Chapter 4

Sound and Music

This chapter describes the basic concepts of sound, hearing and music that you need to know to understand the theories in this book.

The concepts of sound explained here include vibrations, frequency, sine waves and decomposition into harmonic components. These are mathematical concepts, but they also reflect the way that the first stages of human hearing analyse sound.

The relevant concepts of music are pitch, notes, intervals, octaves, consonant intervals, scales, harmony, chords, musical time, bars, time signatures, note lengths, tempo, melody, bass, repetition (free and non-free), lyrics, rhyme and dancing.

4.1 Sound

4.1.1 Vibrations Travelling Through a Medium

Sound consists of **vibrations** that travel through a medium such as gas, liquid or solid. Sound is a type of **wave**, where a wave is defined as motion or energy that moves along (or propagates) by itself. In particular sound is a **compression wave**, which means that the direction of propagation is aligned with the direction of the motion that is being propagated. At sea-level, under average conditions of pressure, the speed of sound through air is 340 metres per second, or 1224 kilometres per hour.

The effect of sound vibrations passing through a given point in space can be characterised as the displacement of the medium from its normal position (the **zero point**) as a function of time, as shown in Figure 4.1.

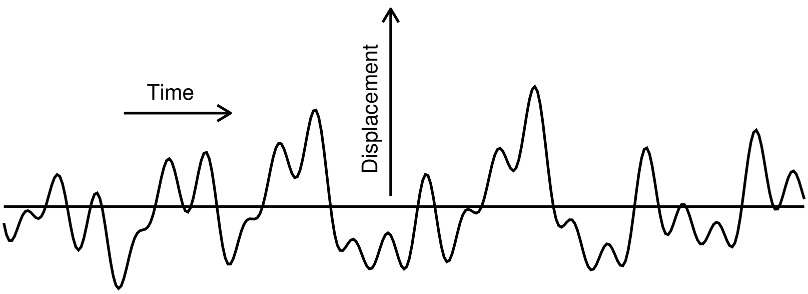

Figure 4.1. A graph of sound waves passing a fixed point, showing displacement as a function of time.

> **Simple Experiment:** Turn on your stereo and play some music moderately loudly. Put your hand on a speaker, and you will be able to feel the speaker vibrating. Now get an empty plastic bottle and hold it in front of the speaker. You will feel the bottle vibrating. The vibrations have travelled from the speaker to the bottle, through the air, in the form of sound waves.

4.1.2 Linearity, Frequency and Fourier Analysis

If two sounds from different sources arrive at a particular point in the medium, the displacements caused by the combined sounds will be the sum of the displacements that would have been caused by the individual sounds. This combination by simple addition is known as **linear superposition** (see Figure 4.2).

If the vibrations that form a sound are regular and repetitive (as in Figure 4.3), we can talk about the **frequency** of the sound. The frequency of a vibration is defined as how many cycles of upward and downward motion occur in a unit of time. Normally vibrations are measured per second. The standard unit of frequency is the **Hertz** (abbreviated **Hz**) which is equal to one vibration per second, e.g. 400Hz = 400 vibrations per second.

The **period** of a vibration is the time it takes to complete one motion from the zero point to a maximum displacement in one direction, back to the zero point, on to a maximum displacement in the opposite direction and back to the zero point again. Period and frequency are necessarily related:

$$\text{frequency} \times \text{period} = \text{unit of time}$$

The human ear can normally detect sounds with frequencies ranging from 20Hz to 20000Hz. The frequency corresponds psychologically to **pitch** which

