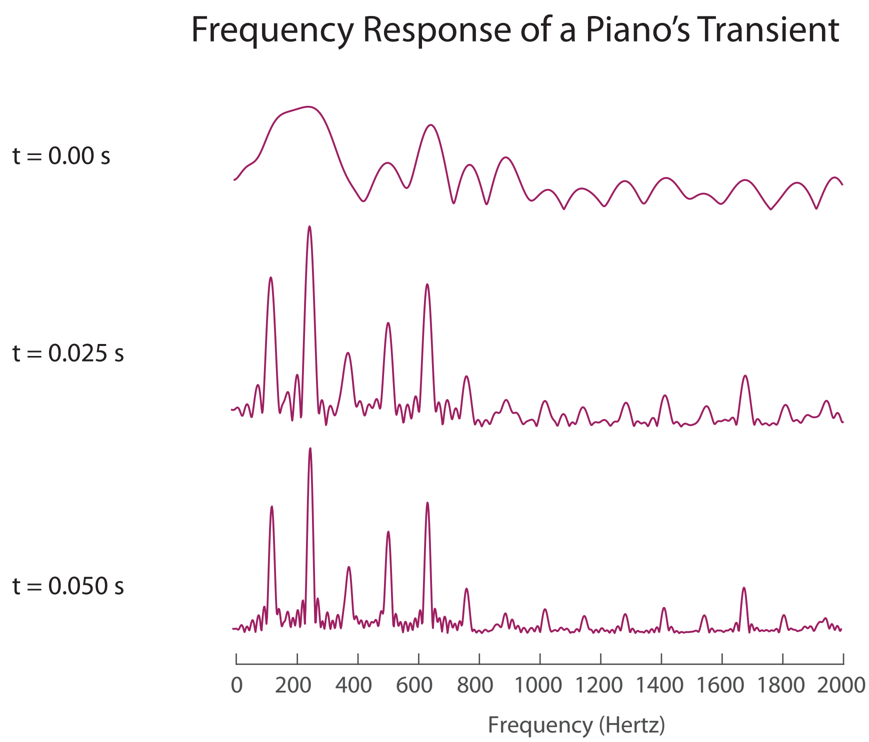


Figure 4.2: The transient characteristics quickly give way to a steady state solution with peaks at integer-multiples of the fundamental frequency.

4.1.3 sample length and frequency resolution

Choosing the time duration of the audio segment constitutes a trade-off between frequency resolution and time resolution. For a sample of length 1/n seconds, the frequency resolution of the Fourier transform is n Hertz. For example, breaking the file into one second clips gives high resolution in the frequency domain (1 Hz) but low resolution in the time domain (1 note per second). Music usually changes pitches much faster than one note per second. However, if the audio file is divided into 10 samples per second, frequency resolution degrades to 10 Hz. This trade-off sets a fundamental limitation in the accuracy of audio transcription. Note that a 10 Hz resolution in frequency may work well for pitches above 1000 Hz (less than 1% error), but will cause large error for pitches in the range of 100 Hz (about 10% error). This implies that transcription will be more accurate at higher frequencies than at low frequencies. The extraction method was tested with sample lengths from one second to one twentieth of a second. Transcription was the most accurate using a third of a second (three samples per second). This value is used throughout the paper.

4.1.4 parameters and optimization

Several parameters can be modified in order to optimize the model's pitch extraction accuracy. One way to increase the time resolution of transcription without losing frequency resolution is to overlap the audio segments, effectively creating a "moving average." A five second audio file with one second samples can be divided into 5 second clips starting on each second, or 9 clips starting on each half second. Transcription accuracy was tested for overlap from a half to a tenth of a sample. The 1/3 sample overlap performed the best. Combined with the third of a second sample length, this means the method takes 1/3 second clips starting every 1/9 of a second.

Additionally, the program must separate a series of extracted root frequencies into distinct notes. The exact frequency of a held-out note can change slightly as it is played. The program registers a new note when the frequency changes by more than some set tolerance. This tolerance is expressed as a percent difference between the fundamental frequencies of the current and previous samples. In other words, it is a "backwards difference" method. In experimentation with tolerances from 0.1% to 5%, the model performed best with a tolerance of 3%. This allows continuous notes to be kept together while separating distinct tones. For example, given the series of frequencies (in Hertz) {100 102 150 151 150}, the program consolidates them into two distinct notes, {100 150}.

4.2 Testing the Model

A test of the model's accuracy is its ability to transcribe music—that is, reproduce the sequence of notes played by the performer. If the model is able to transcribe music accurately, this indicates that it is consistently extracting the correct fundamental frequencies from the recording. Then the extracted frequency versus time information can be used to analyze intonation. In Western music, it is easy to test the accuracy of transcription because performances generally adhere exactly to the written score. Arabic music is largely improvised, so it is not possible to calculate the accuracy of Arabic music transcription by

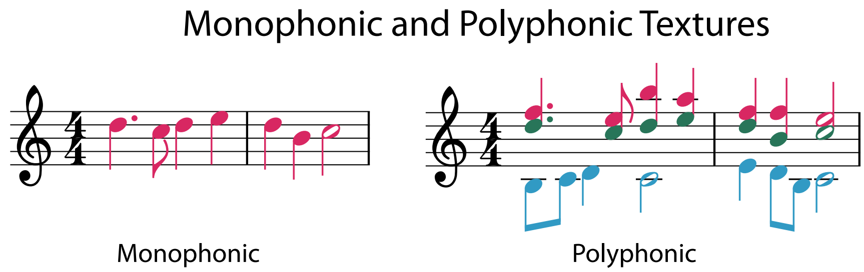


Figure 4.3: Monophonic music has only a single line while polyphonic music has multiple lines occurring at once. The three colors in the polyphonic example represent three musical lines.

comparing to a score.

The accuracy of transcription depends on the musical texture, which can be "monophonic" or "polyphonic." Music is described as being monophonic if only a single musical line is present. That is, only one note is heard at a time. By contrast, polyphonic music has multiple independent musical lines occurring at once. This means multiple notes can be played together or overlap. Figure 4.3 shows an example of monophonic and polyphonic music. The three different colors in the polyphonic music represent three independent musical lines that occur at the same time. These could be played by the same instrument or by different instruments (or voices).

The model assumes that the input is monophonic music. While most music is not monophonic, transcribing polyphonic music is beyond the scope of this paper. Section 7.1 explains why the model can still be an effective tool for analyzing polyphonic music. All models were implemented in Matlab.

Extracting a single series of frequencies from a polyphonic piece of music can, at best, capture the musical line with the highest volume. In reality, if three instruments are playing, the model will extract the pitches from different instruments at different times. Because of this, for polyphonic music, the extracted frequencies will jump around instead of following a single musical line.

4.2.1 monophonic and polyphonic music

Figure 4.4 shows a snapshot of the Matlab algorithm analyzing the Prelude from J.S. Bach's Cello Suite Number 1 in G minor. The piece is slow and monophonic (a single melody without accompaniment), and the method transcribes the piece with greater than 90% accuracy. This means that the notes extracted agree with the notes written in the score of the piece. In the top graph, the live Fourier transform of the signal shows the cello's obvious root with three strong harmonics. The algorithm consistently picks out the root frequency from this data. As a result, the lower plot shows the smooth contour of the cello's melodic line.

Polyphonic music is not as easy to capture. Figure 4.5 shows the analysis of the first movement from Antonio Vivaldi's masterpiece "The Four Seasons." The piece fast and polyphonic, with many different instruments playing interlocking melodies and harmonies.

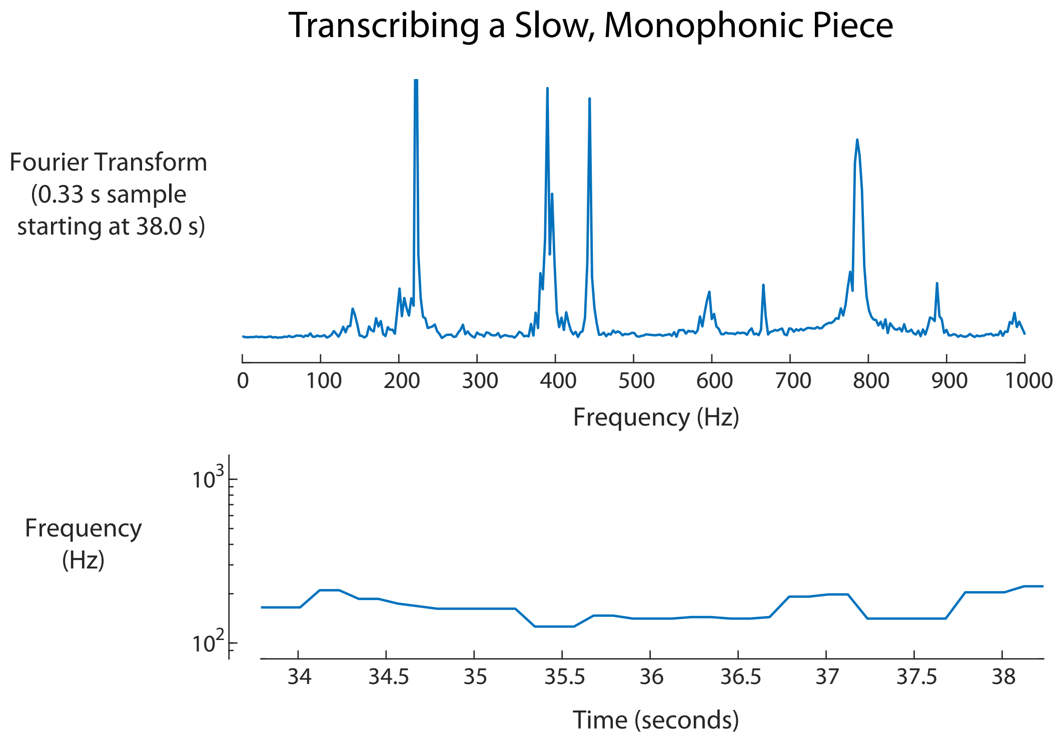


Figure 4.4: Transcribing a slow, monophonic piece results in an accurate curve with smooth transitions between notes.

The Fourier transform at the top of the figure shows that there is no clear fundamental frequency for the melodic line, which is not surprising since several notes are being played at once. As a result, the frequencies extracted over time (bottom graph) jump around between the different musical lines and do not follow any single melodic idea. It does not make much sense to measure the transcription accuracy of a polyphonic piece where a monophonic series of frequencies is extracted. The ideal case would be that the model picks out the primary melodic line and transcribes it, but the model fails to do this. Instead, notes are taken from different parts at different times. In general, the model accurately transcribes slow, monophonic music but fails to transcribe polyphonic music. Section 7.1 will show that despite inaccurate transcription, the model can extract useful information from polyphonic music.

4.2.2 from pitches back to music

Another way to test the accuracy of the extraction process is to use the extracted frequency versus time information to recreate an audio file. Matlab can write an audio file with sine tones at the frequencies indicated by the extracted information. If the recreated audio is played back along with the original recording, it should closely coincide with the performer. Subjectively, the recreated audio does match well and seem to be co-performing the piece. However, by itself the Matlab-produced audio sounds nothing like music. The pure sine tones, lack of dynamics or phrasing, and overall mathematical sound did not constitute an

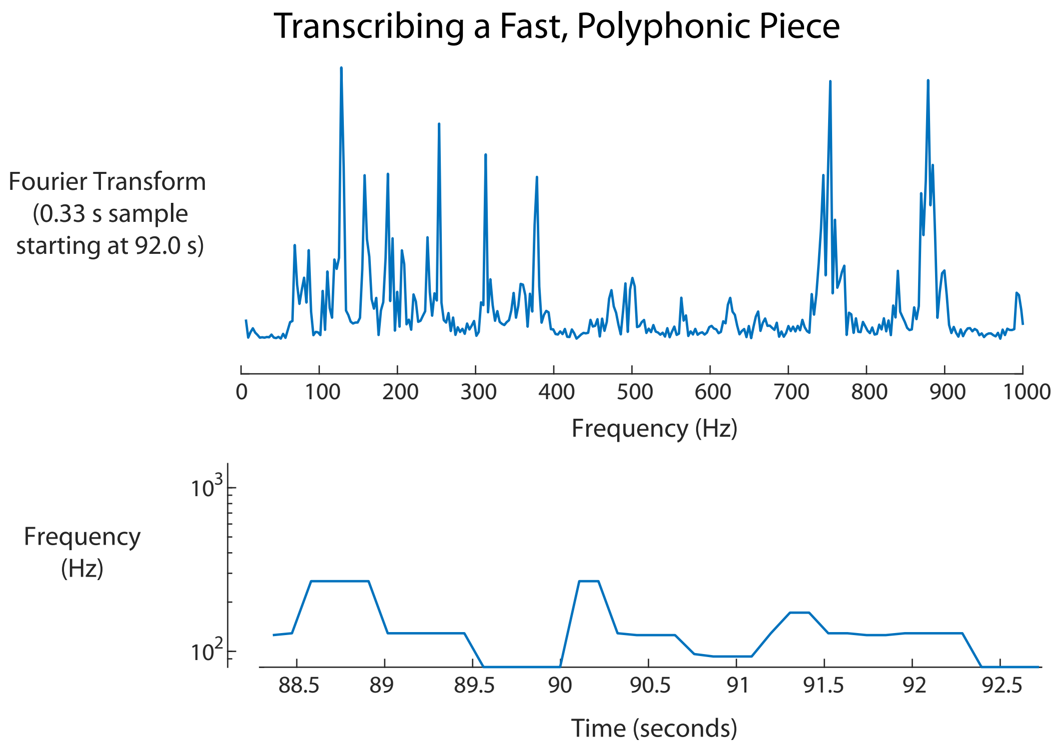


Figure 4.5: Transcribing a fast, polyphonic piece results in an inaccurate curve that jumps between different lines in the music.

electronic performance of the piece. This serves as a reminder that music is much more than a time-series of frequencies. Keeping frequency information while discarding timbre, dynamics, articulation, harmony, polyphony, acoustic space and other factors reflects one aspect of the complex sound that is music. Section 10.3.1 discusses the limitations of analyzing only the frequency aspect of music.

Chapter 5

Applying the Tools to Periodicity and Similarity

The models of Periodicity and Similarity were successful in predicting the diatonic modes used in Western music. Both the Periodicity [4] and Similarity [8] papers claim that their methods apply beyond Western music to microtonal tuning systems. To test these claims, this section characterizes maqamat with discrete sets of intervals, then finds the Periodicity and Similarity scores of the maqamat. Characterizing maqamat as a set of intervals is difficult because the tuning of maqamat is not based on a fixed interval system like Western scales. For a variety of tunings analyzed, Arabic maqamat do not score favorably in either the Similarity or Periodicity method. Neither method is successful in generalizing to Arabic intonation.

5.1 Extracting Scales

A histogram of the fundamental frequencies occurring in an audio file should reveal the most prominent notes used. The histogram has discrete "bins" into which frequencies are sorted. Alternatively, a probability density function could analyze a continuous frequency spectrum. If a musical scale describes a collection of discrete frequencies used in a piece of music, then sharp peaks should arise in the histogram or probability density function at the frequencies in the theoretical scale of the piece.

Using the extraction method developed, histograms of frequencies extracted from both Western and Arabic music showed the emergence of scales. For Western music, the peaks in the histogram are sharp and located at the intervals in the equal temperament scale. Figure 5.1 shows a histogram of frequencies extracted from Beethoven's 9th Symphony, recorded by the London Symphony Orchestra. The symphony is in the key of D minor. The solid green bars in the figure show the theoretical placement of the intervals in the D minor scale, according to equal temperament. The outlined bars show the other notes in the chromatic scale, which do not belong to the key of D minor. The peaks in the histogram (blue) represent the frequencies extracted most often. As expected, these peaks align closely with the theoretical tuning of the D minor and chromatic scale.

