Musical Scales

Questions that have puzzled me include,

(a) Given that there is a continuous sound frequency spectrum, why are there 12 notes (semitones) in an octave? Why not 10 or 16 or any other number? There is, of course, ‘space’ for an indenumerable infinity.

(b) Why does each scale contain just 8 notes (or, rather, 7 notes, excluding the repeated tonic)?

(c) Why are scales so weird? That is, why do different scales have a different number of sharps/flats, and what determines what sharps/flats a given scale should contain?

(d) What is it that determines which combinations of notes constitutes a chord (as opposed from, I presume, a discord?)

(e) Why do minor scales produce music with a different emotional colour from that of major scales?

Let me confess immediately that I intend to be culturally biased. I will be discussing only the Western chromatic scale. The fact that other scales are possible betrays an important fact: that there is no categorical, definitive, unassailably <u>mathematical</u> answer to the above questions. We are in the realms of history and taste when discussing musical scales.

Nevertheless, there are interesting mathematical issues associated with musical scales which at least partly constrain how they may be devised. In the following discussion, bear in mind that the sound frequency spectrum is continuous. The whole issue under discussion is “why are certain frequencies picked out to constitute our notes and chords”…

The (Wrong) 8 Note Scale

Imagine the earliest musicians. For sake of argument, assume they had a stringed instrument. In might be a very primitive affair with just a single string stretched between two fixed points. Having discovered that plucking such a stretched string produced a pleasant sound, what next? They would quickly tire of a single note. It would, quite literally, be monotonous. A new discovery would follow at once: that holding the string fixed at some intermediate point causes the note to change. The most obvious point to choose is the mid-point of the string. Even the least musical person would recognise that there is a clear harmonious relationship between the original note and the new one. The octave is born! This is the first illustration of the interplay between matters of taste and mathematical definition. The motivation for the octave is simply that the new note is pleasing to the ear when compared with the first. Mathematically it corresponds simply to a doubling of the frequency, or halving the string length.

The relation of a note to its octave note is so fundamental that we refer to these two notes as being “the same”, albeit one octave apart. So, our early musician would be no closer to producing complex and emotionally involving music. He would still be stuck with monotony. Nor would halving the length of the string again help. This would simply produce “the same” note again – but one octave higher still.

The answer – and our musician would have discovered it more quickly than it has taken me to write these paragraphs – is to divide the string into parts other than just ‘halves’. Now what we, and our ancient friend, are after is a division of our octave into intermediate notes. If we can do this for our first octave, then we can do it again for every higher (or lower) octave. So, we need to leave *more* than half the string vibrating (that is, we need to hold it at a point *less* than half way along its length). Having already tried the mid-point, the next most obvious point to hold is the "one third" position. This leaves 2/3 of the string vibrating, and hence produces a note of frequency 3/2 times the base (open string) frequency. This is the second note that we have discovered for our octave. I will confound you utterly by telling you that it is known as "the fifth". This is to anticipate that it will turn out to be the fifth, in order of pitch, in the scale we are building – but we don't know that yet. (Similarly, we have as yet no motivation for the term "octave").

Now it gets interesting – or messy, depending upon your point of view. Let us take the obvious approach. The next thing is to hold the string at the ¼ position, leaving ¾ of the string vibrating, and hence producing a note of frequency 4/3 times the base (open string) frequency. In the same way holding at the 1/5 and 1/6 positions gives notes of frequencies 5/4 and 6/5 times the base frequency. We can also hold the string at a position 2/N, providing that this is still less than ½. Doing this for N=4 or 6 just reproduces ½ and 1/3, but for N=5 gives a new note of frequency 5/3 times the base frequency. This is how our attempt at building a musical scale is shaping up so far,

