WENDY'S XENHARMONIC KEYBOARD

JORDAN SCHETTLER

1. Motivation

Wendy Carlos is an American composer and one of my favorite musical pioneers. She helped to popularize and improve the Moog synthesizer with the Grammy winning album *Switched on Bach* in 1968. I recall fondly when my grandfather gave me his copy of that record on vinyl... I didn't know at first if I was playing it at the right speed! Some of Carlos' other best known works are the motion picture soundtracks for Stanley Kubrick's *A Clockwork Orange* in 1971, *The Shining* in 1980, and Walt Disney's *Tron* in 1982.

Here we are interested in Carlos' 1986 original work *Beauty in the Beast* because of the experimentation with various so-called *xenharmonic scales*, i.e., scales other than the familiar 12-tone scale on a standard keyboard. This standard keyboard has equal spacing with 100 cents between consecutive notes. In contrast, the title track on *Beauty in the Beast* uses a 9-tone and 11-tone scale with $\alpha = 77.995...$ and $\beta = 63.814...$ cents, respectively, between consecutive notes. The α and β scales are both "contained" in a scale with $\gamma = 35.097...$ cents between consecutive notes. In her article "Tuning: At the Crossroads" ([Car87]), Wendy Carlos explains how she derived these scales by experimentation. In contrast, we will derive these scales using the theory of continued fractions in the spirit of Dunne and McConnell's article "Pianos and Continued Fractions" ([DM99]). This method will further demonstrate how the containment of the α and β scales inside the γ scale is analogous to the way in which the pentatonic scale (5 black keys) and heptatonic scale (7 white keys) fit into the familiar 12-tone scale found on a standard keyboard. Finally, we will illustrate this analysis by drawing Wendy's γ scale keyboard.

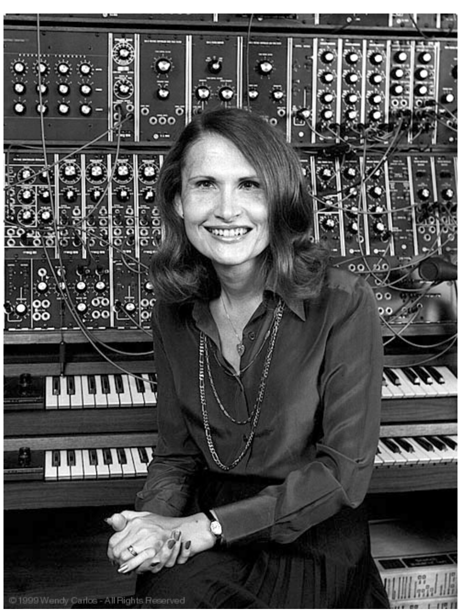

FIGURE 1. Wendy Carlos

2. Background

Before discussing the construction of any scales, we first review some basic physics and functional analysis.

2.1. **Frequency and Pitch.** Given an ideal taut string with fixed endpoints (think of strings on a piano, violin, or guitar), vibrations lead to periodic compression waves in air which we perceive as sound. We define the *fundamental period* T to be the number of seconds in one minimal cycle of these periodic vibrations. The *fundamental frequency* $f = 1/T$ is the number of minimal cycles per second. The units of f here are measured in Hertz:

$$1\,\text{Hz} = 1\,\frac{1}{\text{sec}}.$$

We will not write the units Hz, and we will simply regard f as a number in the interval of positive real numbers $(0, \infty)$. For example, we literally define the frequency of *Concert A* (i.e., the A note above middle C on a piano) to be 440. Various physical properties of the string govern its fundamental frequency as seen in the following formulation of Mersenne's laws (see [Tip78]): if our given string has length L, tension (aka pulling force) F, and linear density (aka thickness) μ, then

$$f = \frac{\sqrt{F/\mu}}{2L}.$$

Either increasing the tension F, decreasing the density μ, or shortening the length L, has the affect of increasing the frequency f which we perceive as an increase in the pitch of the sound. For example, the L and μ for a given string on a guitar remain constant, but we can increase the pitch by adjusting the tuning peg to increase the tension F. As another example, the tensions and lengths of the strings on a given guitar should be similar, but the pitches sound different due to the variation in thickness μ of each individual string; the thinner strings (smaller μ) have a higher pitch.

2.2. **Harmonics and Timbre.** Given the above determination of a pitch based upon only the length, tension, and density of a string, one might ask: Why does a Concert A sound different when played on piano versus a guitar or a synthesizer? The answer lies in the fact that our ideal string may vibrate at multiple frequencies simultaneously not just the fundamental frequency. These other frequencies, which are positive integer multiples of the fundamental frequency f, are called *harmonics*: f, $2f$, $3f$, $4f$, ... One way to see how harmonics influence sound quality is in the context of the periodic functions of a real variable which model the periodic vibrations of the string. Perhaps the best known example of a periodic function is the sine function $\sin(x)$ which has period 2π and corresponds to a sound we would recognize as a dial tone. In general, we use the word *timbre* to indicate the quality of a sound independent of the pitch. For example, $\sin(2\pi fx)$ corresponds to a dial tone with arbitrary frequency f; larger values of f correspond to higher pitched dial tones. The cosine function of frequency f is just a horizontal shift of the sine function, namely $\cos(2\pi fx) = \sin(2\pi fx + \pi/2)$, so this does not correspond to a different timbre. However, it turns out that we can use sines and cosines to describe every sufficiently nice periodic function. Hence any kind of sound made from the vibrations of our idealized string can be built up from combining various dial tones. There is a very elegant interpretation of this which begins with Euler's formula $e^{2\pi ifx} = \cos(2\pi fx) + i\sin(2\pi fx)$ where i is a fixed imaginary unit: $i^2 = -1$. By allowing complex valued functions, we can capture all the harmonics with powers of this exponential function. For instance, the nth harmonic is contained in the expression