Figure 5.3 shows a histogram of the pitches extracted from an Arabic piece by Simon Shaheen on the album "Turath" [37]. The piece is in maqam Nahawand. The arabic maqa-

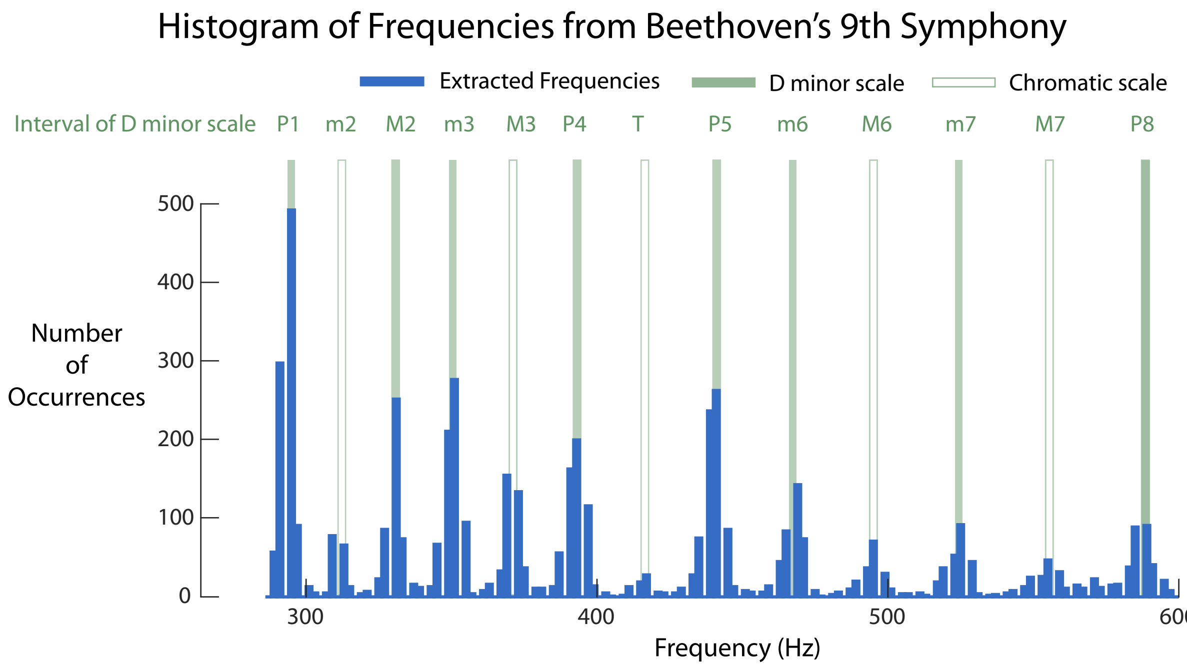


Figure 5.1: The most common frequencies extracted from a recording of Beethoven's 9th symphony in D minor (blue) align well with the theoretical equal-temperament tuning of the D minor scale (green).

mat do not have a definite, exact theoretical tuning. The approximate intonation of maqam Nahawand in Western notation is shown in Figure 5.2. In this case, only the histogram of the actual performance can show the exact intonation of maqam Nahawand. Peaks in the histogram match the theoretical tuning, but there are also smaller peaks and troughs at other frequencies. A large peak also occurrs at the minor second (m2) interval. Notes from outside the theoretical maqam tuning occur because the performer may temporarily modulate (change) to a different maqam [9], or tune the same note differently on different occasions. These variations in tuning make it difficult to define a maqam by an exact set of intervals. Any fixed interval set would be an incomplete description of the maqam.

5.2 Characterizing Maqamat

The tuning of a single maqam can vary significantly based on geography, the performer and even within a performance. The website maqamworld.com has multiple recordings of various maqamat played ascending and descending by oud (Arabic lute), nay (flute) and buzuq (fretted lute). The pitch relationships played in these recordings can be compared to theoretical and experimental models in literature. Figure 5.4 shows the intonation of the notes in maqam Rast for three recordings from maqamworld, as well as other tunings reported in literature. Each row shows the placement of notes in the Rast maqam. For example, the first row shows the tuning of Rast extracted from a recording of maqam Rast from maqamworld.com played on buzuq.

The seventh scale degree, which is theoretically a "neutral third" (between a minor and

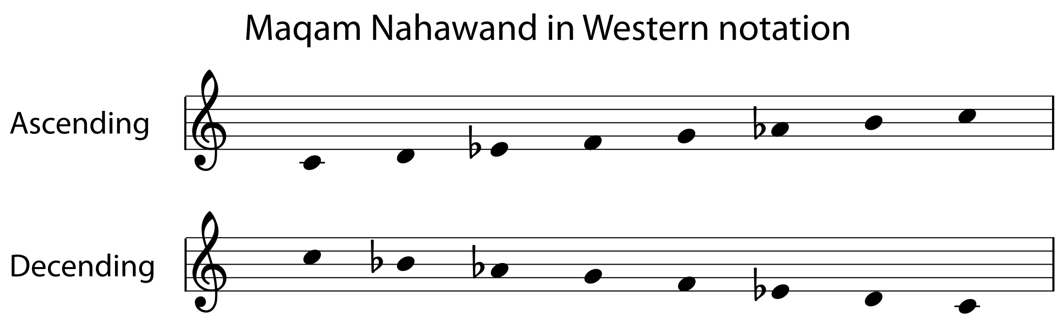


Figure 5.2: The approximate notation of maqam Nahawand in ascending form is the same as the Western harmonic minor scale. In descending form, the seventh degree ('B') is flat, transforming it into the Western aeolian mode [17].

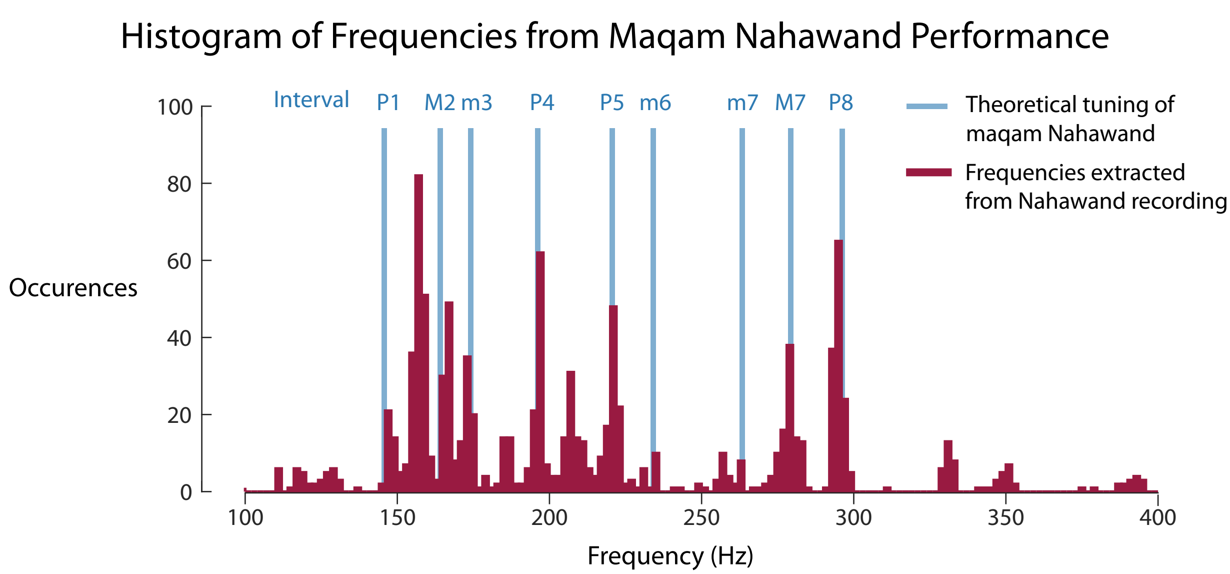


Figure 5.3: The most common frequencies extracted from the piece "Samai Nahawand" by Simon Shaheen [37] align with the theoretical tuning of maqam Nahawand.

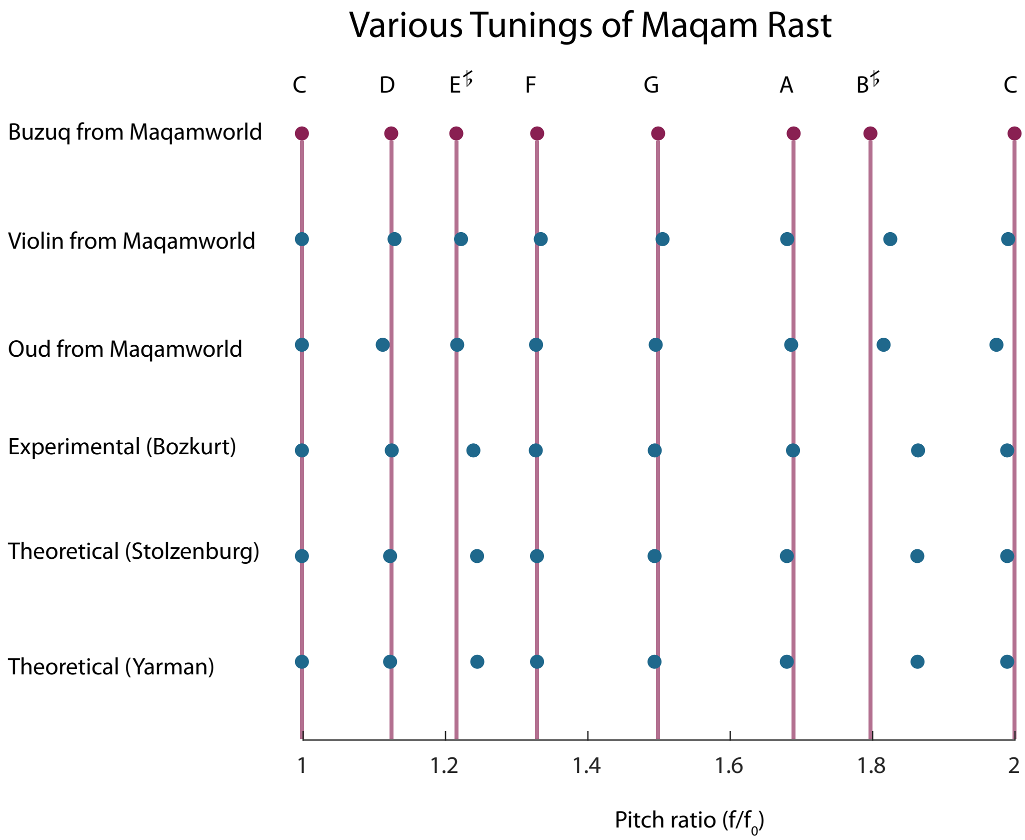


Figure 5.4: Three recordings of maqam Rast from Maqamworld [17] played on buzuq, violin, and oud were analyzed to find the exact ratios between the frequencies played. An experimental result from music analysis by Bozkurt [38] and two theoretical models [4, 39] are shown for comparison.

major third) above the fifth has the most variation between the various tunings. The third degree, a neutral third above the tonic (first scale degree), also has considerable variation. In general, the variation in tuning shows that a single maqam cannot accurately be characterized by a set of interval ratios.

5.3 Periodicity and Similarity of Maqamat

Now that the method can extract the tuning of maqamat as performed in a recording, these extracted scales can now be ranked in the Periodicity method (Stolzenburg) and Similarity method (Gill and Purves). Both authors claim that their methods predict commonly used scales beyond Western music [4, 8]. However, the results do not show favorable rankings for maqamat in either the Periodicity or Similarity methods. Table 5.1 shows how the tunings of the five Turkish maqam tunings used by Stolzenburg are scored by the microtonal implementation of Smoothed Periodicity (SP) and Smoothed Logarithmic Periodicity (SLP), as well as Similarity (SIM). The scores for the Western major and minor scale (in equal temperament) are shown for comparison. A 1% tolerance was used in all the methods.

Table 5.1: Tunings of six maqamat according to values provided by Stolzenburg [11] are scored in the Smoothed Logarithmic Periodicity model (SLP) and Similarity model (SIM).

	1	2	3	4	5	6	7	8	SP	SLP	SIM
Rast	1	1.125	1.248	1.333	1.500	1.688	1.873	2	1411	9.684	0.393
Nihavend	1	1.125	1.185	1.333	1.500	1.580	1.778	2	14652	13.085	0.346
Ussak	1	1.110	1.185	1.333	1.500	1.580	1.778	2	9696	11.521	0.340
Huseyni	1	1.110	1.185	1.333	1.500	1.664	1.778	2	1331	9.423	0.367
Hicaz	1	1.053	1.248	1.333	1.500	1.664	1.778	2	1890	9.616	0.397
Huzzam	1	1.068	1.202	1.283	1.500	1.602	1.803	2	736	9.367	0.333
Western Major	1	1.122	1.260	1.335	1.498	1.682	1.888	2	142	6.362	0.425
Western Minor	1	1.122	1.189	1.335	1.498	1.587	1.782	2	184	6.850	0.391

All of the maqamat score much less favorably than the Western major and minor scale (lower periodicity is favored, while higher similarity is favored). In fact, the scores are similar to the worst ranking scales created with the chromatic scale. In other words, the methods would choose almost any combination of notes in the chromatic scale (even combinations that would very rarely occur in music) before choosing the most common maqamat tunings. This shows that the methods are not successful in predicting the theoretical intonation of the maqamat which were provided by Stolzenburg.

The tunings extracted from Arabic maqamat recordings on maqamworld.com [17] show similar results. Table 5.2 shows the extracted tuning ratios for five common maqamat, again compared to the Western major and minor scale. Maqam Rast, maqam Suznak and maqam Ajam have Similarity scores that are not far from the minor scale, but none of the maqamat score low in the Periodicity method. For comparison, the Western chromatic scale has a Smoothed Periodicity score of 360, lower than all of the maqamat examined. This implies that the Periodicity method would not choose any of the common maqamat before even the most unlikely combination of seven notes from the chromatic scale. Evidently, the methods which so accurately predict the preferred scales in Western music fail to extend to Arabic music.

There are many possible reasons that maqamat do not score well in the models of Periodicity and Similarity. It is possible that classical Arabic music has fundamentally different preferences for combinations of tones. Also, the fact that Arabic music rarely uses harmony [18] may explain why these methods, which developed out of methods for consonant harmony [e.g. 7, 26, 6], fail to predict Arabic music's modal intonation systems.

In reality, though, the models are fundamentally ill-suited for analysis of microtonal music. Both methods rely on approximating frequency relationships as whole-number ratios, and assume that the tuning system will use a fixed set of ratios as the basis for all intonation in the music. This is not an accurate picture of the maqam. The pitch extraction methods show that maqam intonation is flexible, and depends on context and placement. The idea of a fixed pitch set goes against the nature of aurally-learned, improvisation-based music. The following section explores the features in both Arabic and Western music that cannot be accurately depicted by fixed pitch sets.

Table 5.2: The table shows the tunings of maqamat extracted from recorded performances, with their scores in the Smoothed Logarithmic Periodicity (SLP) and Similarity (SIM) Methods. None of the maqamat score well in the Periodicity or Similarity methods.

Scale Degree:	1	2	3	4	5	6	7	8	SP	SLP	SIM
Rast	1	1.129	1.229	1.343	1.500	1.686	1.829	2	1235	8.926	0.351
Suznak	1	1.113	1.217	1.330	1.509	1.604	1.887	2	4512	10.315	0.357
Sikah	1	1.067	1.167	1.317	1.467	1.617	1.750	2	12775	10.735	0.223
Jiharak	1	1.114	1.229	1.314	1.500	1.657	1.829	2	5296	10.307	0.318
Ajam	1	1.132	1.264	1.340	1.509	1.698	1.887	2	614	8.263	0.352
Western Major	1	1.122	1.260	1.335	1.498	1.682	1.888	2	142	6.362	0.425
Western Minor	1	1.122	1.189	1.335	1.498	1.587	1.782	2	184	6.850	0.391

Chapter 6

Non-Discretized Intonation

In reality, music rarely uses a set of discrete frequencies. Some instruments with fixed tuning, like the piano, result in very narrow frequency bands in which each note is tuned. However in many instruments, such as the human voice or a fretless string instrument, tuning is flexible. Cohen [40] use the term "scatter" to describe the width of the distribution in a note or interval's tuning. Western performers tune relatively consistently, but still have measurable scatter. Arabic performers have a much larger scatter. This means that a single note or interval may be tuned in significantly different ways, even by the same performer and in a single performance [40].