First Attempt at a Musical Scale

String Held At	Frequency*	Ratio of Adjacent Frequencies	Modern Name
0	1		C (open string)
1/6	6/5	1.2	D#
1/5	5/4	1.041666	E
¼	4/3	1.066666	F
1/3	3/2	1.125	G
2/5	5/3	1.111111	A
½	2	1.2	C

**these are in multiples of the open string frequency.*

Now our early musician would be playing these notes, probably one after another in the manner of our scale playing today, and judging their sound. It's difficult to do this in print! However, the ratio of adjacent frequencies, given above, is an indication of "how far apart" the adjacent notes of our incomplete scale would sound. For example, the first two and the last two are rather too far apart, having a frequency ratio of 1.2, whereas the second and third are rather too close together, having a ratio of only 1.041666. This 'scale' would sound hopeless (try it – all these notes are on modern instruments – their names in modern notation are given in the last column, assuming an open string tuned to C).

So, based on how our first attempt at a scale sounds, and to make the frequency ratios more uniform, we might want to insert an extra note between the first pair and the last pair, but eliminate the third note which is too close to the second. Hence we might look for something like,

Second Attempt at a Musical Scale

String Held At	Frequency	Ratio of Adjacent Frequencies	Modern Name
0	1		C
0.0871291	$\sqrt{1.2}$	1.0954451	midway [C#, D]
1/6	6/5	1.0954451	D#
¼	4/3	1.1111111	F
1/3	3/2	1.125	G
2/5	5/3	1.1111111	A
0.4522774	$2/\sqrt{1.2}$	1.0954451	midway [A#, B]
½	2	1.0954451	C

(where we have split the adjacent ratios of 1.2 into a product of pairs of ratios, each being $\sqrt{1.2}$ = 1.0954451. The product of all the ratios, must, of course, be exactly 2).

This proposed scale achieves two things; (i)a very good approximation to uniform note "spacing" [by which we mean uniform adjacent note frequency ratios – which is really uniform spacing in log(frequency)], and, (ii)it preserves all but one of the notes based on the simple ratios derived above. Moreover, it has 8 notes – though we did not constrain it to at the outset. So, this is the first intimation of the origin of the 8-note scale. Also note that our first non-octave note, namely frequency x3/2 (G) is indeed the fifth note of this scale, as anticipated above.

However, this is not the scale which Western music has adopted – at any period in history. The two notes which we invented to fill the over-large gap in our first attempt at a scale do not exist in modern (or classical) music. These are the second and seventh notes of our scale, lying respectively mid-way between C-sharp and D and mid-way between A-sharp and B.

Memo: Find some way of trying this scale out. Can't be done on an instrument (unless, say, two guitars were used, with one tuned down to provide the two new notes - but I don't know how I'd tune-in the new notes to the required frequency??). The obvious thing is to use a PC which can supply a tone specified directly by frequency. Try it!!

There are at least four reasons why such a scale was not adopted historically; (a)It does not sound right – is this true? Have we just become conditioned to an alternative? (b)The classical mind would have abhorred the nasty irrational ratios I have used for the two new notes. They liked nice simple ratios. (c)There seems no reason to have left out poor old frequency 5/4 (E) in our scale. If anything 5/4 is 'simpler' than 5/3 or 6/5, so why does it get ditched? No, the classicals insisted on keeping it, thus mucking up our nice simple proposal. (d)A much more subtle reason to do with the way notes in one octave relate to those in higher, and lower, octaves – and the origin of those 12 semitones….

The Chromatic Scale of 12 Semitones

I'm sorry about this, but I just cannot avoid it. You see, it all goes back to those Pythagoreans. They started off trying to devise their musical scale pretty much as described above. That is, they discovered the octave note (frequency x2) and then the 'perfect fifth', i.e. freqency x3/2. But they didn't continue with this approach. Instead they had – to mathematicians – a more elegant method. Their reasoning was (or at least we may surmise) that multiplying any number of factors of 2 is very useful in generating lots of harmonious notes, but using just factors of 2 gets us nowhere in terms of dividing up our octave. Using our 3/2 ratio just once gives us just one more note in our octave. But what happens if we use it twice? Well, we now have just two-thirds of two-thirds of our string vibrating, and the frequency is clearly 9/4 which is greater than 2, and hence lies outside our octave. Shame! But wait a minute. We can also shift things up and down by any number of factors of two and still guarantee to obtain notes which are harmonious. So, we just divide our 9/4 by 2 to give 9/8.