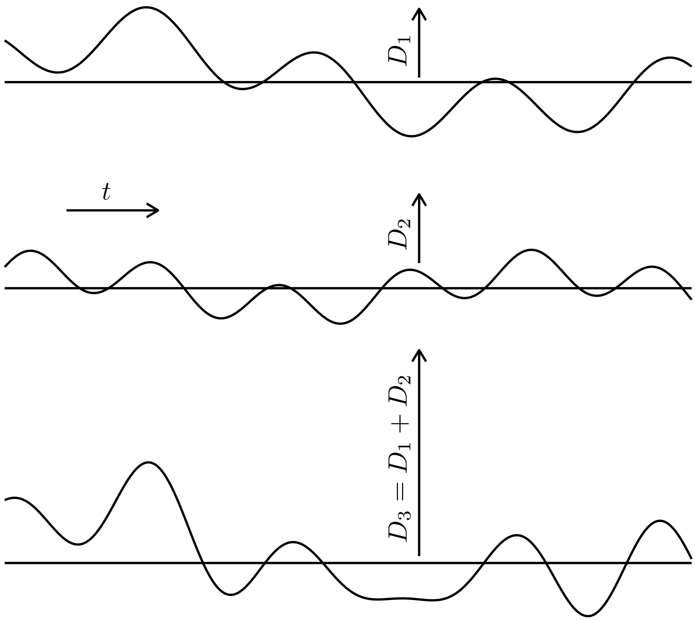

Figure 4.2. Linear superposition. Displacement D is a function of time t. $D_1 + D_2 = D_3$ for each time t. D_1 as a function of time is the displacement at a given point caused by one sound, D_2 is the displacement at the same point caused by another sound, and D_3 is the total displacement caused by the combined effect of those two sounds. (This simple example ignores the complication that if the sounds come from different directions then the displacements will be in different directions, and it will be necessary to use vector addition to add them together.)

represents the listener's perception of how "high" or "low" the sound is. On a piano, lower frequencies are to the left and higher frequencies are to the right.

A regular repetitive sound is completely characterised by its frequency, its **amplitude** and the shape of the vibration. The amplitude is defined as the maximum displacement of the vibration from the zero point, and bears a relationship to the perceived loudness of the sound.[1]

The "shape" of a vibration is the shape that you see if you draw a graph of displacement as a function of time. Psychologically, it corresponds to the perceived quality or **timbre** of a sound. However, perceived timbre is more than just a fixed shape of vibration: it generally corresponds to a shape of

[1] A precise description of this relationship is that perceived loudness is a function of the energy of the wave, and that for a given frequency and shape of vibration, the energy is proportional to the square of the amplitude.

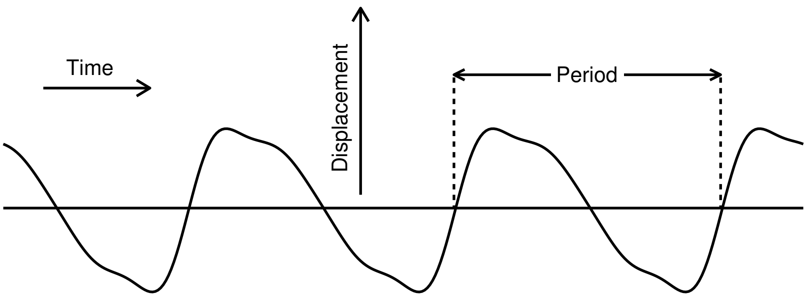

Figure 4.3. Sound consisting of a regular repetitive vibration.

vibration that may change as a function of time (i.e. after initial onset of the sound), and as a function of frequency and amplitude. Vibrations of some instruments, such as the piano, usually change shape and amplitude as time passes, whereas vibrations from other instruments, such as the violin and the saxophone, can be relatively constant in shape and amplitude.

The definition of **period** given above assumes a simple model of vibration consisting of motion upwards to a maximum, downwards to a maximum in the opposite direction, back up to the first maximum, and so on. In practice, a regularly repeating shape of vibration may have smaller upward and downward motions within the main cycle of vibration, as in Figure 4.4. In such cases we measure the period and frequency in terms of the rate of repetition of the total shape.[2]

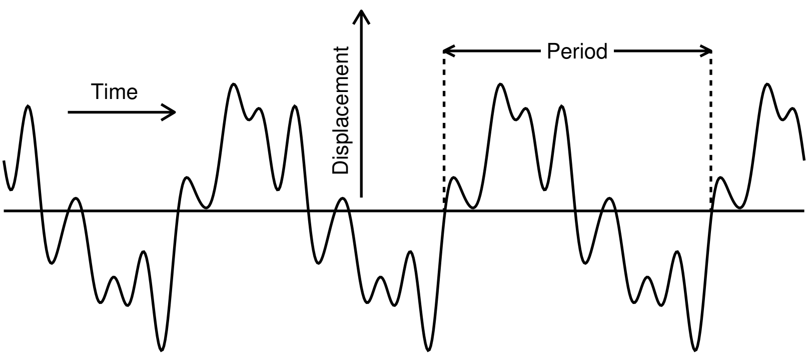

Figure 4.4. Sound consisting of a regular repetitive vibration but with little ups and downs within the main vibration.

