

# Seminar Notes: The Mathematics of Music

Zhou Fan

September 1, 2010



# Contents

<b>Preface</b>	<b>iii</b>
<b>1 Understanding Musical Sound</b>	<b>1</b>
1.1 Sound, the human ear, and the sinusoidal wave . . . . .	1
1.1.1 Sound waves and musical notation . . . . .	1
1.1.2 The human ear and the sinusoidal wave . . . . .	2
1.2 Timbre and resonance harmonics . . . . .	3
1.3 Fourier series and frequency analysis . . . . .	5
1.4 Further topics . . . . .	8
1.4.1 Percussion instruments . . . . .	8
1.4.2 Psychoacoustic phenomena . . . . .	8
<b>2 Chords, Scales, and the Fundamentals of Western Music</b>	<b>9</b>
2.1 Consonance and dissonance . . . . .	9
2.2 Scales and temperament . . . . .	10
2.2.1 Continued fractions and the twelve-tone scale . . . . .	10
2.2.2 Just intonation, Euclidean lattices, and the fundamental domain . . . . .	13
2.2.3 Meantone, irregular, and equal temperament . . . . .	15
2.3 The Javanese gamelan . . . . .	17
<b>3 Musical Structure and Theory</b>	<b>19</b>
3.1 Musical symmetries and transformations . . . . .	19
3.1.1 Groups, generators, and relations . . . . .	19
3.1.2 Intervallic transformations . . . . .	21
3.1.3 Transformations between major and minor triads . . . . .	21
3.1.4 Transformations of motifs and melodic passages . . . . .	23
3.1.5 Symmetries in Debussy's <i>Feux d'artifice</i> . . . . .	24
3.2 The geometry of chords . . . . .	26
3.2.1 Tori of ordered chords . . . . .	26
3.2.2 Orbifolds of unordered chords . . . . .	28
3.2.3 Chromatically descending chord progressions . . . . .	30
<b>Bibliographic Notes</b>	<b>33</b>



# Preface

These notes were created as a set of lecture notes for a three-day seminar on the mathematics of music, intended to introduce topics in this area to the interested reader without assuming strong prerequisites in mathematics or music theory. They are not intended to be a comprehensive and technically rigorous exposition, and an attempt is made to introduce the concepts in an elementary way and to avoid technical details when possible and appropriate. Much has been written on the connections between mathematics and music; this summary draws on a number of such sources, referenced to in the bibliographic notes.



# Lecture 1

## Understanding Musical Sound

### 1.1 Sound, the human ear, and the sinusoidal wave

#### 1.1.1 Sound waves and musical notation

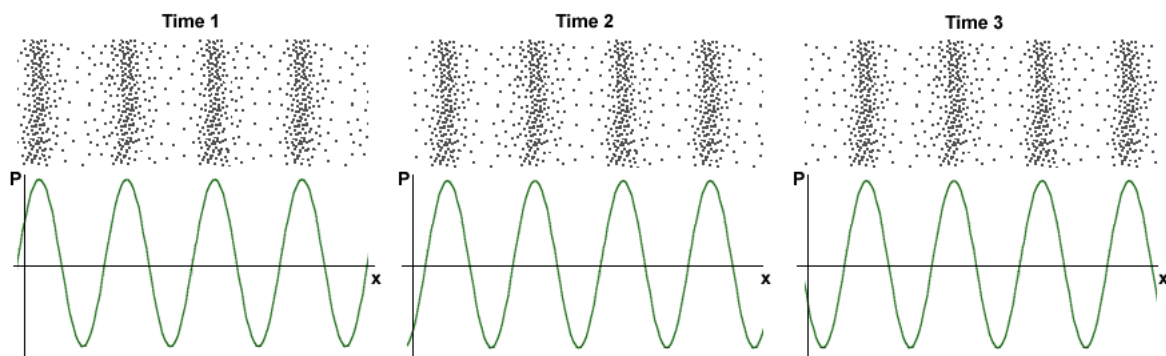
Music is organized sound, and it is from this standpoint that we begin our study. In the world of Western music, notation has been developed to describe music in a very precise way. Consider, for instance, the following lines of music:



A musician who understands this notation would recognize this as the melody of “Twinkle, Twinkle Little Star” (in the key of A major), and the symbols above encode, to great precision, various aspects of the sounds that make up this melody.

Sound travels through air, a gaseous medium of molecules that are constantly in motion and exerting forces against each other. These forces create *air pressure*—when a large quantity of air molecules are compressed in a small space, air pressure is high, and when air molecules are spread out at a great distance from one another, air pressure is low. Physically, sound is a *wave* of alternating high and low air pressure that travels through time and space.

To visualize a sound wave, let us consider a chronological sequence of pictures of the compression of air molecules for a sound wave moving in the rightward direction:

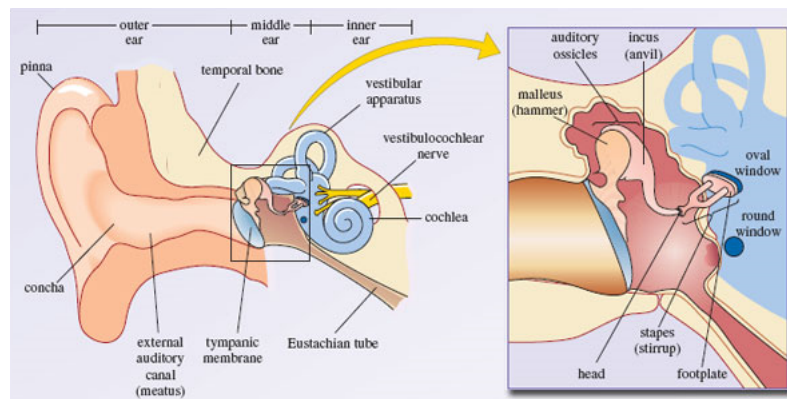


One should note that each individual air molecule moves only a very small distance and does not travel far in the rightward direction, but rather that a rippling effect causes the wave to propagate rightwards. Accompanying these pictures are graphs of the air pressure at each point in space, where we shift our units so that the horizontal axis indicates normal atmospheric air pressure without the perturbation of sound. The shape of the functions in these graphs is called the *waveform*. We have drawn the wave to be *periodic*, meaning that it is a repetition of the same pattern across space, and indeed, a musical pitch corresponds to a periodic sound wave. The distance in space between two repetitions of the pattern is called the *wavelength* of the wave or the *period* of the waveform. If we focus on how the wave moves through a particular point in space, the number of repetitions of the pattern that move through that point in a single second is called the *frequency* of the wave, measured in Hertz (Hz). For waves moving at the same speed, a longer wavelength implies a longer time for a single repetition of the pattern to move across any given point, and hence it implies a lower frequency. Indeed, we have the relationship  $\text{frequency} \times \text{wavelength} = \text{speed of wave}$ . For this particular waveform, known as the sinusoidal wave, there is a clear maximum and minimum amount of air pressure achieved by the wave, and the difference between this and normal atmospheric pressure is called the *amplitude* of the wave.

Returning to our musical example of “Twinkle, Twinkle Little Star”, we may translate the musical notation into statements about the physical waves that make up the sound. The first and last notes of the example are “A above middle C”, which refers to the pitch of the note, a musical term for the frequency of the sound wave. This A above middle C has a frequency of 440Hz, and it is common to see the marking “A440” on digital tuners to represent this pitch. Every musical pitch corresponds to a particular frequency—for instance, the third note of E in the musical example is a frequency of 660Hz—with higher pitches corresponding to higher frequencies and lower pitches corresponding to lower frequencies. The marking *mp* is a dynamic marking to indicate the volume of the sound, which corresponds to the amplitude of the sound wave. Finally, the stems and heads of the notes indicate the duration of each note, which is the amount of time a particular sound wave is present before it changes to the next wave.

### 1.1.2 The human ear and the sinusoidal wave

How does the ear perceive sound and distinguish pitch? A picture of the ear is shown below.

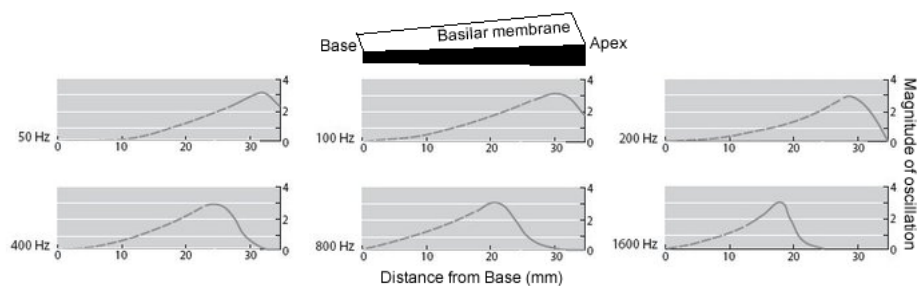


The outer ear, composed of the pinna, concha, and auditory canal, is responsible for amplifying sound and plays an important role in our ability to perceive the location of a sound source. Sound waves passing through the auditory canal create oscillations of the tympanic membrane that separates the outer and middle ear. This membrane is attached to three bones known as the ossicles, which in turn are attached to a liquid-filled tube called the *cochlea* in the inner ear. The ossicles transfer the oscillations of the tympanic membrane to waves in the liquid of the cochlea, and it is the cochlea that plays the main role in pitch perception.





The above image shows a picture of a cross-section of the uncoiled cochlea. Liquid known as perilymph is contained in the regions shaded gray, and waves in this liquid induce oscillations of the *basilar membrane*. The basilar membrane is thin near the base of the cochlea and thick near the apex of the cochlea, and high frequency waves induce oscillations of the thin portion of the basilar membrane while low frequency waves induce oscillations of the thick portion. Some graphs of the magnitude of oscillation along the basilar membrane for different frequencies of sound are shown below:



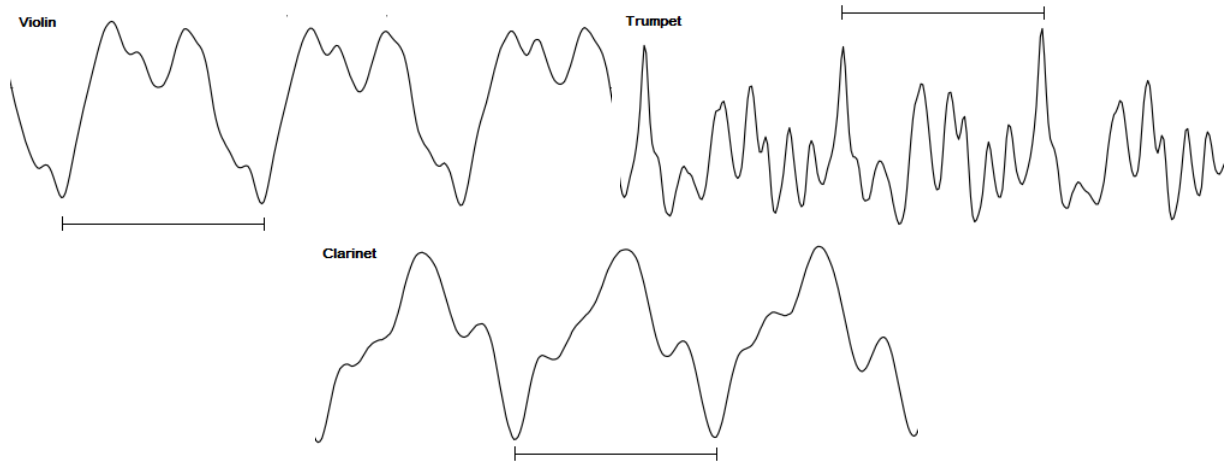
Signals from these oscillations are conveyed by auditory nerves to the brain, and the brain is able to distinguish pitch based on which region of the basilar membrane is oscillating.

In the previous section, we drew a sinusoidal wave as an example of a sound wave, and we defined frequency, wavelength, and amplitude based on this picture. The *sinusoidal wave* is the waveform given by the function  $f(x) = A \sin(\nu x + \theta)$ . The amplitude and frequency are captured in the parameters  $A$  and  $\nu$  respectively, and  $\theta$  corresponds to the phase of the wave, which changes over time to generate wave motion. The sinusoidal wave corresponds to the idea of a “pure” pitch as perceived by the ear and brain, and the reason lies in the physics behind the oscillation of the basilar membrane. (To a first order approximation, we may consider each point of the basilar membrane as an ideal spring. The physical equation governing the displacement  $x$  of an ideal spring is given by  $F = -kx$ , where  $F$  is the force of the spring and  $k$  is the spring constant. The value of this spring constant  $k$  is higher at thicker points of the membrane and lower at thinner points, and it determines the resonant frequency of oscillation at that point. A pure sinusoidal wave in the perilymph inside the cochlea provides a sinusoidal driving force that induces strong oscillations at points of the basilar membrane with resonant frequency close to the frequency of this driving force, and the strength of the oscillation is weaker as the resonant frequency is farther from the frequency of the driving force. This creates the graphs of oscillation magnitude displayed above, which the brain perceives as pure pitches [3].)

## 1.2 Timbre and resonance harmonics

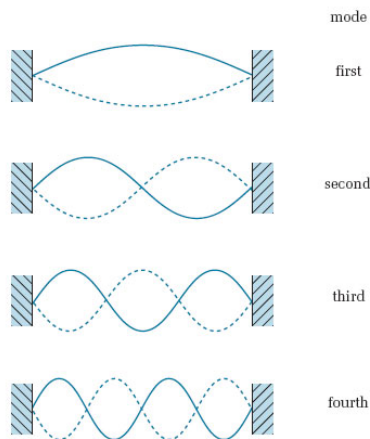
We have attributed the pitch and volume of a musical sound to the frequency and amplitude of the sound wave. This does not, however, explain everything that we hear. Suppose that a violin and a flute play the same pitch at the same volume level, or that an opera singer sings the same note on an “oo” and an “ee” vowel. It is easy to hear the differences between these sounds, but how are they manifested in the physical sound waves?

The answer lies in that, even though our brain perceives the sinusoidal wave as a pure musical pitch, most musical sounds that we hear are not pure pitches. Graphs of the waveforms for a violin, trumpet, and clarinet all playing the note A at 440Hz are shown below:



We observe that each waveform is periodic with the same period, but the shapes of the waveforms are different, and none of them are sinusoidal. Indeed, each wave corresponds to a pure sinusoidal wave of frequency 440Hz, with various waves of higher frequency added to it. (In vocal singing, one can manipulate and change the strengths of these higher frequency sounds by shaping the mouth and throat, a technique that is used in Mongolian and Tibetan overtone singing to allow a single person to sing a melody on top of a bass drone.) Let us, in this section, explore the causes of these superimposed frequencies from a physical perspective for the string and wind instruments, and defer a more mathematical treatment to the subsequent section.

