The Sounds of Music: Science of Musical Scales*

1. Human Perception of Sound

Sushan Konar

Both, human appreciation of music and musical genres transcend time and space. The universality of musical genres and associated musical scales is intimately linked to the physics of sound, and the special characteristics of human acoustic sensitivity. In this series of articles, we examine the science underlying the development of the heptatonic scale, one of the most prevalent scales of the modern musical genres, both western and Indian.

Sushan Konar works on stellar compact objects. She also writes popular science articles and maintains a weekly astrophysics-related blog called *Monday Musings*.

Introduction

Fossil records indicate that the appreciation of music goes back to the dawn of human sentience, and some of the musical scales in use today could also be as ancient. This universality of musical scales likely owes its existence to an amazing circularity (or periodicity) inherent in human sensitivity to sound frequencies. Most musical scales are specific to a particular genre of music, and there exists quite a number of them. However, the 'heptatonic'[1] scale happens to have a dominating presence in the world music scene today. It is interesting to see how this has more to do with the physics of sound and the physiology of human auditory perception than history. We shall devote this first article in the series to understand the specialities of human response to acoustic frequencies.

[1] Having seven base notes.

Human ear is a remarkable organ in many ways. The range of hearing spans three orders of magnitude in frequency, extending from ~20 Hz to ~20,000 Hz (*Figure* 1) even though the sensitivity

Keywords
String vibration, beat frequencies, consonance-dissonance, pitch, tone.

*DOI: https://doi.org/10.1007/s12045-019-0851-z

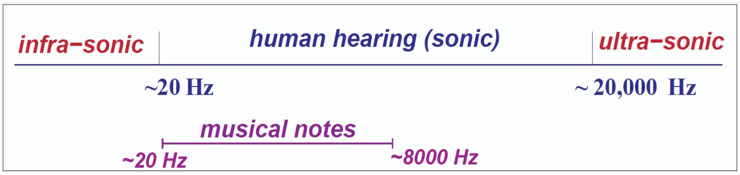

Figure 1. Range of acoustic (sonic) frequencies. Frequencies above and below this range are known as *ultra-sonic* and *infra-sonic* frequencies respectively.

varies widely over this range. Unlike the electromagnetic, sound waves are longitudinal pressure waves. The regions of compression (high pressure) and rarefaction (low pressure) move through a medium with the speed of sound (appropriate for that medium). The separation between pressure peaks (or troughs) corresponds to one wavelength, appropriate for the specific frequency. The modification to the ambient pressure, induced by the propagation of such a pressure wave, is sensed by the human ear and is interpreted as sound. The minimum pressure difference to which the ear is sensitive is less than one billionth ($>10^{-9}$) of one atmosphere, meaning the ear can sense (hear) a wave for which the pressure at the peaks (or troughs) differ from the ambient atmospheric pressure by less than 1 part in a billion. Evidently, human ear is an extremely sensitive instrument even though the sensitivity varies quite a lot from person to person and is not constant over the entire audible frequency range.

Human ear is an extremely sensitive instrument even though the sensitivity varies quite a lot from person to person and is not constant over the entire audible frequency range.

Taken together, the basic characteristics of our ears are as seen in *Table* 1. For ordinary human communication, a frequency span of a decade (ratio of highest to lowest frequency ~10) and a pressure sensitivity ratio of 10^4:1 would suffice. The extremely large range of human auditory sensitivity strongly implies that there must be some other purpose that the human ear is supposed to

Table 1. Sensitivity characteristics of human auditory system.

Human Ear	
Directional coverage	360°
Location accuracy	~5°
Frequency coverage (maximum/minimum)	$\sim 10^3$
Intensity resolution (maximum/minimum)	$\sim 10^9$:1

serve. It is suspected that the ear may have primarily evolved for self-defence (as human beings are endowed with very little in the way of self-defence compared to other big animals). Language and enjoyment of music are likely to be evolutionary byproducts.

1. Sound Waves

1.1 *Standing Waves*

It appears that the modern Western musical scale has its origins in the tuning of a harp-like instrument called the 'lyre' of ancient Greece. This instrument was a 'tetra-chord'[2], and these used to be 'plucked' to create music. The vibrations, thus generated on a string tied at both ends, are *standing* or *stationary* waves as they appear to stand still instead of travelling. In a one dimensional medium, like a string, a standing wave is produced when two waves with the same frequency (therefore, of same the wavelength) and amplitude travelling in opposite directions interfere. Such a standing wave is, therefore, generated in a string by the initial 'pluck', which induces two identical waves that travel in opposite directions and reflect back and forth between the fixed boundary points at two ends. The peak amplitude of a standing wave at any point in space (along the length of the string, in this case) remains constant in time, and the oscillations at different points throughout the wave stay in phase.