[2]Of course we can argue that the smaller vibrations within the larger vibration deserve their own measure of frequency. We will resolve this issue when Fourier analysis is introduced.

A particularly important shape of vibration is the **sine wave**. If we imagine a point on a circle that is rotating evenly at a particular frequency, e.g. 400 cycles per second, then the height of that point above a particular baseline drawn through the centre of the circle, as a function of time, defines a sine wave, as shown in Figure 4.5.

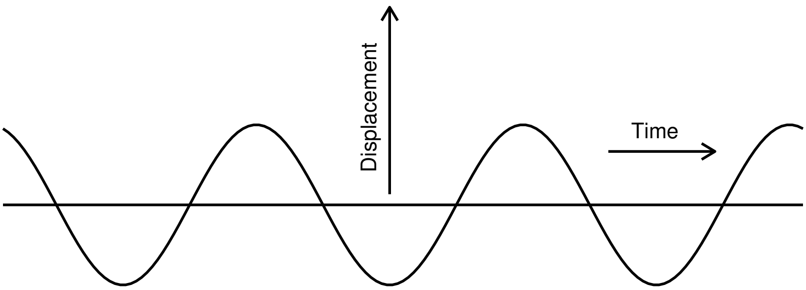

Figure 4.5. Sine wave vibration.

If you remember school-level trigonometry, you may remember sine as being a function of angle. In particular the sine of an angle θ is defined in terms of a right angle triangle, where the angle between two of the sides is θ: the sine is the length of the side opposite the angle θ divided by the length of the hypotenuse (see Figure 4.6).

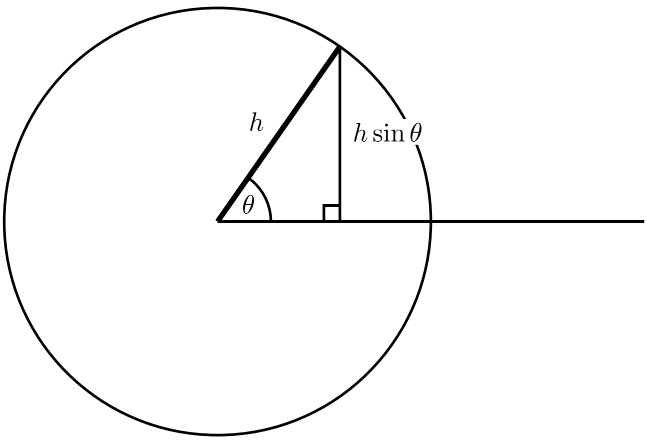

Figure 4.6. Definition of the **sine** function: $\sin\theta$ is the length of the side opposite the angle θ divided by the length of the hypotenuse h ("sin" is the abbreviation for "sine" used in mathematical equations and formulae).

This is the same thing as the definition in terms of a point moving around a circle, as long as we assume that:

- the circle has a radius of 1 unit,
- the point was on the base line at time zero,
- it was travelling upwards at this time, and
- the period of each vibration is mapped to 360 degrees (or 2π radians).

The important thing about sine waves is that *any regular shape of vibration* can be decomposed into a sum of sine wave vibrations, where the frequency of each sine wave vibration is a multiple of the frequency of vibration. For example, any shape of vibration at 100Hz can be decomposed into a sum of sine wave vibrations at 100Hz, 200Hz, 300Hz, and so on.[3] Furthermore, such a decomposition (where it exists) is unique.

Figure 4.7 shows an analysis of a periodic vibration into four sine wave components.

The frequency of the vibration itself is called the **fundamental frequency**, and the multiples of the frequency are called **harmonics** or **harmonic frequencies**. The decomposition of an arbitrary shape of vibration into harmonics is characterised by assigning an amplitude and **phase** to each sine wave component. The phase is the angle of the point on the circle defining that sine wave at time zero.

This decomposition of vibrational shapes into sine waves defines the mathematical topic of **Fourier analysis**. It is important for two main reasons:

1. Sine wave functions have mathematical properties that make them easy to deal with for many purposes. An arbitrary vibrational shape can be analysed by decomposing it into component sine waves, doing a calculation on each sine wave, and then adding all the results back together. As long as the calculation being done is **linear** (which means that addition and scalar multiplication[4] "pass through" the calculation), then this works. It's often even useful when the calculation is almost linear, as long as you have some manageable way to deal with the non-linearities.