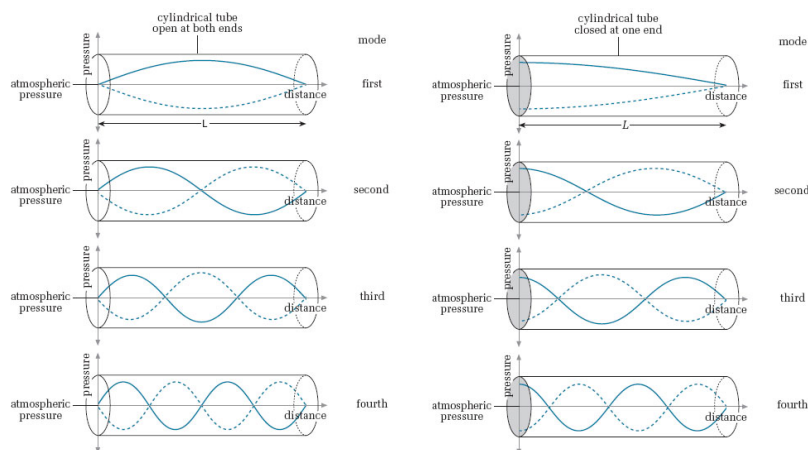
An important group of instruments create sound using a vibrating string. These include the violin, viola, cello, and stringed bass of the Western orchestra, the versatile piano and guitar, as well as a number of traditional Chinese instruments including the zheng, qin, pipa, and erhu. In each of these instruments, the string is stretched taut between two fixed ends, and the middle of the string is allowed to vibrate. This allows the string to produce certain patterns of standing waves known as *modes*.



Letting  $L$  be the length of the string, the first mode of oscillation corresponds to a wavelength of twice this length, or  $2L$ . The  $n^{\text{th}}$  mode corresponds to a wavelength of  $\frac{2L}{n}$ . Translated into frequencies, this implies that if the frequency corresponding to the first mode of oscillation is  $f_1$ , then the frequency of the  $n^{\text{th}}$  mode is  $nf_1$ . When a string vibrates to produce a sound, the standing wave of the string is a superposition of any number of these modes, creating a sound wave that is a superposition of pure sinusoidal waves of varying frequencies.

A similar treatment can be applied to wind instruments. These include instruments of the woodwind family, such as flutes, clarinets, oboes, saxophones, and bassoons, as well as the brass family, such as trumpets, horns, trombones, and tubas. The model of the wind instrument is also a crude but useful approximation for the vocal tract of the human voice. Sound in these instruments is shaped by the reverberation of sound

waves within a tube, which we will model as a cylinder with open or closed ends. Again, the tube allows for modes of oscillation corresponding to standing sound waves inside the tube of varying frequencies. An open end of the tube must correspond to a point of zero pressure difference from normal air pressure, similar to how a fixed end of a string cannot be displaced. If the tube has a closed end, the endpoint of the tube must be a point of maximal pressure difference from normal air pressure. (This is due to the condition that the displacement of the air molecules at the closed end must be fixed at zero and to a relationship between the pressure and displacement.) Hence, the modes of oscillation for tubes with two open ends, such as the flute, or tubes with one closed end, such as the clarinet, are those shown below:



In the case of two open ends, the wavelength of the first mode of oscillation is twice the length of the tube, and the frequency of the  $n^{\text{th}}$  mode is  $n$  times the frequency of the first mode. In the case of one closed end, the wavelength of the first mode of oscillation is four times the length of the tube, and the frequency of the  $n^{\text{th}}$  mode is  $2n - 1$  times the frequency of the first mode.

The frequency of the first mode of oscillation for a string or tube is known as its *fundamental* frequency, or the first *harmonic*, and the frequency that is  $n$  times the fundamental frequency is the  $n^{\text{th}}$  harmonic. That the frequencies of the harmonics are integer multiples of the fundamental frequency has important consequences on scales and temperament in Western music, which we will explore in Lecture 2.

### 1.3 Fourier series and frequency analysis

There are some questions that the preceding discussion has left unresolved. Why must a string oscillate as one of the modes previously depicted, or as a superposition of such modes? Supposing that it does oscillate in this way, how can we examine quantitatively the strength of oscillation of each mode? And if the resulting sound wave is not a pure sinusoidal wave, how does our ear process this sound? To answer these questions, we turn to the mathematical theory of Fourier series.

Recall the notion of a *vector* in  $\mathbb{R}^2$ : It is an ordered pair of real numbers,  $(x, y)$ , that represents a point in the plane. We can take two vectors and add them by adding their coordinates individually, e.g.,  $(2, 3) + (4, 5) = (2 + 4, 3 + 5) = (6, 8)$ . We can multiply a vector by a number and get a new vector, e.g.,  $-2(2, 3) = (-4, -6)$ . The collection of all vectors in  $\mathbb{R}^2$ , with these operations of addition and multiplication by a real number, forms a *vector space*.

We have a third operation for vectors in  $\mathbb{R}^2$  known as the dot product, or *inner product*:

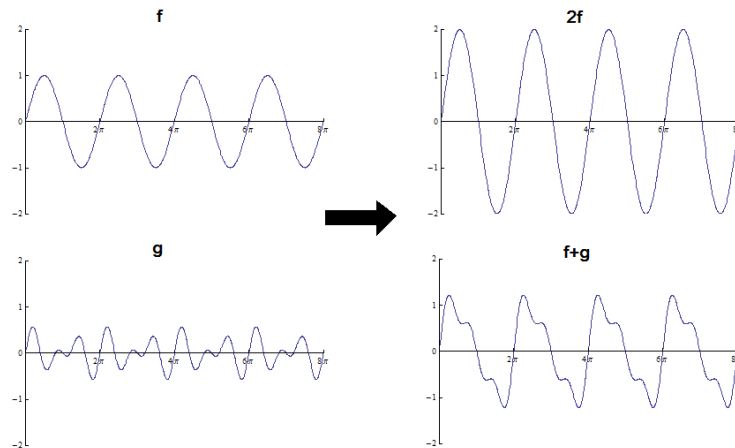
$$\langle (x_1, y_1), (x_2, y_2) \rangle = x_1x_2 + y_1y_2.$$

The inner product of two vectors is a number. Taking an inner product of a vector with itself always returns a nonnegative number, which is the square of the length of that vector. Two vectors are perpendicular if their inner product is zero. A vector space that also has this type of an inner product operation is called an *inner product space*.

In  $\mathbb{R}^2$ , we may define two special vectors  $\mathbf{i} = (1, 0)$  and  $\mathbf{j} = (0, 1)$ , and write any vector as a sum of multiples of these vectors. For instance,  $(2, 3) = 2\mathbf{i} + 3\mathbf{j}$ . For any vector  $\mathbf{v}$ , we have that  $\mathbf{v} = \langle \mathbf{v}, \mathbf{i} \rangle \mathbf{i} + \langle \mathbf{v}, \mathbf{j} \rangle \mathbf{j}$ ; that is, when writing  $\mathbf{v}$  as a sum of multiples of  $\mathbf{i}$  and  $\mathbf{j}$ , the coefficients are given by the inner products of  $\mathbf{v}$  with  $\mathbf{i}$  and  $\mathbf{j}$ . This property is true because  $\mathbf{i}$  and  $\mathbf{j}$  each have length equal to 1 and are perpendicular to each other, and the same property would be true if we take any two other perpendicular vectors of length 1. For instance, if we let  $\mathbf{a} = \frac{1}{\sqrt{2}}(1, 1)$  and  $\mathbf{b} = \frac{1}{\sqrt{2}}(-1, 1)$ , then  $\mathbf{a}$  and  $\mathbf{b}$  have length 1 and are perpendicular, and so  $\mathbf{v} = \langle \mathbf{v}, \mathbf{a} \rangle \mathbf{a} + \langle \mathbf{v}, \mathbf{b} \rangle \mathbf{b}$  for any vector  $\mathbf{v}$ . Any pair of such vectors forms an *orthonormal basis* for  $\mathbb{R}^2$ .

A similar structure holds for  $\mathbb{R}^3$ . Here, a vector is an ordered triple of real numbers. The inner product is defined by  $\langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle = x_1x_2 + y_1y_2 + z_1z_2$ . The vectors  $\mathbf{i} = (1, 0, 0)$ ,  $\mathbf{j} = (0, 1, 0)$ , and  $\mathbf{k} = (0, 0, 1)$  form an orthonormal basis for  $\mathbb{R}^3$ , and we have  $\mathbf{v} = \langle \mathbf{v}, \mathbf{i} \rangle \mathbf{i} + \langle \mathbf{v}, \mathbf{j} \rangle \mathbf{j} + \langle \mathbf{v}, \mathbf{k} \rangle \mathbf{k}$  for any vector  $\mathbf{v}$ . The number of vectors in an orthonormal basis for a vector space is the *dimension* of that space, so  $\mathbb{R}^2$  is a 2-dimensional space and  $\mathbb{R}^3$  is a 3-dimensional space.

Consider any periodic waveform with some period, say  $2\pi$ , given by a function  $f(x)$ . The periodicity condition requires that  $f(0) = f(2\pi) = f(4\pi) = \dots$ . Suppose also that the average value of  $f(x)$  over all  $x$  is 0. We may multiply  $f(x)$  by a number to obtain a new function, or take two such functions  $f_1(x)$  and  $f_2(x)$  and add them to obtain a new function. This new function, in either case, still gives a waveform with period  $2\pi$  and average value 0. Multiplying by a number corresponds to multiplying the amplitude of the wave by that number while keeping the shape of the waveform the same. Adding two waveforms corresponds to the wave obtained by superimposing them on top of each other.



Hence the space of all waveforms with period  $2\pi$  and average value 0 forms a vector space. That is, the “vectors” of this vector space are functions. The problem of decomposing a waveform into a superposition of sinusoidal waves would be solved if the sinusoidal waves, as vectors in this vector space, form an orthonormal basis under some inner product operation.

The inner product operation we seek is the following:

$$\langle f, g \rangle = \frac{1}{\pi} \int_0^{2\pi} f(x)g(x)dx.$$

The claim is that, under this inner product, the functions  $\sin x, \cos x, \sin 2x, \cos 2x, \sin 3x, \cos 3x, \dots$  form an orthonormal basis. It is an exercise in calculus to verify that these functions are orthonormal, i.e. the inner product of each function with itself is 1 and the inner product of any two different functions in this list is 0. That any periodic function can be expressed as a sum of multiples of these functions is given by the following theorem.

**Theorem 1.3.1.** *Let  $f$  be a smooth function on  $\mathbb{R}$  with period  $2\pi$  and average value 0. Let  $a_1(x) = \sin x$ ,  $a_2(x) = \sin 2x$ ,  $a_3(x) = \sin 3x$ , etc., and let  $b_1(x) = \cos x$ ,  $b_2(x) = \cos 2x$ ,  $b_3(x) = \cos 3x$ , etc. Then*

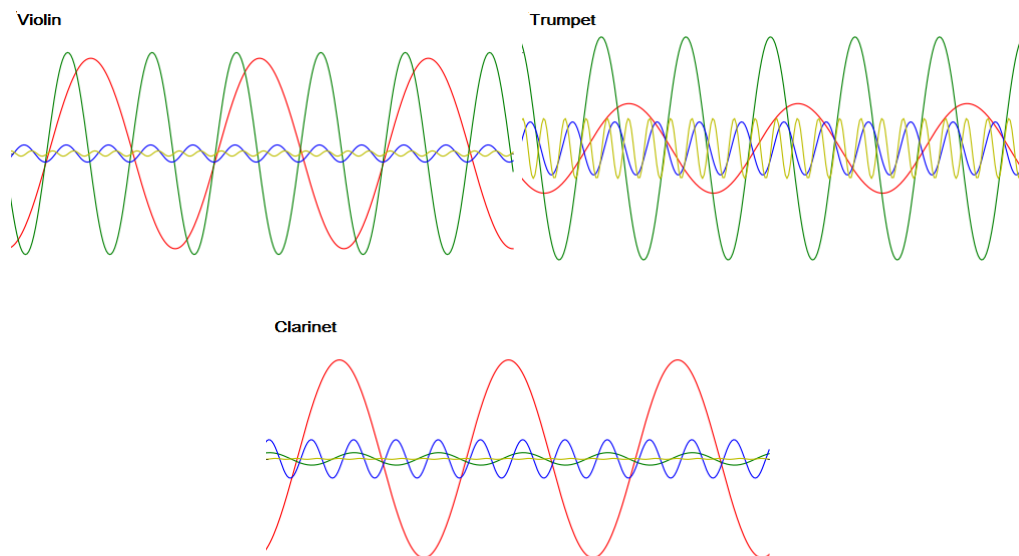
$$f = \sum_{i=1}^{\infty} (\langle f, a_i \rangle a_i + \langle f, b_i \rangle b_i),$$

where the inner product is given by  $\langle f, g \rangle = \frac{1}{\pi} \int_0^{2\pi} f(x)g(x)dx$ .

(Let us note for the sake of correctness that by “smooth”, we mean  $f$  is continuously differentiable, and the equality in the above conclusion holds pointwise. The reader unfamiliar with these terms need not worry about their precise definitions.) In other words, any function with period  $2\pi$  and average value 0 can be written as a sum of sinusoidal functions with periods  $2\pi, \frac{2\pi}{2}, \frac{2\pi}{3}, \frac{2\pi}{4}$ , and so on. The orthonormal basis, in this case, is given by the collection of all sine and cosine functions with periods  $\frac{2\pi}{n}$  for integers  $n$ . We note that, unlike the cases of  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , the orthonormal basis for this space of periodic waveforms has an infinite number of vectors, and hence the space of periodic waveforms is an infinite-dimensional inner product space.

This theorem allows us to answer the questions posed at the start of this section. The modes of vibration of a vibrating string of length  $L$  correspond to sinusoidal waveforms with wavelengths  $2L, \frac{2L}{2}, \frac{2L}{3}$ , etc. The vibrating string does not “know” to vibrate as a superposition of these modes. The possible motions of the string are governed by physical laws relating to the tension of the string, under the boundary conditions that the ends of the string cannot move. These laws require that the displacement of the points of the string create some waveform with periodicity  $2L$  and that the average displacement over all points of the string is zero. (To be more precise, the displacement of the string at point  $x$  and time  $t$  must be of the form  $f(x+ct) - f(ct-x)$ , where  $f$  is a periodic function of periodicity twice the length of the string and mean value zero.) The string can oscillate with any such waveform, and the fact that this oscillation can be written as a superposition of the modes of vibration is a purely mathematical consequence of the above theorem. Similarly, air in a tube does not “know” to vibrate as a superposition of modes, but the fact that any wave that satisfies the physical constraints of the tube can be written as such a superposition is a mathematical truth.