This is a more elegant, and more systematic approach. In terms of where your finger is on the string it means that we are going to generate all our notes from the following operations: where-ever your finger is currently positioned, either, (a)move it one-third along the remaining length of string, or, (b)move it so as to double the length of vibrating string. Clearly if the first operation is carried out r times, and the second n times, (the order is irrelevant) then starting with an open string produces a note of frequency,

$$\left(\frac{3}{2}\right)^r\left(\frac{1}{2}\right)^n$$

Also, it is clear that, for a given value of r there is exactly one value of n which causes this new note to lie within our octave, i.e. having a frequency between 1 and 2 (in multiples of the open string frequency). This particular value of n can be written n_r. The values of n_r for each r are as follows,

Third Attempt at a Scale – The Pythagorean Scale

r	$\mathbf{n_r}$	**Exact Frequency (** $3^r / 2^{r+n}$ **)**	**Decimal Approx**	**Adjacent Ratio**
First Octave				
0	0	**1**	1	-
7	4	2187 / 2048	1.067871094	1.067871094
2	1	**9 / 8**	1.125	1.053497942
9	5	19,683 / 16,384	1.20135498	1.067871094
4	2	**81 / 64**	1.265625	1.053497942
11	6	177,147 / 131,072	1.351524353	1.067871094
6	3	**729 / 512**	1.423828125	1.053497942
1	0	**3 / 2**	1.5	1.053497942
8	4	6,561 / 4,096	1.601806641	1.067871094
3	1	**27 / 16**	1.6875	1.053497942
10	5	59,049 / 32,768	1.802032471	1.067871094
5	2	**243 / 128**	1.8984375	1.053497942
0	1	**2**	2	1.053497942
Second Octave (frequencies divided by 2)				
12	7	531,441 / 524,288	1.013643265	-
19	11	1,162,261,467 / 1,073,741,824	1.0824403	1.067871094
14	8	4,782,969 / 4,194,304	1.1403487	
21	12	10,460,353,203/8,589,934,592	1.2177454	
16	9			
23	13			
18	10			
13	7	1,594,323 / 1,048,576	1.5204649	
20	11			
15	8			
22	12			
17	9			
Third Octave (frequency divided by 4)				
24	14		1.0274727	

The Table covers all values of r from 0 through 24 (just once each), although the entries have been ordered by increasing frequency. This covers two whole octaves and the start of a third. The frequencies in the second octave have been divided by 2 for ease of comparison with the first.

A nice thing about this Pythagorean scale is that the frequency ratio (=logarithmic spacing) between adjacent notes (semitones) is reasonably uniform, so we have achieved our objective of splitting the octave into notes at fairly even spacings. Only two ratios occur: either $3^7 / 2^{11} = 1.067871094$ or $2^8 / 3^5 = 1.053497942$ and these are quite similar. Defining a semitone as the spacing between adjacent notes in our scale, these two ratios are the value of the semitone, depending upon position in the scale. Incidentally, there are 5 of the larger intervals and 7 of the smaller intervals in the octave, so the ratio between the tonic and its octave note is therefore $[3^7 / 2^{11}]^5.[2^8 / 3^5]^7 = 2$, as indeed it must be.