2. Decomposition into sine waves corresponds very closely to how the human ear itself perceives and analyses sound. The point at which sound entering the human ear is translated into nerve signals is the **organ of Corti**. The organ of Corti is a structure which lies on the **basilar membrane** and contains special auditory receptor **hair cells**. The basilar membrane is a membrane which vibrates in response to sounds

[3] This is almost true. Highly sophisticated mathematical concepts were invented by mathematicians trying to completely understand the "almost". It is possible for the reconstruction of a function from its decomposition into sine wave functions to be not quite identical to the original function, but for most purposes this complication can be ignored.

[4] **Scalar multiplication** refers to multiplying something like a function by a simple number—the **scalar** is the simple number.

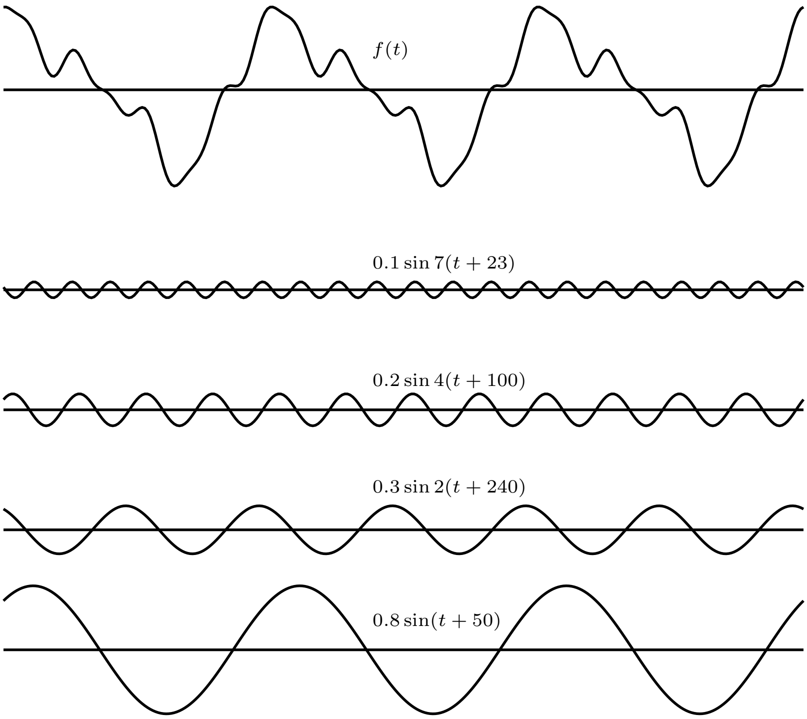

Figure 4.7. The periodic function f can be decomposed into the sum of four sine wave functions: $f(t) = 0.8 \sin(t+50) + 0.3 \sin 2(t+240) + 0.2 \sin 4(t+100) + 0.1 \sin 7(t + 23)$. (Here t is assumed to be measured in degrees.)

that enter the human ear. The shape of the basilar membrane and its position in the ear are such that there is a direct correspondence between the frequency of each sine wave component of a sound and the positions of the hair cells activated by that component. The hair cells become electrically depolarised in response to shearing stress, and this depolarisation activates **spiral ganglion neurons**, which are the next stage in the neural pathway that transmits information about sound from the ear to the auditory cortex.

The human ear and associated auditory processing parts of the brain analyse sound into frequency and amplitude of sine wave components. Each sine wave component also has a phase; but the only major use of phase information appears to be when perceived differences between phases of sounds received by the left ear and the right ear are used to help determine the locations of those sounds. In general phase information appears to play no significant role in the perception of music. One consequence of this is that the manufacturers of stereo equipment must be concerned about preserving the relative phases of the same sounds being processed in the left and right channels (partly because our brains use the phase differences to determine location, and partly because relative phase errors can cause unwanted interference effects), but they do not have to be so concerned about preserving phase relationships between different frequency components of the same sound being processed within one channel.