$$(e^{2\pi ifx})^{\pm n} = \cos(2\pi(nf)x) \pm i\sin(2\pi(nf)x).$$

The physical properties of the string dictate the extent to which various harmonics appear. In other words, the amplitude (aka volume) of certain harmonics will be greater than others depending upon the instrument, and the sum of all these harmonics at various amplitudes determine the timbre. The language of Hilbert spaces neatly packages this viewpoint. For simplicity, we assume for the moment that our period is $2\pi = 1/f$, so all of our possible timbres are of the form

$$\cdots + c_{-2}e^{-i2x} + c_{-1}e^{-ix} + c_0 + c_1e^{ix} + c_2e^{i2x} + \cdots$$

where our coefficients c_n give the amplitudes of the nth harmonic with $n \geq 1$: $a_n = i(c_n - c_{-n})$ for the sines and $b_n = c_n + c_{-n}$ for the cosines. How do we interpret this infinite sum? We define the Hilbert space $L^2([0, 2\pi])$ to be the classes of all complex valued, 2π-periodic functions $h(x)$ with finite norm squared

$$||h(x)||^2 = \frac{1}{2\pi}\int_0^{2\pi} |h(x)|^2\,dx \tag{1}$$

and where another such function $g(x)$ is in the same class as $h(x)$ when $||g(x) - h(x)||^2 = 0$. Here the $|\cdot|^2$ in the integrand denotes the usual complex modulus squared: $|a + ib|^2 = a^2 + b^2$ for $a, b \in \mathbb{R}$. For any $h(x)$ representing a class in $L^2([0, 2\pi])$, we define Fourier coefficients

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} h(x)e^{-inx}\,dx$$

and these give rise to an orthonormal decomposition

$$h(x) = \sum_{n=-\infty}^{\infty} c_ne^{inx}$$

which means

$$\lim_{N\to\infty}\left|\left|h(x) - \sum_{n=-N}^{N} c_ne^{inx}\right|\right|^2 = 0.$$

The functions e^{inx} are analogous to the unit vectors $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, $\mathbf{e}_3 = (0, 0, 1)$ in three-dimensional space $\mathbb{R}^3$, and the above orthonormal decomposition is analogous to writing a vector

$$\mathbf{v} = (c_1, c_2, c_3) = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3.$$

In fact, the three-dimensional Pythagorean theorem

$$c_1^2 + c_2^2 + c_3^2 = ||\mathbf{v}||^2$$

also holds in our infinite dimensional Hilbert space

$$\sum_{n=-\infty}^{\infty} |c_n|^2 = ||h(x)||^2. \tag{2}$$

Those functions $h(x)$ which describe timbres of vibrating strings are real valued, so the complex conjugate satisfies $\overline{h(x)} = h(x)$, which forces $\overline{c_n} = c_{-n}$; this means our amplitudes are real-valued and for $n \geq 1$ satisfy

$$a_n = -2\cdot\mathrm{Im}(c_n) \qquad\qquad b_n = 2\cdot\mathrm{Re}(c_n)$$

where $c_n = \mathrm{Re}(c_n) + i\mathrm{Im}(c_n)$ with $\mathrm{Re}(c_n), \mathrm{Im}(c_n) \in \mathbb{R}$. If we further assume $h(x)$ is odd, i.e., $h(-x) = -h(x)$, then $b_n = 0 = c_0$ for all $n \geq 1$, which implies that we can write our orthonormal decomposition as

$$h(x) = \sum_{n=1}^{\infty} a_n\sin(nx)$$

where a_n is the amplitude of the nth harmonic. For example, we define the "sawtooth" function $s(x)$ on the interval $[0, 2\pi)$ by

$$s(x) = \begin{cases} 0 & \text{if } x = 0 \\ \dfrac{\pi - x}{2} & \text{if } 0 < x < 2\pi, \end{cases}$$

and then we extend $s(x)$ to $\mathbb{R}$ by 2π-periodicity. See Figure 2. The sawtooth function is real-valued, odd, and has a particularly nice orthonormal decomposition

$$s(x) = \sum_{n=1}^{\infty}\frac{1}{n}\sin(nx).$$

The sound corresponding to the sawtooth does not come from some classical stringed instrument, rather is has a "synthesizer" sound. Indeed, some analog synthesizers, like the Moogs played by Wendy Carlos, use the sawtooth function in *subtractive synthesis*. This means that the synthesizer has a built-in approximation to the sawtooth function and can filter out or amplify various harmonics to produce a wide range of sounds from this one function. In contrast, *additive synthesis* would take various sine/cosine

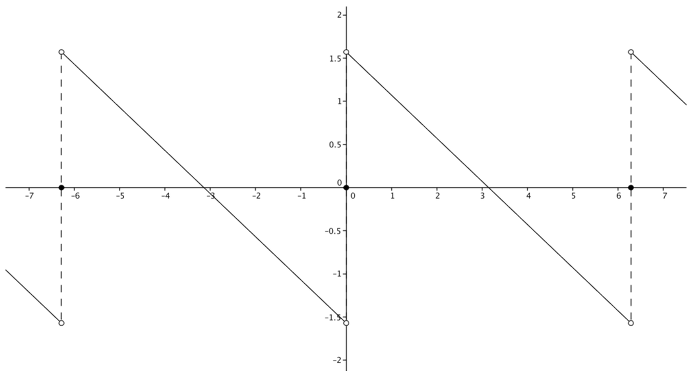

FIGURE 2. Graph of $s(x)$

functions (dial tones) and combine them together with various amplitudes.