[2]Consisting of four strings tied at both ends.

It appears that the modern Western musical scale has its origins in the tuning of a harp-like instrument called the 'lyre' of ancient Greece. This instrument was a 'tetra-chord', and these used to be 'plucked' to create music.

We can express a pair of harmonic waves travelling along the positive and the negative x-axis as:

$$f_1(x,t) = A_0 \sin(\frac{2\pi x}{\lambda} - 2\pi\nu t)\,, \quad (1)$$

$$f_2(x,t) = A_0 \sin(\frac{2\pi x}{\lambda} + 2\pi\nu t)\,, \quad (2)$$

where, x and t are position and time co-ordinates; A_0 is the amplitude and ν is the frequency of the wave. The wavelength, λ, is equal to u/ν where u is the speed of the wave in the medium of propagation. Therefore, the resultant wave, obtained from a

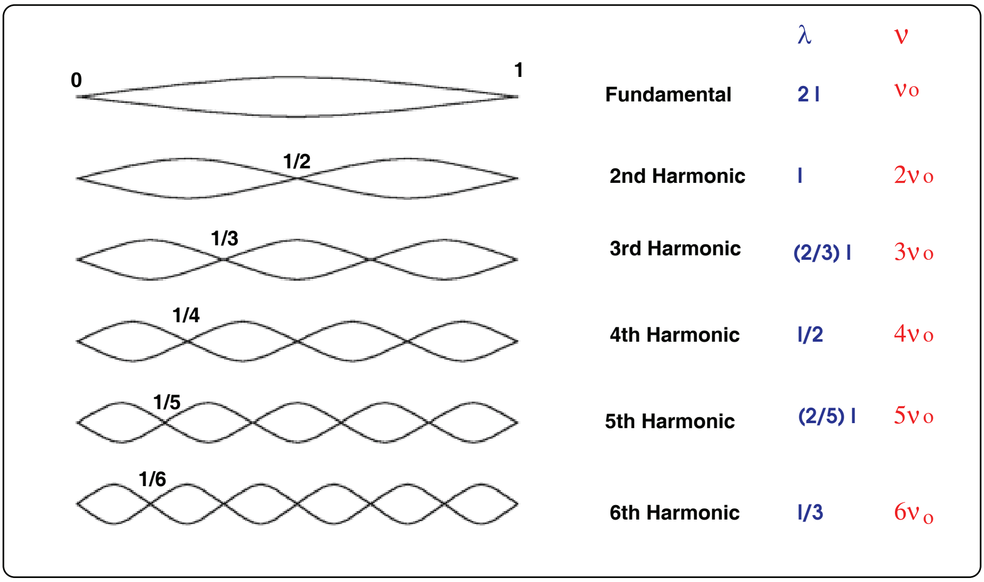

Figure 2. Vibrational modes of a plucked string of length **l**. The wavelengths and the corresponding frequencies of the fundamental (ν_o) and some of the higher harmonics have been illustrated.

superposition of these two is given by,

$$\begin{aligned} f(x,t) &= f_1 + f_2 \\ &= A_0 \sin(\frac{2\pi x}{\lambda} - 2\pi\nu t) + A_0 \sin(\frac{2\pi x}{\lambda} + 2\pi\nu t) \\ &= 2A_0 \sin\left(\frac{2\pi x}{\lambda}\right)\cos(2\pi\nu t)\,. \end{aligned} \tag{3}$$

As can be seen, the amplitude of this wave at a location x, given by $2A \sin\left(\frac{2\pi x}{\lambda}\right)$, is stationary (time-independent). Hence, the name, standing wave. It can also be seen from this that the amplitude goes to zero when,

$$x = \ldots -\frac{3\lambda}{2},\ -\lambda,\ -\frac{\lambda}{2},\ 0,\ \frac{\lambda}{2},\ \lambda,\ \frac{3\lambda}{2},\ldots, \tag{4}$$

and reaches the maximum value when

$$x = \ldots -\frac{5\lambda}{4},\ -\frac{3\lambda}{4},\ -\frac{\lambda}{4},\ \frac{\lambda}{4},\ \frac{3\lambda}{4},\ \frac{5\lambda}{4},\ldots. \tag{5}$$

These points are called 'node's and 'anti-node's respectively.