Theorem 1.3.1 also gives us the method for decomposing any sound wave into components corresponding to the pure sinusoidal frequencies. Revisiting the waveforms of the violin, trumpet, flute, and oboe, we may use this theorem to determine the amplitude of oscillation (i.e. volume) of each of the first four harmonic pitches of these instruments.



We see that the waveform of the clarinet consists primarily of the sinusoidal wave at the fundamental frequency, whereas the waveform of the violin also includes a very strong component at the second harmonic. The waveform of the trumpet includes a second harmonic that is in fact louder than the fundamental pitch, as well as strong third and fourth harmonics.

Finally, when any waveform that is a superposition of sine waves of various harmonic frequencies reaches the cochlea inside the human ear, the different frequency components create oscillations in different portions of the basilar membrane, and we perceive a mix of pitches. So the human ear performs physically what we have done mathematically, to separate the sound into its harmonic pitch components.

## 1.4 Further topics

### 1.4.1 Percussion instruments

In Section 1.2, we discussed the modes of oscillation of string and wind instruments, the latter also serving as an approximation for the human voice. A family of instruments that we did not discuss are the percussion instruments, which can be classified into the two groups of idiophones, such as xylophones and marimbas, and membranophones, such as timpani and drums. The physical law that governed the types of oscillations possible for a vibrating string or a column of air inside a tube was the one-dimensional wave equation, which resulted in periodic waveforms. However, this law does not govern the types of oscillations possible for these percussion instruments, and in fact the waveforms of these percussion instruments are non-periodic.

The space of all waveforms permissible for an idiophone or membranophone is still an infinite-dimensional vector space with an orthonormal basis, but the functions that form this basis are non-periodic and more complicated than simple sinusoidal waves. (The basis functions for the space of waveforms of idiophones involve the hyperbolic sine and cosine functions, and those for membranophones involve Bessel functions of the first kind.) Nonetheless, these basis functions are regular in some way, and it is possible to define a “frequency” of oscillation for each basis function that corresponds to our perceived notion of pitch. For an ideal idiophone, the first four harmonic frequencies are 1.000, 2.758, 5.405, and 8.934 times the fundamental frequency, respectively. For an ideal membranophone, the first four harmonic frequencies are 1.000, 1.593, 2.136, and 2.653 times the fundamental frequency, respectively. We note that this differs from the phenomenon observed in string and wind instruments, where the harmonic frequencies are integer multiples of the fundamental frequency [6].

### 1.4.2 Psychoacoustic phenomena

A number of interesting phenomena regarding our perception of pitch and sound cannot be explained by the discussion of the human auditory system provided in this lecture; let us touch on two of them here. The first phenomenon is that of sum and difference tones. When the ear hears two simultaneous pure pitches corresponding to sinusoidal waves of frequencies  $f_1$  and  $f_2$ , oftentimes weaker notes can also be heard at the frequencies  $f_1 - f_2$ ,  $f_1 + f_2$ ,  $2f_1 - f_2$ ,  $2f_2 - f_1$ , and other combinations of these two pitches. The perception of these pitches is fundamentally different from the perception of the harmonic frequencies of the human voice or a physical instrument, because these pitches are not components of the waveform of the physical sound wave whereas the harmonic frequencies of a real instrument are.

A second phenomenon is that of the missing fundamental. When the ear hears a waveform composed of a number of resonance harmonics, it usually attributes the primary pitch of that waveform to be the pitch at the fundamental frequency of the waveform. However, if the sound is digitally synthesized so that the fundamental frequency is removed but the other resonance harmonics remain, the ear will still perceive the primary pitch to be the missing fundamental frequency. This missing fundamental frequency is perceived even when the second, third, and fourth resonance harmonics are removed, as long as a sufficient number of higher resonance harmonics are still present.

Some of these auditory phenomena are due to oversimplifications of our model for the cochlea and basilar membrane. For instance, an explanation for the perception of sum and difference tones lies in nonlinear oscillations of the basilar membrane in the cochlea that are not governed by simple harmonic motion. Some of the phenomena, including the missing fundamental, are attributable to psychological causes in the way our brain processes combinations of pitch and sound. The detailed study of the physiology and psychology of how we perceive sound is a field known as *psychoacoustics*.

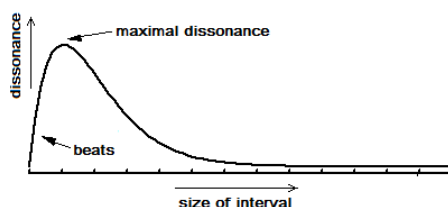
## Lecture 2

# Chords, Scales, and the Fundamentals of Western Music

### 2.1 Consonance and dissonance

Harmony has been an essential element of Western music for over a millenium, and most music today contains combinations of pitches sounding simultaneously. Certain combinations of pitches—for instance, the intervals of the octave and the perfect fifth—sound inherently pleasing, or “consonant”, to the ear, and these intervals were the first to appear in Western harmony. Other combinations of pitches—for instance, the tritone and the major seventh—sound rough and jarring, or “dissonant”. The exact characterization of consonance and dissonance from the music theoretic perspective has changed over time; the intervals of the major third and major sixth that were considered dissonant in Medieval and early Renaissance music would be treated as consonant by the time of Bach and Mozart. Consonance and dissonance from the musical perspective is also dependent on musical context, as the perfect fourth is sometimes treated as a dissonant interval depending on its musical function.

Nevertheless, it is still reasonable to examine consonance and dissonance from a purely “sensory” perspective, irrespective of historical context and musical function, based on how pleasing a combination of sounds is to the ear. Mathematicians since the time of Pythagoras have noticed that the most consonant sounds correspond to small integer ratios of frequencies. For instance, the ratio of frequencies for two pitches an octave apart is  $2 : 1$ , and the ratio of frequencies for two pitches a fifth apart is  $3 : 2$ . They offered as a reason for our perception of consonance the inherent simplicity of small integer ratios, and this was the prevalent explanation for a long time. In 1965, Plomp and Levelt put this to the test by conducting a study in which they played various combinations of two pure pitches (having perfect sinusoidal waveforms) to people with no musical training, and asked them to rate how “pleasant” the intervals sounded [10]. The aggregation of results from this study is summarized in the following Plomp-Levelt curve.

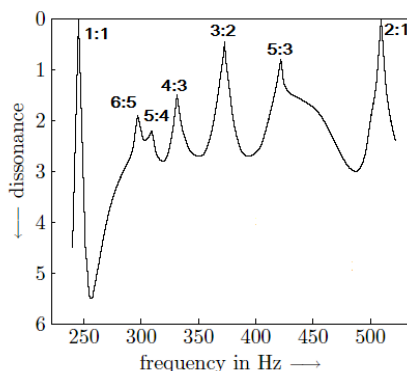


The results of Plomp and Levelt’s study show that pitches very close in frequency are perceived not as dissonant, but rather as a single pitch with the presence of beats. Such beats can be heard, for instance, when two instruments of a musical ensemble play the same note slightly out of tune from each other. Sensory dissonance increases quickly as the two pitches are moved farther apart so that they are perceived as a musical interval, and then decreases and reaches a stable level when the interval becomes large. What may be surprising, to a trained musician, is that the interval of a tritone is roughly as dissonant as the perfect fifth,

and the interval of a major seventh is roughly as dissonant as the perfect octave. This seems contradictory to our actual experience and to any music theoretic notions of dissonance.

The reason for the discrepancy lies in that Plomp and Levelt chose to use pure sinusoidal pitches in their study, whereas the musical sounds that we hear are (usually) produced by real musical instruments. As we explored in Lecture 1, these instruments produce more complex waveforms, consisting of a number of superimposed frequencies. The reason that a major seventh sounds dissonant on Western instruments is not that the interval of the major seventh is fundamentally dissonant, but rather that the second harmonic of the lower note (which is a perfect octave above its fundamental pitch) is very close to the fundamental pitch of the higher note, and that a number of higher resonance harmonics of both sounds clash as well. Similarly, the third harmonic of the lower note clashes against the second harmonic of the higher note in a tritone. We may test this theory of dissonance using a digital synthesizer, by creating an “instrument” in which the resonance harmonics of each note are slightly higher than exact integer multiples of the fundamental pitch. On this digital instrument, perfect octaves and fifths sound very dissonant, but an interval slightly larger than the perfect octave that matches the harmonics of the instrumental sound becomes consonant.

We can thus modify the Plomp and Levelt dissonance curve in the following way: For any pitch, consider the sound with that pitch as its fundamental frequency, plus a number of natural resonance harmonics at integer multiples of that frequency. The dissonance of two notes can then be expressed as the sum of the dissonances of all pairs of pitches from the two sounds. The resulting graph is shown below, and this time, the graph corresponds to our intuition with high consonance (low dissonance) at the intervals of the fifth ( $3 : 2$ ) and octave ( $2 : 1$ ).



Since the primary pitch-producing instruments of Western music are string and wind instruments, which have resonance harmonics at precisely the integer multiples of the fundamental frequency, this curve is a good model for the concept of consonance and dissonance in the Western musical tradition.

This brings us full circle to the explanation that small integer ratios of frequencies produce consonant sounds. This is true for string and wind instruments because the strongest resonant frequencies of these instruments are small integer multiples of the fundamental pitch. Indeed, the relative minima of the above graph occur when the frequency ratios of the two pitches are expressible as ratios of small integers, and we may take these ratios to be the definition of the most consonant intervals: the perfect octave is a ratio of  $2 : 1$ , the perfect fifth  $3 : 2$ , the just major sixth  $5 : 3$ , the perfect fourth  $4 : 3$ , the just major third  $5 : 4$ , and the just minor third  $6 : 5$ .

## 2.2 Scales and temperament

### 2.2.1 Continued fractions and the twelve-tone scale

We discussed in Lecture 1 that a pitch is a frequency of oscillation for a sound wave. This frequency can take on any positive real value, and hence there are an infinite number of possible pitches for sound. Many physical instruments, however, can play only a finite number of pitches. Let us take, as an example, the standard modern piano, which can play 88 different pitches. Grouping pitches into their pitch classes, the



piano can play notes from 12 different pitch classes, corresponding to the seven white keys and five black keys within any octave.

If we were to design the piano, how can we choose a finite set of pitches that the piano can play? We would like the piano to be able to play consonant chords, and hence let us use consonance and dissonance as a starting point. Suppose that we add a pitch with a frequency of 1 unit to the piano. (1 unit can correspond to 440Hz, for instance, so that the pitch is A440.) Since the octave is the most consonant interval, a reasonable thing to do would be to add all pitches that are octaves away from the starting pitch in some specified range. The frequency ratio of two pitches in an octave is 2 : 1, so we would have the pitches with frequencies of 2, 4, 8, 16, etc., as well as the pitches with frequencies of  $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$ , etc. If we started with the pitch A440, we would now have the note A in all octaves.

The second most consonant interval, after the octave, is the perfect fifth at a frequency ratio of 3 : 2. Let us add the criterion that if we have pitches in a particular pitch class, then we also need the pitches from the pitch class that is a perfect fifth above it. Starting with the pitch of frequency 1, this would give us a pitch of frequency  $\frac{3}{2}$  and all pitches in that pitch class: 3, 6, 12, etc., and  $\frac{3}{4}$ ,  $\frac{3}{8}$ ,  $\frac{3}{16}$ , etc. We would then need to add the pitch that is a perfect fifth above this, or  $\frac{3}{2} \times \frac{3}{2} = \frac{9}{4}$ , and pitches in this pitch class as well:  $\frac{9}{2}$ , 9, 18, etc., and  $\frac{9}{8}$ ,  $\frac{9}{16}$ , etc. We may continue this process infinitely many times. The problem we encounter is that we will never return to our starting pitch class. That is, multiplying by  $\frac{3}{2}$  repeatedly will never bring us to a power of 2, because the only integer solution to the equation  $(\frac{3}{2})^m = 2^n$  is  $m = n = 0$ . Each time we add a pitch class to our piano, we add one new pitch in the octave between frequencies of 1 and 2 units:  $1, \frac{3}{2}, \frac{9}{8}, \frac{27}{16}, \frac{81}{64}, \frac{243}{128}, \dots$ . This process will never end, so we will end up with an infinite number of pitches to each octave of the piano.

This seems paradoxical when we consider the actual piano, because each pitch class on the actual piano has the pitch class that is a fifth above it, and proceeding upwards by the interval of a fifth 12 times brings us back to our original pitch class. For instance, starting at the note F and proceeding upwards by fifths 12 times, we obtain the sequence F-C-G-D-A-E-B-F $\sharp$ -C $\sharp$ -G $\sharp$ -D $\sharp$ -A $\sharp$ -F, returning to where we started. What is occurring is that the piano is making an approximation in this last step of the cycle. Let us return to the sequence of pitches  $1, \frac{3}{2}, \frac{9}{8}, \frac{27}{16}, \frac{81}{64}, \frac{243}{128}, \dots$  in the octave between frequencies of 1 and 2 units obtained by progressing upwards in fifths, and write this sequence in decimal format: 1, 1.5, 1.125, 1.688, 1.266, 1.898, 1.424, 1.068, 1.602, 1.201, 1.802, 1.352, 1.014.... The 13<sup>th</sup> pitch in this sequence, 1.014, is extremely close to the first pitch of 1, so close that we can make the choice of not including this 13<sup>th</sup> pitch class and, instead, using the starting pitch class of 1 as the pitch class a fifth above the 12<sup>th</sup> pitch. This is what is done on the modern piano, and hence the piano has 12 pitches to each octave. To be musically precise regarding pitch classes, the pitch class that is a perfect fifth above A $\sharp$  is E $\sharp$ , not F, and the sequence proceeds as A $\sharp$ -E $\sharp$ -B $\sharp$ -F $\sharp$ -.... The frequency ratio of E $\sharp$  to F, B $\sharp$  to C, or F $\sharp$  to G is 1.014, and this ratio is called a *Pythagorean comma*. Similarly, progressing downwards in fifths from F, we have the pitch classes F-B $\flat$ -E $\flat$ -A $\flat$ -D $\flat$ -G $\flat$ -C $\flat$ -F $\flat$ ..., and the frequency ratio of, say, G $\sharp$  to A $\flat$  or B to C $\flat$  is also 1.014. Pitches that differ by the Pythagorean comma of 1.014 are called *enharmonic*, and on most modern instruments, the distinction between enharmonic pitches is lost.

