The birth of the blues: how physics underlies music

To cite this article: J M Gibson 2009 Rep. Prog. Phys. 72 076001

View the article online for updates and enhancements.

Related content
- The Everyday Physics of Hearing and Vision: Hearing: the perception of sound
  B Mayo
- Musician's and physicist's view on tuning keyboard instruments
  Martin Lubenow and Jan-Peter Meyn
- The physics of music
  S R Hoon and B K Tanner
The birth of the blues: how physics underlies music

J M Gibson

Argonne National Laboratory, 9700 Cass Avenue, Argonne IL 60439, USA
E-mail: jmgibson@aps.anl.gov

Received 15 May 2008, in final form 12 February 2009
Published 30 June 2009
Online at stacks.iop.org/RoPP/72/076001

Abstract
Art and science have intimate connections, although these are often underappreciated. Western music provides compelling examples. The sensation of harmony and related melodic development are rooted in physical principles that can be understood with simple mathematics. The focus of this review is not the better known acoustics of instruments, but the structure of music itself. The physical basis of the evolution of Western music in the last half millennium is discussed, culminating with the development of the ‘blues’. The paper refers to a number of works which expand the connections, and introduces material specific to the development of the ‘blues’. Several conclusions are made: (1) that music is axiomatic like mathematics and that to appreciate music fully listeners must learn the axioms; (2) that this learning does not require specific conscious study but relies on a linkage between the creative and quantitative brain and (3) that a key element of the musical ‘blues’ comes from recreating missing notes on the modern equal temperament scale. The latter is an example of ‘art built on artifacts’. Finally, brief reference is made to the value of music as a tool for teaching physics, mathematics and engineering to non-scientists.

1. Introduction

The ‘blues’ is a unique 20th century musical form that emerged in America, birthed jazz and rock music, and continues to hold audiences today. If there are three essential characteristics of the blues, they are the ‘life is tough but I’ll survive’ lyrics; the ‘swing’ rhythm and the melodic ‘bent’ blue notes on the electric guitar. The name of the ‘blues’ has attached to these characteristics—the ‘blues’ is a feeling and the ‘blues’ is a musical scale. The blues emerged from African and Western music melting in the American pot. In this rich cultural history, there are key players whose names are rarely mentioned—physics and mathematics. The palette on which blues is written—the very blue note itself, while emerging as a mysterious rich sound to the American audiences familiar with classical 19th century music, was a simple solution to the...
simplest equation of all. To understand this, we will need to review the origin of western musical scales and harmony, and in doing so reveal the overpowering role of mathematics and physics in musical interpretation.

Examining the development of western musical scales, the review starts with Pythagoras, covers some other tunings and ends at the ubiquitous modern-day equal temperament scale. These scales were developed to bring out straightforward harmonic relationships between notes. Harmonic consonance is based on the overlapping of overtones shared by two or more notes, which often means that the frequency ratio of the notes is close to a simple rational fraction. The desire of composers to have a wider palette of tones and harmonies with which to compose is seen as the driving force for the evolution of scales. In the process, one can point to interesting musical developments and artifacts. Many artifacts became accepted as part of the treasured musical heritage. The blues, at least as implemented on the piano, is an example of this.

It appears that music involves the development of a set of rules shared between the composer and the listener. These rules may not be consciously known, but appear to be the context in which the listener can appreciate the composer’s art. In this sense music is alike to mathematics, and the rules play the role of axioms. It is difficult to understand how this might work in the listener without assuming a strong connection between the analytical and the creative parts of the brain (no longer clearly associated with the ‘left’ and the ‘right’ sides (Fink et al 1996)).

Music is not only dependent on mathematics and physics, but it makes an excellent tool for teaching these subjects to non-scientists. In the final section of the review this subject is explored briefly, specifically as music can be related to the wave-particle duality, quantum mechanics and other phenomena through the use of harmonic analysis.

2. Musical scales

Music seems to have existed in human culture for at least ten thousand years (Hoffecker 2005). This review does not address the origins of music, but it is necessary to identify some key aspects which underpin the attributes of modern music. When suitably under-damped objects are struck or plucked they can give out clear sounds. Today we consider sounds to be ‘notes’, e.g. C or D, when they have a well-defined frequency, known in music as ‘pitch’. If when the object is struck there is an unclear resonance or if the damping is high, then the tone may not have recognizable pitch. Early in human history, it presumably became clear that when two notes were played together, certain combinations were more appealing than others. It is now understood that appealing combinations are in most cases those whose overtones are coincident in frequency. A musical scale is a sequence of notes, and harmony occurs when two or more notes are played at the same time. The notes on the scale are chosen to favor good harmonic combinations.

To understand the choice of notes for a musical scale, we will examine a vibrating string. Figure 1 illustrates the simplest overtones of a string. An idealized plucked string which is fastened at both ends contains overtones at integer multiples of the fundamental frequency $\omega$. This particular string is shaped so that it contains only the first three harmonics—the fundamental and two overtones. Imagine now that we pluck two strings with different fundamental frequencies at the same time. The trivial case when each string has the same fundamental frequency is known as ‘unison’. When the two strings have fundamental frequencies $\omega$ and $2\omega$ they sound very consonant, and comprise an ‘interval’ (a musical term for the relationship between any two notes) that is called the octave. The strong consonance is associated with the common audible overtones. Adding (or subtracting) an octave to (or from) a note is thus equivalent to doubling (or halving) its frequency. If two strings with fundamental frequencies $\omega$ and $3\omega$ are played together, they will also sound harmonic. By halving the frequency of the upper note ($3\omega$) so that it falls into the first octave, $3\omega$ one obtains an interval with frequency ratio $3/2$. This interval is known in music as a perfect fifth and is considered the most profound and simple harmony after unison and the octave. For this interval the 1st overtone of one note matches the 2nd of the other. Consonances like the perfect fifth are the basis of harmony, which is the focus of this review. Whereas harmony requires several notes played at the same time, melody represents the relationship between notes played in sequence. Melody and harmony are very closely connected, for example, the melodic notes from the opening of Mozart’s ‘Eine Kleine Nachtmusik’ (figure 2) when played together comprise the harmonic major triad of G (G–B–D).