Now, for a string of length **l** fixed at both ends, the fundamental frequency is given by the vibration whose nodes are at the two ends of the string. Therefore, the wavelength of the fundamental frequency is **2l**. The fundamental frequency is then given by,

$$\nu = \frac{u}{2L}, \tag{6}$$

where u is the velocity of wave propagation which can be obtained as,

$$u = \sqrt{\frac{T}{\mu}}. \tag{7}$$

where T is the tension in the string and μ is the linear mass density of the string. This implies that (a) the shorter the string, the higher the frequency of the fundamental, (b) the higher the tension, the higher the frequency of the fundamental, (c) the lighter the string, the higher the frequency of the fundamental. Therefore, it is possible to modify the fundamental frequency of a string by varying any one of these three quantities (length, mass per unit length, tension). Moreover, the n-th harmonic is obtained when the wavelength is $\lambda_n = 2L/n$. Therefore, the frequency of the n-th harmonic is obtained to be,

$$\nu_n = \frac{nu}{2L} = \frac{n}{2L}\sqrt{\frac{T}{\mu}}, \tag{8}$$

as seen in *Figure* 2. It is obvious that the set of possible vibrational modes supported by such a string is quantised in nature, as only a discrete set of possible wavelengths are allowed. This quantisation property comes from the boundary condition that requires the string to have zero vibrational amplitude at each end.

Once the sound is produced, it travels through the intervening medium to reach the human ear. When two such waves, nearly equal in frequency, are sounded together they produce beats.

1.2 *Beat Frequency*

Once the sound is produced (for example in the string discussed above, or by the human voice or by any other agent), it travels through the intervening medium (the air surrounding us, in most cases), to reach the human ear. When two such waves, nearly equal in frequency, are sounded together they produce beats.

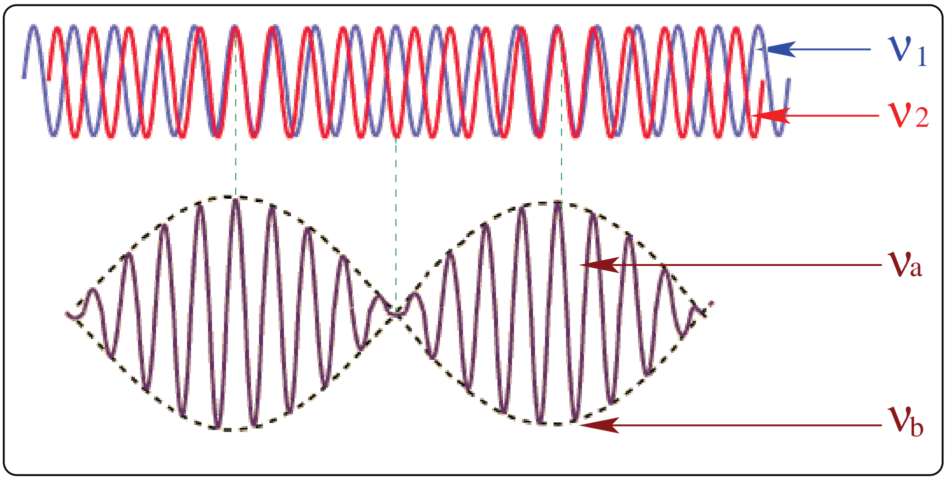

Figure 3. Superposition of two waves with frequencies ν_1 and ν_2. The resultant is a wave of frequency ν_a $(= (\nu_1 + \nu_2)/2)$ modulated by a wave of frequency ν_b $(= (\nu_1 - \nu_2)/2)$.

Whatever may have been the evolutionary logic, the human ear has been endowed with excellent sensitivity to sound. This sensitivity comes accompanied with an appreciation for harmonic relationships between the frequencies present in a given sound.