Looking at the sequence of pitches 1, 1.5, 1.125, 1.688... again, we observe possible places, other than after the 12<sup>th</sup> pitch, where we may choose to stop. The sixth pitch of 1.898, when brought down an octave, is 0.949, which is fairly close to our starting pitch of 1. Hence we may choose to stop after adding only 5 different pitch classes, and the resulting pitches that we obtain form the five notes of the pentatonic scale. (If we started at F, the notes would be F-C-G-D-A.) A number of traditional Chinese instruments, such as the zheng and qin, have strings tuned to play exactly the five pitches of the pentatonic scale. The eighth pitch of 1.068 in our sequence, likewise, is very close to 1, and hence we may also choose to stop after adding the first 7 pitch classes. The resulting pitches that we obtain form the diatonic major scale. (If we started at F, we would obtain the seven pitches of the C major scale.)

We may frame this question of how many pitch classes to include mathematically as the question of for what integers  $m$  and  $n$  is  $\frac{(\frac{3}{2})^m}{2^n}$  approximately equal to 1. Taking the base-two logarithm, this is the question of when  $m \log_2(\frac{3}{2}) - n$  is approximately zero, or equivalently, how to approximate the irrational number  $\log_2(\frac{3}{2})$  by a rational number  $\frac{n}{m}$ . One method of doing this is by the method of continued fractions [3]: We

may write any irrational number  $r$  as

$$r = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}},$$

where we choose  $a_1$  as the largest integer less than  $r$ , then  $a_2$  as the largest integer less than  $\frac{1}{r-a_1}$ , then  $a_3$  as the largest integer less than  $\frac{1}{\frac{1}{r-a_1}-a_2}$ , and so on. We have the following theorem:

**Theorem 2.2.1.** *If*

$$r = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\dots}}}$$

*is the continued fraction expansion of  $r$  and*

$$\frac{p_n}{q_n} = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\dots + \frac{1}{a_n}}}}$$

*is the truncation of the expansion to  $a_n$  (with  $p_n$  and  $q_n$  relatively prime), then  $\left| r - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}$ . In addition,  $\frac{p_n}{q_n}$  is the closest approximation to  $r$  among all fractions  $\frac{p}{q}$  with  $q \leq q_n$ .*

This theorem tells us that we obtain increasingly closer approximations to any irrational number  $r$  as we write out its continued fraction expansion to increasingly more terms. The continued fraction expansion for  $\log_2\left(\frac{3}{2}\right)$  is

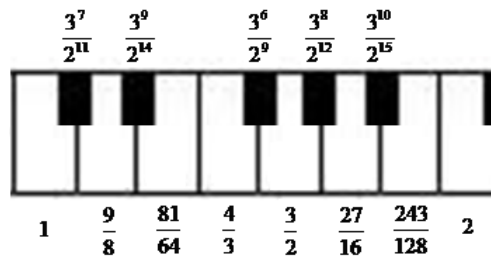
$$\log_2\left(\frac{3}{2}\right) = \frac{1}{1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{3 + \frac{1}{1 + \frac{1}{5 + \frac{1}{2 + \dots}}}}}}}}}$$

The first few rational approximations given by this continued fraction expansion are  $\log_2\left(\frac{3}{2}\right) \approx 1, \frac{1}{2}, \frac{3}{5}, \frac{7}{12}, \frac{24}{41}$ , and  $\frac{31}{53}$ . The approximation of  $\frac{3}{5}$  indicates that  $\frac{(\frac{3}{2})^5}{2^3} \approx 1$ , i.e. going up by five fifths and then down by three octaves brings us close to our starting pitch, and this is the rationale for stopping at the pentatonic scale. The approximation of  $\frac{7}{12}$  indicates that  $\frac{(\frac{3}{2})^{12}}{2^7} \approx 1$ , i.e. going up by twelve fifths and then down by seven octaves brings us close to our starting pitch, and this is the rationale for stopping at twelve pitch classes for the piano keyboard. These are approximations that we have already discussed. The next two approximations of  $\frac{24}{41}$  and  $\frac{31}{53}$  are even closer approximations to  $\log_2\left(\frac{3}{2}\right)$  and indicate that, if we are unsatisfied with treating pitches at a pitch ratio of 1.014 as the same pitch, then 41 or 53 different pitch classes are the next natural places to stop. Indeed, in 1876, Robert Bosanquet designed a “generalised keyboard harmonium” with exactly 53 pitches to each octave [3].



### 2.2.2 Just intonation, Euclidean lattices, and the fundamental domain

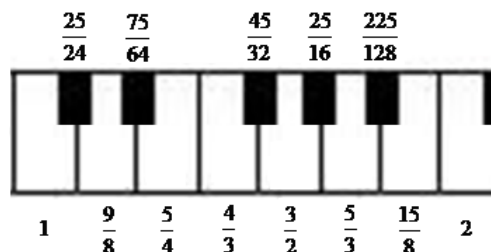
That the piano and many other Western instruments restrict themselves to twelve different pitch classes has an important consequence: it is impossible to tune such instruments so that every interval is perfectly in tune. Indeed, we have observed that, moving upwards by fifths twelve times from a starting pitch class, it is not possible to return to that pitch class unless an approximate fifth is used. One method of tuning the twelve pitch classes, known as Pythagorean tuning, would be to tune each of these fifths as a perfect 3 : 2 pitch ratio, except for the very last fifth. If we denote the pitch C as a frequency of 1 unit and tune fifths perfectly starting at F and moving upwards, then this would give the twelve pitches of the octave from C to C the following frequencies:



The final fifth from A $\sharp$  back to F has frequency ratio  $\frac{2^{18}}{3^{11}} \approx 1.480$  rather than  $\frac{3}{2} = 1.5$  (where  $1.5/1.480 = 1.014$  is our Pythagorean comma).

One problem with this method of tuning is the lack of a perfectly tuned major triad. We discovered in Section 2.1 that the frequency ratio of 5 : 4 is a particularly consonant interval, called the *just major third*, and this interval can be used along with the perfect fifth to form the *just major triad*. The pitch ratio of the root to the just major third above the root to the perfect fifth above the root is 4 : 5 : 6 in the just major triad, making it a particularly consonant version of the major triad chord that is used ubiquitously in Western music. The intervals of the major thirds C-E, F-A, and G-B in the above Pythagorean tuning, on the other hand, have frequency ratio 81 : 64, which differs from the just major third by a factor of  $\frac{81}{80} = 1.0125$ . This frequency ratio of 1.0125, obtained by moving a pitch class upwards by four perfect fifths and downwards by one just major third, is called the *syntonic comma*. It is large enough for the difference between the just major triad and the major triad under Pythagorean tuning to be clearly audible to the trained musician.

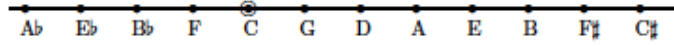
A number of tuning methods address this problem by making certain intervals of the major third into just major thirds rather than the Pythagorean major third, and these tuning methods are collectively known as *just intonation*. Since the major triads I, IV, and V are particularly important in compositions in a major key, most just intonation systems ensure that these are just major triads. The pitches in the I, IV, and V chords include the entire major scale, so most just intonation systems differ only in the remaining five chromatic pitches. One simple version of just intonation is Euler's monochord; in the key of C major, it is given by the following frequencies:



This tuning method has the property that the C major, E major, F major, G major, A major, and B major triads are all just major triads, and in particular these include the I, IV, and V chords in the key of C major.

Just intonation may be viewed as a mathematical generalization of Pythagorean tuning in the following way [3]. Suppose that we begin with a pitch class and consider all pitch classes that can be obtained by

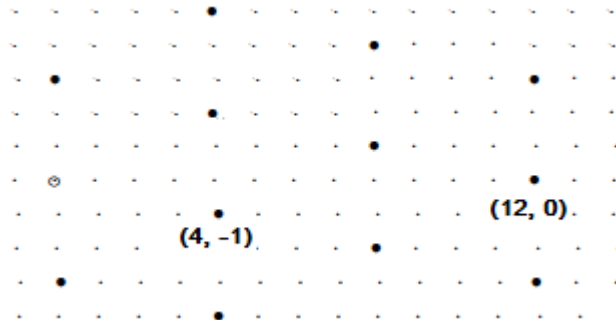
moving upwards and downwards from this pitch class by perfect fifths. We may represent each such pitch class by a unique signed integer to indicate the number of fifths from the starting pitch class. For instance, if we begin at C, then the pitch class C is represented as 0, the pitch class G that is a perfect fifth above C is represented as 1, the pitch class D that is a perfect fifth above G is represented as 2, and so on. The pitch class F that is a perfect fifth below C is represented as  $-1$ , the pitch class B $\flat$  that is a perfect fifth below F is represented as  $-2$ , and so on.



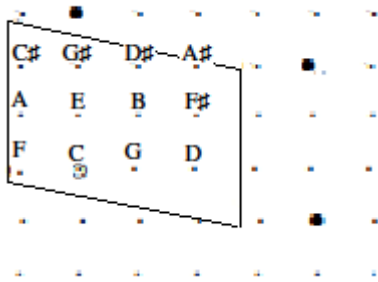
The set of all such pitch classes, represented in this way, forms a *one-dimensional lattice*  $\mathbb{Z}$ . Pythagorean tuning follows from the realization that the actual frequency distance of two pitch classes twelve steps from each other on this lattice is very small, so that we may identify any two pitches separated by twelve steps as being the same pitch. Making this identification defines what is called an *equivalence relation* on the lattice—points that are twelve steps away are “equivalent” pitches. We may take any twelve consecutive points on this lattice to define our Pythagorean tuning, and any other point on the lattice is equivalent to one of these twelve points. If we represent C as 0 and take the points  $-1, 0, 1, \dots, 10$ , we obtain the Pythagorean tuning for F, C, G, ... A $\sharp$  that was previously discussed.

To extend this idea to just intonation, let us start with a pitch class and consider all pitch classes that can be obtained by moving upwards and downwards from this pitch class by any combination of perfect fifths and just major thirds. We may now represent each such pitch class by an ordered pair of integers  $(a, b)$ , where  $a$  and  $b$  are the numbers of perfect fifths and just major thirds that the pitch class lies above the starting pitch class. For instance, if we begin at C and represent it as  $(0, 0)$ , then the G that is a perfect fifth above C is represented as  $(1, 0)$ , the E that is a just major third above C is represented as  $(0, 1)$ , and the B that is a perfect fifth followed by a just major third above C is represented as  $(1, 1)$ . The set of these pitch classes now forms a *two-dimensional lattice*  $\mathbb{Z}^2$ . Just intonation follows from the realization that points on this lattice separated by the vectors  $(4, -1)$  or  $(12, 0)$  are very close in frequency— $(4, -1)$  corresponds to moving up four perfect fifths and down one just major third, which gives the syntonic comma, and  $(12, 0)$  corresponds to moving up twelve perfect fifths, which gives the Pythagorean comma. Identifying pitches that are separated by syntonic and Pythagorean commas defines an equivalence relation on this two-dimensional lattice.

If we take our starting pitch of  $(0, 0)$  and consider the set of all points equivalent to it, these points form a *sublattice*. Any point on this sublattice is a point of the form  $m(4, -1) + n(12, 0)$  for some integers  $m$  and  $n$ , and hence we say that the vectors  $(4, -1)$  and  $(12, 0)$  *generate* the sublattice.



There are many different pairs of vectors, other than  $(4, -1)$  and  $(12, 0)$ , that generate the same sublattice; for instance,  $(4, -1)$  and  $(0, 3)$  is another such pair. The parallelogram that is formed by any pair of vectors that generate the sublattice is called a *fundamental domain* for the equivalence relation defined by the sublattice. If we move a fundamental domain parallelogram to anywhere on the two-dimensional plane, any point on the lattice will be equivalent to exactly one point in the parallelogram. We may take the points in the parallelogram to define our just tuning system. For instance, if we take the following fundamental domain formed by the vectors  $(4, -1)$  and  $(0, 3)$ , we obtain Euler’s monochord in C major as previously discussed.



We note that under this terminology, the single number 12 generated the sublattice in the one-dimensional Pythagorean tuning lattice, and any set of 12 consecutive points of the lattice formed a fundamental domain.

Thus, just as we arrived at Pythagorean tuning by considering the single interval of the perfect fifth and approximating the Pythagorean comma by unison, we may arrive at just intonation by considering the two intervals of the perfect fifth and just major third and approximating both the Pythagorean comma and the syntonic comma by unison. If we wish for our tuning system to have another perfect interval, we may generalize this process to three dimensions. For instance, the dominant seventh chord is another commonly used chord in Western music, and the most consonant version of this chord has pitch ratios  $4 : 5 : 6 : 7$ . We may start at any pitch class and consider pitch classes that can be obtained from it by moving in intervals of the  $3 : 2$  perfect fifth,  $5 : 4$  just major third, and  $7 : 4$  minor seventh. We may represent all such pitch classes, then, as an ordered triple of integers, which forms a three-dimensional lattice  $\mathbb{Z}^3$ , and we may choose three vectors that represent pitch ratios close to unison to define an equivalence relation on this lattice and generate the corresponding sublattice. We may keep the two vectors  $(0, 3, 0)$  and  $(4, -1, 0)$  from before, and a natural third vector to use is  $(2, 0, 1)$ . This corresponds to moving up by two perfect fifths and then a minor seventh, and the pitch ratio between the  $(2, 0, 1)$  and  $(0, 0, 0)$  pitch classes is  $\frac{64}{63} \approx 1.016$ , known as the septimal comma. The sublattice generated by these three vectors has a three-dimensional fundamental domain, and we may take the points within this domain to define our tuning.

We conclude with a theorem that tells us how many lattice points lie within any fundamental domain.

**Theorem 2.2.2.** *Let  $v_1, \dots, v_k \in \mathbb{Z}^k$  be  $k$   $k$ -dimensional vectors with integer coordinates. The number of lattice points contained within the fundamental domain defined by these vectors is given by the unsigned matrix determinant*

$$\begin{vmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{vmatrix}.$$

Recall that the unsigned determinants of a  $2 \times 2$  matrix and  $3 \times 3$  matrix are

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = |ad - bc| \text{ and } \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = |aei + bfg + cdh - ceg - bdi - afh|.$$

Hence we may compute

$$\begin{vmatrix} 12 & 4 \\ 0 & -1 \end{vmatrix} = 12, \begin{vmatrix} 0 & 4 \\ 3 & -1 \end{vmatrix} = 12, \text{ and } \begin{vmatrix} 0 & 4 & 2 \\ 3 & -1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 12.$$

So each of the fundamental domains that we have considered contains exactly 12 distinct pitch classes, and hence we may tune according to these domains without introducing additional pitch classes to the piano keyboard.