In addition to melody and harmony, rhythm is a third fundamental aspect of music. I will not explore rhythm much in this paper, but the subject is also closely connected to the physics of resonance. For example, the natural rhythm of swinging a tool drove the ‘swing’ which is the bedrock of jazz. African slaves originated the blues, as they were forced to work in the fields swinging heavy tools. That a major part of the African contribution to jazz was rhythmic is not
surprising, since African music is widely recognized as the most sophisticated polyrhythmic music (Schuller 1968).

Consonance is based on overlap of overtones, which typically means that intervals are in low integer fractional ratios, since strings and pipes have integer harmonics. Even the human voice has integral overtones, which may explain the pervasiveness of integral resonances in music around the world. However, there are cultures where instruments use metal bars or solid pieces of wood as resonators, for example the Indonesian ‘Gamelan’ orchestra. The overtones of a rectangular bar are non-integer in the sequence 1 : 2.76 : 5.49 : 8.93. The scales and intervals of Gamelan music reflect these ratios (Sethares 2005).

The overlapping overtones aspect of harmony is exploited when tuning instruments. The beats between overtones enables a precise way of tuning two notes separated by a harmonic interval—estimates are that a relative accuracy of better than 0.5% can be attained by ear (Olson 1957).

2.1. Pythagorean scale

Having understood the simple harmonic intervals, we now turn our attention to constructing a musical scale. The mathematical nature of the modern western musical scale can be traced back to Pythagoras around 500 BC. The Greek philosopher and mathematician was the first known to have experimented with the relationship between different musical sounds. He found that strings with the same tension but integer ratios of their lengths sounded harmonic when plucked together. We understand today that this is due to the coincidence of overtones of the waves produced by the strings (thanks to Mersenne and Helmholtz (1636, 1863), the latter being the first to explain using waves). However, Pythagoras, not understanding waves or the direct relationship between pitch and frequency, thought that harmony arose from a more fundamental aspect of the mathematical ratio of the lengths of strings. The ratios 2/1, 3/2, 2/3 and 4/3 were the key harmonic intervals identified by Pythagoras. Pythagoras experimented with an instrument called the monochord—a single string which could be divided into two by a movable bridge and the two sides independently plucked to give two notes. Tension is constant and does not affect relative pitch of the notes made by the two parts of the monochord string. The pitch of each part of the monochord is inversely related to its length. Pythagoras explored the harmony obtained by plucking the two parts together. When the string is divided into two, then the notes are the same (unison). When the ratio is 2 : 1, one creates an octave. The ratio 3 : 1 produces a fifth plus an octave. The ratio 3 : 2 makes a perfect fifth.

Figure 3 shows notes on a piano keyboard and on the treble stave for reference. The sequence C, D, E, F, G, A, B, the white notes on the piano, makes the diatonic scale of C major (‘do, re, mi, fa, so, la, ti’). We have explained that when two notes are juxtaposed they make a musical ‘interval’, but we must explain the naming convention for intervals. Intervals are named by the number of notes of the scale between the lower and upper notes, counting both notes themselves. Therefore the smallest interval, e.g. C–D is a second. The interval C–E (or D–F) makes a third, C–F a fourth and C–G a fifth. The qualifiers perfect, major, minor, diminished and augmented which are used to preface interval names will be explained later (see also the appendix).

One can generate, through the approach used by Pythagoras, the diatonic 7-note scale that is the foundation of Western music. The simple algorithm begins with the note F whose frequency we denote as $v_F$ (a practical value for this frequency is identified in table 1). One makes the perfect 5th of F, which has a frequency of $3/2 v_F$. This note is an octave higher than we want, so we will divide by 2 to give the 2nd note C on our scale, with $v_C = 3/4 v_F$. Continuing, a perfect fifth from C gives our third note G at $9/8 v_F$. By using this process seven times, and dividing by 2 as necessary to keep the notes within the first octave of the note C, we get the sequence of notes shown in table 1 under the column marked Pythagorean. Even though we started with F, the first note in this series is actually C because that is the only starting note which gives the sequence of intervals of what we call a major scale.

2 The peculiarity of interval naming is attributed to the lack of zero in the counting system of the ancient Greeks.
It is interesting to observe the ratio (step) between adjacent notes on the Pythagorean scale. There are only two distinct step values, the ‘tone’ (T), which has a value of 9/8, and the ‘semitone’ (S), which has a value of 256/243. Observe that the semitone step is a little less than half-size of the tone step (to make this comparison we subtract 1 from each before taking the ratio). The ratio between the semitone and half of the tone is 1 + 139/2187. Another important and related ‘error’ in the Pythagorean tuning system is called the ‘Pythagorean comma’, which is 312/392. Observe that the three major fifths, the nearest and the octave, are where they are. They split the tones in the key of C (see locations in table 1) into approximate semitones, about half the attractive number at which to stop, where as we have said the difference between (3/2)n, and a power of 2 is about 1% when n is 12 (this is the Pythagorean comma), but this difference grows again until n = 19 and then n = 29 (Hartmann 1987). Imagine playing a piano with 29 notes in the octave!

As we discuss further the evolution of the scale and harmony, it will be useful to see the location on the piano keyboard of the overtones of the note C. Figure 4 shows the full list of overtones up to the eighth order, marked on the piano keyboard, showing the fifth G and third E reduced into the first octave.

2.2. Pentatonic scales

If one takes just the first five notes generated through the sequence of fifths described above, beginning with C (these are marked in bold on table 1): C, G, D, A, E (in pitch order that is C, D, E, G, A) then one has the pentatonic scale in
Figure 4. The overtones of the note C and their approximate locations relative to the piano keyboard. The notes E, G and B flat discussed in the text are also shown reduced into the first octave.

Figure 5. An example of pentatonic music from the first two bars of ‘Scotland the Brave’ (trad), in the key of F sharp (in this key the pentatonic scale comprises the black notes on the piano keyboard).

the key of C. This scale is found extensively in the music of many cultures, from Celtic to Chinese. The ratios of the notes are favorable in almost any combination [always reasonable integer ratios], using the just intonation scale where the third is perfect. An example of pentatonic music is shown in figure 5 from traditional Scottish music. The pentatonic scale is also widely used in jazz, one assumes because of its lovely harmonies and its connection with African music.

The black notes on the piano also form a pentatonic scale (in the key of F sharp). This is not a coincidence, because the notes can be thought of as a complete set of notes obtained from consecutive fifths.