6.1 Pitch Distribution and Scatter

Any interval will have a distribution of the exact frequency ratios at which it is played. In Arabic music, this distribution tends to be very wide. Cohen and Katz [41] call the width of this distribution "scatter." Their study found that Arabic musicians use large scatter, especially in intervals of seconds and thirds. In contrast, Western musicians generally tune with less scatter.

6.1.1 case study: the neutral third

The large scatter of intonation in Arabic music does not imply that musicians are imprecise in their tuning. Conversely, Arabic musicians distinguish between slight variations in tuning that might go unnoticed in a Western context [9]. For example, Arabic violinist and composer Sami Abu Shumays demonstrates in an instructional recording that in the context of twelve different maqamat, the E (or E flat or E half flat) is tuned in twelve distinct ways [9]. The differences in the tunings of these notes result from the structure of the maqam in which they occur. Therefore, in order to appreciate the differences in the intonation of these maqamat, the frequency resolution should capture twelve distinct notes between E and Eb.

In the Periodicity method, the intonation is allowed 1% error. By taking the frequencies of these twelve notes and mapping them to whole number ratios (to a fixed note C) with tolerance of 1%, several of these frequencies are mapped to the ratio 11/9. However, the differences in the intonation of these E's are exactly what give each maqam a unique character. For example, Shumays explains that in maqam Jiharak, "the third scale degree

Table 6.1: When the tolerance for tuning ratios is decreased from 1% to 0.66% (the tolerance necessary to differentiate between various maqam thirds), the periodicity method no longer finds the common Western scales. Each Western mode is ranked and scored with the Smoothed Logarithmic Periodicity (SLP) method, using 1% tolerance and then 0.66% tolerance.

Scale	SLP 1% Rank	SLP 1% Score	SLP 0.66% Rank	SLP 0.66% Score
Major (Ionian)	1	6.453079015	346	11.23089366
Lydian	2	6.584155849	331	11.21911335
Mixolydian	3	6.606909814	352	11.24267396
Dorian	4	6.615323198	356	11.25639209
Minor (Aolian)	5	6.766593725	376	11.27068192
Phrygian	6	6.777567382	382	11.28317661
Locrian	7	6.790062074	384	11.29567131

is just barely higher than Rast, but not quite high enough to be Ajam" [42]. In Shumays' demonstration of Jiharak on violin, the third is played at a frequency 1.228 times the root. In maqam Rast, he plays the third at a ratio of 1.220. In Stolzenburg's algorithm, using a tolerance of 1%, both of these ratios are mapped to 11/9. This indicates that the method does not capture the nuances in intonation that distinguish one maqam from another.

The tolerance required to distinguish between these notes is about 0.66%. However, if Stolzenburg's Periodicity method is implemented with only 0.66% tolerance for finding tuning ratios, the method is no longer effective. Table 6.1 shows that when the tuning tolerance is only 0.66%, the Periodicity method does not give good scores to any of the Western diatonic modes. This implies that the method is unable to accommodate the small tolerances necessary to describe the intonation of maqamat.

6.1.2 case study: the leading tone in Western music

Table 6.2 shows the most prevalent frequencies in a Bach Cello Suite performance, along with the theoretical tuning of each note in equal temperament. While most of the notes conform well with the expected equal temperament tuning, the F# is tuned with a large amount of variance. In fact, the distribution of frequency for F# seems to be bimodal. Sometimes it is at the expected tuning of the major seventh, and other times it is tuned very sharp (at a higher frequency than the standard tuning of the note). The context of the F# explains this bimodal behavior. When the F# is followed by the neighboring G, the first note creates tension before resolving upward to the tonic. In this case, the F# is functioning as the "leading tone." A leading tone is a note that creates tension, then resolves up by a half step to a place of rest [15]. When a note functions as the leading tone, it is common practice for the note to be played sharp.

The same phenomenon occurs less prominently with the third scale degree leading to the fourth (B to C in this case), which explains the occurrence of a sharp third in row 9 of Table 6.2. The variable and contextual tuning of notes in Western music is an important reminder that fixed pitch systems and equal temperament do not truly describe musical

Table 6.2: The most common frequencies extracted from the Sarabande movement of J.S. Bach's Cello Suite number 1 in G are shown with their number of occurrences (Num), frequency ratio with the root (Ratio), scale degree, the just intonation ratio (Just Ratio), and the perecent difference from the just ratio (% Dif).

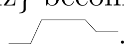
Num	Frequency	Ratio	Scale Degree	Just Ratio	% Dif
46	167.566	0.8400	Major 6	0.8333	0.792
36	199.493	1.0000	Root	1.0000	0.000
34	297.862	1.4931	Perfect Fifth	1.5000	0.461
31	165.840	0.8313	Major 6	0.8333	0.243
28	188.275	0.9438	Major 7	0.9375	0.667
22	170.155	0.8529	Major 6 (Sharp)	0.8333	2.325
18	223.654	1.1211	M2	1.1250	0.346
17	296.136	1.4844	Perfect Fifth (flat)	1.5000	1.042
15	252.129	1.2638	Major 3 (sharp)	1.2500	1.102
13	112.341	0.5631	Major 2	0.5625	0.113
13	147.720	0.7405	Perfect 5 (flat)	0.7500	1.278
13	265.935	1.3331	Perfect 4	1.3333	0.021
11	197.767	0.9913	Root	1.0000	0.869
11	247.815	1.2422	Major 3	1.2500	0.624
11	379.836	1.9040	Major 7 (leading tone)	1.8750	1.535
10	299.588	1.5017	Perfect 5	1.5000	0.116

performance. It also suggests that tuning may depend heavily on context, and that the relationships between notes and their musical contexts are indispensable (as with the leading tone). It is possible that, even without any absolute tuning system, links in intonation between cultures could emerge through the contextual tuning of consecutive notes.

6.1.3 issues with fixed pitch set

No fixed set of pitch ratios can capture the intonation of a maqam. Regional variation in tuning, scatter in performance, non-equal temperament, and the importance of context in intonation all render the idea of a fixed ratio set for a maqam essentially inadequate. The maqam forms the melodic structure of Arabic music, but it may be counterproductive to label it as a "scale" in the western sense of a set of pitch ratios. The examples above have shown that describing the maqam using western measurement systems often discards essential information about intonation, not to mention other aspects of the maqam such as melodic phrases, large scale form, and character that go beyond intonation. Even in western music, expressing intonation as a fixed set of pitch ratios is an incomplete picture of true musical practice. In reality, musicians with the ability to produce continuous or flexible pitch on their instruments instruments use musical context to mold intonation beyond a fixed equal temperament tuning.

6.2 Shape Grammars: Words and Vocabularies

In order to understand intonation in the context of music, rather than as isolated occurrences of various notes, it is necessary to examine a series of frequencies in time. The series of frequencies may contain two, three, or more distinct pitches. In this method, the ratios between a set of consecutive frequencies from a piece of music is called a "word." Each ratio in the word is called a "syllable." These musical words are conceptually similar to the idea of a shape grammar. They describe a feature in a musical piece that may reoccur frequently, and may be related to the style, composer, function or cultural background of the piece. A word describes the relative intonation of a series of frequencies, and is described by the ratio of each frequency to the first frequency. For example, the series of frequencies {100 Hz, 200 Hz, 150 Hz} becomes the three-syllable word {1/1, 2/1, 3/2}. In frequency-time space, this looks like . Since only the ratio relationships between the pitches are considered, words are equivalent through transposition. That is, making all of the frequencies higher or lower by a common factor does not change the word. For example, the frequencies {300 Hz, 600 Hz, 450 Hz} also form the word {1/1, 2/1, 3/2}. Two syllable words simply describe an interval, while words with more than two syllables describe a melody or melody fragment.

6.2.1 finding words

Finding series of frequency ratios follows naturally from the methods used to extract the frequencies from an audio segment. The program simply takes a series of distinct ratios and divides by the first frequency to create the ratios of the word. A word is created starting on each extracted frequency. When the amplitude of the audio signal drops below a fixed threshold, the program interprets a "rest" (the musical term for the absence of a note) and no frequency is taken. When the same note occurs before and after a rest, the program merges them into one note so that the same ratio will not occur twice in a row in any word.

6.2.2 vocabulary

The set of all words used in a selection of music is called the 'vocabulary'. Each word in the vocabulary is ranked by its frequency of occurrences in the selection. Vocabularies indicate how certain pitch relationships, intervals, and melodic ideas reoccur in a selection of music. Because the frequencies extracted from the original audio samples are not discretized (for example, to a twelve-tone equal temperament scale), the words provide detailed information about intonation in a temporal context. For example, a Western string player's sharp leading tone will appear in the context of its upward resolution, and will not be "auto-tuned" to the equal temperament value of the note. This continuous resolution is critical for retaining the intonation of Arabic music, which does not adhere to discrete intervals.

Part III

Results, Analysis, and Conclusions

Chapter 7

Accuracy of Vocabulary Extraction

The method of extracting words and vocabularies provides interesting insight into the intonation of Western, Arabic and Chinese music. In Part III, the method of vocabularies is used to analyze intonation of various recordings of Western, Arabic and Chinese music. (Appendix C provides a list of citations for all musical recordings analyzed in this research.) The following chapters analyze the reliability of the vocabulary method, and then explore how it can be used to learn about intonation in different cultures.

The compiled vocabulary of a selection of music can be represented as a histogram of words. This type of vocabulary histogram will be used to compare the common words found in various composers, styles and genres. Figure 7.1 shows an example of the vocabulary extracted from J. S. Bach's Brandenburg Concerto Number 1. The horizontal dimension represents the relative prevalence of the word compared to all words occurring in the piece.

Analyzing two recordings of the same piece gives insight into the accuracy of the vocabulary extraction. If the method is reliable, it should extract the same (or similar) vocabulary from two recordings of the same piece (assuming the piece is played from a written score, and not improvised). For two recordings of a piece played according to a score, any differences in the extracted vocabulary are due to error in the model. Various factors including the model parameters (samples per second, tolerances), the style of music (Western, Arabic, Chinese), the tempo (fast or slow) and the musical texture (monophonic or polyphonic) can affect the accuracy of the vocabulary extraction process.

7.1 Tempo and Polyphony

The tempo (speed) and polyphony (one note at a time or many notes at once) affect the accuracy of transcription and extraction of individual notes and words in the method. This section examines the impact of these factors on the overall vocabulary found by the model.

7.1.1 slow, monophonic music

As discussed in Section 4.2, the model accurately transcribes slow, monophonic music such as Stravinsky's "Three pieces for Clarinet Solo." As expected, the vocabularies extracted from two recordings of this piece are very similar (Figure 7.2). This suggests that the method accurately captures the vocabulary. The vocabulary extraction is accurate for short segments

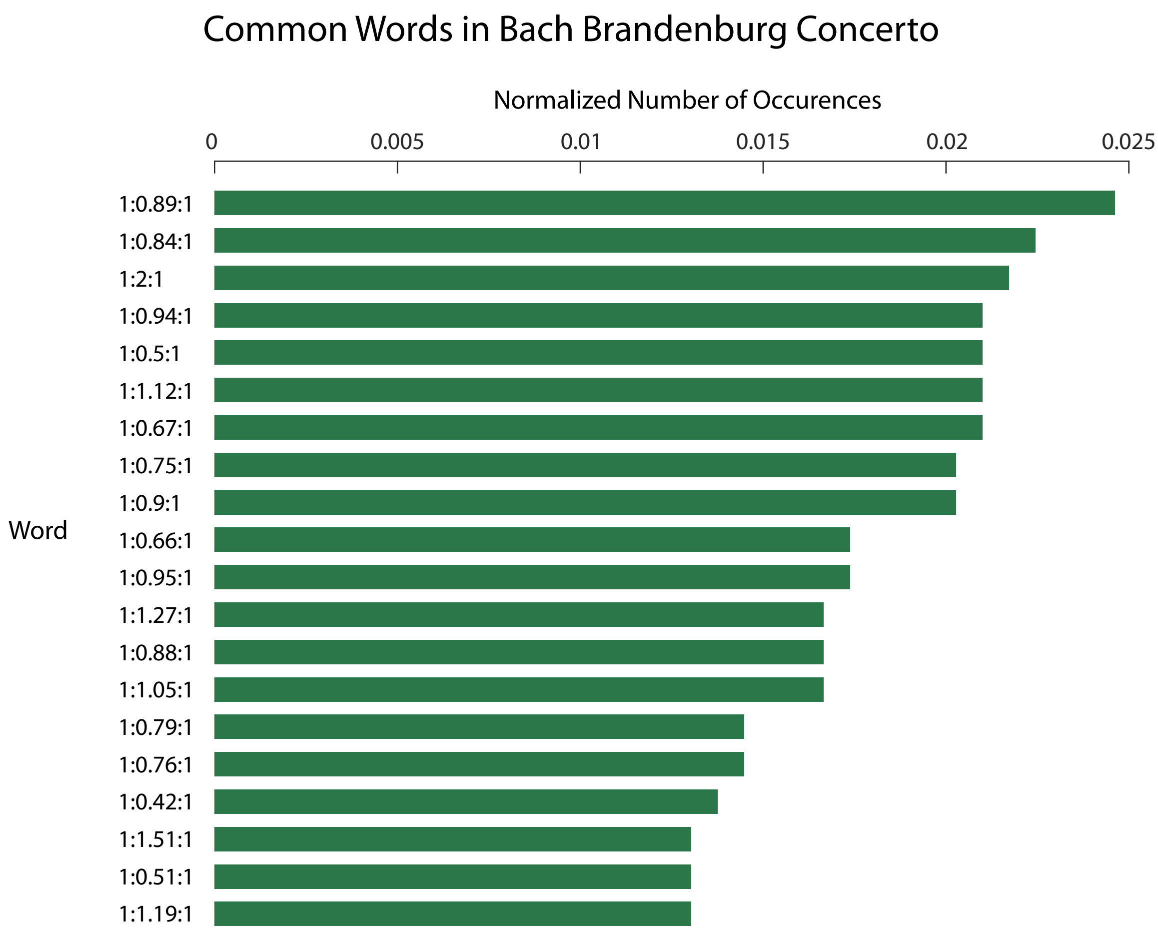


Figure 7.1: The most common three-note words in Bach's Brandenburg Concerto Number 1.