### 2.2.3 Meantone, irregular, and equal temperament

We developed just intonation in the preceding section to ensure that the I, IV, and V chords in a major key are just major triads. The *just minor third* is an interval with frequency ratio  $6 : 5$ , and using it we may

analogously define the *just minor triad* as the chord composed of a just minor third and a perfect fifth above the root note. Just as the important major triads are I, IV, and V in a major key, the important minor triads are ii, iii, and vi. Under Euler's monochord just intonation in C major, we see that the E minor iii chord and A minor vi chord are already just minor triads. Ideally, we would like a system of just intonation in which the ii chord is a just minor triad as well.

Unfortunately, this is not possible. Let us consider the common chord progression I-vi-ii-V-I in C major and suppose that C has frequency of 1 unit. To make the starting I chord a just major triad, we must tune E and G to frequencies of  $\frac{5}{4}$  and  $\frac{3}{2}$  units. We may then make the A minor vi chord a just minor triad by tuning A to  $\frac{5}{3}$  units, and the D minor ii chord a just minor triad by tuning D to  $\frac{10}{9}$  units and F to  $\frac{4}{3}$  units. Unfortunately, the interval between G and D is now no longer a perfect fifth. If we were to re-tune G so that the G major V chord is a just major triad, we would need to tune it as  $\frac{40}{27}$  units. To tune the final C major I chord as a just major triad, we would need to re-tune C as  $\frac{80}{81}$  units. Hence we have drifted downwards in pitch by a syntonic comma.

This drift of a syntonic comma along the I-vi-ii-V-I progression is unavoidable in just intonation, in the same way that the drift of a Pythagorean comma was unavoidable when we followed the circle of fifths twelve times in Pythagorean tuning. Unfortunately, both the difference of a syntonic comma and the difference of a Pythagorean comma are large enough to be noticeable to the careful listener, and hence under just intonation, one of the chords in the I-vi-ii-V-I progression will sound noticeably out of tune. To remedy this situation, musicians have played with a variety of tuning systems that spread these commas over a number of different notes, so that the error in tuning is divided into smaller errors across different chords rather than aggregated in one chord. Methods of tuning that begin with just intonation and spread the syntonic comma are collectively known as *meantone temperaments*, and methods of tuning that in addition spread the Pythagorean comma are known as *irregular temperaments*. For example, one method of C major meantone temperament begins with C major just intonation and sets each of the four fifths C-G, G-D, D-A, and A-E at a quarter of a syntonic comma smaller than the perfect fifth, so that the I-vi-ii-V-I progression returns to the same C major I chord.

Just intonation, by focusing on the commonly used chords in a particular key, is key specific. C major just intonation works well for pieces in C major and is useable for pieces in the keys of F and G major that are nearby on the circle of fifths, but it would sound terribly out of tune for a piece written in, for instance, C $\sharp$  major. Hence, in practice, just intonation must be re-tuned for each key. Certain irregular temperaments remedy this situation by spreading the comma across notes in a way that allows for a larger number of keys to sound in tune. It is most likely that J. S. Bach's *The Well-Tempered Clavier*, a collection of 24 preludes and fugues in all 24 major and minor keys, was written for a particularly versatile version of irregular temperament. Despite their versatility, irregular and meantone temperaments do not treat all keys equally, and it is from these differences in tuning that, historically, each key was associated with a particular "character" or "mood"—D major is the key of triumph and rejoice while D-flat major is a key of grief and rapture.

Taking this idea of spreading an error across different notes one step further brings us to equal temperament, the tuning system used in most modern-day settings. Under equal temperament, each half-step has pitch ratio of precisely  $2^{1/12} : 1$ , so that twelve half-steps create a perfect octave. When discussing equal temperament, it is useful to convert frequency ratios to a logarithmic scale of *cents*, as cents =  $1200 \times \log_2(\text{frequency ratio})$ . Hence each half-step in equal temperament is exactly 100 cents. The Pythagorean comma is about 23.46 cents, and the syntonic comma is about 21.51 cents. To elucidate the difference between C major Pythagorean tuning, C major just intonation, and equal temperament, we may compute the value in cents of the difference between each note of the C major scale and the tonic pitch of C:

	Equal temperament	Pythagorean tuning	Just intonation
C	0	0	0
D	200	203.910	203.910
E	400	407.820	386.314
F	500	498.045	498.045
G	700	701.955	701.955
A	900	905.865	884.359
B	1100	1109.775	1088.269

We see that fifths in equal temperament are very slightly narrower than the perfect fifth, and major thirds in equal temperament are 13.686 cents wider than the just major third. From equal temperament, we gain the versatility of having each major and minor key equally in tune, at the cost of making no interval perfectly consonant.

## 2.3 The Javanese gamelan

We explained the origins of the Western twelve-tone scale and various temperaments of the scale using the concepts of consonance and dissonance, and we explained consonance and dissonance using the resonance harmonics of string and wind instruments. That the scale has twelve tones, that we tune for perfect fifths and just major and minor thirds—all of this is dependent on the fact that the resonance harmonics of most pitch-producing instruments have frequencies that are small integer multiples of the frequency of the fundamental pitch. What would happen if we consider instruments with different resonance harmonics?

Let us take a dive into the world of musicology and consider the *gamelan*. The gamelan is an ensemble of mostly percussion instruments found in the Javanese tradition of central Indonesia, typically including the xylophone-like saron and gambang, the vibraphone-like gender, the kettle-shaped bonang, the gong, and the drum-like kendang. As discussed in Section 1.4.1, percussion instruments such as these have very different resonance harmonics from string and wind instruments and hence very different concepts of consonance and dissonance. Therefore, we would expect that the types of musical scales used in gamelan music differ from the Western twelve-tone scale as well.

Each Javanese gamelan is tuned differently, but most systems of tuning use some variant of a five-note *slendro* scale or seven-note *pelog* scale. In the same way that many notes of the Western scale fall at or near points of relative minima on the Plomp-Levelt dissonance curve, we may observe that the notes of the *slendro* and *pelog* scales fall near points of relative minima of appropriately defined dissonance curves as well. The five notes of the *slendro* scale fall near minima of the dissonance curve generated by the bonang and a harmonic sound, whereas the seven notes of the *pelog* scale fall near minima of the dissonance curve generated by the saron and a harmonic sound. The harmonic sound, in the case of the gamelan, can be the human voice or one of several less common harmonic instruments including the lute-like rebab or recorder-like suling. Hence, in the same way that we may view the Western scale as a method of accommodating consonant intervals between harmonic sounds, we may view the gamelan scales as methods of accommodating consonant intervals between a harmonic sound and an inharmonic instrument in the gamelan [11].





## Lecture 3

# Musical Structure and Theory

### 3.1 Musical symmetries and transformations

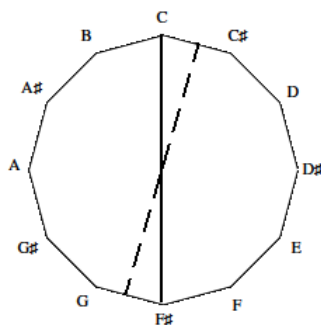
#### 3.1.1 Groups, generators, and relations

Musical compositions are (usually) not random collections of sound. Oftentimes a single melody line, harmonic structure, motif, or idea occurs multiple times in a musical work, and it may undergo various transformations as the music progresses. Consider, as an example, the following melody excerpt from Bach's first two-part invention for keyboard:



We recognize the pattern of the first seven notes, A-G-F-E-G-F-A, as a melodic unit that repeats multiple times in this excerpt, in the form F-E-D-C-E-D-F and then again in the form D-C-B-A-C-B-D. If we look and listen more closely, we may also recognize that the sequences E-G-F-A-G-F-E and C-E-D-F-E-D-C in the excerpt are this same pattern backwards and upside-down. We obtain these repetitions of the opening melodic unit by applying musical symmetries or transformations analogous to the geometric symmetries of “translation” and “reflection”. Accordingly, we may discuss and understand musical transformations such as these using the mathematical language of symmetry groups [7].

As a simple example of a symmetry group, let us consider the set of twelve distinct pitch classes on the piano keyboard and place them on the vertices of a regular dodecagon as follows:



The dodecagon has a number of rotational and reflectional symmetries. For instance, we may consider rotation clockwise by  $30^\circ$ . This would send the vertex C to  $C^\sharp$ ,  $C^\sharp$  to D, D to  $D^\sharp$ , etc., and corresponds to musical *translation* or *transposition* upwards by a half-step. We may consider rotation counterclockwise by  $30^\circ$ , which would send C to B,  $C^\sharp$  to C, D to  $C^\sharp$ , etc., corresponding to musical translation downwards by a half-step. If we apply the clockwise rotation by  $30^\circ$  and then the counterclockwise rotation by  $30^\circ$ , each

pitch class returns to its original position. We may also consider the reflection about the axis connecting C and F $\sharp$ , shown as the solid line in the above figure, which sends C to itself, C $\sharp$  to B, B to C $\sharp$ , D to A $\sharp$ , A $\sharp$  to D, etc. If we apply this reflection twice, we are again back where we started. We may likewise consider the reflection about the axis between C-C $\sharp$  and F $\sharp$ -G, shown as the dashed line in the above figure, which sends C to C $\sharp$ , C $\sharp$  to C, B to D, D to B, etc. Again, if we apply this reflection twice, we are back where we started. These musical reflections are known as *inversions*.

Giving labels to some of these transformations, let us denote the upward translation of a half-step as  $x$  and the inversion about C and F $\sharp$  as  $y$ . Let us denote the application of  $x$  followed by  $x$  as  $x \cdot x$ , or  $x^2$  for short. Similarly, let us denote the repeated application of  $x$   $k$  times as  $x^k$ . Let us denote the application of  $y$  followed by  $x$  as  $y \cdot x$ . We may use this  $\cdot$  operation of combining transformations to obtain new transformations different from  $x$  and  $y$ . For instance,  $x \cdot x$  is a clockwise rotation of  $60^\circ$  or translation upwards by a whole step. We may verify that  $y \cdot x$  sends C to C $\sharp$ , C $\sharp$  to C, B to D, D to B, and so on. In fact,  $y \cdot x$  is precisely reflection about the axis between C-C $\sharp$  and F $\sharp$ -G that was previously discussed. For any transformation, we may denote the transformation that undoes it with a superscript  $-1$  symbol. For instance,  $x^{-1}$  would be the downward translation of a half-step that undoes  $x$ . We observe that  $x^{-1} = x^{11}$ , and  $y^{-1} = y$ . Since  $y \cdot x$  is also a reflection, we know that  $(y \cdot x)^{-1} = y \cdot x$ .

Mathematically, we have constructed a *group*. A group  $G$  is a set of elements together with an operation that can combine any two elements to give another element. In the above example, the elements of the group are the set of all symmetry transformations of the regular dodecagon; we labeled two particular elements of the group as  $x$  and  $y$ . The operation is the  $\cdot$  operation, and any combination of  $x$ 's and  $y$ 's linked with the  $\cdot$  operation is also an element of the group. Every group must have an *identity* element, labeled 1, such that  $1 \cdot g = g$  and  $g \cdot 1 = g$  for any element  $g$  of the group. In the above example, the identity element is the “transformation” that simply leaves the dodecagon in place. Finally, every element  $g$  of a group must have an *inverse* element  $g^{-1}$  in the group such that  $g \cdot g^{-1} = 1$  and  $g^{-1} \cdot g = 1$ .

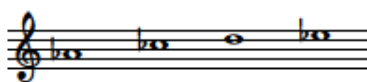
The number of distinct elements in a group is called the *order* of the group. The particular group of symmetries of the regular dodecagon has 24 distinct elements, 12 corresponding to the rotations of  $0^\circ, 30^\circ, 60^\circ, \dots, 330^\circ$  (where the  $0^\circ$  rotation is the identity) and 12 corresponding to reflections about the symmetry axes. This group is called the *dihedral group* of order 24, denoted as  $D_{12}$ . The group of reflection and rotation symmetries of any regular  $n$ -gon is the dihedral group  $D_n$ . (In some books, these are denoted instead as  $D_{2n}$ ; we will not use this notation.) Dihedral groups are not the first examples of groups we have encountered. Indeed, the lattices from Section 2.2.2 were also groups. The points of the lattice were the group elements, the group operation was vector addition, and the identity element was the origin. Whereas the dihedral groups  $D_n$  have a finite number of elements, lattices are groups with infinite numbers of elements. Other examples of infinite groups in mathematics are the set of all integers with group operation addition, the set of all real numbers with group operation addition, and the set of all positive numbers with group operation multiplication. (The set of all real numbers with operation multiplication is not a group because the number 0 does not have a multiplicative inverse.) Another example of a finite group in mathematics is the set of all integers modulo some integer  $n$ , with group operation addition mod  $n$ .

Returning to the group  $D_{12}$  of symmetries of the regular dodecagon with our labeled elements  $x$  and  $y$ , we can verify that any of the 24 symmetry elements in the group can be obtained as a composition of  $x$ 's and  $y$ 's using the  $\cdot$  operation. Such a set of elements, where any group element can be generated as a combination of things in this set and their inverses using the group operation, is called a *generator set* for that group. Hence  $\{x, y\}$  is a generator set for  $D_{12}$ . We note that  $\{x\}$  alone is not a generator set for  $D_{12}$  because we cannot obtain any reflections by using just  $x$ , and  $\{y\}$  alone is not a generator set for  $D_{12}$  because we can obtain only the elements  $y$  and 1 using just  $y$ . In  $D_{12}$ , there are a number of *relations* involving  $x$  and  $y$  that hold true—for instance,  $x^{12} = 1$ ,  $y^2 = 1$ , and  $y \cdot x = x^{11} \cdot y$ . (We may verify the last relation—that reflection using  $y$  followed by rotation using  $x$  is the same as 11 rotations using  $x$  followed by reflection using  $y$ —by checking that it is true when applied to each vertex of the dodecagon.) From these, we may derive other relations— $x^{-1} = x^{-1} \cdot 1 = x^{-1} \cdot x^{12} = x^{11}$ ,  $y^{-1} = y^{-1} \cdot 1 = y^{-1} \cdot y^2 = y$ ,  $y \cdot x \cdot y \cdot x = x^{11} \cdot y \cdot y \cdot x = x^{11} \cdot y^2 \cdot x = x^{11} \cdot x = 1$ , etc. In fact, the generators  $x$  and  $y$  and the relations  $x^{12} = 1$ ,  $y^2 = 1$ , and  $y \cdot x = x^{11} \cdot y$  completely define the group  $D_{12}$ .

In the remainder of this section, we will follow the method of analysis developed by music theorist David Lewin of applying the language of symmetry groups to study transformations on various sets of musical objects.