A beautiful connection of this sawooth function to number theory comes from plugging $s(x)$ into Equations 1 and 2, which gives

$$\frac{1}{2}\sum_{n=1}^{\infty}\frac{1}{n^2} = \sum_{n=-\infty}^{\infty}|c_n|^2 = ||s(x)||^2$$
$$= \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{\pi - x}{2}\right)^2 dx = \frac{\pi^2}{12}.$$

This recovers Euler's famous solution of the Basel problem: $1/1^2 + 1/2^2 + 1/3^2 + \cdots = \pi^2/6$.

3. The Topological Group of Notes

Now we want to classify frequencies according to the *notes* they determine... but what's a note? Given a frequency f, we have all the pitches coming from harmonics, which correspond to positive integer multiples of f as above, but some of these pitches sound more related to f than others. In particular, the pitches corresponding to f and $2f$ really do sound alike, even though $2f$ corresponds to a higher pitch. In fact, they sound so similar that we consider them to be the same note. For example, if $f = 220$, then $2f = 440$ is our Concert A, so f is also an A note, just with a lower pitch. In general, we say $2f$ is one *octave* higher than f. Likewise, we say the frequency $2^{-1}f$ is one octave lower than f In this way, all the frequencies of the form $2^n f$ for $n \in \mathbb{Z}$ determine the same note. The converse is true as well, which means the frequency of a pitch that we would recognize as determining the same note as f is some number of octaves higher or lower than f. For example, since humans can typically only hear dial tones of frequencies in the range from 20 to 20000 (see [Tip78], section 23-5), the frequencies giving an A note that are audible to humans are precisely 27.5, 55, 110, 220, 440, 880, 1760, 3520, 7040, and 14080.We get a natural equivalence relation on frequencies $f, g \in (0,\infty)$:

$$f \sim g \Leftrightarrow f = 2^n g \text{ for some integer } n.$$

We can now define a note to be the equivalence class of a frequency $[f] := \{g \in (0,\infty) : f \sim g\}$. Thus [440] IS the A note in the sense that this class is literally the set of frequencies which determine an A note. We define the note space $N_2 := \{[f] : f \in (0,\infty)\}$ where the subscript 2 indicates that two frequencies are related precisely when their quotient is an integer power of 2. There are infinitely many notes in N_2 of course, since every note is of the form $[f]$ for some unique $f \in [1,2)$.

Our note space N_2 has more relevant structure than just being a set of notes. The interval of positive real numbers $(0,\infty)$ is a group under multiplication, i.e., the products and multiplicative inverses of positive real numbers are positive real numbers. The set $\langle 2\rangle$ of integer powers of 2 is a subgroup of $(0,\infty)$, so our note space is a quotient group $N_2 \cong (0,\infty)/\langle 2\rangle$ with the product of notes given by $[f]\cdot[g] = [f\cdot g]$.

We also want to make sense of the distance between two notes. Since N_2 is in one-to-one correspondence with the interval $[1,2)$ where $[1] = [2]$, we should think of N_2 as a line segment with its endpoints identified. In other words, N_2 should be a circle. Indeed, there is a way of making this precise. The unit circle $\mathbb{S}^1$ in the complex plane $\mathbb{C}$ is a nice model of a circle since it is naturally a group under the usual multiplication of complex numbers and it can be described very simply as the set of complex numbers of the form $e^{i\theta}$ with $\theta \in [0,2\pi)$. These $e^{i\theta}$ are those complex numbers with distance one from the origin and which make an angle of θ with the positive real axis. Actually, we can take this θ to be any real number since the function e^{ix} is periodic in x with period 2π, i.e., adding or subtracting $2\pi = 360^\circ$ to the angle of a point on $\mathbb{S}^1$ does not change the point. If $[f] = [g]$, then $g = 2^n f$ for some $n \in \mathbb{Z}$, so the properties of logarithms imply

$$\log_2(g) = \log_2(2^n f) = n + \log_2(f),$$

and the periodicity of the exponential further gives

$$e^{2\pi i \log_2(g)} = e^{i(2\pi n + 2\pi \log_2(f))} = e^{2\pi i \log_2(f)}.$$

Therefore we have a well-defined map from our note space to the unit circle

$$N_2 \longrightarrow \mathbb{S}^1 : [f] \mapsto e^{2\pi i \log_2(f)}.$$

Moreover, this map is an isomorphism, meaning that it is a one-to-one correspondence which additionally preserves the group structures of N_2 and $\mathbb{S}^1$:

$$e^{2\pi i \log_2(f\cdot g)} = e^{2\pi i \log_2(f)} e^{2\pi i \log_2(g)}.$$

There is an obvious notion of distance in $\mathbb{S}^1$ obtained by simply considering the distance of points in the complex plane. We wish to consider a different (though topologically equivalent) measure of the distance between two points, namely, we define the distance between P and Q on $\mathbb{S}^1$ to be the radian measure θ of the non-obtuse angle formed by the segments joining P, Q with the origin. Without loss of generality, we have $P = e^{i\theta_1}$ and $Q = e^{i\theta_2}$ with $0 \le \theta_1 \le \theta_2 < 2\pi$, so the distance between P and Q is the minimum of $\theta_2 - \theta_1$ and $2\pi + \theta_1 - \theta_2$ as in Figure