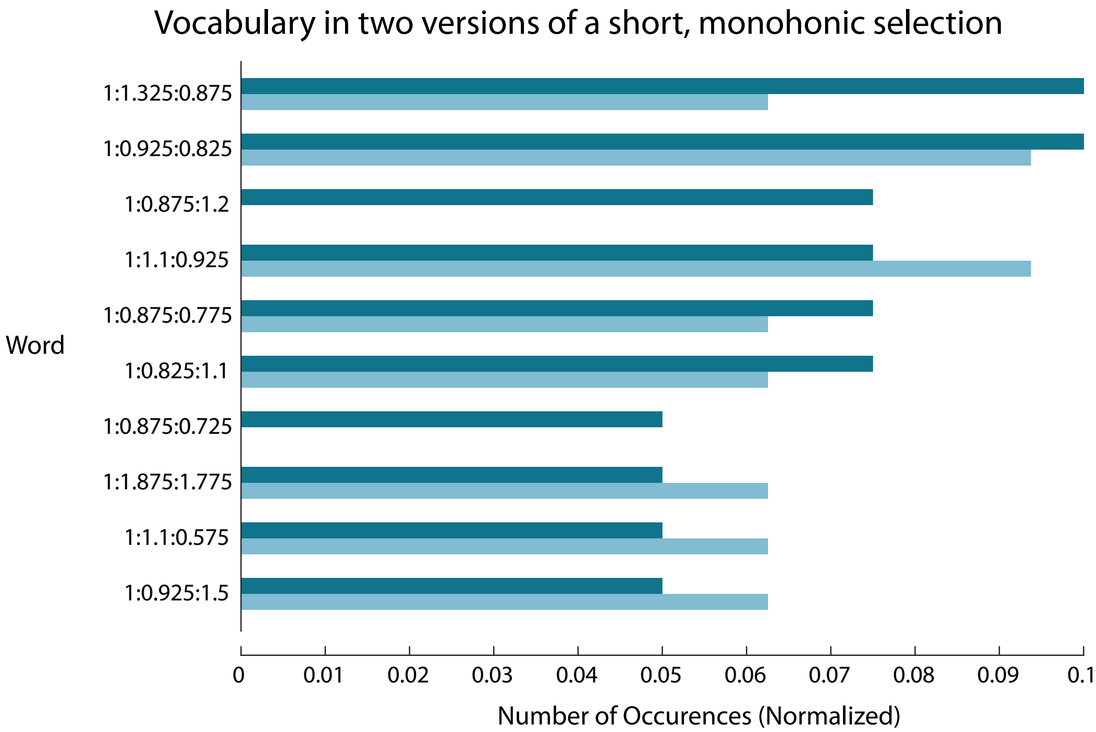


Figure 7.2: The vocabularies extracted from two versions of a slow, monophonic piece are very similar, even when the selection is less than two minutes long. The segment analyzed for this example was a ninety second clip from Igor Stravinsky's "Three Pieces for Clarinet Solo."

of the music (less than thirty seconds) as well as over the course of a longer piece (several minutes to hours).

7.1.2 fast, polyphonic music

For fast, polyphonic music, the transcription was not accurate. However, this does not mean that the vocabulary extracted from these pieces is meaningless.

Using filtering and a large sample set, the model can accurately find the vocabulary of fast, polyphonic pieces. The algorithm uses two types of filters. First, a bandpass filter discards frequencies outside of a set range. This not only removes low background noises, but also focuses the transcription on a specific frequency range. If the frequency bandwidth is customized for an individual song, it can focus on one musical line by isolating frequencies in that line's frequency range.

For a short segment of music, on the order of seconds to a few minutes, the noise in transcription dominates the small sample size of the extracted vocabulary. Two recordings of a single piece have dissimilar vocabularies at this close-up time scale. Figure 7.3 shows the vocabulary extracted from two recordings of a thirty second selection from Beethoven's Ninth Symphony. The dissimilar vocabulary is a result of errors and noise in the extraction process. In general, the vocabulary extraction is not accurate for short selections of polyphonic music.

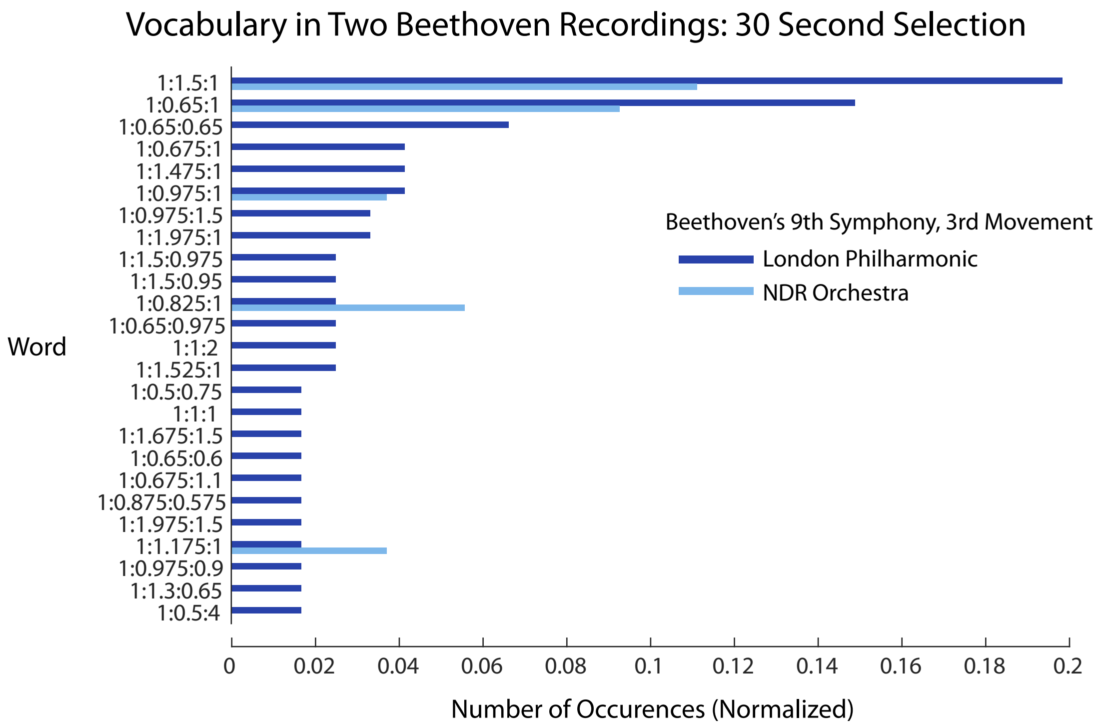


Figure 7.3: Vocabulary was extracted from a thirty second selection of Beethoven's Ninth Symphony recorded by two different orchestras. The sharp differences in vocabulary show that, for short segments of polyphonic music, the vocabulary extraction method is inaccurate.

However, over a long piece of polyphonic music, the vocabulary can converge and the influence of noise is reduced. For example, for an hour-long piece like Beethoven's Ninth Symphony, two recordings converge to similar vocabularies (Figure 7.4). Using a tolerance of 0.025 for distinct ratios shifted the focus from exact intonation to towards the general notes played. In general, using a tolerance of 0.01 will give more detailed insight into the intonation, but using 0.025 in this case shows that the extraction method over a large polyphonic sample can converge.

Still, for some long selections the vocabulary does not converge. For example, the method extracted dissimilar vocabularies from two recordings of a set of twenty-four Chopin preludes. This likely occurs because of the chordal texture where many lines occur at once. Still, in the music analyzed, the vocabulary converged for over 90% of long selections of Western music. Because Arabic music is improvised rather than played note-for-note from a score, it is not possible to do the same type of comparison for Arabic music.

7.1.3 file formats

Wave files store much higher resolution audio information than a compressed, lossy mp3 format. Figure 7.5 shows that converting a wav file to mp3 and m4a formats did not significantly affect the vocabulary extracted. However, the mp3 files were processed much faster than the much larger wav files. Choosing to use the smaller mp3 format increased

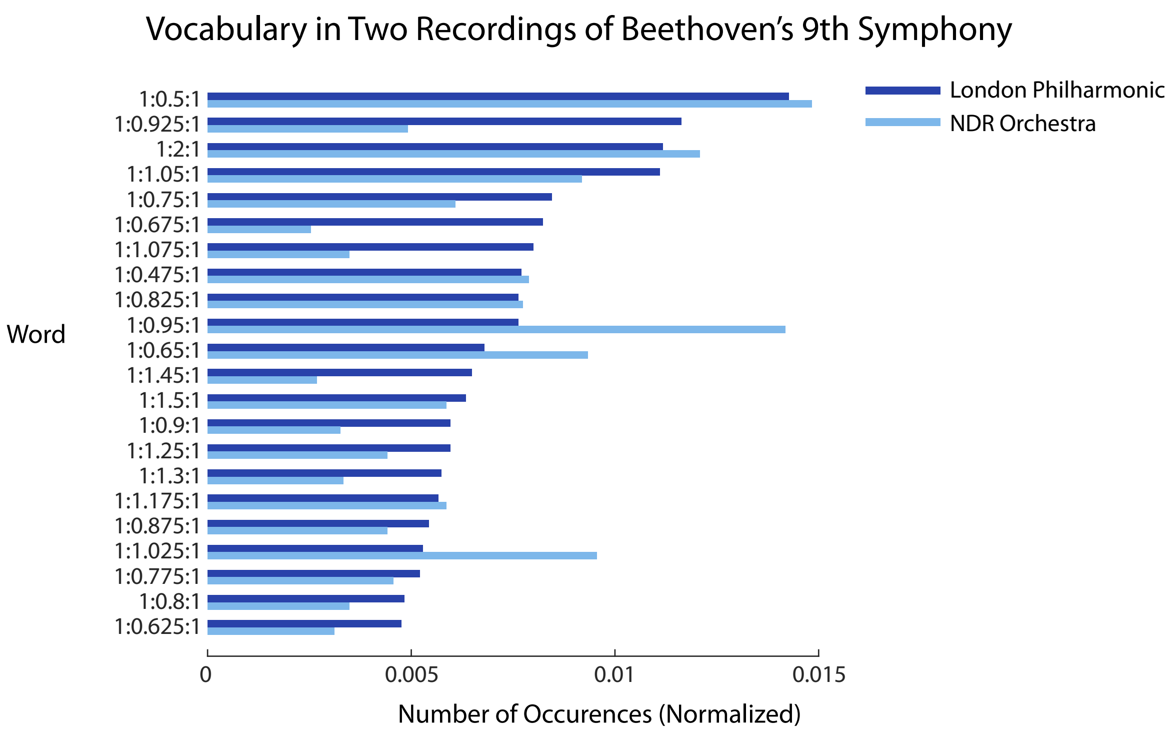


Figure 7.4: Theoretically, the vocabulary should be nearly identical in two recordings of the same piece. In this case, the method finds predominantly similar vocabularies in two recordings. The recordings are by the North German Radio (NDR) Symphony Orchestra and the London Philharmonic.

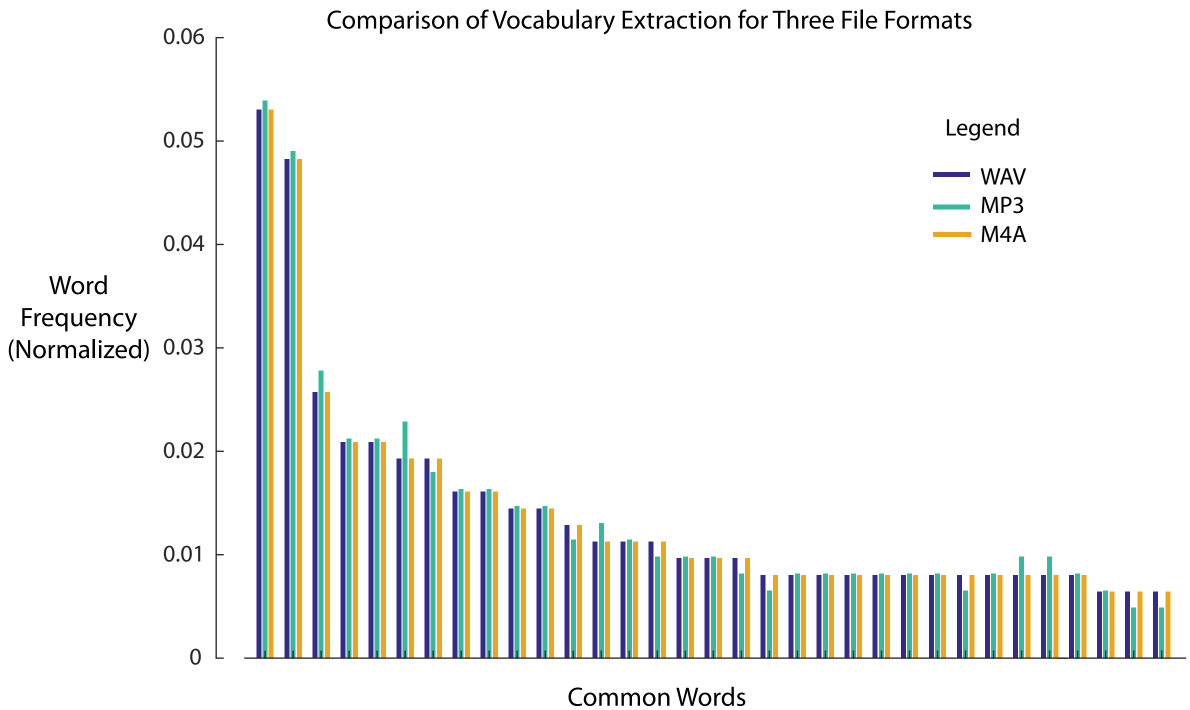


Figure 7.5: The vocabulary extracted from the piece "Spring on a Moonlit River" [43] is very similar when the original .wav files are compressed into .mp3 or .m4a formats.

efficiency without compromising the vocabulary extraction. For this reason, mp3 files were used for all analysis.

7.2 Symmetry of Three-Syllable Words

Irrespective of genre, almost all of the most common three-syllable words found by the method return to the note they started on. That is, they are of the form {1:R:1} where R is the interval moving to some other pitch, before returning back to the same note. The pattern is so pervasive that it is necessary to ask whether this phenomenon an accurate representation of the patterns in music, or an error resulting from the methodology.

There are at least two possible ways that the method could falsely over-report these symmetrical words of the form {1:R:1}. First, consider two simultaneous frequencies with similar amplitudes. They might be from two separate notes or from the harmonic series of the same note. It is possible that as their amplitudes and harmonic spectra change with the duration of the notes, the program chooses first one, then the other frequency, and back to the first. This "flip-flopping" between two choices for the fundamental frequency would result in these alternating-note patterns.

Second, a similar error could occur even with a monophonic piece. When notes are played legato (connected) or have a long reverberation time, two frequencies may be present at once, which again could cause an alternating choice of root frequency. However, after manually analyzing the transcription from monophonic and polyphonic pieces, neither of these errors were common. This suggests that the prevalence of the {1:R:1} word form is an accurate representation of common three-note phrases in music.

Though vocabulary extraction for polyphonic music is inaccurate on a small time scale, it converges to the true vocabulary with a larger sample size. This makes it useful for analyzing large selections of music, such as the works from an entire album, a compilation of a composer's work, or music from an entire genre.

Chapter 8

Comparing Vocabularies

The power of the vocabulary is in the ability to compare and contrast the words and types of words used in different musical selections. This chapter first compares the vocabulary of two selections from a single composer, then compares two different composers' vocabulary, and finally looks at the vocabulary of all of Western music. In the following chapter, the vocabulary of Western music is compared to Chinese and Arabic music.

8.1 Vocabulary of Composers

Figure 8.1 shows the top twenty words in the vocabulary from two album's of J.S. Bach's music. The first sample is a recording of Brandenburg Concerto No. 1, and the second is an album of various selected masterpieces [43, 44]. Many of the most prevalent words in the Brandenburg Concerto are also common in the other works contained in the Masterpieces album. This indicates that for a single composer, vocabulary can be consistent across multiple works.

Next, it is necessary to compare Bach's vocabulary, of the Baroque era, to that of composers in different eras of Western music. Figure 8.2 shows that the vocabularies in Bach's and Beethoven's music are very different. This is not entirely surprising because the Baroque Era (Bach) and late Classical Era (Beethoven) are stylistically very different. When comparing the more stylistically similar works of Beethoven and Brahms (Romantic period), the vocabularies are much more similar (Figure 8.3). This suggests that the vocabularies can characterize stylistic genres. More in depth analysis of vocabulary comparisons across stylistic periods would be necessary to confirm this hypothesis.