### 3.1.2 Intervallic transformations

The simplest example of a musical transformation group is  $D_{12}$ , whose elements are transformations of the 12 pitch classes as discussed in the previous section. In accordance with our previous notation, let  $x$  denote an upwards translation of a half-step, and let  $y$  denote inversion about C (or  $F\sharp$ ). We may view any musical interval as a transformation on this space of pitch classes that takes one note of the interval to the other, and any melody line or musical motif as a sequence of such transformations in which we apply a transformation to each pitch to obtain the next pitch. Consider, for example, the Zauber motif that appears repeatedly in Wagner's opera *Parsifal*:



We may view this motif as a sequence of two minor-thirds and a half step, or in the language of  $D_{12}$ , two applications of  $x^3$  and an application of  $x$ . The motif presented above begins on the note  $A_b$ , but it occurs throughout the opera with different starting pitches. We may view all occurrences of this motif as the same sequence of elements of  $D_{12}$ , applied to these different starting pitches.

Viewing musical intervals as transformations of pitch classes can apply not only to successive notes within a melody line, but also to successive tonicizations in a musical passage. In the Prelude to the opening act of Wagner's same opera, we hear the following passage:



The passage tonicizes, in sequence, the notes  $A_b$ ,  $C_b$ ,  $D$ , and  $E_b$ . The intervals between these temporary tonic pitches are precisely the intervals of the Zauber motif, corresponding to the same sequence of group elements:  $x^3$ ,  $x^3$ ,  $x$ . Hence the intervallic transformations of the Zauber motif have been manifested at the very start of the opera, much before the Zauber motif appears in melody form.

### 3.1.3 Transformations between major and minor triads

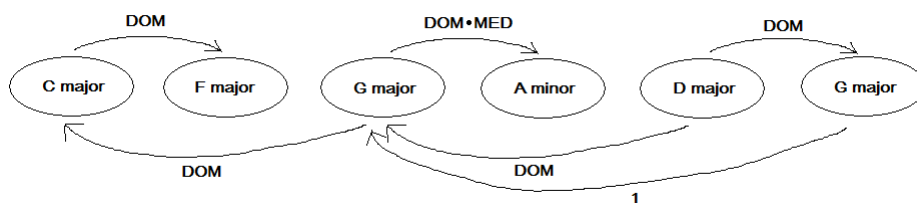
Let us consider the set of all 12 major triads and all 12 minor triads. Certain relationships between these triads have particular musical importance, including the relationship of a major I chord to its dominant V chord, the relationship of a major I chord to its mediant vi chord, and the relationship of a minor i chord to its mediant VI chord. We may view these relationships in terms of transformations on the set of triads. Let us consider a transformation DOM that sends any major triad to its dominant major triad. Thus, DOM sends C major to G major, F major to C major, etc. We must define the action of DOM on all 24 triads in our set, including the minor triads, so suppose that DOM sends C minor to G minor, F minor to C minor, etc. (These minor chord relationships do not have strong musical significance, as the dominant chord to the tonic i chord in a minor key is the major V chord, not the minor v chord, but we define DOM in this way so that it is invertible.) Let us also consider a transformation MED that sends a triad to its mediant triad. So MED sends C major to A minor, G major to E minor, A minor to F major, E minor to C major, etc.

Compositions of DOM and MED generate a group of transformations between these 24 major and minor triads. We note that applying MED twice to a major or minor chord is the same as applying DOM to that chord, so  $MED^2 = DOM$ . This implies that  $\{MED\}$  is a single-element generator set for this group, since DOM can be generated using MED. Groups that can be generated by a single element are known as *cyclic groups*. Repeatedly applying DOM to a major or minor chord cycles through all twelve major or minor chords along the circle of fifths, so  $DOM^{12} = 1$ . Thus  $MED^{24} = 1$ , and this group is the cyclic group of order 24, denoted  $C_{24}$ .

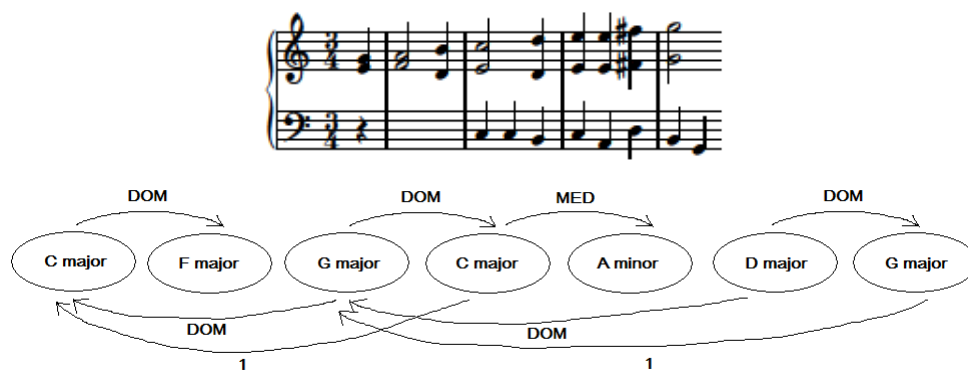
We may use the elements of this group to label the relationships between harmonies in a musical passage, as an alternative to using Roman numeral notation. Consider the opening to Beethoven's First Symphony:



Treating dominant seventh chords for this purpose as simple major chords, we may lay out the chord progression and highlight the transformational relationships between these chords:



We may do the same for a reduction of the first few measures of the third movement of the symphony:



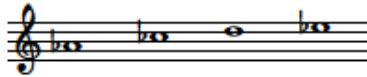
Here, we have condensed G major-C major-G major-C major to just a single repeat of G major-C major in the network. (We have also followed the analysis in [7] of omitting a bass C pedal in the first complete measure of music, understanding the two chords in that measure as F major and G major.) We see that the underlying chord progressions are very similar variations of a shared transformational network. A similar harmonic progression in a different key would also yield the same network of elements of the transformation group, in the same way that it would yield the same traditional Roman numeral analysis.

We may add to the  $C_{24}$  transformation group a third transformation called PAR that sends each major chord to its parallel minor chord and vice versa, e.g. PAR would send C major to C minor, C minor to C major, etc. We may verify that the parallel minor chord of a major chord is obtained by repeatedly taking the mediant eight times, but that the parallel major chord of a minor chord is obtained by repeatedly taking the mediant sixteen times. Hence the PAR transformation on the entire set of major and minor triads is not equivalent to any combination of MED transformations, so it does not lie in the  $C_{24}$  group generated

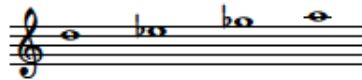
by MED. If we consider the group generated by  $\{\text{MED}, \text{PAR}\}$ , we have the additional relations  $\text{PAR}^2 = 1$  and  $\text{PAR} \cdot \text{DOM} = \text{DOM} \cdot \text{PAR}$ . This second relation may be written as  $\text{PAR} \cdot \text{MED}^2 = \text{MED}^2 \cdot \text{PAR}$ . We note that  $\text{MED} \cdot \text{PAR}$  sends any major chord to the major chord whose root is a minor third lower, and it sends any minor chord to the minor chord whose root is a major third lower. Since moving downwards by three major thirds or four minor thirds returns to the same pitch class,  $(\text{MED} \cdot \text{PAR})^{12} = 1$  is another relation. The generators  $\{\text{MED}, \text{PAR}\}$  and relations  $\text{MED}^{24} = 1$ ,  $\text{PAR}^2 = 1$ ,  $\text{PAR} \cdot \text{MED}^2 = \text{MED}^2 \cdot \text{PAR}$ , and  $(\text{MED} \cdot \text{PAR})^{12} = 1$  fully define the group. One can show that any element of this group can be uniquely written as  $\text{MED}^j (\text{MED} \cdot \text{PAR})^k$  for  $0 \leq j < 24$  and  $0 \leq k < 11$ . Hence this group has order 288.

### 3.1.4 Transformations of motifs and melodic passages

Let us consider a sequence of notes forming a musical motif or melody line, for instance the previously discussed Zauber motif from Wagner's *Parsifal*:



Composers commonly repeat a motif in various forms throughout a work of music, oftentimes transposing it to different keys, inverting it, or retrograde inverting it. Let us consider here the transformations of transposition and retrograde inversion. Define the transformation T to shift an entire motif upwards by a half-step, hence  $T^{12} = 1$ . Define a second transformation RI that sends a motif to its *retrograde inversion*—the same motif backwards and upside-down—such that the first note of the retrograde inversion is the second-to-last note of the original motif. For instance, RI applied to the above Zauber motif is the following:



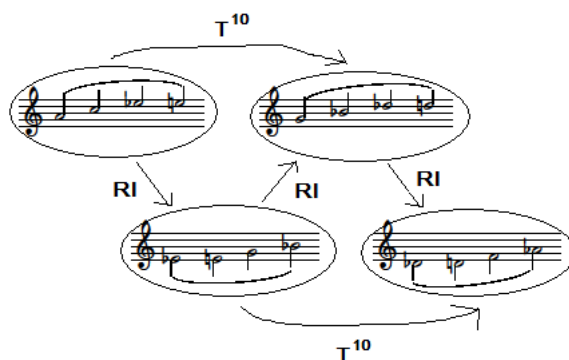
We note that the sequence of intervals between notes of a retrograde inversion of a motif is the same sequence of intervals of the original motif in reverse.

T and RI generate a group of transformations between all transpositions and retrograde inversions of the Zauber motif. We note that applying RI twice to the Zauber motif transposes the motif upwards by the interval of a minor seventh, or ten half steps. Similarly, applying RI twice to a retrograde inversion of the Zauber motif transposes that retrograde inversion upwards by ten half steps. Hence  $\text{RI}^2 = T^{10}$ . We also note that interchanging the order of T and RI does not affect the final result, i.e.  $T \cdot \text{RI} = \text{RI} \cdot T$ . The generators  $\{\text{T}, \text{RI}\}$  and relations  $T^{12} = 1$ ,  $\text{RI}^2 = T^{10}$ , and  $T \cdot \text{RI} = \text{RI} \cdot T$  fully define the group. One can show that any element of this group can be uniquely written as  $T^j \text{RI}^k$  for  $0 \leq j < 12$  and  $0 \leq k < 2$ , so this group has order 24. We note that in deriving the relation  $\text{RI}^2 = T^{10}$ , we were considering the application of  $\text{RI}^2$  on only translations and retrograde inversions of the Zauber motif. If we apply  $\text{RI}^2$  to a different melody line, it may correspond to a different interval of transposition; hence the structure of the group of transformations generated by RI and T is dependent on the melody being transformed. (If we were to consider the group of transformations between all possible melody lines generated by RI and T, there would be no relation between RI and T other than  $\text{RI} \cdot T = T \cdot \text{RI}$ . The resulting group would have the same structure as the 2-dimension lattice from Section 2.2.2 and is called the free abelian group with two generators.)

By defining the RI operation such that the first note of the retrograde inversion is the second-to-last note of the original motif, the last two notes of the motif are always the first two notes of RI of that motif. Composers use this technique to create chains of RI transformations. Returning to Wagner's *Parsifal*, a passage of Act I ("Vom Bade kehrt...") and the transformation music) has the following sequence of local tonics:



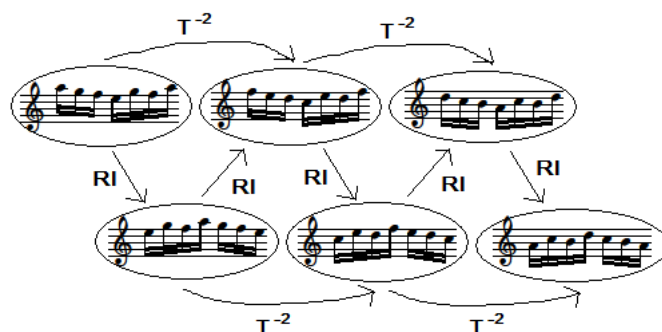
We see in this sequence the repeated use of the Zauber motif and its retrograde inversion, linked together in the following network of T and RI transformations:



As a final example, let us return to the passage from Bach's first two-part invention mentioned at the start of this section, and consider the repeated melody unit:



Here, let us consider only the seven distinct pitch classes of the C major scale, rather than all twelve pitch classes, and let our space be all translations and retrograde inversions of the above melody unit within the key of C major. Define T to translate all elements of this set upwards by one scale degree, so that  $T^7 = 1$ . Define RI of any translation of the melody unit to be its retrograde inversion starting on its second-to-last note, as before, but define RI of any retrograde inversion of the melody unit to be the original melody unit starting on its fourth-to-last note. Then applying RI twice to any translation or retrograde inversion of the melody unit corresponds to a translation downwards by two scale degrees, i.e.  $RI^2 = T^{-2}$ , and we again have  $T \cdot RI = RI \cdot T$ . The generators  $\{T, RI\}$  and relations  $T^7 = 1$ ,  $RI^2 = T^{-2}$ , and  $T \cdot RI = RI \cdot T$  fully define a transformation group of order 14. The melody sequence from Bach's invention makes repeated use of the melody unit and its retrograde inversion in the following network of transformations:



### 3.1.5 Symmetries in Debussy's *Feux d'artifice*

We have thus far used mathematical transformations to analyze specific passages of music within longer works. These transformations can form the structural basis of musical works on a larger scale as well. Let us take *Feux d'artifice*, the last of Debussy's twenty-four preludes for piano, as an example [8]. The full piano score is available in public domain at [1].

The opening measure of *Feux d'artifice* immediately introduces an inversion transformation—we repeatedly hear the sequence of notes F-G-A-B $\flat$ -A $\flat$ -G $\flat$ , where the last three notes are an inversion of the first three notes. Using our previous analogy of the dodecagon, this inversion is about the symmetry axis between G-A $\flat$ , or equivalently, between D $\flat$ -D. The inversion sends F to B $\flat$ , G to A $\flat$ , and A to B $\flat$ . Let us notate

this inversion transformation as  $I$ . We note, in addition, that the first three notes F-G-A themselves exhibit a symmetry under inversion about G, and that the last three notes B $\flat$ -A $\flat$ -G $\flat$  exhibit a symmetry under inversion about A $\flat$ . Let us notate inversion about G, or equivalently about D $\flat$ , as  $J$ , and inversion about A $\flat$ , or equivalently about D, as  $K$ . We will see that these three inversions recur throughout the piece.

Denoting transposition upwards by a half-step as  $T$ , we observe that  $I \cdot K = J \cdot I = T$ . That is, transforming any note by  $I$ -inversion followed by  $K$ -inversion shifts the note upwards by a half-step, and  $J$ -inversion followed by  $I$ -inversion also shifts the note upwards by a half-step. Similarly,  $K \cdot I = I \cdot J = T^{-1}$ . From these, we may derive the additional relations  $K = I \cdot T$  and  $J = T \cdot I$ . Hence, as a consequence of the application of the above inversions, many  $T$  and  $T^{-1}$  relations are heard throughout the prelude as well.