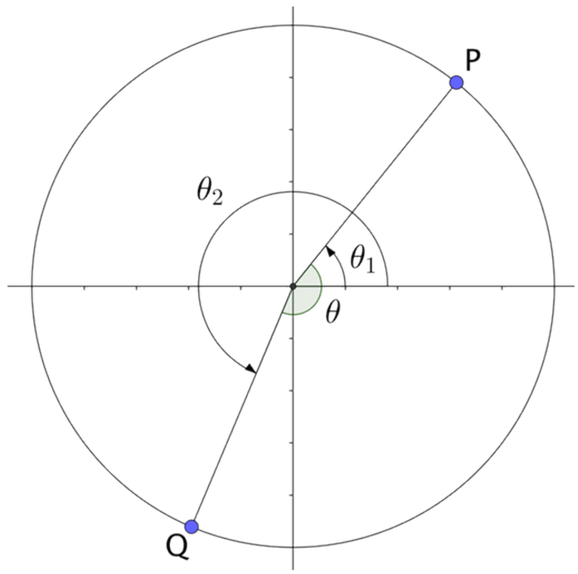

FIGURE 3. Distance between P and Q is θ

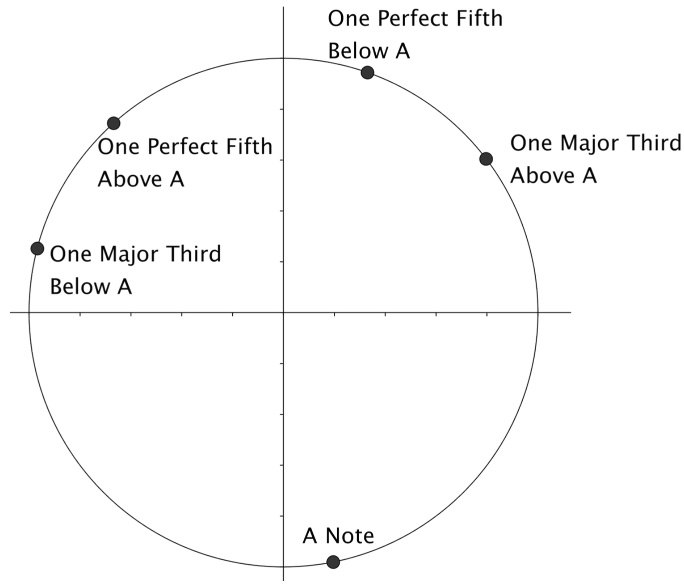

FIGURE 4. Perfect Fifth and Major Third

3. This distance turns $\mathbb{S}^1$ into a topological group, which means that multiplication and inversion are continuous operations: given convergent sequences on $\mathbb{S}^1$, say $P_n \to P$ and $Q_n \to Q$ as $n \to \infty$, then

$$P_n \cdot Q_n \to P \cdot Q \quad \text{and} \quad P_n^{-1} \to P^{-1}$$

as $n \to \infty$. We can give our note space this inherited distance, and then N_2 becomes a topological group which is isomorphic to $\mathbb{S}^1$.

4. THE PERFECT FIFTH

We know from the above that every note has a unique frequency in the range $[1, 2)$ and that this range represents one octave or, equivalently, one time around the circle $\mathbb{S}^1$. In fact, any octave range $[f, 2f)$ represents one time around the circle, and we may regard multiplication by the note $[f]$ as a counterclockwise rotation of the circle through an angle of $2\pi \log_2(f)$ radians. What about the non-octave harmonics? For example, neither $3f$ nor $5f$ are octaves of f, but both are natural frequencies to consider along with f. We have $[3f] = [(3/2)f]$ and $[5f] = [(5/4)f]$ with $3/2, 5/4 \in [1, 2)$. Thus $(3/2)f$ and $(5/4)f$ are frequencies lying strictly within the octave range $[f, 2f)$, and both sound pleasant when played with f. We say $(3/2)f$ is one *prefect fifth* above f and $(5/4)f$ is one *major third* above f, so we get two "new" notes. We could also consider the frequencies $(4/3)f$, $(8/5)f$ which are a perfect fifth and major third below $2f$. Again we get two more new notes.We have plotted these notes on the unit circle in Figure 4 in the case that $[f] = [440]$ is A. What about higher harmonics? For example, $[6f] = [(3/2)f]$, $[10f] = [5f] = [(5/4)f]$, etc., are not a new notes,but $[7f] = [(7/4)f]$, $[11f] = [(11/8)f]$, etc., are new notes. Of course, there are infinitely many notes coming from harmonics of f since there are infinitely many primes. These more complex ratios $(7/4)f$, $(11/8)f$, ..., sound less consonant when played with f. For this reason, we first consider the problem of trying to construct a scale using only octaves and perfect fifths. In other words, we want a musical scale (collection of frequencies) such that if f is in the scale, then so are the octave $2f$ and the perfect fifth $(3/2)f$ because these are the two most pleasant sounding ratios.

One problem with this idea is that the number of new notes in N_2 we will obtain is infinite! To see why this is so, let's begin with the frequency of $f = 1$; we may do this since we could later rotate the circle $\mathbb{S}^1$ so that our scale begins with any frequency we like. Taking octaves gives a sequence 1, 2, 4, 8, ..., which all represent the same note, but taking perfect fifths gives a sequence

$$\begin{array}{lll} (3/2)^0 = 1, & (3/2)^1 = 3/2, & (3/2)^2 = 9/4, \\ (3/2)^3 = 27/8, & (3/2)^4 = 81/16, & \ldots \end{array}$$

which we will now show all represent distinct notes. Two of these frequencies $(3/2)^m$ and $(3/2)^n$ for nonnegative integers $m < n$ represent the same note if and only if the corresponding points on the circle are equal, i.e.,

$$e^{2\pi i m \log_2(3/2)} = e^{2\pi i n \log_2(3/2)}. \tag{3}$$

Equation 3 is equivalent to $(n-m)\log_2(3/2)$ equaling some integer k, but this leads to $3^{n-m} = 2^{k+n-m}$, contradicting the fundamental theorem of arithmetic which states that positive integers have unique prime factorizations up to ordering of factors. This also proves that $\log_2(3/2)$ is an irrational number. This, in turn, is equivalent to the fact that the subgroup $\langle [3/2] \rangle$ of the note space N_2 generated by $[3/2]$ is infinite. More generally, $\langle [f] \rangle$ is infinite if and only if $\log_2(f)$ is irrational.