The shortcoming of the Pythagorean scale is, however, immediately apparent. The second octave should start at a frequency of exactly 2 (or 1 after re-scaling by 2, as in our Table). It is very close! The start of the second octave misses 1 by just 1.364…%, that is by a factor of $3^{12} / 2^{19}$ = 1.0136433. It is this near numerical coincidence that gives the Pythagorean approach credibility. Nevertheless, it is flawed. The gap between the true octave and the Pythagorean 'close miss' is known as the 'Pythagorean comma' and the associated note, i.e. that at a frequency of $3^{12} / 2^{19}$ times the octave note, is known as the "wolf note". Does it sound that bad?

The Pythagorean scale does two things; it shows why it may be natural to split the octave into 12 semitones. This is just how many notes there are within one octave which can be made from the operations stated above, i.e. from the formula $3^r/2^{(r+n)}$. Or rather, it would be if it were not for the awkward fact that, with r = 12, n = 7, we get another new note in our original octave, a mere 1.0136433 times greater in frequency than our tonic note. We have no need for this embarrassing note, virtually indistinguishable from the tonic. Indeed the whole of the original 12 notes (semitones) are reproduced for r in the range 12 to 23 – except with an 'error' of x1.0136433 (=$3^{12} / 2^{19}$). The pragmatic way forward is just to ignore the unwelcome repetition. We call it a day after we have defined the first set of twelve notes. For our present purposes we are content that we have found a reasonable motivation for having 12 semitones per octave.

Finally in regard to the Pythagorean scale we note that 7 of the 12 notes are distinguished in having the small 'adjacent frequency' ratio. Curiously these same 7 notes also have the simplest frequencies, i.e. the frequencies defined by ratios of the smallest integers (shown in bold in the Table). Perhaps this is the first inkling that musical scales may be composed most harmonically out of a sub-set of 7 of the possible 12 semitones? Also, the spacing of these notes, in terms of semitones, is 2, 2, 2, 1, 2, 2, 1, which is tantalisingly similar to the spacing of notes in the major Western chromatic scales. The later are actually 2, 2, 1, 2, 2, 2, 1. If adopted, the former spacing would give the scale of C as C, D, E, F#, G, A, B, C, which most definitely does not sound right. There is no Western chromatic scale with these notes. Alternatively, if we interpret the tonic of the Pythagorean scale as F, then the note sequence is F, G, A, B, C, D, E, F – just the notes of C-major (or A-minor), but with the wrong tonic, and which consequently doesn't sound right either. The actual scale of F is F, G, A, A#, C, D, E, F.

The Equal Tempered Scale

The equal tempered scale is that in most common use in Western music. This has been the case since around the time of J.S.Bach. It is usual to attribute the widespread introduction of the scale to Bach, if not its actual invention. However, this is contentious. The scale is simple to define. Rather than agonise further about perfect harmonic integer ratios, it is simply taken as a given that there will be 12 notes within the octave and that we wish to have them equally spaced [in log(frequency)]. Since an octave is a factor of 2 on frequency, it follows immediately that all 12 semitones will be simply a factor of $2^{1/12} = 1.0594631$ above the last. This resulting equal tempered semitone notes are very close to those of the Pythagorean scale,

Note	Pythagorean	Equal Tempered	Difference (%)
Tonic	1	1	0.00
1	1.0679	1.0595	0.79
2	1.125	1.1225	0.23
3	1.2014	1.1892	1.01
4	1.2656	1.2599	0.45
5	1.3515	1.3348	1.23
6	1.4238	1.4142	0.68
7	1.5	1.4983	0.11
8	1.6018	1.5874	0.90
9	1.6875	1.6818	0.34
10	1.8020	1.7818	1.12
11	1.8984	1.8877	0.56
12 (octave)	2	2	0.00

The equal tempered scale was devised for very practical reasons. It is self-similar. That is, if we shift our base note (or tonic) up by one semitone (or any number of semitones, for that matter) the shifted scale still has exactly the same frequency ratios to the new base note.