We observe this already in the first six notes of the piece—not only are the sets of notes {F,G,A} and {B $\flat$ , A $\flat$ , G $\flat$ } related by  $I$ -inversion, but also the second set is the  $T$ -translation of the first set. Indeed, we note that the set of notes {F,G,A} inverts to itself under  $J$ -inversion; let us say that such a set is  $J$ -invariant. Similarly, {B $\flat$ , A $\flat$ , G $\flat$ } is  $K$ -invariant, meaning that it inverts to itself under  $K$ -inversion. If  $S$  is any  $J$ -invariant set, then  $I$  applied to  $S$  is the same as  $J \cdot I$  applied to  $S$ , which is  $T$  applied to  $S$  or  $I \cdot K$  applied to  $S$ . That is, the  $I$ -inversion of  $S$  is the same set of notes as the  $T$  transposition of  $S$ , and the  $I$ -inversion of  $S$  is also  $K$ -invariant. Hence it is not a coincidence that {B $\flat$ , A $\flat$ , G $\flat$ } is both the  $I$ -inversion and  $T$ -transposition of {F,G,A}, or that it is  $K$ -invariant; these are mathematical statements that may be derived from the  $J$ -invariance of {F,G,A}.

In the context of these inversions, the D and A $\flat$  sparks in measures 3-14 have special significance as the centers of  $K$ -inversion. In measures 7-10, the D spark is joined by the additional note C, which sounds together with the F-G-A sequence. The set of five notes {C,D,F,G,A} forms a pentatonic scale, a sound commonly employed in Debussy's compositions, and here it carries additional significance because the set is  $J$ -invariant. The black-key glissando of measure 17 consists of the notes {D $\flat$ ,E $\flat$ ,G $\flat$ ,A $\flat$ ,B $\flat$ }, which is both the  $T$ -transposition and  $I$ -inversion of {C,D,F,G,A}. As such, it is also  $K$ -invariant.

Measures 25-46 present three variants of what can be considered the primary “melodic” motif of the prelude; they are reproduced below:



We hear  $T$ -transposition relations in these variants, for instance from {C,A,G} to {C $\sharp$ ,A $\sharp$ ,G $\sharp$ } to {D,B,A} in the first variant, and {F,E $\flat$ } to {G $\flat$ ,F $\flat$ } and {B $\flat$ ,G $\flat$ ,F $\flat$ } to {B,G,F} in the third variant. We may also hear the shift from {F,D $\flat$ ,C $\flat$ } to {G,E $\flat$ ,D $\flat$ } in the second variant as the composition of two  $T$ -transpositions. The harmony of the first variant, measures 25-34, consists primarily of the notes {E,D,C,B $\flat$ ,G}, which is again a  $J$ -invariant set. This set contains G, one of the centers of  $J$ -inversion, and is extended to include C $\sharp$ , the other center of  $J$ -inversion, in measures 33 and 34. The notes of the harmony undergo two  $T$ -transpositions in measure 30 to match the  $T$ -transpositions of the melodic motif. The harmony of the second variant, measures 35-38, consists of the notes {G,D $\flat$ ,E $\flat$ ,F,C $\flat$ }, which has a distinctive whole-tone sound and can be considered a transition to measure 39. The harmony of measures 39 and 40 consists of the notes {F,E $\flat$ ,D $\flat$ ,C $\flat$ ,B $\flat$ ,A $\flat$ ,G $\flat$ }, which is  $K$ -invariant, and the  $K$ -inversion symmetry is highlighted by the A $\flat$  notes in the bass that are a center of this symmetry. This transitions to a harmony consisting of the notes in the D $\flat$  major diatonic scale in measures 41-43, which actually is invariant under a new symmetry about the notes A and E $\flat$ . This is the harmony to the third variant of the melodic motif.

The subsequent section from measure 46 to measure 64 exhibits a number of symmetries. The recurring note-set {B,A,F $\sharp$ ,D $\sharp$ ,C $\sharp$ } in measures 46-48 is the  $T^{-1}$ -transposition of the  $J$ -invariant note-set {E,D,C,B $\flat$ ,G}

from measure 25, whereas the note-sets  $\{C\sharp, B, G\sharp, E\sharp, D\sharp\}$  and  $\{A, G, E, C\sharp, B\}$  in measures 47 and 48 are its  $T$ -transposition and  $T^{-2}$ -transposition. The notes of measures 53-56 form the whole-tone set  $\{A\flat, B\flat, C, D, E, F\sharp\}$  which, being six equally spaced notes of the chromatic scale, is invariant under all inversions.  $T$ ,  $T^2$ ,  $T^{-1}$ , and  $T^{-2}$  transpositions are heard in the melody line  $C\sharp-B\sharp-C\sharp-D-F\sharp-E-F\sharp-G\sharp$  in measures 57-60. The subsequent sequences of major triad chords  $C-F\sharp-E-B\flat$  and  $E\flat-A-G-C\sharp$  in measures 61-64 exhibit  $K$ -inversion symmetry, in the sense that  $\{C, F\sharp\}$   $K$ -inverts to  $\{E, B\flat\}$  and  $\{E\flat, A\}$   $K$ -inverts to  $\{G, C\sharp\}$ .

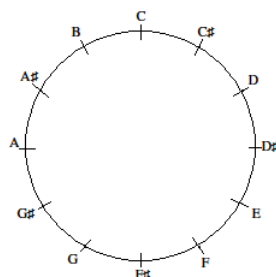
In measures 65-70, we hear a reprise of the melodic motif from measures 25-46, and this is harmonized by the note-set  $\{D\sharp, C\sharp, G\sharp, B, E\sharp\}$  that is the  $T$ -transposition/ $I$ -inversion of the  $J$ -invariant set  $\{E, D, C, B\flat, G\}$  from measure 25. Hence this note-set is  $K$ -invariant. The subsequent measures 71-78 transition to a second reprise of the melodic motif in measures 79-86, which is harmonized by the subset  $\{B\flat, C, D\}$  of the original  $J$ -invariant set. The build-up of measures 85 and 86 is a sequence of  $T$ -transpositions, which explodes into the double-glissando of measure 87. We note that both the white-note portion of the double-glissando, as the C major diatonic scale, and the black-note portion of the double-glissando, as the black-key pentatonic scale, are  $K$ -invariant. Measures 88 and 89 of the coda reprise the very opening of the prelude, with its short-scale  $I$ ,  $J$ , and  $K$  inversions. The snippet of the French national anthem, the *Marseillaise*, in measures 92-94 highlight the notes G and D; G here pairs with the low  $D\flat$  in the bass as the two centers of  $J$ -inversion, while D pairs with the low  $A\flat$  in the bass as the two centers of  $K$ -inversion. We hear a final  $T$ -transposition in measures 95-96 within the last reprise of the melodic motif, and the piece concludes with  $D\flat$  sparks that are the  $T^{-1}$ -transposition and  $I$ -inversion of the original D sparks in the opening.

This *Feux d'artifice* prelude is an example of an atonal composition—there is no clear concept of key, of modulation between keys, or of a tonic. Chord and note relationships are, instead, established by the  $I$ ,  $J$ , and  $K$  inversion symmetries used extensively throughout the piece. The large-scale structure of the piece is not easily explained by traditional analyses of musical form, but it is a macrocosm of the very first measure of the prelude and is governed by the same symmetry transformations. In the transition from the  $J$ -invariant note-set in measure 7 to its  $K$ -invariant  $I$ -inversion in measure 17, from the  $J$ -invariant harmony accompanying the first variant of the melodic motif in measure 25 to its  $K$ -invariant  $I$ -inversion accompanying the third variant of the melodic motif in measure 39 and the reprise of the motif in measure 65, and from the  $D\flat$  spark of measure 3 to its  $I$ -inversion D spark of measure 97, we see that the same mathematical relations governing the first six notes of the piece form the basis for the composition's large-scale structure.

## 3.2 The geometry of chords

### 3.2.1 Tori of ordered chords

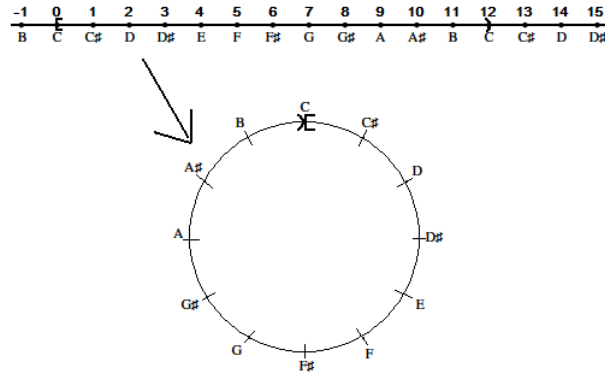
In Section 3.1, we placed the twelve chromatic pitch classes on the vertices of a regular dodecagon; this allowed us to visualize the musical symmetries of transposition and inversion as rotations and reflections of this dodecagon. Recall from Lecture 1, though, that a pitch is simply a frequency of oscillation, and hence the space of all possible pitch classes is actually a continuous space. It is thus natural to place the twelve chromatic pitch classes as equally spaced points on a circle. Points on the circle between two labeled pitch classes then represent the continuous range of pitch classes lying in between the two labeled pitch classes. This is the starting point from which we will investigate the geometry of chord spaces [13].



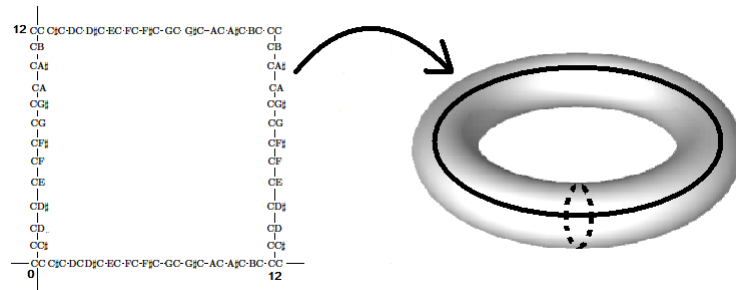
The mathematical construction of this pitch class circle relies on the concept of an equivalence relation, explored in Section 2.2.2. We may represent the space of all musical pitches as the real-number line  $\mathbb{R}$ , where



middle C is represented by the value 0, the C $\sharp$  a half-step higher by the value 1, the D a whole-step higher by the value 2, etc. If we wish to group pitches that are an octave apart into the same pitch class, we impose the equivalence relation  $x \sim x + 12$  for all values  $x$  on this line, i.e. each value is equivalent to the value 12 higher. The range of pitches from 0 to 12 thus forms a fundamental domain for this equivalence relation, and each pitch class is represented by a point in this fundamental domain. If we consider a pitch class that moves rightwards along this line, it will loop back around to 0 when it reaches 12. Hence we may glue the points 0 and 12 together to form a circle, and the space of pitch classes is this circle.



What about the space of chords? Musically, we define the type of a chord based on which pitch classes are present in the chord, ignoring the absolute pitches in the chord and the spacing of the chord. (The C major triad is a chord consisting of notes from the pitch classes C, E, and G; it does not matter to which octaves these notes belong.) We would like our space of chords to capture this intuition. Let us consider a two-note chord. We may represent the absolute pitches of the two note chord as a point  $(x, y)$  in the plane  $\mathbb{R}^2$ , where the coordinates  $x$  and  $y$  represent the individual pitches. Let us then impose the equivalence relations  $x \sim x + 12$  and  $y \sim y + 12$  for all points  $(x, y)$  in the plane. The square  $0 \leq x < 12, 0 \leq y < 12$  forms a fundamental domain for this equivalence relation, and each two-note chord is represented by a point in this fundamental domain. If we slide the first note of the chord upwards in pitch, our point moves in the positive  $x$ -direction and loops back around to 0 when it reaches 12. Likewise, if we slide the second note of the chord upwards in pitch, our point moves in the positive  $y$ -direction and also loops back around to 0 when it reaches 12. Hence we may glue the left and right edges of this square together to form a cylinder, and then glue the top and bottom edges of this cylinder together to form a donut.



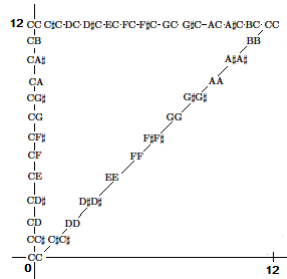
The space of two-note chords is precisely the space of all points on the surface of this donut. We may represent each such point as a pair of coordinates from a circle, where the first coordinate represents the position along the solid circle and the second coordinate represents the position along the dotted circle in the above picture. These coordinates correspond to the individual pitch classes of the chord.

We may construct the spaces of three-note and four-note chords in a similar way, although they become more difficult to visualize. The space of three-note chords is obtained by taking the cube  $0 \leq x < 12, 0 \leq y < 12, \text{ and } 0 \leq z < 12$ , and gluing together the top and bottom faces, the left and right faces, and the front and back faces. This is only possible to visualize in four-dimensions. The space of four-note chords is

obtained by taking a four-dimensional hypercube (the space of all points  $(x, y, z, w)$  where each coordinate is in the range from 0 to 12) and again gluing together opposite pairs of faces. Mathematically, the structure of these spaces is known as a *torus*. The circle of pitch classes is a one-dimensional torus, denoted  $T^1$ . The donut of two-note chords is a two-dimensional torus, denoted  $T^2$ . The space obtained by starting with an  $n$ -dimensional hypercube and gluing together opposite faces is an  $n$ -dimensional torus, denoted  $T^n$ . The  $n$ -dimensional torus is a space in which, in a small enough region around each point, we may represent points in that region using  $n$  real-valued coordinates. In mathematics, such a space is called an  $n$ -dimensional *manifold*.