8.2 Vocabulary of a Genre

Compiling vocabulary from a wide selection of Western music gives a picture of the intonation for the entire Western musical culture. Figure 8.4 shows the overall interval makeup from a compilation of Western music. The the equal-tempered major (solid green) and chromatic (outlined green) scales are shown for reference. The fact that peaks in intonation match closely with the theoretical values from equal temperament adds confidence that the model is performing well. In Section 5.1, histograms of frequency showed that the frequencies used in a Western piece align with the theoretical equal temperament scale. In Figure 8.4,

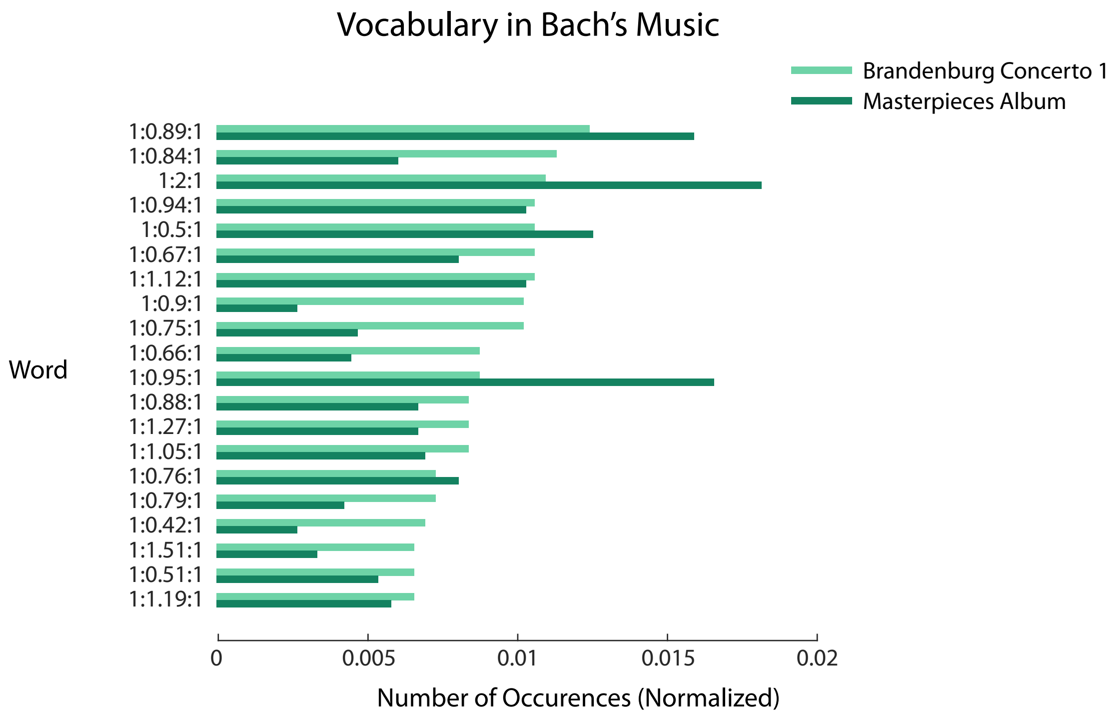


Figure 8.1: Most common words in two albums of Bach's music

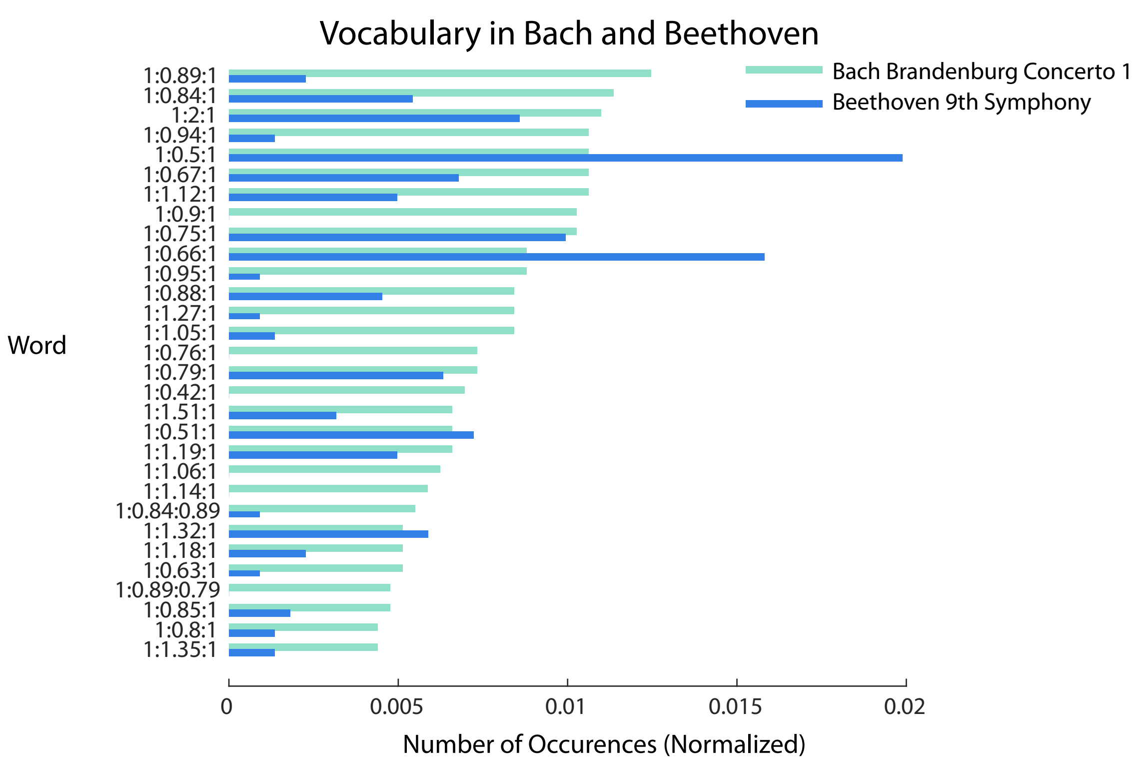


Figure 8.2: Most common words in pieces by Bach and Beethoven

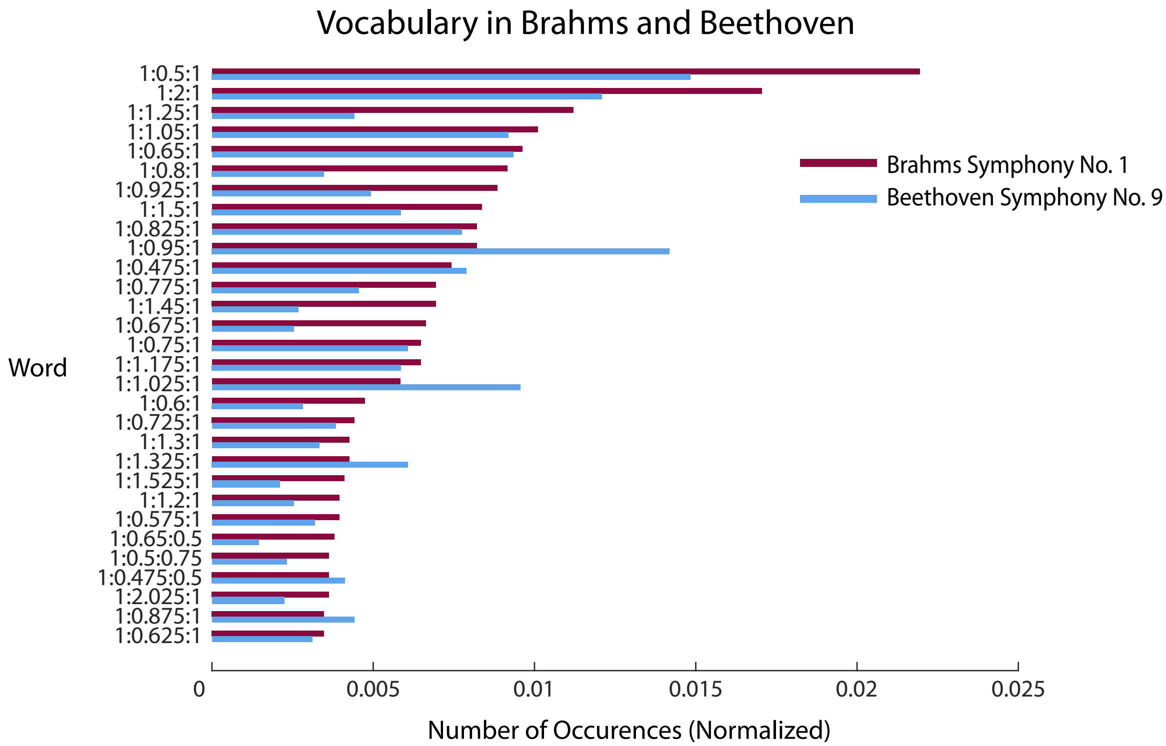


Figure 8.3: Most common words in pieces by Brahms and Beethoven

the alignment of peaks with the theoretical scale reveals that the *intervals* between notes are also discretized at intervals in the theoretical scale.

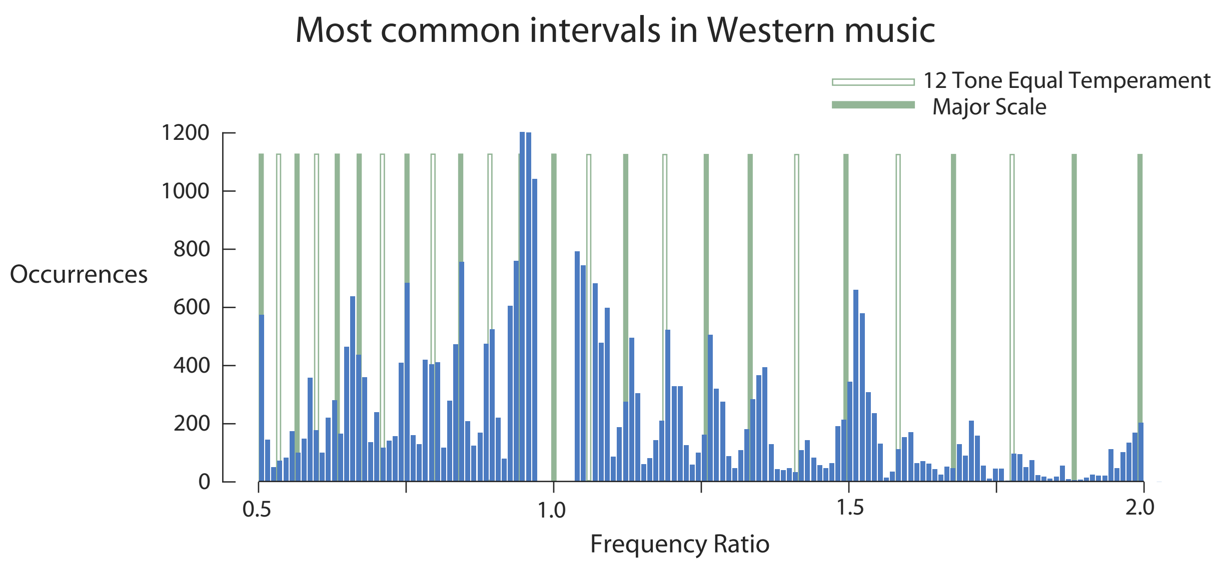


Figure 8.4: A histogram of the most common intervals in Western music (blue) aligns well with the theoretical equal temperament tuning of the major (solid green) and chromatic (outlined green) scales.

Chapter 9

Vocabularies Across Cultures

Finally, the vocabulary method can be used to compare works across musical cultures. This comparison reveals similarities and differences between the intonation used in Western, Chinese and Arabic music.

9.1 Intervals and Words

Figure 9.1 shows the ten most common words from a selection of Western and Arabic music. The tables list the words ranked by their frequencies in the music, and how far they are from the nearest interval in twelve tone equal temperament. Below the tables, the words are notated on a staff. Western words, as well as Arabic, have some variance from the theoretical tunings. This shows that both Western and Arabic musicians use intonation that does not always follow the exact intervals in the equal temperament tuning system.

9.2 Vocabulary of Three Genres

Figure 9.2 shows the vocabulary for large selections of work from Western, Arabic and Chinese culture individually. (Appendix C lists the recordings used for analysis.) The gap in intervals around 1.0 occurs because the program assumes intervals within a small tolerance of the current note are continuations of the note. The vocabularies for the three musical cultures are very distinct, and give insight into the differences between each culture's intonation.

In Western music (Figure 9.2, top), the most common intervals occur at ratios in the major scale and chromatic scale. This is expected because Western music generally moves in intervals defined by twelve-tone equal temperament. The figure shows the equal-temperament ratios in the major scale (solid green bars) and chromatic scale (outlined green bars). The peaks in the observed vocabulary align well with this theoretical framework, yet there is significant variance in intonation where intervals do not fall directly on an equal temperament ratio. For example, the sharp leading-tone phenomenon discussed earlier is also apparent here. This is seen in the prevalence of intervals between 1.89 and 2.0. When the music moves by a major seventh, musicians tend to play the higher note sharp, for instance at a ratio of 1.95 instead of 1.89. This creates extra tension before the note resolves to the octave at a ratio of 2.0, a place of rest and resolution.

Common 3-Note Words in Western and Arabic Music

Western

Rank	Word			Spelling	12TET 2nd note	% Dif	12TET 3rd note	% Dif
1	1	0.89	1	C Bb C	0.8910	0.11	1.0000	0.00
2	1	1.12	1	C D C	1.1225	0.22	1.0000	0.00
3	1	0.75	1	C G C	0.7490	0.13	1.0000	0.00
4	1	1.33	1	C F C	1.3348	0.36	1.0000	0.00
5	1	1.34	1	C F C	1.3348	0.39	1.0000	0.00
6	1	1.25	1	C E C	1.2599	0.79	1.0000	0.00
7	1	0.83	1	C A C	0.8410	**1.32**	1.0000	0.00
8	1	1.14	0.8	C Bb Ab	1.1220	**1.59**	0.7940	0.75
9	1	1.14	1	C D C	1.1220	**1.59**	1.0000	0.00
10	1	0.88	1	C Bb C	0.8910	**1.24**	1.0000	0.00

Arabic

Rank	Word			Spelling	12TET 2nd note	% Dif	12TET 3rd note	% Dif
1	1	0.94	1	C B C	0.9440	0.42	1.0000	0.00
2	1	1.06	1	C Db D	1.0590	0.09	1.0000	0.00
3	1	1.07	1	C Db C	1.0590	**1.03**	1.0000	0.00
4	1	0.95	1	C B C	0.9440	0.63	1.0000	0.00
5	1	0.93	1	C B C	0.9440	**1.49**	1.0000	0.00
6	1	1.05	1	C Db C	1.0590	**0.85**	1.0000	0.00
7	1	0.94	0.89	C B Bb	0.9440	0.42	0.8910	0.11
8	1	0.91	1	C Bb C	0.8910	**2.11**	1.0000	0.00
9	1	0.94	1.06	C B Db	0.9440	0.42	1.0590	0.09
10	1	1.05	0.95	C Db B	1.0590	**0.85**	0.9440	0.63

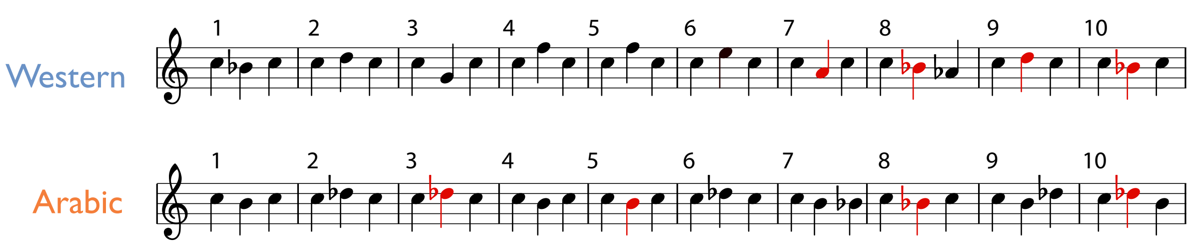


Figure 9.1: The 10 most common words in Western and Arabic music vary are compared to the closest ratios from twelve tone equal temperament (12TET). The percent difference (% Dif) shows how far the actual tuning is from the theoretical tuning. Ratios with greater than 0.8% difference compared to their equal temperament value are in bold red. The staves show the same words in Western notation.

In general the intonation is spread out over the entire range of intervals rather than at twelve discrete ratios. This suggests that intonation in Western music is a dynamic and flexible element of performance, not a strict set of mathematical ratios.