Let us consider two such travelling waves with frequencies ν_1 and ν_2 and equal amplitude. The superposition of these is given by,

$$\begin{aligned} F(t) = f_1(t) + f_2(t) &= A\cos(2\pi\nu_1 t) + A\cos(2\pi\nu_2 t) \\ &= 2A\cos(\pi(\nu_1 + \nu_2)t)\cos(\pi(\nu_1 - \nu_2)t) \quad (9) \end{aligned}$$

Clearly, these appear as two correlated waves (see *Figure* 3), one modulating (or acting as an 'envelope' for) the other; where one of the frequencies, ν_a, is equal to the average of the original frequencies $((\nu_1 + \nu_2)/2)$ and the other, ν_b, is equal to half of the difference $((\nu_1 - \nu_2)/2)$. When $\cos(\pi(\nu_1 - \nu_2)t) = 1$, the two (original) waves are in phase and interfere constructively. When this quantity vanishes, the waves are out of phase and interfere destructively.

It should be noted that in the modulation pattern, every second burst in the modulation pattern is inverted. Each peak is replaced by a trough and vice versa. Since the human ear is not sensitive to the phase of a sound, only to its amplitude or intensity, the frequency of the envelope appears to have twice the frequency of the modulating wave. Therefore, the audible beat frequency is given by,

$$\nu_{\text{beat}} = 2\nu_b = \nu_1 - \nu_2 \,. \quad (10)$$

2. Human Auditory Perception

Whatever may have been the evolutionary logic, the human ear has been endowed with excellent sensitivity to sound. This sensitivity comes accompanied with an appreciation for harmonic relationships between the frequencies present in a given sound. This, in turn, manifests as two special characteristics of human hearing, as described below.

2.1 *Consonance and Dissonace*

Human perception of harmony between two tones[3] is surprisingly arithmetic, meaning that the closer the two tones are to having a simple ratio of frequencies, the more 'harmonious' (or consonant) they sound. This happens because, in any musical instrument, instead of a pure tone, we generally have a harmonic series comprising the fundamental frequency and its higher harmonics. In general, the fundamental frequency (say, ν) is expected to appear with the maximum amplitude while the higher harmonics $(2\nu, 3\nu, 4\nu..)$ would appear with progressively smaller amplitudes. In presence of these harmonics, the notes (musical frequency) which differ by a ratio of a/b will share many of the early (and therefore stronger) harmonics, when a and b are two small integers. For each integer n the nb-th harmonic of one will be the na-th harmonic of the other. That is,

$$\nu_1 = \frac{a}{b}\,\nu_2 \quad \Rightarrow \quad nb\,\nu_1 = na\,\nu_2\,, \tag{11}$$

implying that tones whose fundamentals are ν_1 and ν_2 would sound sweet, or consonant, together.

[3] A tone is the sound corresponding to one single frequency.

Recent studies have shown that the human brain[4] has two separate centres for processing consonant and dissonant sounds which are associated with different emotions. This explains why humans associate 'pleasure' and 'displeasure' with the consonant and dissonant sounds. Interestingly, natural sounds evoking negative emotional responses, such as avalanches, gales/high winds, tornadoes, or human/animal sounds of somewhat negative nature

[4] See Nandini Chatterjee Singh and Hymavathy Balasubramanian, The Brain on Music, *Resonance*, Vol.23, No.3, pp.299–308, 2018.

(groans, shrieks, howls, roars), typically have non-integer-related harmonic content; whereas human singing voices are inherently consonant in nature (i.e. have integer-related harmonic content).

2.2 *Tone vs. Pitch*

Human beings can hear a large range of acoustic frequencies, but our perception of sound has a curious periodicity to it, and therein lies the concept of *pitch* – a subjective, perceived (by humans) aspect of sound, associated with musical tones.

Human beings can hear a large range of acoustic frequencies, but our perception of sound has a curious periodicity to it. Therein lies the concept of *pitch* – a subjective, perceived (by humans) aspect of sound, associated with musical tones. A musical tone is defined by its frequency. The pitch of a tone is closely related to this frequency. In fact, the frequency and the pitch of a tone can effectively be thought of as being 'proportional' (but not quite). Let us consider two pure tones having the same frequency ν. When these two are heard together, the sound is said to be in 'unison' – both the notes are the same. Let us increase one of the frequencies, such that there is now a non-zero difference in the frequencies ($\delta\nu \neq 0$). From the discussion in the previous section, it can be seen that now we should have a sound wave of base frequency equal to $\nu_a = \nu + \delta\nu/2$, modulated by a beat frequency of $\delta\nu$. Evidently, the base frequency is very close to the original frequency for small values of $\delta\nu$. For $\delta\nu \lesssim 12$ Hz, human ear only detects the sound combination as a single tone (known as a 'fused' tone), because the amplitude modulation for such cases is too slow to be detectable by the human auditory system.