We know that we cannot stack perfect fifths indefinitely, but when should we stop? Again beginning with a frequency $f = 1$, we get notes

$$[1], [3/2], [9/8], [27/16], [81/64], \ldots$$

where we have chosen frequency representatives to lie in the interval $[1, 2)$. The first five notes with these representatives satisfy the inequalities

$$1 < 9/8 < 81/64 < 3/2 < 27/16 < 2 \tag{4}$$

and the distances (i.e., angle measures) between the consecutive notes on $\mathbb{S}^1$ which these frequencies determine range from about 1.06767 to 1.54. Something interesting happens with the representative of the sixth note:

$$[(3/2)^5] = [243/128] \text{ with } 9/8 < 243/128 < 2.$$

This means the sixth note is closer to the first note than any of the previous four. The distance on $\mathbb{S}^1$ between the first and sixth note is around 0.472417 which is significantly less than the gaps between the first five notes. The sixth note does not represent a great approximation to first note because humans can distinguish between two notes which are a distance of more than 0.026 apart (see [Loe06]), although the ability to discern different pitches varies from person to person and also depends on both amplitude and timbre.

We can proceed in two ways. First, we could just pretend the sixth note agreed with the first, i.e., we are pretending that the frequency $243/128 = 1.8984375$ is 2. Then we would just repeat the arrangement of the frequencies from Eq. 4 in every octave to create our scale:

$$\ldots < 3/4 < 27/32 < 1 < \ldots < 2 < 9/4 < 81/32 < \ldots$$

Although we have true perfect fifths above some frequencies like 1 since (3/2 is in our scale), we do not for others like $81/64$ (since $243/128 \neq 2$ is not in our scale). Additionally, we cannot always translate a piece of music written in this scale from one key to one another.

Alternatively, we could distribute the error between the sixth note and the first among all the five prior notes. One way to do that is to make the gaps between consecutive notes the same. In other words, each note on $\mathbb{S}^1$ would be a distance of $2\pi/5 \approx 1.256637$ from its neighboring notes. Five equally spaced notes are of the form $e^{2\pi i(k/5)} \leftrightarrow [2^{k/5}]$ for $k = 0, 1, 2, 3, 4$. Now the sixth note $e^{2\pi i(5/5)} = 1 = e^{2\pi i(0/5)}$ is the same as the first, but the approximation to the perfect fifth $[3/2] \overset{?}{\approx} [2^{3/5}]$ is not great since the distance on $\mathbb{S}^1$ is

$$2\pi|\log_2(3/2) - 3/5| \approx 0.0944834,$$

which is above the lower bound 0.026 where humans can distinguish notes. This equal distribution of notes, called an *equal temperament*, has a major added benefit: we can transpose any piece of music in any way we like. This means that a piano score written in one key can easily be shifted to another key while preserving all the relative

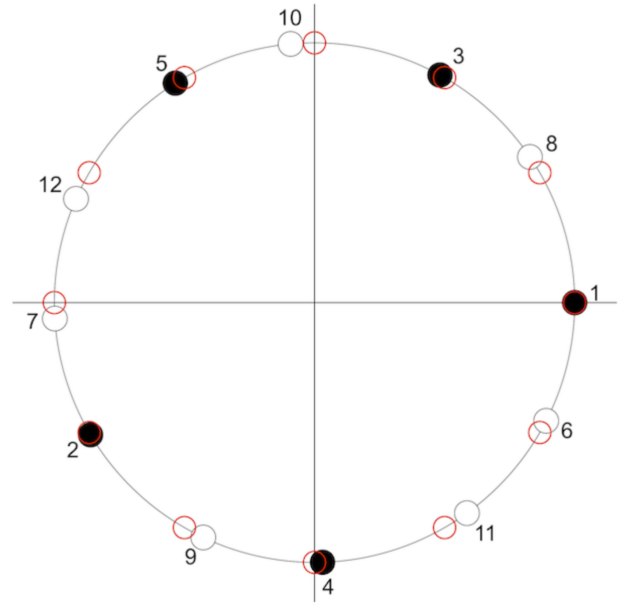

FIGURE 5. Stacked Perfect Fifths (black and white dots) and Equally Spaced Approximations (red circles)

distances of notes in the work. As long as $f = 1$ is in our scale (by appropriately scaling our units of frequency), this transposition amounts to multiplying every frequency in the score by a fixed frequency in our scale. In algebraic terms, this means that our set of five equally spaced notes forms a finite subgroup of the note space N_2. In particular, the notes are invertible as well:

$$e^{2\pi i(1/5)}e^{2\pi i(4/5)} = 1 \qquad e^{2\pi i(2/5)}e^{2\pi i(3/5)} = 1,$$

so we can go up or down any number of approximate perfect fifths we like. Scales with five notes per octave are called *pentatonic*. We have just described two pentatonic scales, one equally tempered and the other not. The non-equally tempered scale had true perfect fifths but was not transposable, while the equally tempered scale was transposable but did not have a great approximation to the perfect fifth. One way to get the best of both worlds is to increase the number of notes in our equally tempered scale to get a better approximation to the perfect fifth.