In Chinese music (Figure 9.2, middle), which is based primarily on pentatonic scales, the observed vocabulary also agrees with theory. The various inversions of the pentatonic scale emerge as peaks in the common frequency ratios. The figure shows the theoretical Chinese twelve-tone scale built from perfect fifths (outlined bars) and one common inversion of the pentatonic scale (solid bars). The pitch relationships extracted from Chinese music primarily align with the theoretical scales. There are some less-common ratios at other intervals, but the intonation is primarily focused at discrete intervals. This results in sharp peaks at the pentatonic intervals in the histogram.

In Arabic music (Figure 9.2, bottom), the results are entirely different. Rather than having peaks emerge to reveal a scale, the intervals are concentrated around 1.0. Most of the intervals used are between 0.8 and 1.2. This implies that Arabic music moves primarily in small intervals rather than larger intervals, or “leaps.” There is a small peak at 1.5 (the perfect fifth), but there is no “scale” or set of peaks like in the Western or Chinese cases. It is evident that Arabic musicians use many different ratios between 0.8 and 1.2. In fact, there is a continuum rather than a discrete set of intervals. Categorizing the intervals as only minor and major seconds would discard most of the interesting variation in intonation. In Arabic music, the intonation is dominated by a continuous spectrum of pitch relationships inside a ratio of 1.2.

Another difference between the vocabularies is the use of the leading tone. The leading tone is used frequently in Western and Arabic music to create tension, but infrequently in Chinese music. Both the Western and Arabic histograms have very common intervals just below 1.0 (around 0.95), while Chinese music usually moves down by a larger interval of a whole step (around 0.89).

9.3 Returning to Periodicity

It is now possible to examine how the most common words of each genre are ranked in the methods of Periodicity and Similarity discussed earlier. These measures could be a useful way to quantify the differences between the words used in Western, Chinese and Arabic cultures. Table 9.1 shows Stolzenburg’s Smoothed Logarithmic Periodicity (SLP) score and Gill’s Similarity (SIM) score for the most common intervals in Western, Arabic and Chinese music. Recall that in the Periodicity method, low scores are preferred, while in the similarity method, high scores are preferred. Chinese intervals score better in both methods. This is not surprising, because Chinese music uses larger leaps to intervals in the pentatonic scale, which is built from whole number ratios. The Arabic and Western intervals, which are predominantly small, do not score well in the Periodicity or Similarity methods. This suggests that horizontal motion in music (melody) is not governed by the same interval preferences as vertical combinations of tones (harmony). This result agrees with music theory [15] and will be obvious to a musician. Harmony tends to spread out notes rather than cluster them. Melody tends to move in small steps, mostly avoiding larger leaps. This creates continuity in the melodic line and makes it easier to sing [15].

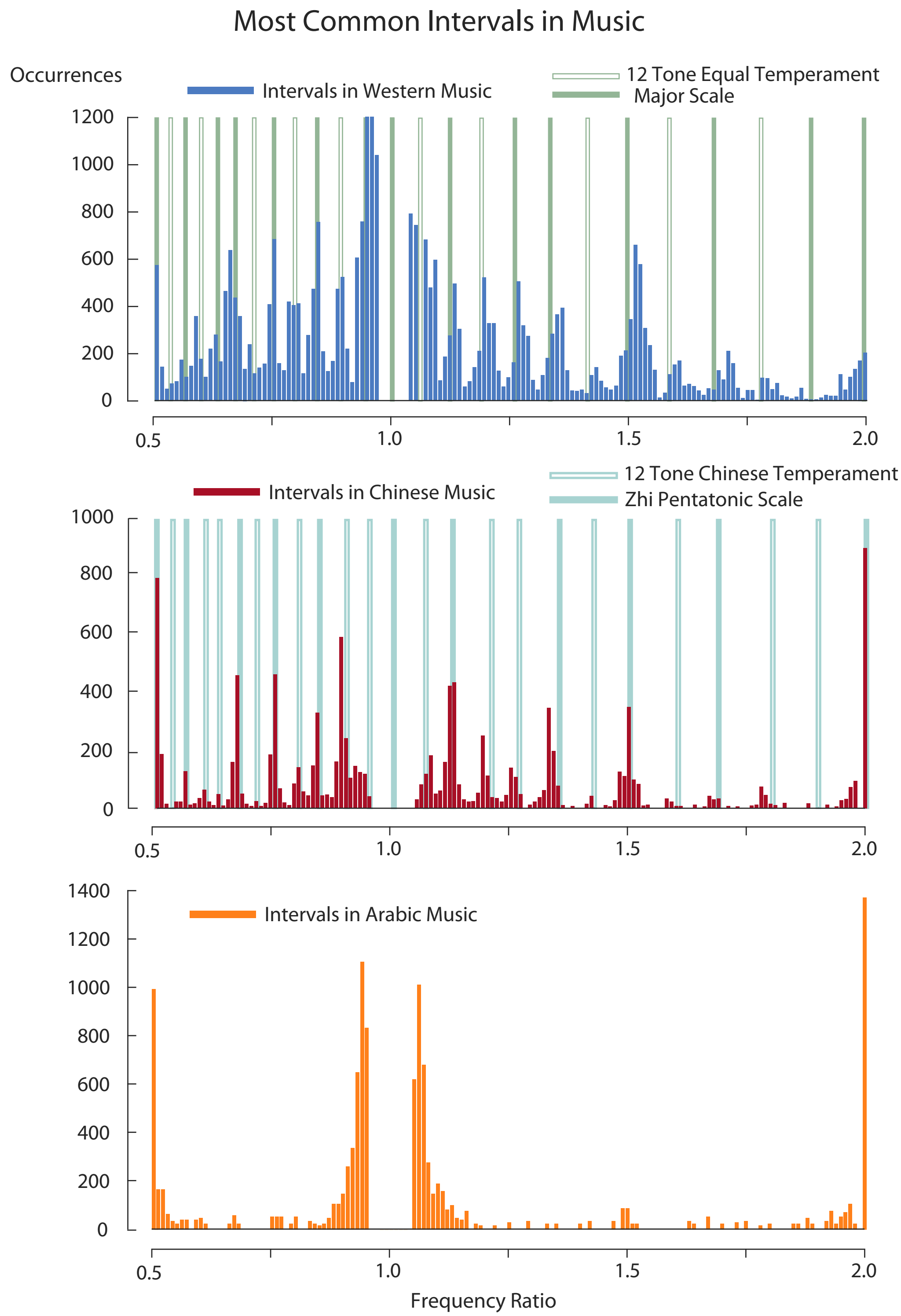


Figure 9.2: The figure shows the relative prevalence of intervals in Western, Arabic and Chinese music.

Table 9.1: The Smoothed Logarithmic Periodicity (SLP) and Similarity (SIM) are calculated for the ten most common intervals for Western, Arabic and Chinese music. In both methods, the common Chinese intervals score better than Western and Arabic intervals.

	Western			Arabic			Chinese		
Rank	Interval	SLP	SIM	Interval	SLP	SIM	Interval	SLP	SIM
1	1.05	13	0.095	0.94	10.5	0.118	0.89	8.5	0.222
2	0.94	10.5	0.118	1.06	11.5	0.111	0.75	2.5	0.500
3	0.95	12	0.100	0.95	12	0.100	0.67	2.5	0.667
4	0.96	15	0.080	1.07	10	0.133	1.13	8.5	0.222
5	1.03	19	0.059	0.93	9	0.143	1.12	6.5	0.222
6	0.93	9	0.143	1.05	13	0.095	1.5	2.5	0.667
7	0.84	5	0.333	0.92	8.5	0.154	1.33	2.5	0.500
8	1.04	15.5	0.077	0.94	10.5	0.118	0.84	5	0.333
9	0.48	7.5	0.120	1.08	9	0.154	1.19	4	0.086
10	0.75	3	0.500	0.91	7.5	0.182	0.9	7	0.200
Average		10.95	0.162		10.15	0.131		4.95	0.362

Figure 9.3 plots the Similarity Score versus Periodicity Score for the common intervals of the three genres. Arabic and Chinese music show some separation with Chinese intervals having the more favorable scores (high Similarity, low Periodicity). The common Western intervals are more spread out. This helps demonstrate the differences in intonation between the three cultures. Chinese music uses the pentatonic scale with consonant, mathematically favorable whole-number pitch ratios. Arabic music, on the other hand, uses small, harmonically-dissonant intervals like leading tones that create tension and form a smooth melodic line. Western music incorporates both types of intervals, using larger leaps at simple ratios to create consonance, as well as small ratios to create tension and smooth melodic lines.

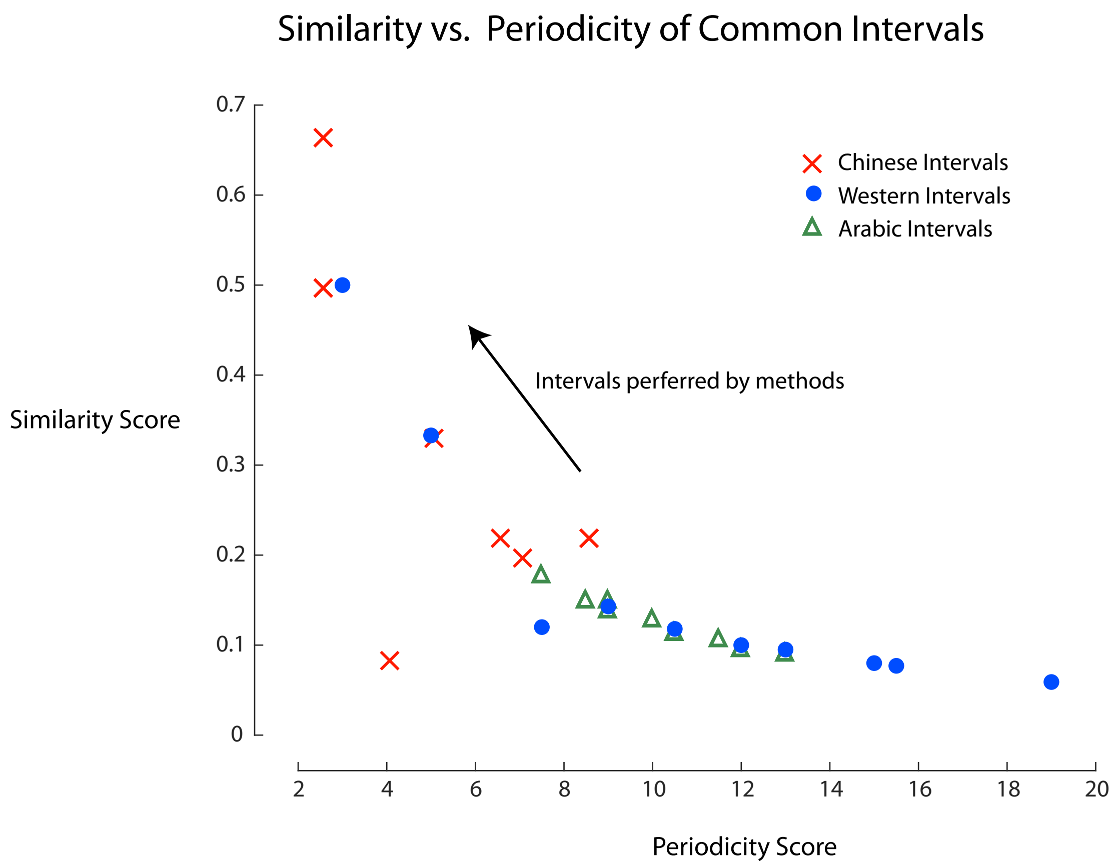


Figure 9.3: Similarity and Periodicity of common words in Western, Arabic and Chinese music. The upper left of the plot represents "favorable" scores while the lower right represents less favorable scores.

Chapter 10

Conclusions

The method of analyzing the contextual intonation in music as words and vocabularies has yielded interesting insights into the intonation of Western, Arabic and Chinese music. Both Western and Arabic musicians use flexible intonation that is based on musical context, not purely on theoretical values of an interval. Still, intonation in Arabic music is fundamentally different from Western and Chinese traditions. While Western and Chinese intonation are based on discrete sets of frequency intervals taken from a theoretical scale, Arabic intonation employs a continuous range of small intervals. Chinese and (to a lesser extent) Western music use large leaps at intervals with favorable harmonic relationships. In Arabic music, these interval motions are replaced by small melodic movements between a flexible collection of pitches. The fundamental differences in intonation between cultures suggest that intonation is not based on universal principles, contrary to the claims of models such as Stolzenburg [4] and Gill and Purves [8]. Some interesting results and implications of this research are discussed below.

10.1 Words as Building Blocks

The building blocks of music are described differently depending on the perspective of analysis (Section 1.1). From the vertical perspective, chords form the building blocks of music. From the modal perspective, scales or discrete pitch sets describe the musician's color palate. The method developed in this research takes the horizontal perspective, analyzing frequency ratios within melodic fragments. From this perspective, short melodic phrases, or "words," form the building blocks of the musical language.

This is certainly not a new or original view of musical analysis. Arabic musicians learn key melodic fragments aurally rather than reading notated music [9]. Arabic musician Sami abu Shumays explains, "I do not spontaneously generate those shorter phrases from a set of rules about how notes go together in the maqam system—I have learned them over years of imitating and repeating, in the same way that I learned the words of spoken language, and each one has its own identity." [9]

An important feature of the method developed in this research is that it retains true frequency ratios from a musician's performance, rather than mapping pitches or ratios to a fixed tuning system. Avoiding discrete measurement systems is necessary and natural for Arabic and other microtonal music. It also provides valuable insight into Western music, where intonation is not as discrete as a pianist might imagine.

10.2 Discrete Versus Continuous Measurement Systems

Western measurement systems may not be appropriate for describing artifacts and traditions from other cultures. In the case of Arabic music, the twelve-tone equal temperament system is inadequate for describing the intonation of the Arabic maqam. Arabic music is an oral tradition in which pitch relationships are arbitrary intervals learned by ear rather than exact mathematical relationships [9]. Even the added convention of "half-sharps" and "half-flats" does not fully describe the microtonal intonation of Arabic music. By mapping maqamat onto the Western system of notation (the staff) and intonation (equal temperament), distinguishing features of the maqamat are lost. For example, the ratio of E𝄳 to the C is distinct for maqam Rast and maqam Bayati. In Western notation, both intervals were mapped to a "neutral third" of E𝄳 to C, with no distinction.

The discretized nature of Western music notation is an issue for Western music, as well as non-Western music. Western musicians are expected to shape intonation away from equal temperament when appropriate. For example, the leading tone may be played sharp, or thirds may be tuned in just intonation [10]. Despite this expectation, the score does not provide instructions for intonation beyond equal temperament.

Using discrete, named notes used to describe intonation certainly has its advantages. Discrete tuning systems are extremely useful because they allow a scale or musical idea to be standardized and easily expressed as a series of notes. There is a fundamental trade-off between the resolution of the measurement system and its simplicity. Sometimes, it makes sense to use a low-resolution discrete sizing system for the sake of simplicity. For instance, drinks at a fast food restaurant come in small, medium and large rather than a continuous spectrum of sizes. Other times, however, a continuous sizing system is necessary. The amount of a chemical fluid needed for a precise experiment is not simply small, medium or large. In this case, a graduated cylinder may be appropriate. In order to obtain the correct results, a high degree of precision—on a continuous measurement scale—is required.

In the case of music, sharps and flats add to the complexity (and resolution) of the basic measurement system containing only the notes A, B, C, D, E, F and G. For Arabic music, the resolution can be extended to quarter-tones with half-sharps and half-flats. However, this is still an inadequate system for describing the maqam. To fully capture the nature of Arabic music, a continuous perspective of intonation is necessary. Even then, intonation will have a temporal dependence on the musical context.