2.3. Chords

A chord is a combination of at least three harmonized notes that do not contain unisons or octaves. The simplest chord is a triad made up of three notes. The triad can be viewed as a combination of two intervals between the lowest and middle and the middle and highest notes. The designation major (M) for an interval means that there are 4 semitones (two tones) in it. Minor (m) means only 3 semitones (one tone plus a semitone). Diminished means that an interval is reduced by one semitone (augmented is the opposite). The fifth, exemplified by C–G, is neither major nor minor (it is present in both simple major and minor triads) so it is called perfect.

Exploring harmony begins with enumerating the triads spaced by thirds on the diatonic scale. The three note triads in the key of C are shown in figure 6, using only the notes on the scale and skipping a white note to make each interval a third. The chord on C has a major third (TT) followed by a minor third (ST). This triad is called a major chord (C major). The next chord on D is a minor 3rd (TS) followed by a major 3rd (TT). This triad is called a minor chord (D minor). (The C minor chord would involve C–E flat–G, and since it contains the note E flat is not in the key of C major.) Looking at the patterns in the key of C major made up of the seven white notes on the piano, there are three chords of the seven which are major (C, F and G, sometimes called the root (I), sub-dominant (IV) and dominant (V)). There are also three minor chords (D (II), E (III) and A (VI)) and one ‘special chord’ the diminished triad on B (a minor 3rd and a second minor 3rd). (The Roman numeral system for denoting chords, which has a long history back to the ‘figured bass’ of ancient organ music, is obviously based on the position of the chord’s root in the scale.)

The chord structure of the major diatonic scale (exemplified here in the key of C) leads to the common use of the major chords C, F and G in folk music, popular music and in the blues. These three chords provide the underpinning of a great deal of simple music, and of the twelve-bar blues (I IV I V IV I V I V I V).

The term ‘mode’ describes a sequence of seven consecutive notes. A mode is characterized by the steps (intervals) between adjacent notes. For example, the major diatonic mode is TTSTTTS. We can find most of the modes that are in common use on the C major notes. The sequence C, D, E, F, G, A, B is the diatonic mode. This is the only major mode. The sequence D, E, F, G, A, B, C is called the Dorian mode. It has a root that is minor. The next are the Phrygian (E), Lydian (F), Mixolydian (G), Aeolian (A) and Locrian (B). These modes are sometimes known as the ecclesiastical modes because of their use in medieval church music, but they are also used in modern times in jazz. In general, the major (C) and the Aeolian minor (A) are the most widely used modes. The reason is that these are the only modes
which have a tonic, sub-dominant and dominant, that are of the same character: major (C) or minor (A) (Zweifel 2005). This relationship leads to simple recognizable harmonic structure on which to build music.

3. Baroque music and harmony

By the time of the baroque period of music, beginning in \(\sim 1600\), just intonation produced commonly used scales based on perfect fractions. As we described, in the just tuned instruments, the tonic intervals (i.e. thirds and fifths) are ‘good’ in the basal key, and closely related keys can be acceptable. Most of the fifths in the major diatonic scale are good with the exception of the fifth B–F (in the key of C). During the early baroque period a piece of music was typically confined to a single key, and so the Just tuning system was adequate, in fact the results could sound very harmonic with pure intervals. To demonstrate this effect one can listen to music such as Pachelbel’s canon in D (figure 7), which uses only the simple triads and sounds very good when played on a just tuned instrument in the correct key, but sounds very far out of tune if played in a remote key unless the instrument is retuned.

3.1. Modulation and transposition

With 12 notes to the scale, there are 12 major keys that can be constructed on each of the notes. A piece of music written in the key of C, for example, can be simply transposed or switched to the key of F\# or B flat. If an instrument is tuned by the Pythagorean or just tuning into a particular key, the music will not sound harmonious if transposed into another key unless the instrument is retuned. This is not practical, so the just tuning limited, for example, the ability to transpose music to suit the range of a singer. But there is a more significant limitation to just tuning that has to do with modulation. Modulation is the temporary changing of key within a piece. Modulation allows composers to mix chords from different keys and so add complexity and richness to music. It opens up the ‘palette’ of the composer and allows him or her to produce more complex music following new axioms. Johann Sebastian Bach was an early protagonist for this when he composed the famous Das Wohl-Temperirte Clavier (‘The Well Tempered Clavier’) of 48 preludes and fugues: a prelude and a fugue in each of the 12 major and 12 minor keys for the clavier (a precursor of the piano). The opening bars from one of these pieces, the prelude in C major, is shown in figure 8. This piece is in the key of C major and the notes in the first several bars satisfy that ‘rule’. However, at the place marked by the (A), Bach modulates into the key of D briefly, and returns to C in a few bars. Later he flits into E major at (B). Such modulation would not be palatable without equal temperament or an equivalent tuning. Not only does such tuning allow clear modulation, but also the creative use of ambiguity where the listener is not sure which key the composer intends. A pretty color added to the artist’s palette, and used so beautifully by Bach.

3.2. Wolves and the origin of equal temperament

The just intonation was an improvement over Pythagorean tuning, and led to chords that sounded quite harmonic provided one stayed in the same key. However, in different keys the problem was severe, with scales sounding clearly out of tune. To overcome this, for hundreds of years musicians experimented with means to ‘temper’ tuning, by making it imperfect to some degree, but more perfect in a greater number of keys. (As mentioned above this is equivalent to recipes for distributing the Pythagorean comma.). There were many approaches to this, ‘Meantone Temperament’ for example, and...
Figure 7. The opening of ‘Canon in D’ by Johann Pachelbel (arranged for piano by Robert Schultz (Warner Bros., Miami)) presents a beautiful example of baroque music well-suited for a just tuned instrument in the key of D. (©1992 Beam Me Up Music all rights controlled and administered by Alfred Publishing Co., Inc. All rights reserved. Used by permission of Alfred Publishing Co., Inc.) Three sound files to support this figure are available in the multimedia content associated with this paper, which is available online at stacks.iop.org/rop/72/076001. In the sound file ‘PureMajorD.mp3’ one can hear the Pachelbel’s Canon (figure 7) played in the key of D on an instrument that is tuned in the just tuning for that key. As baroque music, described in the text, it sounds very good played on this scale. Keeping the same tuning, the piece was transposed up one step and played in E Flat major (‘PureMajorEFlat.mp3’). It is atrociously dissonant because the intervals in this key are far from perfect. The third file contains renditions in both keys (D first, then E Flat) in the equal temperament tuning (‘EqualTemperament.mp3’). In this case, while the intervals are very slightly imperfect, the piece sounds good in both keys.