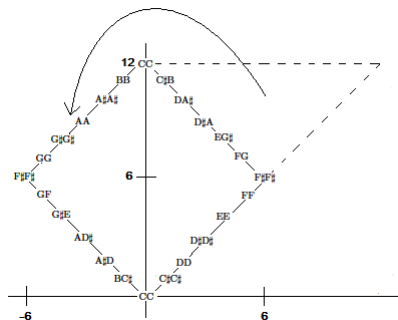
### 3.2.2 Orbifolds of unordered chords

One problem with these spaces of chords is that they are spaces of ordered pitches. The chord C-E and the chord E-C are represented by two different points on  $T^2$ , because we constructed the space by treating the ordered pairs  $(0, 4)$  and  $(4, 0)$  as different points. Likewise, the chords C-E-G, G-C-E, and E-G-C are different points on  $T^3$ . Oftentimes in musical analysis, we do not wish to make the distinction between these chords, and we wish to label C-E-G, G-C-E, and E-G-C all as simply a “C major chord”. Hence, mathematically, we need to define a new equivalence relation on our chord spaces. For  $T^2$ , we define the relation that  $(x, y)$  is equivalent to  $(y, x)$ . Visually, this means that each point of the plane  $\mathbb{R}^2$  is equivalent to its reflection across the line  $y = x$ , and this shrinks the fundamental domain of our equivalence relations from the square  $0 \leq x < 1$  and  $0 \leq y < 1$  to the following triangle:



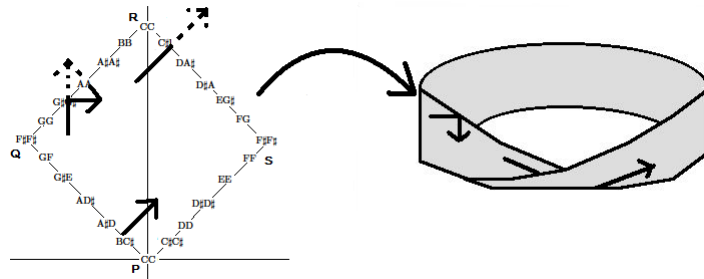
Each chord of two unordered pitches is represented by a point in this triangle. The resulting space of chords is mathematically denoted  $T^2/S_2$ . Similarly, we may impose the equivalence relation that  $(x, y, z)$  is equivalent to  $(x, z, y)$ ,  $(y, x, z)$ ,  $(y, z, x)$ ,  $(z, x, y)$ , and  $(z, y, x)$  on  $T^3$ —that is, each point is equivalent to its reflection about the  $x = y$ ,  $x = z$ , and  $y = z$  planes. This shrinks the fundamental domain from a cube to a tetrahedron, and the resulting space of unordered chords is denoted  $T^3/S_3$ . In general, the space of points of  $T^n$  under the additional equivalence relations of permutations of the coordinates is denoted  $T^n/S_n$ . These spaces are examples of mathematical structures known as *orbifolds*.

How can we visualize the orbifold  $T^2/S_2$ , in the same way that we visualized  $T^2$  as the surface of a donut? We may begin at the above triangular fundamental domain and glue the left and top edges together, in the same way that we visualized  $T^2$ . It turns out that it is, instead, easier to visualize  $T^2/S_2$  by beginning with a different choice of fundamental domain, shown as the tilted square in the below figure:



This fundamental domain can be interpreted by noting that if we take any two-note chord in our space, then sliding the chord parallel to the line  $y = x$ , i.e. in the direction of the vector  $(1, 1)$ , corresponds to transposing this chord while keeping the interval between the two notes fixed. On the other hand, the chords lying on the line  $y = -x$  perpendicular to  $y = x$  are those whose coordinates sum to 0 modulo 12, and all the possible intervals between two pitch classes are found along this line. At  $(0, 0)$ , we have the unison interval C-C. As we slide the point upwards and to the left along the line  $y = -x$ , we obtain increasingly wider intervals until we reach the tritone A-E $\flat$ , represented as  $(-3, 3)$ , and then increasingly narrower intervals until we return to the unison interval F $\sharp$ -F $\sharp$ , represented as  $(-6, 6)$ . (Any interval wider than a tritone is equivalent to the octave minus this interval, which is narrower than a tritone.) If we take the line segment from  $(0, 0)$  to  $(-6, 6)$  and slide it parallel to the line  $y = x$ , we pass over the points representing all transpositions of possible intervals between two notes. When we reach the line segment from  $(6, 6)$  to  $(0, 12)$ , we note that the chords on this segment are the same as the chords on our starting segment, and hence the tilted square bounded by  $(0, 0)$ ,  $(-6, 6)$ ,  $(0, 12)$ , and  $(6, 6)$  forms our fundamental domain. Let us label these points as  $P$ ,  $Q$ ,  $R$ , and  $S$ .

We note that the line segments  $QR$  and  $PS$  are reflective line segments—when we take a point inside the fundamental domain and slide it across one of these segments, the portion of the path outside of the fundamental domain is mapped to its reflection across the segment. On the other hand, when we move a point across  $RS$ , it reappears at the segment  $PQ$ .

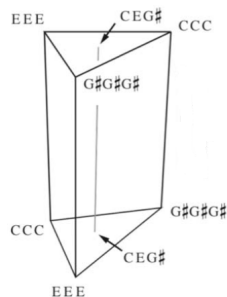


Hence we may glue the segments  $PQ$  and  $RS$  together, but we note that their orientations are reversed, and so we must glue them together with a twist. The resulting structure is a Möbius strip, shown above, and this is the space of unordered two-note chords.

We may visualize the space of unordered three-note chords using a similar procedure. Sliding a chord parallel to the line  $x = y = z$ , i.e. in the direction of the vector  $(1, 1, 1)$ , corresponds to transposing this chord while keeping the intervals between the notes of the chord fixed. All possible types of three-note chords can be found in the plane perpendicular to this line. The unison chords of C-C-C, E-E-E, and G $\sharp$ -G $\sharp$ -G $\sharp$  form a repeating triangular lattice over this plane, with the chords within each triangle repeated in all of the triangles.



We obtain a fundamental domain by taking one such triangle and sliding it parallel to the line  $x = y = z$ , until we reach the same set of chords as in our original triangle. This fundamental domain is a triangular prism.



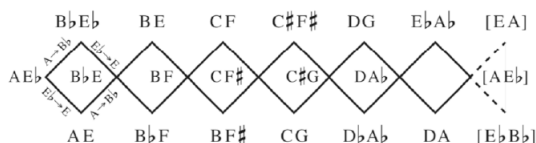
The three rectangular faces of the prism are reflecting faces, whereas the top triangular face connects to the bottom triangular face with a  $120^\circ$  twist. Gluing these faces together, the resulting structure is the following:



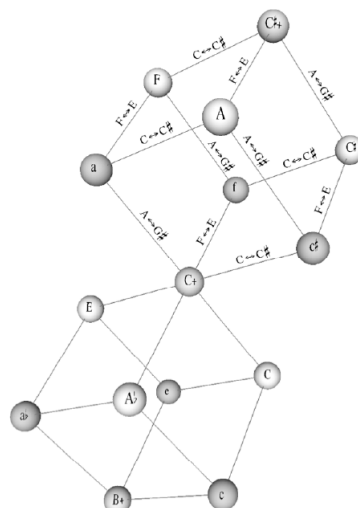
Likewise, the fundamental domain for the space of unordered four-note chords is a four-dimensional prism over a tetrahedron with unison chord vertices  $C-C-C-C$ ,  $E_b-E_b-E_b-E_b$ ,  $F\sharp-F\sharp-F\sharp-F\sharp$ , and  $A-A-A-A$ . The top and bottom tetrahedral “faces” are connected with a twist that sends  $C-C-C-C$  to  $E_b-E_b-E_b-E_b$  to  $F\sharp-F\sharp-F\sharp-F\sharp$  to  $A-A-A-A$  and back to  $C-C-C-C$ , and the remaining “walls” are reflecting. The structure is analogous in higher dimensions, for chords with more than four notes.

### 3.2.3 Chromatically descending chord progressions

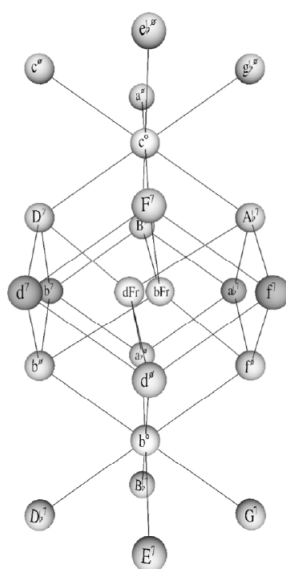
Considering these geometric spaces of chords gives us a way of visualizing chord progressions. Returning to the Möbius strip  $T^2/S_2$ , we see that along the center of the strip lies the interval of the tritone, which divides the octave into two equal halves. To move from the tritone chord  $C-F\sharp$  to the tritone chord  $B-F$  that is a half-step lower by moving one note at a time, we may either first lower  $C$  to  $B$  and then lower  $F\sharp$  to  $F$ , or vice versa. The intermediate chords  $B-F\sharp$  or  $C-F$  form a square with the original chords  $C-F\sharp$  and  $B-F$ , and they are both the interval of a fourth/fifth. We may link a sequence of these squares together at the tritone chords to form a chain, and a path through this chain corresponds to a progression of two-note chords.



Moving to the orbifold  $T^3/S_3$ , we see that along the center of the triangular prism lies the augmented triad, which divides the octave into three equal parts. To move from the augmented triad chord  $C-E-G\sharp$  to the augmented triad chord  $B-E_b-G$  that is a half-step lower by moving one note at a time, we can lower each of the three notes  $C$ ,  $E$ , and  $G\sharp$  in any order. Lowering any of the three notes yields a major triad, either  $C-E-G$ ,  $A_b-C-E_b$ , or  $E-G\sharp-B$ . Lowering a second note yields a minor triad, either  $C-E_b-G$ ,  $G\sharp-B-D\sharp$ , or  $E-G-B$ . These three major triads, three minor triads, and two starting and ending augmented triads  $C-E-G\sharp$  and  $B-E_b-G$  form a cube, of which the augmented triads are diagonally opposite vertices. We may link a sequence of these cubes together at the augmented triad chords to form a chain, and a path through this chain corresponds to a progression of three-note chords.



Finally, considering the orbifold  $T^4/S_4$ , along the center of the four-dimensional prism lies the diminished seventh chord, which divides the octave into four equal parts. To move from the diminished seventh chord C-Eb-Gb-Bbb to the chord B-D-F-Ab that is a half-step lower by moving one note at a time, we can lower each of the four notes C, Eb, Gb, and Bbb in any order. Lowering one note gives one of four possible dominant seventh chords. Lowering a second note gives one of four possible minor seventh chords or two possible French sixth chords. Lowering a third note gives one of four possible half-diminished seventh chords. These start and end diminished seventh chords are the diagonally opposite vertices of a four-dimensional cube, known as a *tesseract*, and these intermediate dominant seventh, minor seventh, French sixth, and half-diminished seventh chords are the remaining vertices of the tesseract. We may link a sequence of these tesseracts together at the diminished seventh chords to form a chain, and a path through this chain corresponds to a progression of four-note chords.



These types of chord progressions, in which chords transition from one to the next by small chromatic motions in the individual notes, were often used by Romantic composers such as Chopin, Brahms, and Wagner. Consider the progression of chords from the opening of Chopin's Prelude in E Minor.



The chords labeled (a) are a sequence of diminished seventh chords, successively lowered by a half-step. The transitions between these diminished seventh chords correspond to paths on the tesseract between these chords. We have labeled as (b) the dominant seventh chords that are obtained by taking a first step along the tesseract, as (c) the minor seventh and French sixth chords that are obtained by taking a second step along the tesseract, and as (d) the half-diminished seventh chords that are obtained by taking a third step along the tesseract. We note that in the chord transitions, Chopin takes one or two steps on the tesseract chain at a time, and we may visualize the chord progression as a path along the tesseract chain. The order in which notes are lowered from each diminished seventh chord to the next is somewhat arbitrary, with the exception of the first step. To move from each diminished seventh chord to the subsequent dominant seventh chord, Chopin always chooses to lower the note that was the fifth of the previous dominant seventh chord, which becomes the root of the new dominant seventh chord. Mathematically, this means that the first step after each diminished seventh chord alternates between the vectors  $(0, 0, -1, 0)$  and  $(-1, 0, 0, 0)$ . Musically, this means that the sequence of dominant seventh chords appearing in this progression descends by the circle of fifths.

# Bibliographic Notes

The presentation in these notes draws heavily on Dave Benson's text *Music: A Mathematical Offering* [3]. This text provides a more detailed discussion of many of the topics covered in these notes, particularly those from the first two lectures. An overview of resonance harmonics and psychoacoustic phenomena from a less mathematical perspective can be found in Howard and Angus's *Acoustics and Psychoacoustics* [6], and a more rigorous treatment of the physics of sound, the auditory process, and musical instruments can be found in [3]. The notes for Lecture 1 draw extensively from these two sources. The proof of Theorem 1.3.1 can be found in Stein and Shakarchi's text [12], which is also a good introductory text on Fourier analysis.

The material from Lecture 2 draws from [3] and Sethares' text [11]. The discussions of continued fractions and the just intonation lattice come from the former, which also describes in much more detail the various systems of just intonation and meantone and irregular temperament. The latter provides a more thorough discussion on the relation between scales and resonance harmonics, and it also provides examples with respect to the Indonesian gamelan and several other non-Western musical traditions. The proof of Theorem 2.2.1 can be found in Hardy and Wright's text [5] on elementary number theory.

The material on musical groups and transformations, including the musical examples contained therein, from Lecture 3 is drawn exclusively from David Lewin's *Generalized Musical Intervals and Transformations* [7], with the analysis of Debussy's *Feux d'artifice* highlighting some of the main points from the analytical essay in [8]. The concepts of groups, equivalence relations, and fundamental domains can be found in standard texts on abstract algebra, for instance [4] and [2]. In particular, [2] devotes a chapter to discussing the relationship between groups and symmetry. The material on the geometry of chords and the discussion of Chopin's prelude are drawn from Dmitri Tymoczko's preprint [13], which uses this geometric perspective to examine principles of voice leading. Some of the geometric ideas are taken from algebraic topology and can be found in Munkre's topology text [9].





# Bibliography

- [1] International music score library project, August 2010. <http://imslp.org>.
- [2] M. Artin. *Algebra*. Addison-Wesley, second edition, 2010.
- [3] D. Benson. *Music: A Mathematical Offering*. Cambridge University Press, 2006.
- [4] D. S. Dummit and R. S. Foote. *Abstract Algebra*. John Wiley and Sons, third edition, 2003.
- [5] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, fifth edition, 1980.
- [6] D. M. Howard and J. A. S. Angus. *Acoustics and Psychoacoustics*. Elsevier, fourth edition, 2009.
- [7] D. Lewin. *Generalized Musical Intervals and Transformations*. Yale University Press, 1987.
- [8] D. Lewin. A transformational basis for form and prolongation in Debussy’s “Feux d’artifice”. In *Musical Form and Transformation: Four Analytic Essays*. Yale University Press, 1993.
- [9] J. Munkres. *Topology*. Prentice Hall, second edition, 2000.
- [10] R. Plomp and J. M. Levelt. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38:548–560, 1965.
- [11] W. A. Sethares. *Tuning, Timbre, Spectrum, Scale*. Springer-Verlag, second edition, 2005.
- [12] E. M. Stein and R. Shakarchi. *Fourier Analysis: An Introduction*. Princeton University Press, 2003.
- [13] D. Tymoczko. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. preprint.