10.2.1 aural versus written traditions

While Western musicians learn music by reading notes from a score, Arabic musicians learn by ear. From this perspective, it is not surprising that Western music performance tends to use discrete and fixed intervals while Arabic music uses much more flexible and continuous intonation. Whether the learning style leads to the tuning style, or the other way around, is hard to guess. The use of instruments with discrete pitches in Western music, like the piano, certainly forces performance in the direction of discrete intonation. The adoption of keyboard instruments and Western notation into Arabic music may threaten the flexible and continuous style of intonation that give maqam music a unique and extremely expressive sound.

10.2.2 uncertainty: a quantum physics analogy

Because it is convenient (for transposition and analysis) to have an equal-temperament system, some theories for maqam music use equal temperament of a large number of tones which give resolution better than even quarter tones (24 tone equal temperament). These models may use 54 (or more) tones in an equal tempered scale [9]. At this point, it seems that the central problem is not truly being addressed. The intonation of a single note in a maqam fluctuates by much more than 1/54 of an octave, and precisely for this reason the 54 tone equal temperament cannot describe a maqam. The maqam is simply more than a discrete set of intervals.

Given the concept of scatter (the statistical variance of each note's tuning), a maqam might be described as a discrete set of intervals with a probability distribution around each interval. However, this suggests that the intonation of the note is performed randomly to create a probability distribution. In fact, the variance in intonation is not random at all. It is very intentional and depends on musical context.

Therefore, there seems to be an uncertainty principle in the intonation of Arabic music. Attempting to pin down a note's exact tuning with more and more precision distances the model from the melodic line, the before and after, the musical phrase where the note actually occurs. By focusing completely on finding the frequency (position) of a single event, the bigger picture of the musical direction (momentum) is completely lost. It is more informative to retain an idea of the melodic line as well as a looser description of the position of a single note. Essentially, the approach in this paper was to retain some musical context in the analysis of intonation by taking a sequence of notes as the elementary building block of music.

10.3 The Origins of Scales

One finding of this study was that Chinese intonation is primarily based on intervals with high degrees of tonal consonance, while Arabic intonation relies on small intervals with less consonance. Western intonation uses a combination of both types of intervals. This raises the question of how scales and intonation patterns are developed Huron [34] suggests that Western scales originated from the desire to maximize tonal consonance, but their continued use throughout history is a result of their familiarity and "cultural inertia" [34]. Many factors can influence the development of musical cultures and traditions. It seems unlikely that any perceptual or mathematical model can predict universal principles of intonation across all musical cultures.

10.3.1 Music Is More Than Pitch

Music is much more than a series of frequencies in time. In Section 4.2.2, a Matlab script created an audio file using the extracted frequency versus time information. The computer-recreation accurately recreated the pitches in the piece, but did not sound like the original piece or even like music at all. Pitch is just one element of music. Without rhythm, tone color, dynamics, articulation and phrasing, the pitch does not really sound like music. Thus, studying intonation only gives us a glimpse at one aspect of how radically distinct the music of other cultures can be. Analysis of rhythm, form, or any number of other aspects of music could provide new insights into the differences between musical traditions.

Even within the realm of pitch, this research suggests that intonation is a rich and complex feature of music. Recreation of the equal temperament notes indicated on a score is just the beginning of intonation in Western music, and hardly even a beginning for Arabic music. A computer program "performing" music from an electronic score would most likely play notes according to their equal temperament values. In advanced notation programs like Sibelius, the specific tuning of a note can be adjusted manually. However, for the computer to make informed choices about intonation would require an understanding of the many factors a musician automatically considers while performing. These include musical context, stylistic conventions for the composer, an understanding of the intonation's aesthetic and most importantly an emotional intention for the musical phrase. A human musical performance is influenced by many factors and shaped by the performer's own personality. While computers may eventually be able to model many of the factors that influence intonation, it is the complex and spontaneous emotional aspect of musical performance that is difficult to recreate, but ever present in live music.

10.4 Further Research

In many ways, this research has only touched the surface of the problem of modeling intonation in non-Western (and Western) musical cultures. The concept of avoiding discrete intonation and focusing on contextual, relative intonation in the form of musical words may be useful in applications beyond this paper. This research used a limited data set for analysis, and could benefit from analysis of a larger body of music. Additionally, certain aspects of the analysis such as the root finding algorithm were built from scratch for simplicity, but more advanced algorithms could be more accurate and efficient. For example, the YIN root finding algorithm [45] may be more accurate than the algorithm used in this research for finding fundamental frequencies in an audio recording. Increased accuracy or increased time resolution could lead to better transcription, finding more words and words longer than three syllables.

This research only analyzed words with three syllables, because the number of extracted reoccurring words with four or more syllables was small, even for large samples of music. However, given large enough sample sizes, analysis of words with more than three syllables could give new insight into the intonation and melodic shapes used in different cultures.

Only Arabic music and some Chinese music were analyzed in this study. There are many musical cultures whose intonation systems are radically different from anything in Western and Arabic music, and which may require new forms of analysis. Also, once vocabularies are extracted, there are many possible ways to analyze and compare them. Creating histograms of words was just one approach for trying to understand the characteristics of intonation in each genre. Principal Component Analysis of the vocabularies was used to graph the vocabularies of pieces from the three cultures examined, but did not show any separation between the cultures. Other approaches to visualizing and comparing vocabularies may yield interesting results.

Appendices

Appendix A

Human Hearing: From Pressure Waves to Neural Signals

This appendix provides an introduction to the way humans hear physical events as sound. It is intended for readers who are not familiar with the process of hearing. The summary highlights the aspects of hearing that are important to understand in the context of harmony perception.

Introduction

Sound is a construct of the mind. Outside of the brain, what we refer to as "sound" is simply energy propagating as a pressure wave. The ears transform the energy into a signal that the brain interprets as sound. The mind pays attention to many attributes of the signal—such as the frequency (oscillations per second of the pressure wave), amplitude (magnitude or strength of the pressure wave), and envelope (how the sound's volume begins and decays over time).

There are constantly many sources of sound around us, and we hear them as distinct entities. The music playing in the background, the voice of a person, the buzz of an air conditioner, and the chirping of birds sound as separated and unrelated to us as they are in real life. Incredibly, everything we hear is actually captured as a single waveform superimposing all of the signals from around us (Figure A.1). The mind must take this single stream of information and separate it into the distinct sources from the world. This decoupling and parsing of an audio stream is a fascinating ability which present day computers can only mimic at a rudimentary level [e.g. 46].

Physical waves are transformed into neural signals in six stages:

1. Sound moves through air as pressure waves
2. The pinna collects sound waves
3. The tympanic membrane is vibrated by pressure changes
4. Mechanical motion is amplified and transmitted through the inner ear
5. The cochlear membrane is stimulated according to specific frequencies
6. Hair cells in the cochlea send a neural signal to the brain

This document will describe each of these stages. Figure A.2 shows the parts of the outer, middle and inner ear. The six stages progress through the diagram from left to right.

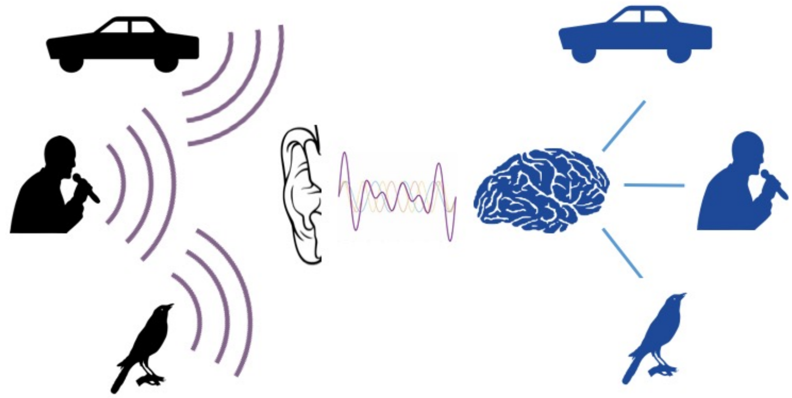


Figure A.1: The brain recreates real-world sources of sound from a single stream of information.

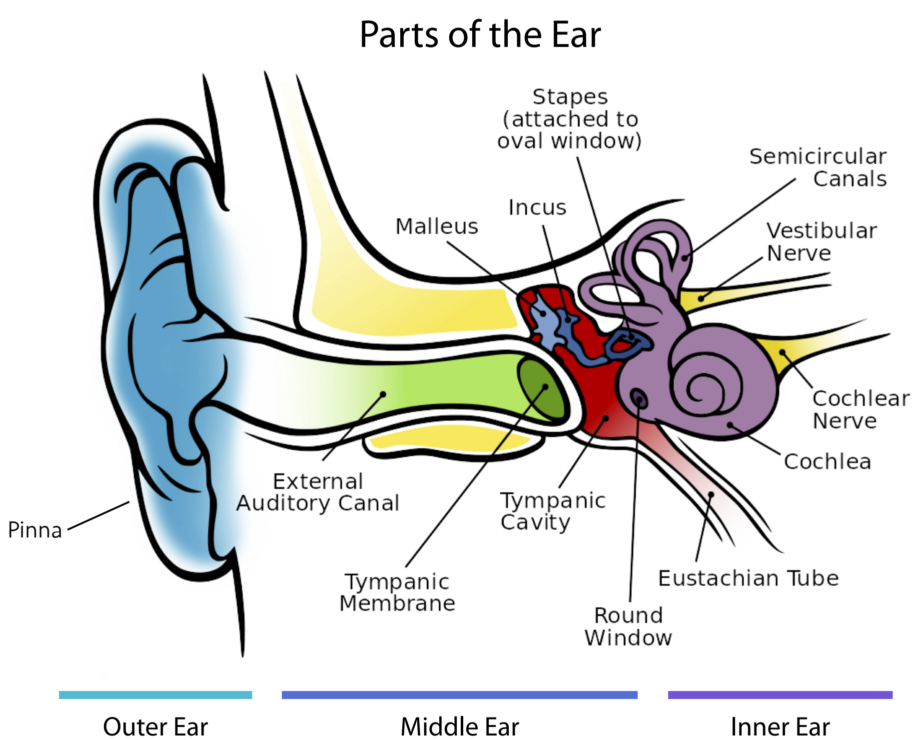


Figure A.2: Parts of the outer, middle and inner ear. Adapted from [47].

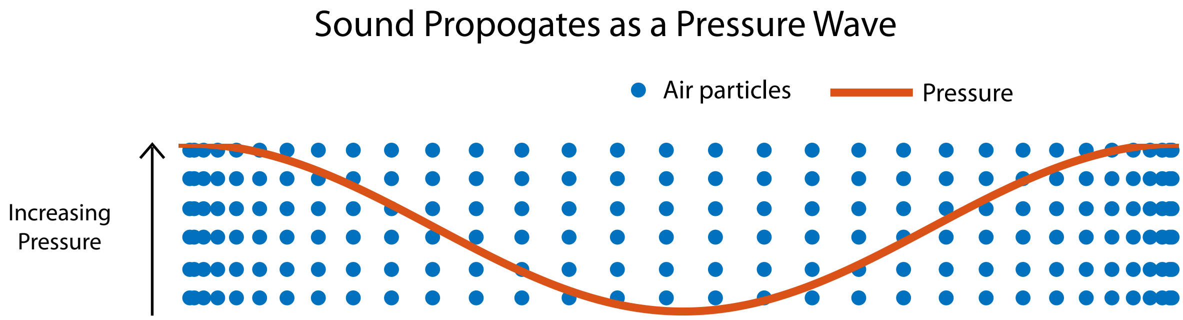


Figure A.3: In the real world, sound exists as pressure waves. The dots represent particles moving closer together (high pressure) and further apart (low pressure).

Sound in the World

Stage 1: Sound moves through air as pressure waves

In the world, 'sound' exists as longitudinal pressure waves propagating through a physical medium. Humans interpret the frequency of the wave as pitch, and the amplitude as loudness. Figure A.3 shows how particles move as a sound wave propagates.

Sound in the Outer Ear

Stage 2: The pinna collects sound waves

The part of the ear you can see, the outer ear, is called the pinna. The pinna is essentially a cone that captures sound waves and sends them into the middle ear. The shape of the pinna helps to capture sounds from all around us, but also helps us distinguish which direction a sound came from. People can hear pitches from 20 Hz to 20kHz, but the pinna is especially effective for sounds in the frequency range of the human voice.

Sound in the Middle ear

In the middle ear, pressure waves in air are transformed into mechanical motion. Use Figure A.2 as a visual reference for the parts of the middle ear.

Stage 3: The tympanic membrane is vibrated by pressure waves

Pressure fluctuations in the external auditory canal reach the tympanic membrane (the eardrum). The eardrum is a membrane in the middle ear which vibrates in response to pressure waves.

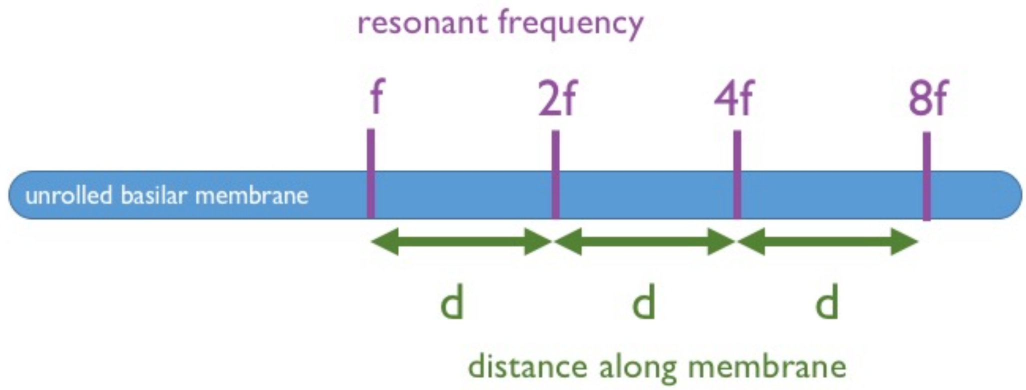


Figure A.4: On the basilar membrane, frequency scales logarithmically with distance.

Stage 4: Mechanical motion is amplified and transmitted

On the inner side of the eardrum are the three smallest bones in the human body: the malleus, incus and stapes. These bones act as levers, mechanically transmitting the vibrations of the eardrum to the end of the stapes. The stapes is attached to another membrane called the oval window. The mechanical advantage of the three-bone lever system comes from their tiny size: the cross-sectional area of the stapes is much smaller than the area of the eardrum, and transmits twenty-two times the pressure. Thus, the middle ear transmits and amplifies the mechanical vibrations picked up from the environment.

Sound in the Inner Ear

In the inner ear, the mechanical oscillations of the stapes are parsed into specific frequencies and sent to the brain as electrical signals.

Stage 5: The cochlear membrane is stimulated according to specific frequencies

When the stapes pushes on the oval window, fluid inside a spiral-shaped cavity called the cochlea is compressed. The cochlea takes the time-domain signal and transforms it into the frequency domain, creating an amplitude versus frequency signal. This may sound exactly like a Fourier transform, because it is. The mechanism is easier to understand if you take the cochlea (usually rolled up in a spiral) and unroll it. The frequency-sensitive strip is called the basilar membrane. Each position along the length of the basilar membrane has a specific resonant frequency. The resonant frequency is logarithmically proportional to position, so that moving a constant distance along the basilar membrane will multiply the resonant frequency by a constant factor. For example, in Figure A.4 the distance ‘d’ corresponds to a doubling of the resonant frequency on the basilar membrane. The ability of the ear to analyze sound in the frequency domain is essential to understanding human perception of music.

Stage 6: Hair cells in the cochlea send a neural signal to the brain

When fluid in the cochlea resonates, the part of the cochlear membrane with the correct resonant frequency will resonate sympathetically. The vibrations stimulate nearby specialized cells on the cochlea called "hair cells," which send an electrical signal to the brain. The amplitude of the vibrations will determine the strength of the neural signal. Hair cells in various locations on the cochlea are constantly firing, creating a flow of information to the brain. After this, the brain interprets these signals as sounds.