Figure 8. The first few bars of Johann Sebastian Bach’s prelude No 1 in C major from the ‘Well Tempered Clavier vol II’. At the location marked A Bach introduces the chord of D major into C major by modulation, and at B the chord of E major. For these chords to sound harmonic the equally tempered tuning is desired. (©Breitkopf & Härtel KG. Reprinted by permission.)
a good summary is given by McLean (2001). Most such efforts, while improving things for some keys, led to some very poor tunings of intervals in other keys—extremes were known as ‘wolves’ because of the dissonance. To overcome the problems of wolves, musicians argued for centuries over the right temperament. There was even a philosophical problem in that people were extremely reluctant to give up on integer ratios because it was believed that these were essential (as concluded by Pythagoras). It is important to recall at this stage that a particular interval, such as a major third, should represent the same ratio of frequency for its two component notes, irrespective of those notes. The reality is that there is no tuning system, except the equal temperament system, in which this can be always true for all pairs of notes. The problem for the equal temperament system is that this equality is obtained by sacrificing, to a varying extent, the purity of the harmony of all intervals.

The optimum answer was first proposed to our knowledge by Werckmeister in the 1690s (1983 (reprint)). The solution to the problem became clear after the invention of logarithms by Napier. All notes should be evenly spaced on a logarithmic scale. As a result, the interval of a semitone is $2^{1/12}$. Happily, this leads to many intervals being quite close to perfection (table 1), which is likely why it is still used. For example all the fifths are within 0.1% of the correct ratios. The worst of the highly harmonic intervals is the major third which is off by almost 1%. We will come back to this when we discuss the blues.

The beats between overtones in the equal temperament scale provide a powerful method for precise tuning used by piano tuners. In order to use this on a piano the tuner must limit the action to striking only one string per note, done using a piece of damping felt (there are three strings per note in the mid range of the piano keyboard). The fifth is used as the primary tuning interval. She first tunes one note (usually A) from a tuning fork. Then she tunes each fifth and listens for the beats. For the fifth C/G, for example, the difference in an overtone frequency is about 1.7 Hz, which means about one beat every 0.6 s (see table 1). (The ‘sign’ of the difference can be detected only through the direction in which the mistuning occurs relative to the correct tuning.) This method of tuning is very precise since it is a vernier method. In practice, hearing the beats is the easy part, and can be readily learned. From personal experience, I can say that the challenge is setting the string. That is, you must ‘set’ the tension of the string stably in the correct position where it will hold the note for a reasonable period of time. Each time you hit a key the string tension relaxes by a tiny amount, but if the string is well set it will take months to become audibly out of tune. For this and other reasons, a piano tuner has to be a skilled artisan.

One reason that the piano needs three strings per note (at least in the middle region) is because of these beats. With only one string the audible beats could be disturbing. With three strings in the middle regions, each will inevitably be at a slightly different pitch, the beats are complex and inaudible. The ‘timbre’ of the piano sound is still considerably affected by the beats, even though they are not directly recognizable (Kirk 1959). One example is that a certain ‘swell’ in the sound can be induced by a clever tuner if the strings are deliberately mistuned by a very slight amount. The slowly swelling beat intensity can compensate, for a fraction of a second, the natural and rapid amplitude decay of the notes after the ‘attack’ phase when the strings are struck.

4. Music in the era of equal-temperament

Throughout the history of music, scales were developed and the harmonic palette which was available to composers changed, and with it the rules under which music was composed also evolved. It appears as if these rules were a form of axioms that the listener learned, perhaps intuitively, and used to interpret music of that era. More modern scales allowed more complex rules because they permitted more sophisticated harmony from other keys to be accessible during performance. In the baroque period, for example, Pachelbel’s Canon uses only the classic chords I, III, IV, V, VI. As it became possible to change keys, an important axiom emerged, based on the dominant seventh chord.

4.1. The dominant seventh chord

If one takes the triad based on V, which is a major chord (i.e. it contains a major third followed by a minor third), and adds another minor third to it one gets a very harmonic sounding chord—the dominant 7th, shown in figure 9. This is our first example of a four-note chord. In shorthand it is often written as (e.g.) C7. It is to be distinguished from, although it is closely related to, the minor 7th chord, which has the same minor 7th upper interval built on a minor triad base. There is also a major 7th chord, which involves adding a major third on top of a major triad. On C the major 7th is C–E–G–B (natural), an impressionistic chord that was used from the time of Debussy and is common in jazz. The remaining chord, the diminished 7th, will be discussed in detail later.

The dominant 7th chord is particularly harmonic sounding because the added note, which is B flat for the C chord, making
Figure 10. The circle of fifths connecting the 12 keys. For each key, the major triad is shown in the treble clef with the root in the base.

an interval of a minor 7th with the root C, is very close to a natural overtone (the seventh harmonic) of C reduced into the first octave (see figure 4). Thus the chord sounds quite harmonic. However, the chord also sounds unresolved because the note B flat is not in the major scale of C. So while we hear the pleasant harmony of the dominant 7th, we are dissatisfied to hear a note not on our scale. We have been 'trained' at the resolution of this musical problem. From C7 we go to F major. The key of F major (figure 10) has only one changed note relative to the key of C—the note B translates into B flat. Tah-dah! (a sound associated with this famous musical resolution, e.g. C7 F). The dominant seventh is the foundation of a simple technique, known as the circle of fifths, to change keys or modulate in a composition.

4.2. The circle of fifths

The circle of fifths (figure 10) dates from the 18th century. In the direction that I just described with the dominant seventh, it goes C–F–B flat–E flat–A flat–D flat–G flat–B–E–A–D–G–C. This represents clockwise travel on the circle depicted in the figure.

The clockwise direction is the most popular use of the circle, and is perceived as relieving tension. But this direction should be called the circle of fourths. The anticlockwise direction (counter clockwise) is actually the true direction of the circle of fifths, since G is the fifth of C etc. This counterclockwise direction is perceived as building tension. A simple popular example of the circle is in the Beatles song ‘Yesterday’, shown in figure 11.