If we continue stacking true perfect fifths beyond the sixth, then the next time we get closer to where we started is on the thirteenth note, which has a distance on $\mathbb{S}^1$ of around 0.12284 from the first note. Now we distribute this error among all the twelve prior notes by considering an equally tempered scale with 12 notes. Here our notes on $\mathbb{S}^1$ are all of the form $e^{2\pi i(k/12)}$ for $k = 0, 1, \ldots, 11$, and our approximation to the perfect fifth becomes

$$2\pi|\log_2(3/2) - 7/12| \approx 0.01023636$$

which is now below the lower bound 0.026 where humans can distinguish notes. In other words, we would perceive the pitches $\log_2(3/2)$ and $7/12$ to be the same, so this is an excellent approximation. In Figure 5, we have plotted

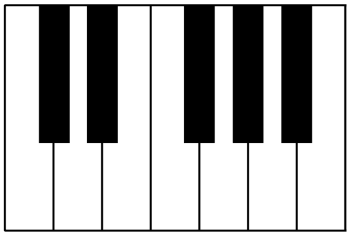

FIGURE 6. One Octave on a Standard Keyboard

the 12 notes from this equally tempered scale alongside the stacked true perfect fifths that each note represents. Notice how close note number 2 (true perfect fifth) is to its approximation. This first five stacked perfect fifths are colored black, and the last seven are colored white. Coloring the notes this way allows us to see that our equally tempered 12 note scale contains an approximation to the non-equally tempered pentatonic scale we constructed earlier. We also get a non-equally tempered seven note, or *heptatonic*, scale from the white notes. If we cut and straighten our note space/circle $N_2 \cong \mathbb{S}^1$, then this arrangement of black and white notes corresponds precisely to the keys in one octave of a standard keyboard as seen in Figure 6. Since the 12 note equal temperament scale is so prevalent, the distance between notes or frequencies is often measured in *cents*:

$$1200 \text{ cents} = 2\pi.$$

These units of cents are chosen so that consecutive frequencies in this scale are exactly 100 cents apart. In these units, humans can not distinguish frequencies less than 5 cents apart, and the 12 note equally tempered scale has a decent, but not great, approximation to the major third (see note number 5 in Figure 5):

$$2\pi|\log_2(5/4) - 4/12| \approx 0.07166 \approx 13.686 \text{ cents.} \tag{5}$$

What if we wanted better approximations to the perfect fifth and major third? We could add more notes. The next time our perfect fifth stacking gets closer to the first note than all of the previous is on the forty-second note. This would give us 41 notes in each octave! Where do these numbers of notes 5, 12, 41 come from? Could we have found them in a simpler way?

5. Continued Fractions

We want an algorithm to determine the numbers n of stacked perfect fifths such that $[(3/2)^n]$ is closer to $[1]$ than any previous note. For a positive integer b, the distance between $[(3/2)^b]$ and $[1]$ is

$$2\pi|b\log_2(3/2) - a| \tag{6}$$

where a is the unique positive integer which ensures this quantity is less than π, i.e.,

$$|b\log_2(3/2) - a| < 1/2,$$

so such an a minimizes the expression 6. Here a/b is a rational approximation to the irrational number $\log_2(3/2)$ which becomes arbitrarily better by choosing a sufficiently large denominator b or, equivalently, choosing more notes per octave. Ideally, we would like to do this in an efficient way, which means getting a good approximation while using the minimal number of notes. Equivalently, we want to find positive integers n and m such that

$$|b\log_2(3/2) - a| > |n\log_2(3/2) - m|$$

whenever $0 < b < n$ and a is any integer. In this way, m/n will be the "best" rational approximation to $\log_2(3/2)$ having denominator less than or equal to n. Such best approximations are characterized as being convergents of continued fractions. As we will see, these continued fractions and their convergents can be successively computed in a systematic way.

For a real number a_0 and positive real numbers a_1, ..., a_k, we define the *finite continued fraction*

$$[a_0; a_1, a_2, \dots, a_k] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots \cfrac{1}{a_k}}}.$$

We define an *infinite continued fraction* to be the limiting value of finite continued factions:

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}} := \lim_{k\to\infty} [a_0; a_1, a_2, \dots, a_k].$$

It turns out that this limit exists if and only if the infinite sum $\sum_{i=0}^{\infty} a_i$ diverges (see [Khi92]). We assume now that all the a_i are integers, so the infinite continued fraction always exists in this case. In fact, much more is true.

Theorem 1. *Let α be an irrational real number. There is a unique infinite* continued fraction expansion

$$\alpha = [a_0; a_1, a_2, \dots]$$

such that a_0 is an integer and $a_1, a_2, \dots$ are positive integers.

The uniqueness claim in this theorem is not difficult to show. For example, one might proceed as follows: $a_0 = \lfloor \alpha \rfloor$ must be the *floor* of α, i.e., the greatest integer which is less than α since we would have

$$\alpha - a_0 = [0; a_1, a_2, \dots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}} \in (0, 1).$$

Likewise, we must also have

$$a_1 = \left\lfloor \frac{1}{\alpha - a_0} \right\rfloor \qquad a_2 = \left\lfloor \cfrac{1}{\cfrac{1}{\alpha - a_0} - a_1} \right\rfloor$$

and so on. Hence the real content of the theorem is proving that this sequence of a_i's so-defined actually converges to the irrational α you started with. That is indeed the case (see [Khi92]), and so these digits a_i in the continued fraction expansion are somewhat like decimals in a decimal expansion in the sense that computing more digits gives you a progressively better rational approximation. Here are a few examples of continued fraction expansions of irrational real numbers:

$$\begin{aligned}\sqrt{11} &= [3; 3, 6, 3, 6, 3, 6, 3, 6, 3, 6, 3, ...]\\ e &= [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, ...]\\ \pi &= [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, ...]\end{aligned}$$

The suggested patterns in the expansions of $\sqrt{11}$ and e continue forever, but there is no known pattern in the continued fraction expansion of π.