When $\delta\nu \gtrsim 12$ Hz, the human ear begins to perceive two separate tones with some roughness (like 'buzzing'). This perception continues till the difference is large enough for the human ear to perceive two separate tones with clarity. It needs to be noted that the value of $\delta\nu$, at which this clarity is achieved, depends strongly on ν_a. For audio frequencies near the upper limit of human hearing, $\delta\nu$ could be as large as 400 Hz for the two notes to be perceived separately.

If we now continue to increase the frequency difference, the human ear would continue to perceive the two tones as two separate ones, but only up to a point. As the second tone approaches twice the frequency of the first, its 'pitch' returns to its starting point. This gives rise to the situation when two tones, whose frequencies are a factor of two apart, are perceived to be the same musical 'note', although their frequencies are different. These two notes are said to be one *octave* apart in Western musical tradition, an octave being the interval between one musical note and another with half or double its frequency. In frequency space, the interval between the fundamental and second harmonic (or between any successive harmonics) of a harmonic series, is an octave.

All the notes in between unison and the octave are said to have different *chroma* or colour, and notes with the same chroma but separated by one or more octaves are said to have different tone height. While chroma applies to a continuous variation in frequency, pitch refers to tones with specifically defined values of frequency modulo an integral multiple of octaves. This perception of a unique note across octaves is a natural phenomenon that has been referred to as the 'basic miracle of music', the use of which is common in most musical systems. Because the notes that are an integral multiple of octave apart 'sound' same, they are given the same name in musical notations. This is called the *octave equivalence* of musical frequencies[5] and is a part of most contemporary musical cultures.

All the notes in between unison and the octave are said to have different *chroma* or colour, and notes with the same chroma but separated by one or more octaves are said to have different tone height.

[5]The assumption being that the frequencies one or more octaves apart are musically equivalent.

Other similar relations have been seen to exist which have also been of great importance to our musical tradition. For example, two notes which differ in frequency by a factor of 3/2 sound extremely consonant together, and this relation is known as the *consonance of the fifth*. We shall see that this relation (or the more general one described by (11)) is actually a consequence of the octave equivalence, which can be understood in terms of combination tones (or beat frequencies).

A musical fifth comprises two basic frequencies ν_1 and ν_2 where $\nu_2 = 3/2\nu_1$. The superposition of these two generates the beat frequency $\nu_b = \nu_2 - \nu_1 = \nu_1/2$, which is exactly one octave below ν_1. Therefore, the tones generated by a musical fifth $(\nu_1/2, \nu_1, 3/2\nu_1)$ constitute the first three terms in a harmonic series, where ν_b is the 'fundamental', and ν_1 and ν_2 are the second and the third harmonics. Thus when two tones constituting a musical fifth are played, the periodicity of their superposition curve enables the auditory system to sense the presence of a fundamental tone that is not actually present and is, therefore, called a 'missing fundamental'. In other words, a musical fifth is perceived as a harmonic series by the human auditory system. Similarly, for a sequence of harmonics consisting of even and odd integral multiples of the fundamental, each pair of successive harmonics create a combination tone with a periodicity equal to that of the fundamental (which is quite remarkable). Therefore, the fundamental is perceived even when it is absent from the original sequence.

This particular ability of the human auditory system to identify the consonant sound and sense the *pitch* rather than the absolute frequency (tone) of a note is actually at the root of our musical scales. We shall see in the next article how it gave rise to our familiar western classical scale of music.

Suggested Reading

[1] R E Berg and D G Stork, *The Physics of Sound*, Prentice Hall, New Jersey, 1995.

[2] D Benson, *Music: A Mathematical Offering*, Cambridge University Press, 2006.

[3] W C Elmore and N A Heald, *Physics of Waves*, McGraw-Hill, 1969.

[4] K Z Gill and D Purves, A Biological Rationale for Musical Scales, *PLOS One*, 4(12), e5144, 2009.

[5] E G Schellenberg and S E Trehub, Frequency Ratios and the Perception of Tone Patterns, *Psychonomic Bulletin & Review*, 1(2), pp.191–201, 1994.

Address for Correspondence
Sushan Konar
National Centre for Radio Astrophysics
Tata Institute of Fundamenta Research
Pune University Campus, Ganeshkhind
Pune 411 007 India.
Email: sushan@ncra.tifr.res.in