A classic example of the tension-resolving use of the counter-clockwise direction is found in the cadence. A cadence is a terminating harmonic sequence in music. There are many cadences, for example the harmonic sequence VII—III—VI—I, which are often used in hymns and popular music. In a cadence we do not usually change keys, as a result of which many of the chords are not major chords. But the circle is a classic tool for modulating into different keys and making the listener feel comfortable that the rules are being followed.

4.3. The diminished chord

The diminished (seventh) chord is a very interesting and strange sounding chord. It is simple—it is made of three intervals on top of each other—each of which is a minor third. The chord of C diminished (7) is shown in figure 4. While each of these intervals is itself harmonic, yet they lead to clashing when played together, for example, C–E flat sounds harmonic. E flat-G flat sounds harmonic, but C G-flat (a diminished fifth) sounds very dissonant (so much so that it was known in the medieval times as the ‘devil’s’ interval). So when put together the diminished triad (or seventh, but the extra note is not essential) has an odd sound, both harmonic and dissonant. This sound is unique to its use, which is well represented by Chopin’s music, as shown in figure 13, and by an excerpt from Beethoven’s Moonlight Sonata (figure 14).

Diminished chords were a hallmark of the ‘romantic’ music period.

There is an additional usefulness to the diminished chord and its application to the musical palette. It is degenerate. With 12 notes on the chromatic scale, there should be 12 diminished seventh chords. In practice, there is a significantly smaller number of distinct chords. Because the chord is based on stacked minor third intervals, the diminished chords built on any of the notes in the diminished chord C dim7 are identical (they are in fact ‘inversions’, i.e. the notes are the same but in a different order). So C dim7, E flat dim7, G flat dim7 and A dim7 are all the same chord. Based on this logic, there are only 3 distinct diminished chords C, C# and D. This limited set provides a tool to quickly change to remote keys. In the key of C you play C dim, which could be considered as G flat diminished, leading to the key of G flat—presto! On the circle of fifths it would have taken six steps to get there. So the introduction of diminished chords allowed more complexity through modulation into remote keys.

It should be becoming clear that the ‘rules’ or axioms in music are based on sound physical principles, and require aural training for the listener to understand and become familiar with them. As time goes on the rules evolve because the composer demands a new palette of musical colors. The composer can create on top of the rules, and often takes advantage of ambiguity. As a simple example of ambiguity, if only two notes are played, then there is ambiguity about which triad or chord is implied, e.g. minor or major if the two notes are the tonic and the fifth. This allows for some interesting creative tension.

Many more harmonic rules in which four-note and five-note chords are widely used can be found in textbooks on harmony (Mehegan 1959, Kosta and Payne 2004). Jazz itself is quite complicated harmonically, but its roots in the blues
are simpler to understand. We will discuss some of these later.

There has been extensive mathematical analysis of harmony and scales. For example, Sevgen has analyzed possible scales in the context of complexity and has shown that the common scales and intervals can be derived from least complexity using group theory (Sevgen 2000). Honing and Bod have analyzed scales and harmony in terms of topology, and have shown that preferred scales are ‘convex’ (Honing and Bod 2005). Karp (1984) used a matrix approach to analyzing tuning systems. Zweifel (2005) used a group theory approach to understand scales and intervals. In general these approaches, while differing in the mathematical basis, demonstrate that harmony and scales are chosen because they form a closed set of relationships. These approaches show that underlying simplicity and symmetry is the origin of the favored musical scales, since these lead to more interconnections between chords and scales which provide more rather than less flexibility. They also suggest that the analytical part of the brain plays a role in appreciating music.

5. Artifacts in western music from equal temperament

Sometimes things are well planned and logical, but culture and art also flourish on artifacts. An excellent example is the Marshall amplifier. When heavy rock guitarists in the 60s were looking for amplification to play in large arenas, they turned to high power vacuum tube amplifiers made by the British company Marshall. These amplifiers, typical of vacuum tube non-linearity, had a serious limitation—when driven too hard the plate voltage sagged and the sound was distorted. These defects in the amplifier have now become much-loved effects from musicians like Jimi Hendrix (Shadwick 2003) and Eric Clapton. Today some try to digitally reproduce the sound of these defective analog amplifiers.
Similarly many artifacts have emerged from the historic tunings. For example, the major third in equal temperament tuning is significantly sharp of the ideal just tuning ratio of 5/4, leading to the ‘bright’ sound of the major triad on the pianoforte. One can recognize these brilliant major chords throughout piano music, for example the crashing opening chords of Beethoven’s Emperor Concerto. This aspect of brightness in the major scales seems to have emerged since the time of equal temperament. In contrast, minor keys have become associated with somber music. The major-third ‘problem’ in equal temperament directly led to the piano blues style as we will soon discuss.

A particular mystery has been why some keys sound brighter than others, when in equal temperament the keys should have identical harmonic characteristics, given that harmony is based on ratios. Several explanations for this are found in the literature. Firstly, the absolute frequencies make a difference to the tone of sound. This is partly due to the structure and response of the ear. The inner ear is a complex active organ. It maps different frequencies to specific sets of active auditory nerve cells. This feat is performed by spatial dispersion of frequencies along the basilar membrane, a pseudo-resonant tapered coil with varying damping, width and stiffness characteristics. A second explanation is that in orchestral pieces, many instruments do not actually play in equal temperament, and the notes they play could be different in different keys. For example, many brass instruments have natural harmonics in the keys of B flat and E flat, and...
would be likely closer to the natural scale in some keys than in others. From informal conversations with orchestral players, we learned that an orchestra uses ‘adaptive’ tuning (Sethares 1994) depending on the type of music and the type of instruments. If you are playing music of the baroque period which does not venture into remote keys, then you may play much more in the just or natural tuning. When a piano is involved in the orchestra, there may be more effort on the part of other instruments to accommodate equal temperament, and all this would be dependent on the key in which music is played. Finally, a very interesting analysis about how a piano might sound different in different keys has been expounded (Barbour 1947). Barbour pointed out that most real pianos are out of tune. Once the note is set by the tuner, its inevitable path is to get flatter with time due to forces on the string/system, especially each time that the string is hit. As mentioned earlier, the practical skill of the tuner comes not only in hearing the beats to set the notes for equal temperament but also in the setting of the note so that the pin will not relax quickly after she has gone and the instrument is played. Still, some flattening with time and playing is inevitable. Barbour’s analysis considers that if the piano is played most often in the familiar keys of C, F, G etc so that the notes in these keys are struck most often, the flattening of notes leaves keys like E flat and B flat to become relatively well tuned to the just intonation, and Bmaj, F#, etc to be most out of tune, possibly accounting for the more mellow timbre of the flat keys.