In other words, if we start with a base note of frequency f, the equal tempered scale will consist of the notes f, $2^{1/12}$ f, $2^{2/12}$ f, $2^{3/12}$ f, ….Now shifting to a new base with $f' = 2^{p/12} f$ the scale is clearly just $2^{p/12} f$, $2^{p+1/12} f$, $2^{p+2/12} f$, $2^{p+3/12} f$ ….= f', $2^{1/12} f'$, $2^{2/12} f'$, $2^{3/12} f'$, …. which is just the same as the original scale with f replaced by f'.

This self-similarity is not achieved by any other scale – because it is a consequence of having perfectly uniform semitone intervals, and there is clearly only the one such scale – the equal tempered scale. So, the very important corollary is that with all other scales many instruments would only be able to play in one key without complicated re-tuning**.** Indeed, a fretted string instrument, like the guitar, would require, even for a fixed key, fret spacings different for the different strings. The frets could no longer simply span the whole fret-board.

With the equal tempered scale, however, the fret spacing on all strings becomes simply,

$$x = 1 - 2^{-\frac{n}{12}}$$

where x is the fraction of the length of the string down which the n^{th} fret is positioned.

"Just Intonation" Scales

"Just Intonation" scales are any scales obtained from rational factors (that is, integer ratios) of the tonic frequency. The equal tempered scale is therefore <u>not</u> a 'just intonation' scale, because of the irrational factor of $2^{1/12}$ which defines the semitone intervals. The Pythagorean scale <u>is</u>, strictly speaking, a "just intonation" scale. However, this is not what is really meant in practice by "just intonation". The term is used to mean a scale devised using small integers in the ratios (in stark contrast to the very large integers in the Pythagorean scale), but consistent with achieving a scale which is close to the Pythagorean or equal tempered scales. The rationale behind the use of simple ratios, i.e. small integers, is that it is more likely to produce harmony when playing multiple notes (chords) and less likely to produce audible beats. In any case, the aim is one of which the classicals would have thoroughly approved.

Such just intonation scales are not unique. However, going back to our first attempt at a scale, we recall that there was a 'central core' of ratios that seemed simplest and most desirable, namely 6:5, 5:4, 4:3, 3:2, and 5:3. All the most common versions of just intonation scales deploy these particular ratios. The greatest number of variants is for the first note after the tonic, i.e. the first semitone. The various suggestions are,

21 / 20	= 1.05	
256 / 243	= 1.0534979	$(=2^8/3^5)$
16 / 15	= 1.0666666	(diatonic semitone)
2187 / 2048	= 1.0678711	$(=3^7/2^{11}$, the Pythagorean semitone)

and these are to be compared with the equal tempered interval of $2^{1/12} = 1.0594631$.

An example of a just intonation scale is…

Note	Name	Just Intonation	Equal Tempered	Difference (%)	Modern Name
Tonic	**unison**	**1**	**1**	0.00	C
1	Semitone	21:20 = 1.05	1.0595	-0.90	C#
2	**Major second**	**9:8 = 1.125**	**1.1225**	0.22	D
3	Minor third	6:5 = 1.2	1.1892	0.91	D#
4	**Major third**	**5:4 = 1.25**	**1.2599**	-0.79	E
5	**Perfect fourth**	**4:3 = 1.333333**	**1.3348**	-0.11	F
6	Tritone*	7:5 = 1.4	1.4142	-1.00	F#
7	**Perfect fifth**	**3:2 = 1.5**	**1.4983**	0.11	G
8	Minor sixth	8:5 = 1.6	1.5874	0.79	G#
9	**Major sixth**	**5:3 = 1.666666**	**1.6818**	-0.90	A
10	Minor seventh	9:5 =1.8	1.7818	1.02	A#
11	**Major seventh**	**17:9 = 1.88888 OR 15:8 = 1.875**	**1.8877**	0.06 -0.67	B
12 (octave)	**octave**	**2:1**	**2**	0.00	C

**so-called because it is 6 semitones = 3 tones. Also known as the Devil's Interval, and reputably beloved of Black Sabbath.*

Hence, this just intonation scale differs from the equal tempered scale by around the same amount as the Pythagorean scale, namely by around 1% or less. There is a tendency amongst musical purists to regard the just intonation scales as being the best. The equal tempered scale is regarded by some as being essentially "out of tune" and is tolerated only because of its practical convenience.