We define the *kth convergent* of a real irrational

$$\alpha = [a_0; a_1, a_2, ...] \tag{7}$$

to be the value of the finite continued fraction

$$[a_0; a_1, a_2, \ldots, a_k].$$

The 0th convergent is $[a_0] = a_0/1$, the first convergent is $[a_0; a_1] = (a_1 a_0 + 1)/a_1$, and the second convergent is

$$[a_0; a_1, a_2] = \frac{a_2(a_1 a_0 + 1) + a_0}{a_2 a_1 + 1}.$$

In general, one can prove that the kth convergent equals p_k/q_k where

$$\begin{aligned}p_k &= a_k p_{k-1} + p_{k-2}\\ q_k &= a_k q_{k-1} + q_{k-2}\end{aligned} \tag{8}$$

with the initial conditions

$$p_0 = a_0, \quad p_1 = a_1 a_0 + 1, \quad q_0 = 1, \quad q_1 = a_1. \tag{9}$$

These numerators p_k and denominators q_k of convergents give us precisely the kind of best approximations we are looking for.

Theorem 2. *Given an irrational real number α, the convergents p_k/q_k (as determined by Eq.s 7, 8, and 9) are in lowest terms and represent best approximations in the following sense:*
For any integers a, b with $0 < b < q_k$, we have

$$|b\alpha - a| > |q_k\alpha - p_k|.$$

Conversely, if m, n are integers with $n > 0$ such that

$$|b\alpha - a| > |n\alpha - m|$$

whenever a, b are integers with $0 < b < n$, then $m = p_k$ and $n = q_k$ for some k.

This theorem (proven in [Khi92]) tells us that the number n of notes we should divide the octave into will be the denominator q_k of some convergent in the continued fraction expansion

$$\log_2(3/2) = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{3 + \cdots}}}}}$$

In general, the convergents of α satisfy the ineqaulities

$$\frac{p_2}{q_2} < \frac{p_4}{q_4} < \frac{p_6}{q_6} \ldots < \alpha < \ldots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1},$$

and for $\alpha = \log_2(3/2)$ we have

$$\frac{1}{2} < \frac{7}{12} < \frac{31}{53} < \ldots < \log_2(3/2) < \ldots < \frac{24}{41} < \frac{3}{5} < \frac{1}{1}.$$

Here we see the denominators 5, 12, 41, 53, ..., reflecting the numbers of notes needed per octave in order to get the successively next best approximation to the perfect fifth using equal temperament. The numerators tell us which note in the equally tempered scale approximates the perfect fifth. For example, the convergent $7/12$ tells us that the approximation to a perfect fifth in the equally tempered 12 note scale occurs 700 cents above a given frequency. Likewise, the convergent $24/41$ tells us that the approximation to a perfect fifth in the equally tempered 41 note scale occurs $1200 \cdot (24/41) = 702.\overline{43902}$ cents above a given frequency.

What about an equally tempered heptatonic scale? Our equally tempered 12 note scale contains a non-equally tempered heptatonic scale (7 white keys), but 7 is not the denominator of a convergent for $\log_2(3/2)$, so there is no reason to expect an equally tempered heptatonic scale to be desirable. However, it is a somewhat desirable scale since 7 is a denominator in one of the intermediate fractions

$$\frac{0 \cdot 3 + 1}{0 \cdot 5 + 2} = \frac{1}{2} < \frac{1 \cdot 3 + 1}{1 \cdot 5 + 2} = \frac{4}{7} < \frac{2 \cdot 3 + 1}{2 \cdot 5 + 2} = \frac{7}{12}$$

In general, an *intermediate fraction* is of the form

$$\frac{tp_{k-1} + p_{k-2}}{tp_{k-1} + p_{k-2}} \qquad \text{where } 0 \le t \le a_k$$

These are also best approximations, but in a weaker sense than Theorem 2: if r, s are integers with $s > 0$ such that

$$|\alpha - a/b| > |\alpha - r/s|$$

whenever a, b are integers with $0 < b < s$, then r/s is an intermediate fraction. For this to be true as stated we need to include the index $k = -1$: $p_{-1} = 1$, $q_{-1} = 0$.

We should pay attention to the fact that our decent major third approximation in Eq. 5 was essentially coincidental. We focused solely on approximations to the perfect fifth, but there are techniques involving so-called ternary continued fractions which give constructions for finding simultaneous rational approximations to two or more irrational real numbers. See [Bar48] for details on the applications to music.

6. Xenharmonic Scales

Why did we have to start with an octave interval $[1, 2)$? In other words, why must we identify the frequencies f and $2f$ as representing the same note? We could choose, for instance, to identify f with higher harmonics like $3f$ or $5f$, or we could work with intervals smaller than one octave. Scales which are not based on the octave interval are called *xenharmonic*. For example, in the article [Car87], Wendy Carlos started instead with the perfect fifth interval $[1, 3/2)$, and then considered dividing this into equal pieces. Wendy's approach for dividing this perfect fifth interval was largely experimental. She picked some notes (including the major third) that she wanted to be well approximated and then gradually incremented step sizes and computed (minus) the total squared deviations between the ideal frequencies and the approximations thereof. She found the following desirable divisions

step sizes	number of notes
$\alpha = 77.995...$ cents	9
$\beta = 63.814...$ cents	11
$\gamma = 35.097...$ cents	20