Another artifact that we have not yet considered is that the harmonic series as a series of integers applies only for infinitely thin strings. Real strings and real vibrating instruments exhibit some non-linearity. The finite thickness and stiffness of strings means that the higher harmonics are progressively flatter than the integer harmonics. When a tuner tunes a piano he or she knows this, and deliberately flattens the highest notes so that they will sound in tune with the middle notes, and sharpens the lower strings for the same reason.

6. African music and the blues

The blues originated from the mixing of African and European musical traditions in America. African slaves were brought into Northern and Central America in the 17th and 18th centuries. The musical traditions that they brought were very different from those of the Native Americans or the colonists of European descent whom they encountered. The juxtaposition of these cultures led to unique musical styles. It has been suggested that in Latin America, the blending was more immediate, leading to Latin styles of music today that are more rhythmically different than traditional western styles. On the other hand, the juxtaposition in North America was a harsher one, given the segregation of the slave trade, and led to much later ‘blending’ of the music to form jazz and the blues. The fact that the blending occurred at a later time after each group had assimilated the other’s music to some extent led to new styles of music, very different from Latin music. For example, the simple blues harmonies that are found in the 12 bar blues (I, IV, V) are likely inherited from the hymns which were sung on slave plantations.

Schuller (1968) chooses to understand the complex origins of jazz less from cultural history and more from musical analysis, using a scientific approach. There are certain characteristics of African music which he emphasizes. Most important of all is rhythm—as Duke Ellington immortalized ‘It Don’t Mean a Thing if It Ain’t Got that Swing’. The meaning of ‘swing’ is a regular beat (rhythm)—where the vertical patterns of the music lose ground to the horizontal. (The vertical characteristics mean the groups of notes played simultaneously, i.e. harmony, whereas a horizontal characteristic is the time structure of notes.) According to Schuller, African music is the most complex rhythmically in the world—so-called polyrhythmic. Despite this rhythmic complexity, there are one or more regular beats and syncopation associated with the pounding of a tool, for example. Music was an integral part of life in Africa and not just sat down and listened to. When blues began in America, it was most often used by slaves as they worked.

Although Western music is rhythmical, its construction focuses more on the vertical harmony than on the horizontal line, and the ‘rhythm’ is often subordinate and not very regular—in musical terms ‘rubato’. Whereas in jazz and the blues the rhythm is like clockwork—that is swing. The technique of ‘syncopation’, i.e. missing the beat, is very effective in jazz ONLY because the beat is so well defined. Even though syncopation is strongly associated with African music, it did exist in Western music. An early use of syncopation is found in the final movement of Beethoven’s ninth symphony.

African music, such as that played on the Malian n’goni (a forerunner of the banjo) brought in the ancient pentatonic scales which became components of jazz and the blues. If one wishes to improvise on a chord, then a simple way to accomplish this is to learn the associated pentatonic scale for the equivalent key. There is a blues scale which is based on the (minor) pentatonic scale. The blues scale in the key of C major, is shown in figure 15. (If you would like to read a good set of musical references, especially for piano blues and jazz, then look at the four books by John Mehegan of the Julliard School in New York (Mehegan 1959).)

In addition to the pentatonic scale, there are important additional notes in the blues scale. While excluding some notes on the diatonic scale, the blues scale includes B flat and E flat and these are sometimes called the blue notes. We discussed B flat under the section on the dominant seventh, where we saw that it is very close to a natural harmonic of C at a ratio of 7/8. The ‘blue note’ shown as E flat is also very harmonic. Actually these notes, if played on a horn or on an instrument where the notes can be bent to suit, would be played precisely at the correct harmonic value which is 5/4 (the major third). In the just scale this note is mathematically correct on the keyboard, but in the equal tempered scale the third has been adjusted to be quite sharp of the blue note. Figure 16 depicts a linear frequency scale showing the notes on the piano, equally tempered, and the correct minor 3rd (6/5) and major 3rd (5/4)—the latter is close to the blue note.

The major 3rd is an interval with strong harmonic characteristics, since the lower integer number means that
Barrelhouses were not houses of good repute, and more primal things than crushing notes were going on in the back rooms, so the crass sound might have been appropriate. But, in St Louis, the young Scott Joplin, a few years earlier, was trying to get a similar effect in a less crass way. Even though he also played in bars, his music made it to the more sophisticated salons, and was called ragtime. It also had the swing of the blues—the word ragtime refers to the regular rhythm of such dances. But Joplin was also reaching for the blue note. Instead of crushing the notes, he used accidentals (added notes off the scale) which were played in sequence, e.g. E and E flat, in runs to create the sensation of the intermediate note. The example in figure 18 is from his ‘Cascades Rag’.

The sound that we associate with ragtime today intrinsically includes this clever hint of the blue note. The note that ought to be there on the scale, but is not. We can speculate that if just intonation had been preserved we would not have such treasured styles of music. It is amazing how cultural richness emerges in some case from artifacts, and in all cases musical culture is built on a rock of physically based axioms for harmony and melody.

Modern jazz has come a long way from the blues, but is based on it, and the richness of the pure unvarnished blues is still there for the taking (especially if you live in a city like Chicago, where the electric blues blossomed in the mid 20th century).

7. The musical particle and quantum mechanics

This paper has focused on what one can learn about music from physics and mathematics. Even within this broad field I have limited my review to the physics of music itself, and not to the instruments which play it. That area of musical acoustics is a better known and widely applied field. An excellent summary of this field can be found in the book by Donald Hall (2002). However, there is another side of the connection between physics and music which I would like to explore briefly in the remainder of this paper, and which I believe has not been greatly exploited, and that is how one can use music to teach lay people the essence of modern physics and engineering. I offer an apology that I have not done an extensive review of the literature on music in physics education for this section, but instead propose some of my own ideas that follow from the concepts discussed in the paper.