Overview

The process of hearing is fascinating and useful to study. For every noise humans hear, pressure waves from the surroundings are captured, transformed into mechanical vibrations, then transformed into frequency-dependent electrical signals. The mechanisms by which the brain interprets these signals are complex and not well understood, and are beyond the scope of this report. However, understanding the physical and physiological nature of hearing gives insights into human perception of sound. For example, the logarithmic organization of basilar membrane explains the logarithmic perception of pitch, and the exact characteristics of its stimulation can be useful for studying the quality of harmony Helmholtz calls roughness. The mechanical solutions life has developed for accurately translating and transforming a wide range of frequencies are some of the most incredible biological systems. Understanding the process of hearing is invaluable for studying music perception.

Appendix B

Whole Number Ratio Algorithm

The following algorithm presented by Stolzenburg [33] approximates a rational number x as a ratio of integers a/b, with maximum relative deviation (error) d. It is written in pseudocode.

$x_{min} = (1 - d)x;$
$x_m ax = (1 + d)x;$
$a_l/b_l = floor(x)/1;$
$a_r/b_r = (floor(x) + 1)/1;$
loop
$a_m/b_m = (a_l + a_r)/(b_l + b_r);$
if x_{min} a_m/b_m x_{max} then
    return $a_m/b_m;$
else if $x_{max} < am/bm$ then
$k = floor((a_r - x_{max}b_r)/(x_{max}b_l - a_l));$
$a_m/b_m = (a_r + ka_r + kb_l);$
    $a_r/b_r = a_m/b_m$
else if $a_m/b_m < x_{min}$ then
    $k = floor((x_{min}b_l - a_l)/(a_r - x_{min}b_r));$
    $a_m/b_m = (a_l + ka_r)/(b_l + kb_r);$
    $a_l/b_l = a_m/b_m$
end if
end loop

Appendix C

Works Cited: Music Recordings

Aal, Bashir Abdel, perf. *SAUDI ARABIA Bashir Abdel Al: The Art of the Arabian Flute - The Nay*. Naxos Digital Services US Inc., 1998. CD.

Anthology of World Music: China. Rounder, 1998. CD.

"Arabic Maqam World." *Arabic Maqam World*. N.p., n.d. Web. 2 Feb. 2017.

Bach, Jo Sebastian. *The 6 Unaccompanied Cello Suites Complete*. Perf. Yo-Yo Ma. Yo-Yo Ma. Masterworks, 2010. CD.

Bach, Johann Sebastian. *Bach: Cello Suites*. Perf. Mstislav Rostropovich. Mstislav Rostropovich. EMI Classics, 1955. CD.

Bach, Johann Sebastian. *Bach, J.S.: Cello Suites Nos. 1-6 (Maisky)*. Perf. Mischa Maisky. Naxos Digital Services US Inc., 1994. CD.

Bach, Johann Sebastian. *Brandenburg Concertos*. Perf. Max Pommer. New Bach Collegium Musicum Leipzig. Naxos Digital Services US Inc., n.d. CD.

Bach, Johann Sebastian. *Hilary Hahn Plays Bach*. Perf. Hilary Hahn. Sony, 1997. CD.

Bach, Johann Sebastian. *The World's Greatest Masterpieces: Johann Sebastian Bach*. Perf. Marga Scheurich-Henschel. Southwest-Studio Orchestra. Madacy, 1999. CD.

Beethoven, Ludwig Van, Ann Murray, René Pape, Lucia Popp, Anthony Rolfe Johnson, and Klaus Tennstedt. *BEETHOVEN, L. Van: Symphony No. 9, "Choral" (Popp, Murray, Rolfe-Johnson, Pape, London Philharmonic Choir and Orchestra, Tennstedt)*. Naxos Digital Services Ltd., 2009. CD.

Beethoven, Ludwig Van. *BEETHOVEN, L. Van: Piano Concerto No. 1*. Naxos Digital Services US Inc., 2014. CD.

Beethoven, Ludwig Van. *BEETHOVEN: Symphony No. 5 and 6*. Naxos Digital Services US Inc., 2002. CD.

Brahms, Johannes. *SYMPHONIE N °I EN UT MINEUR OP*

Appendix D

Listening Tracks

The tracks listed below are included as multimedia content to accomapny this paper. Tracks 00-05 were created by the author. All other audio files are from Maqamworld [17].

Tones, Intervals and Scales

[00] Minor Second Interval (piano)
[01] Perfect Fifth Interval (piano)
[02] C Major Scale (piano)
[03] C Minor Scale (piano)
[04] Pure (sine) tone, 440 Hz
[05] Violin 'A,' 440 Hz
[06] Maqam Rast on C (oud)
[07] Maqam Bayati on C (buzuq)
[08] Maqam Sikah on E half flat (violin)

Arabic Music Selections

[10] Shaheen, Simon. Maqam Nahawand. "Taqsim on oud."
[11] Bey, Sami. Maqam Ajam. "Samai Ajam."
[12] Maqam Bayati: Fairuz. "Ya Rayt Minhoun." *Fairuz Sings Philemon.* Voix De L'Orient, 1993. CD.
[13] Makam Sikah: Abraham Salman. Taqsim Qanun on Rast. NADA Records, 1997. CD.
[14] Maqam Rast: Darwish, Sayyed. "Muwashah ya Eudhaib." *Orientalia collection: muwashah & songs.* Usek, 1993. CD.

Bibliography

[1] Inbal Shapira Lots and Lewi Stone. Perception of musical consonance and dissonance: an outcome of neural synchronization. *Journal of The Royal Society Interface*, 5(29): 1429–1434, 2008. ISSN 1742-5689. doi: 10.1098/rsif.2008.0143. URL `http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2008.0143`.

[2] Lienhard. Ancient Chinese Bells, 2002. URL `http://www.uh.edu/engines/epi1676.htm`.

[3] Rudolf Rasch. Tuning and temperament. In *The Cambridge History of Western Music Theory*, chapter 7. Cambridge University Press, 2006.

[4] Frieder Stolzenburg. Harmony perception by periodicity and granularity detection. In *Proceedings of 12th International Conference on Music Perception and Cognition and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, volume 29, pages 958–959, 2012.

[5] Norman D Cook and Takashi X Fujisawa. The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon. *Empirical Musicology Review*, 1(2):106–126, 2006.

[6] William Hutchinson and Leon Knopoff. *The acoustic component of western consonance.* Swets & Zeitlinger, Amsterdam, 1978.

[7] Akio Kameoka and Mamoru Kuriyagawa. Consonance Theory Part II : Consonance of Complex Tones. *Journal of the Acoustical Society of America*, 1969.

[8] Kamraan Z Gill and Dale Purves. A Biological Rationale for Musical Scales. *PLoS ONE*, 4(12), 2009. doi: 10.1371/journal.pone.0008144.

[9] Sami Abu Shumays. The Fuzzy Boundaries of Intonation in Maqam: Cognitive and Linguistic Approaches. 2009.

[10] Ross W. Duffin. *How equal temperament ruined harmony (and why you should care)*. 2007. ISBN 0393075648.

[11] Frieder Stolzenburg. A PERIODICITY-BASED APPROACH ON HARMONY PERCEPTION INCLUDING NON-WESTERN SCALES. 2010.

[12] Hermann von Helmholtz and A. J. Ellis. *On the sensations of tone as a physiological basis for the theory of music.* New York: Dover Publications, 1954.

[13] Phil N. Johnson-Laird, Olivia E. Kang, and Yuan Chang Leong. On Musical Dissonance. *Music Perception: An Interdisciplinary Journal*, 30(1):19–35, 2012. ISSN 07307829. doi: 10.1525/mp.2012.30.1.19.

[14] David Huron. Tonal Consonance versus Tonal Fusion in Polyphonic Sonorities. *Source: Music Perception An Interdisciplinary Journal*, 9(2):135–154, 1991.

[15] Steven G. (Steven Geoffrey) Laitz. *The complete musician : an integrated approach to tonal theory, analysis, and listening*. Oxford University Press, 2012. ISBN 0199742782.

[16] David Huron. Interval-Class Content in Equally Tempered Pitch-Class Sets: Common Scales Exhibit Optimum Tonal Consonance. *Music Perception*, 11(3):289–305, 1994.

[17] Maqam World, 2007. URL `maqamworld.com`.

[18] Saed Muhssin. Some Notes on harmony, 2007. URL `http://www.saedmuhssin.com/harmony.html`.

[19] Joseph Needheam, Wang Ling, and Lu Gwei-Djen. Science and civilisation in China. *Physics and physical technology*, 4:563, 1971. ISSN 0717-6163. doi: 10.1007/s13398-014-0173-7.2.

[20] Roger Kamien. *Music: An Appreciation*. McGraw Hill, New York, 2008. ISBN 978-0-07-340134-8.

[21] Richard Parncutt and Graham Hair. Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies. *Journal of Interdisciplinary Music Studies*, 5(2):119–166, 2011. doi: 10.4407/jims.2011.11.002.

[22] Keith Mashinter. Calculating Sensory Dissonance: Some Discrepancies Arising from the Models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff. *Empirical Musicology Review*, 2006.

[23] Johann Joseph Fux and Alfred Mann. *The study of counterpoint : from Johann Joseph Fux's Gradus ad parnassum*. W.W. Norton, 1971. ISBN 0393002772.

[24] Christopher Hasty. Segmentation and Process in Post-Tonal Music. *Music Theory Spectrum*, 3(Spring):54–73, 1981.

[25] Richard Parncutt. Commentary on "Calculating Sensory Dissonance: Some Discrepancies Arising from the Models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff". *Empirical Musicology Review*, 1(2):65–84, 2006. ISSN 1559-5749.

[26] R. Plomp and W. J. M. Levelt. Tonal Consonance and Critical Bandwidth. *The Journal of the Acoustical Society of America*, 38(4):548–560, 1965. ISSN 00014966. doi: http://dx.doi.org/10.1121/1.1909741.

[27] James K. Wright and Albert S. Bregman. Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, 2(1):63–92, 1987.

[28] Mark Ballora. *The Science of Music*. Cognella Academic Publishing, 2003.

[29] Akio Kameoka and Mamoru Kuriyagawa. Consonance Theory, Part I. *Journal of the Acoustical Society of America*, 1969.

[30] Josh H McDermott and Andrew J Oxenham. Music perception, pitch, and the auditory system. *Current opinion in neurobiology*, 18(4):452–63, aug 2008. ISSN 0959-4388. doi: 10.1016/j.conb.2008.09.005.

[31] Josh H McDermott, Alan F Schultz, Eduardo A Undurraga, and Ricardo A Godoy. Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature Publishing Group*, 535(7613):547–550, 2016. ISSN 0028-0836. doi: 10.1038/nature18635.

[32] Marion Cousineau, Josh H McDermott, and Isabelle Peretz. The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences of the United States of America*, 109(48):19858–63, nov 2012. ISSN 1091-6490. doi: 10.1073/pnas.1207989109.

[33] Frieder Stolzenburg. Harmony Perception by Periodicity Detection. *Journal of Mathematics and Music*, 9(March 2013):37–41, 2015. doi: 10.1080/17459737.2015.1033024.

[34] David Brian Huron. *Sweet Anticipation: Music and the Psychology of Expectation - David Brian Huron - Google Books*. MIT Press, Cambridge, MA, 2006. ISBN 0262083450.

[35] G Langner, M Sams, P Heil, and H Schulze. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoecephalography. *J. Comp. Physiol. A.*, 181(6):665–676, 1997.

[36] Greg Hunter. Arabic Tone System. URL `https://dubsahara.wordpress.com/maqam/overview/`.

[37] Simon Shaheen. *Turath*. CMP Records, 1992.

[38] Barış Bozkurt, Ozan Yarman, M. Kemal Karaosmanoğlu, and Can Akkoç. Weighing Diverse Theoretical Models on Turkish Maqam Music Against Pitch Measurements: A Comparison of Peaks Automatically Derived from Frequency Histograms with Proposed Scale Tones. *Journal of New Music Research*, 38(1):45–70, 2009. doi: 10.1080/09298210903147673. URL `http://www.tandfonline.com/action/journalInformation?journalCode=nnmr20http://dx.doi.org/10.1080/09298210903147673`.

[39] Ozan Yarman, Edilen Ortak, and Bir Izgara. A Comparative Evaluation of Pitch Notations in Turkish Makam Music. *Journal of Interdisciplinary Music Studies*, 1(2):43–61, 2007.

[40] Albert Cohen. A History of Consonance and Disonance. *Music & Letters*, 71(2):226–228, 1990. URL `http://www.jstor.org/stable/736439`.

[41] Dalia Cohen and Ruth Katz. *Palestinian Arab Music: A Maqam Tradition in Practice*. University of Chicago Press, 2004. ISBN 9780226112992.

[42] Sami Abu Shumays. Maqamlessons.com, 2013. URL `http://maqamlessons.com/`.

[43] Zhang Tasen. Spring On A Moonlit River. In *Phases Of The Moon: Traditional Chinese Music*. CBS Records, 1981.

[44] Johann Sebastian Bach, Christiane Jaccottet, Dubravka Tomsic, Georg Egger, Gunter Holler, and Marga Scheurich-Henschel. *The World's Greatest Masterpieces: Johann Sebastian Bach*. Madacy, 1999.

[45] Alain De, Cheveignéhideki Kawahara, Alain De Cheveigné, and Hideki Kawahara. YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(43), 2002. doi: 10.1121/1.1458024 .

[46] Erich Schroger, Alexandra Bendixen, Susan L Denham, Robert W Mill, @bullet Tamás, Ohm @bullet, and István Winkler. Predictive Regularity Representations in Violation Detection and Auditory Stream Segregation: From Conceptual to Computational Models. *Brain Topography*, (Special issue: Mismatch Negativity), 2013. doi: 10.1007/s10548-013-0334-6.

[47] Brockmann Chittka. A diagram of the anatomy of the human ear, feb 2009. URL `https://en.wikipedia.org/wiki/File:Anatomy_of_the_Human_Ear_en.svg`.

Academic Vita

Sam Lapp

sam.m.lapp@gmail.com
samlapp.com

Objective	Seeking a job where I can use my technical skills and hands-on project experience to help push the boundaries of human innovation and exploration		
Education	B.S. in Engineering Science The Pennsylvania State University, Schreyer Honors College		May 2017
Experience	**Front-end Developer**		Summer 2016
	Human Computer Interaction Department	Carnegie Mellon University	
	Explored computational analogies for research papers Built user interfaces to explore research analogies		
	Internet of Things Programmer		Summer 2015
	Wipro Technologies	Bangalore, India	
	Developed a flexible, generic parser for processing data Built a graphical interface to control the parser		
	Virtual Reality Remote Driving Researcher		Summer 2015
	OPEN Design Lab	Penn State University	
	Explored possibility of virtual reality for driving vehicles Integrated Oculus Rift virtual reality with remote driving systems		
Projects	Thesis in Engineering Science and Music		Spring 2016-present
	Examining tuning systems of non-Western culture Evaluating computational models for predicting perception of dissonance		
	Helping Hand: Biomechanics Design Project		Spring 2014
	Designed a glove to restore function to customer's injured hand Optimized mechanical and aesthetic properties for real-world use Chosen as best design by the customer		
	LightMood: Smart-home Design Project		Spring 2014
	Smart-home lighting system to treat Seasonal Affective Disorder Awarded "Most Innovative Design" at Design Showcase		
Leadership	Founder, Vice-President	Music Therapy Club	2014-2015
	President	Jazz Club	2014-2015
Honors	Matthew George Workman Scholarship in ESM Vernon H. Neubert Dynamics Award in ESM The Evan Pugh Scholar Junior Award The Provost Award for Academic Excellence		

Courses	Fluid Mechanics	Skills	C++
	Thermodynamics & Heat Transfer		Java
	Mechanics: Statics, Dynamics & SOM		MatLab
	Electronic Properties of Materials		HTML
	Wave, Quantum & Particle Physics		JavaScript