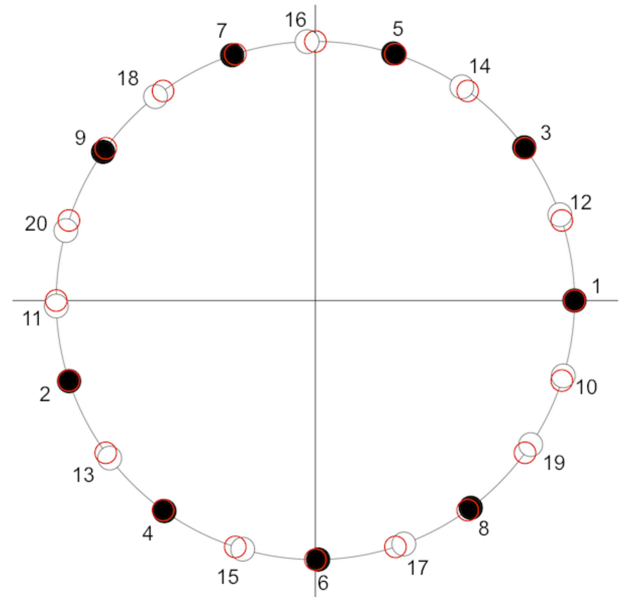

Figure 7. Stacked Major Thirds (black and white dots) and Equally Spaced Approximations (red circles)

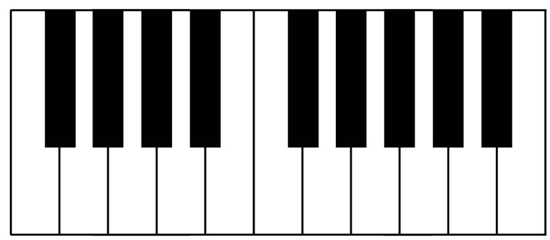

Figure 8. One Perfect Fifth on Wendy's Keyboard

All three of these scales fit beautifully into our continued fraction apparatus if we simply replace the octave and perfect fifth with the prefect fifth and the major third, respectively. To see why this is, we first note that our equivalence relation now becomes

$$f \sim g \Leftrightarrow f = (3/2)^n g \text{ for some integer } n,$$

so we get a new note space $N_{3/2}$ consisting of these equivalence classes $[f]$. Again, the note space is isomorphic to the unit circle $\mathbb{S}^1$ as topological groups:

$$N_{3/2} \longrightarrow \mathbb{S}^1 \colon [f] \mapsto e^{2\pi i \log_{3/2}(f)}.$$

Now every note $[f]$ has a unique representative in $[1, 3/2)$. The most significant frequency lying strictly within the interval $[1, 3/2)$ is $5/4$, coming from the major third. To use the tools of Section 5, we must consider rational approximations of $\log_{3/2}(5/4)$ coming from the continued fraction expansion

$$\log_{3/2}(5/4) = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{2 + \cfrac{1}{6 + \cdots}}}}}.$$

In this case, we have the inequalities of convergents

$$\frac{1}{2} < \frac{11}{20} < \frac{82}{149} < ... < \log_{3/2}(5/4) < ... < \frac{71}{129} < \frac{5}{9} < \frac{1}{1}.$$

Immediately, we see the α and γ scales coming from the denominators 9 and 20, respectively. The β scale can be seen in the denominator 11 of an intermediate fraction:

$$\frac{0 \cdot 5 + 1}{0 \cdot 9 + 2} = \frac{1}{2} < \frac{1 \cdot 5 + 1}{1 \cdot 9 + 2} = \frac{6}{11} < \frac{2 \cdot 5 + 1}{2 \cdot 9 + 2} = \frac{11}{20}$$

Therefore this $9 + 11 = 20$ division of the perfect fifth is in striking analogy to the $5 + 7 = 12$ division of the octave. It is worth mentioning that in any of the equally tempered worlds α, β, or γ, the perfect fifths are true perfect fifths, we get good (and sometimes great) approximations to the major third, but the octave is no longer a priority! For instance, the error in the approximation to the octave in the gamma scale is given by ≈ 6.6765 cents, which is not great at all considering we are using around 34 notes per octave.

If we follow our approach in Section 4, we can stack major thirds around a perfect fifth circle. In analogy with the equally tempered 12 note scale, we color the first 9 notes black, and the remaining 11 notes white. The result is seen in Figure 7 Again, we can cut and straighten our note space/circle $N_{3/2} \cong \mathbb{S}^1$, and this arrangement of black and white notes corresponds precisely to the keys in one perfect fifth of Wendy Carlos' γ scale keyboard as seen in Figure 8.

REFERENCES

[Bar48] J. M. Barbour, *Music and ternary continued fractions*, The American Mathematical Monthly **55** (1948), no. 9, 545–555.

[Ben08] Dave Benson, *Music: A mathematical offering*, 2008, Available at http://www.maths.abdn.ac.uk/, pp. 517+xiv.

[Car87] Wendy Carlos, *Tuning: At the crossroads*, Computer Music Journal **11** (1987), no. 1, 29–43.

[DM99] Edward Dunne and Mark McConnell, *Pianos and continued fractions*, Mathematics Magazine **72** (1999), no. 2, 104–115.

[Khi92] A. Ya. Khinchin, *Continued fractions*, third ed., Dover, 1992.

[Loe06] D.B. Loeffler, *Instrument timbres and pitch estimation in polyphonic music*, Master's thesis, Georgia Tech, April 2006, Department of Electrical and Computer Engineering.

[Tip78] Paul E. Tippens, *Applied Physics*, second ed., McGraw-Hill, 1978.