The key to using music to teach modern physics, in my mind, is harmonic analysis. Harmonic analysis is nothing other than Fourier analysis, a tool which is highly valued in physics and engineering. It is not widely known by the non-scientist, yet it is easy to explain in the musical context—when someone looks at a graphic equalizer on their stereo they are looking at a real-time Fourier amplitude analysis of the music that is being played. One can explain how the synthesis of harmonics affects the timbre of musical sound—how the coefficients in a Fourier series can describe different real functions. A very simple example is the comparison between the organ pipe and the piano string. The ‘stopped’ organ pipe has only odd-numbered harmonics and the piano has all harmonics. (Because the ‘stopped’ organ pipe has a pressure node at one
Figure 17. A drill demonstrating the crushing of blue notes called ‘Blue notes in barrelhouse style’ by Aaron Blumenfeld from ‘The Blues, Boogie and Barrelhouse Piano Workbook’ (Ekay Music, Katonah). The crushed blue notes are shown by ‘accidentals’ marked as D♯ played with E repeatedly in the treble clef. A video showing a different, but typical, example of the ‘crushing’ of notes is included in the multimedia content associated with this paper, which is available online at stacks.iop.org/rop/72/076001.

Figure 18. Opening bars from ‘The Cascades—A Rag’ by Scott Joplin with an example of a run around the blue note (marked with an asterisk) which was typical of ragtime piano music. (©Dover Publications, Inc. Reprinted by permission.)
end and a pressure antinode at the other, whereas a string has nodes at both ends.) The different symmetry of the boundary conditions accounts for a major aspect of the difference in the timbre of the organ and the piano. Another persuasive example is the effect of ‘tinning’ the felt hammers in a piano to create a ‘honky-tonk’ effect. By adding metal to the hammers you increase the transverse speed and curvature of the string deflection and so more high frequency harmonics are produced, giving the music a ‘tinny’ sound. This effect is used more subtly in the ‘regulation’ of pianos. A tuner will poke the felt tips of hammers with pins to soften them and restore an appropriate timbre by controlling the balance of harmonics. A rigid plectrum plucking a guitar also gives a ‘tinny’ effect by appropriate timbre by controlling the balance of harmonics. A felt tips of hammers with pins to soften them and restore an subtly in the ‘regulation’ of pianos. A tuner will poke the

giving the music a ‘tinny’ sound. This effect is used more subtly in the ‘regulation’ of pianos. A tuner will poke the felt tips of hammers with pins to soften them and restore an appropriate timbre by controlling the balance of harmonics. A rigid plectrum plucking a guitar also gives a ‘tinny’ effect by introducing high harmonics. More detailed analysis of these effects in various instruments can be found in Fletcher and Rossing (1998).

Quantum mechanics is one of the most significant revolutions in physics of the last few centuries. At the heart of quantum mechanics is the wave-particle duality. This duality is expressed by the Heisenberg Uncertainty principle, which in the frequency domain can be written as $\Delta f \Delta t \geq 1$ or equivalently in the spatial domain as $\Delta k \Delta x \geq 2\pi$. (Note that the equality holds for Gaussian-like wavepackets.) We understand that this uncertainty relationship is simply the result of the Fourier principle that in order to have a well-defined frequency or wavelength many oscillations of the wave must be measured. The error in determining the frequency of the wave is inversely related to the time over which it can be measured.

Musically, this is very simple to demonstrate by using a musical wave packet. Figure 19 shows an example of a wave packet of frequency 100 Hz and attack and decay time of 2 ms.

Audible wavepackets can be generated in at least two ways. An appropriate Mathematica™ program (or equivalent software package, such as Matlab™) on a computer is capable of synthesizing complex wavepackets that can be played on the computer’s sound card. Equivalently, a musical synthesizer is specifically designed for such tasks. The former is easiest for the classroom, but the latter is much more dramatic to demonstrate. The fundamental frequency is associated with the pitch of a note. Hence when one ‘hears’ a note and identifies its pitch, one is measuring the frequency—at least the relative frequency to other notes. Hearing in this way does not require absolute pitch, where one’s brain must be detecting and identifying the absolute frequency—a talent possessed by few. But relative pitch measurement is essential to hearing a melody at all.

Not surprisingly, one’s ability to hear a note and separate its pitch from an adjacent note is determined by the Fourier uncertainty principle. For example, consider middle C on the piano and the next note on the scale D. In the equal tempered, concert scale, the frequency of middle C is 261.6 Hz. That of D is 293.7 Hz. The difference of \(~30\) Hz, could be measured if the uncertainty in the pitch of each was better than \(~15\) Hz. The duration of the notes would have to be longer than \(~80\) ms (more than \(~20\) oscillations to get a pitch accuracy of \(~5\%)). In order for you to recognize a melody such as EDC EDC... (‘three blind mice’) the note durations would have to exceed \(~80\) ms. Try it! With the synthesizer you can adjust the ‘attack’ and ‘decay’ of notes to create a wave packet, whose length you can smoothly change with a slider bar as you are playing. When the length exceeds \(~80\) ms, the audience can recognize the melody. When the notes are much shorter, they hear only ‘clicks’. These ‘clicks’ are musical particles. Voila—quantum mechanics explained! (Well, I agree there is more to it—for example human perception of sound could increase the time required, but this is a teaching tool with which you can intuitively explain the meaning of the particle-wave duality.)

Musical instruments with rapid decay times on the scale of the vibration frequency sound ‘percussive’, the opposite, such as the piccolo, are ‘melodic’. From experience I have found this demonstration of a musical particle to be very effective and intriguing to an audience.

7.1. Harmonic analysis

Beyond the simple concept of a musical particle, music can be used to demonstrate Fourier analysis. For example:

- the modes of electrons in a one-dimensional atom are the integer multiples of the basic ‘vibration frequency’ of a string. This can be extrapolated to explain atomic spectra and the ‘bandgap’ in semiconductors and insulators.
- The differing colors of nanoscale semiconductor particles as a function of size are equivalent to the different tones of organ pipes as a function of length. It is a good chance to talk about nanotechnology, comparing the organ pipe lengths and the wavelength of sound (\(\sim m\)) to the equivalent for electrons (\(\sim n m\)).
- X-ray diffraction is a form of harmonic analysis. Imagine trying to reconstruct the sound wave from a map of its frequency spectrum. That is the challenge of solving structures from diffraction patterns. It works well of course for periodic objects. By analogy with sound one can explain the concept of spatial periodicity as a complement to the more familiar concept of temporal periodicity. I often use this to introduce waves as having periodicity in time as in space, which is the key physical and mathematical property of wave phenomena.