The difficulty with tuning a guitar to this just intonation scale can be illustrated as follows. The guitar is usually tuned as follows,

E	A	D	G	B	E	(Open Strings)
F				C	F	(First Fret)
	B	E	A			(Second Fret)
G	C	F		D	G	(Third Fret)

Now, for a guitar with frets which are continuous across the neck of the instrument, this is only possible if the interval between the notes on the different strings are all the same. Thus, we require the following equalities in terms of interval,

EF = BC	(one semitone = one fret)
AB = DE = GA	(two semitones = two frets)
EG = AC = DF = BD	(three semitones = three frets)

These equalities are guaranteed in the equal tempered scale. For the above just intonation scale we have,

EF = BC = 1.06666666
AB = DE = 1.125, but GA = 1.111111
EG = AC = DF = 1.2 but BD = 1.185185

Consequently, it is impossible to tune a conventional guitar to this just intonation scale. You would need either a fretless guitar, and the use of subtly different finger positioning on the different strings, or a guitar with different frets for the different strings. In the latter case, you would need a different guitar to play in different keys. The virtual necessity of adopting the equal tempered scale in practice is apparent.

OK, But What About the Different 8 Note Scales? And Why Are They So Weird?

Western music, be it classical, folk, blues, pop, rock or jazz is all based on certain sub-sets of the 12 note diatonic scale. These are the 8 notes shown as bold in the above Table, and named "major" or "perfect". In terms of semitones, the intervals between successive notes of this 8-note scale are 2, 2, 1, 2, 2, 2, 1. This is reflected by the black notes on a piano, which occur in alternating groups of two and three.

In the scale of C, the 8 notes obtained in the above manner are called D, E, F, G, A, B, C. This is the diatonic scale (i.e. the white notes on a piano).

All major scales are obtained from the same semitone sequence, i.e. 2, 2, 1, 2, 2, 2, 1, starting from the tonic note. Hence…

8-Note Major Scales (read downwards from tonic at the top)

C	**C#**	**D**	**D#**	**E**	**F**	**F#**	**G**	**G#**	**A**	**A#**	**B**
D	D#	E	F	F#	G	G#	A	A#	B	C	C#
E	F	F#	G	G#	A	A#	B	C	C#	D	D#
F	F#	G	G#	A	A#	B	C	C#	D	D#	E
G	G#	A	A#	B	C	C#	D	D#	E	F	F#
A	A#	B	C	C#	D	D#	E	F	F#	G	G#
B	C	C#	D	D#	E	F	F#	G	G#	A	A#
C	C#	D	D#	E	F	F#	G	G#	A	A#	B

NB: C#= Dflat; D#=Eflat; F#=Gflat; G#=Aflat; A#=Bflat

So, what are the minor scales? These are generated from the same semitone intervals, but with the last two notes shifted to the beginning. Thus, the semi-tone sequence is 2, 1, 2, 2, 1, 2, 2. Consequently, for each major scale there is a minor scale (the relative minor) which consists of exactly the same notes – but with the tonic (the starting note) shifted down by two notes (3 semi-tones). Thus, the relative minor of C-major is A-minor, and the relative minor of E-major is C#-minor, etc.

If the major scale has the same notes as the relative minor scale, why do they sound different? Because the tonic, the starting note, is different, and because the accentuation within the tune reinforces the tonic (which is said to give a strong ***tonal centre***). Also, the chord progressions in the major and relative minor keys are different. So it ain't just the notes which make the key.

So what about chords? That will have to wait for another day.