I would like to provide more about the use of music for teaching science, but I feel that is better done in a textbook and not a review paper.
8. Conclusion

While many see physics and mathematics as having much in common, the differences are well understood by practitioners. Albert Einstein said ‘Insofar as mathematics is about reality, it is not certain, and insofar as it is certain, it is not about reality’. Mathematics is abstract and ideal, yet it is the language on which physics can build a model of the universe, a model which will always be an approximation. So what does this have to do with music? Paradoxically music is directly connected to both physics and mathematics, not only through the connection physics has with math. Obviously, the sounds and sensations are rooted in physics. We can describe them as mathematical objects, but only approximately. On the other hand, composition of music deals with axioms—with rules. Mathematics can be thought of as the discipline of developing all the possible logical consequences of a set of mutually consistent axioms. In music, a composer works with a set of musical rules, not necessarily written down but nonetheless implicitly understood and agreed upon by the listener, as the framework on which he/she creates. I consider these musical rules to be like a palette of colors for a painter. The most simple of the musical axioms is the scale—the notes on which the composition is rendered. Beyond this there can be many other rules, for example which notes can occur in combination (harmony) or in sequence (melody and countrypoint).

In this paper the connections between art and science have been emphasized through music. The sensation of harmony and the related melodic development are rooted in physical principles that can be understood with simple mathematics. This review focused on the science of music itself, rather than the related melodic development. All scales (e.g. the diatonic major) can be shifted up or down by up to 12 semitones to create the 12 keys. The key signature is the combination of sharps and flats needed for a particular scale. All scales (e.g. the diatonic major) can be

Acknowledgments

This work was supported by the Department of Energy under Contract DE-AC02-06CH11357. Thanks to Faye Gibson for the graphics.

Appendix A. Some musical definitions

Chord—three or more notes played together which are harmonic, not including intervals of unison or the octave.
Interval—intervals are named by the number of notes between the lower and upper (piano) keys, counting both keys themselves. So the smallest interval, e.g. C–D is a second. A third would be C–E (or D–F), a fourth C–F and a fifth C–G.
The terms perfect, major, minor, diminished and augmented are also used to preface intervals.
(Interval) unison—when two notes which are identical are played together. This is the most harmonic combination.
(Interval) octave—two notes with a ratio of twice the frequency. This is also an eighth. Because of the strong basic harmony, these are considered the same note, separated by an octave, sometimes designated with a prime, e.g. C–C′, of G–G′.
(Interval) major third—this interval contains two tones (TT). (Interval) minor third—this interval contains a tone and a semitone (TS—the order does not matter).
(Interval) perfect fifth—this interval comprises three tones and a semitone (TTST), the sum of a major and minor third. It is called perfect because of its very low integer ratio (3/2) in natural tuning, and thus its strong harmony.
(Interval) diminished fifth—diminishing is accomplished by removing a semitone from an interval. So the diminished fifth in C is the combination of C and Gflat. An augmented fifth is the reverse (C–G#). Diminishing a major third give a minor third.
Key (signature)—the key is labeled by the fundamental note for a particular scale. All scales (e.g. the diatonic major) can be shifted up or down by up to 12 semitones to create the 12 keys. The key signature is the combination of sharps or flats needed in the appropriate key for the appropriate scale, notated at the beginning of a musical stave. For example, the key C major (whose key signature has no sharps or flats) when shifted up 3 semitones makes the key of E flat major (whose key signature has three flats). The use of sharps or flats in key signatures is chosen so that each key signature has a unique number of sharps or flats, which are never mixed.
Mode—a sequence of seven consecutive notes in a given key.
Modulate—change the key during a piece of music, not necessarily involving either transposition or recognition with a new key signature.
Overtone—the Fourier components (most often discrete) contained in a musical tone, and above the base frequency of a tone. These are also sometimes called harmonics, but for clarity, in this text we restrict the word ‘harmonic’ to refer to the pleasant sound of two notes played simultaneously, which comes from overlapping overtones.
Pitch—this is the identification of a musical note, which is dependent on the frequency of the fundamental vibration.
Scale—a sequence of notes on which music is performed. In western music there are most often 7 notes in a scale all lying within a single octave. A scale can be a mode, but not all modes are recognized as scales.
Temperament—a system in which the values of notes are adjusted slightly (‘tempered’) to create more harmonic intervals in some cases.
Tonic—first note in a scale, which also gives its name to the musical key.

Transpose—change music in whole from one key to another.

Triad—three notes representing two successive steps of a third. The triad is the simplest chord.

References

Barbour J 1947 Bach and the art of temperament Musical Q. 33 64

Fink G R et al 1996 Where in the brain does visual attention select the forest and the trees? Nature 382 626–8

Fletcher N H and Rossing T D 1998 The Physics of Musical Instruments (Berlin: Springer)

Hall D 2002 Musical Acoustics (Singapore: Brooks Cole (Thomson Learning))


Honigh A and Bod R 2005 Convexity and the well-formedness of musical objects J. New Music Res. 34 293

Karp C 1984 A matrix technique for analyzing musical tuning systems Acustica 54 209–16

Kirk R E 1959 Tuning preferences for piano unison groups J. Acoust. Soc. Am. 31 1644


McLean D 2001 The science behind Europe’s music scale Interdiscip. Sci. Rev. 26 211–21

Mehegan J 1959 The Jazz Improvisation Series (4 volumes) (New York: Watson-Guptill)

Mersenne M 1636 Harmonie Universelle (Paris: CNRS)

Olson H F 1957 Acoustical Engineering (Princeton, NJ: Van Nostrand-Reinhold)


Sethares W A 2005 Tuning, Timbre, Spectrum, Scale (London: Springer)


Shadwick K 2003 Jimi Hendrix, Musician (Milwaukee, WI: Hal Leonard)

von Helmholz H 1863 On the Sensation of Tone (New York: Dover) (reprint 1954)