Very few natural sounds consist of completely regular repeated vibrations. But many sounds can be regarded as close enough to regular over a limited time period or **window** (see Figure 4.8). Thus one can analyse sound into frequency components as a function of time by performing analysis of the sound in a sliding window, where the window is centred on the current point in time. The amplitude of each frequency at each moment of time is then defined to be the amplitude of the frequency component of the sound contained within the window at that time. In practice we use a window that is much larger than the period of the vibrations being perceived (which in the human case is never more than 1/20 of a second) and much smaller than the period of time over which we are tracing the evolution of the characteristics of the sound. The result of this analysis is a **spectrogram**. A variety of computer software is available that can be used to create spectrograms. The software I used to generate the spectrograms in Figures 4.9 and 4.10 is PRAAT. PRAAT is licensed under the GNU General Public License, and it can be downloaded from **http://www.praat.org/**.

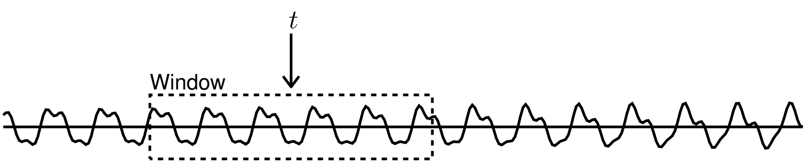

Figure 4.8. Vibration analysed inside a sliding "window". A window size is chosen such that the pattern of vibration is approximately constant within the window. Frequency analysis at each time t is based on analysis of vibration within the window centred on that time.

Figure 4.9 shows a spectrogram of some speech, and Figure 4.10 shows a spectrogram of part of a song. Even looking at these small fragments, you can see that the song has more regularity in both pitch and rhythm. The harmonics are clearly visible in the vowel portions of the syllables. The consonants tend to show up as an even spread of frequencies at the beginnings of syllables, reflecting their "noisy" nature.

Although a sound can have an infinite number of harmonics, the human ear cannot normally hear sounds over 20000Hz. If a sound has a fundamental frequency of (for instance) 1000Hz, it can have harmonics for all multiples of 1000Hz going up to infinity, but any harmonics over 20000Hz will make no

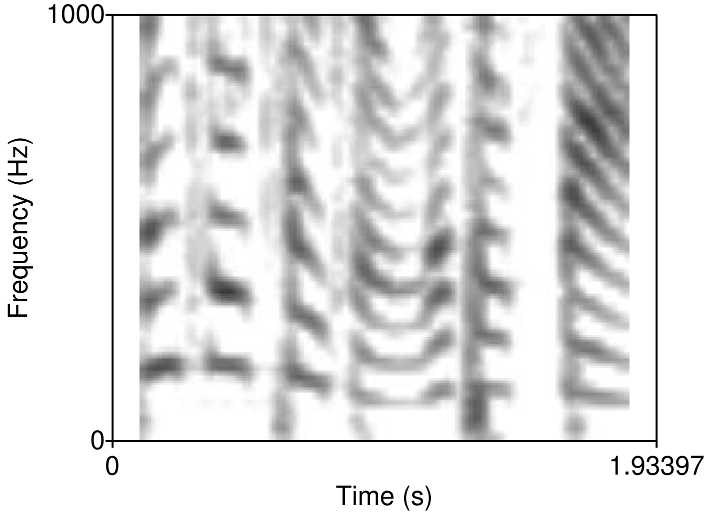

Figure 4.9. A spectrogram of the author saying "Twinkle Twinkle Little Star".

difference to our perception of that sound.

4.2 Music: Pitch and Frequency

4.2.1 Notes

A fundamental component of music is the **note**. A note consists of a sound that has a certain unchanging (or approximately unchanging) frequency and a certain **duration**. Notes are generally played on **instruments** (which can include the human voice). The shape of vibration of a note will depend on the **timbre** of the instrument which will determine the shape as a function of elapsed time, frequency and amplitude. (In cheap electronic instruments the shape will be constant regardless of frequency, amplitude and elapsed time. In proper instruments the shape will vary according to elapsed time, frequency and amplitude in a manner which is pleasing to the ear and which contributes to the musicality of the music.)

In musical contexts, frequency is referred to as **pitch**. Strictly speaking, pitch is a perceived quantity that corresponds *almost* exactly to frequency—variables such as timbre and amplitude can have a small effect on perceived pitch, but mostly we can ignore these effects.

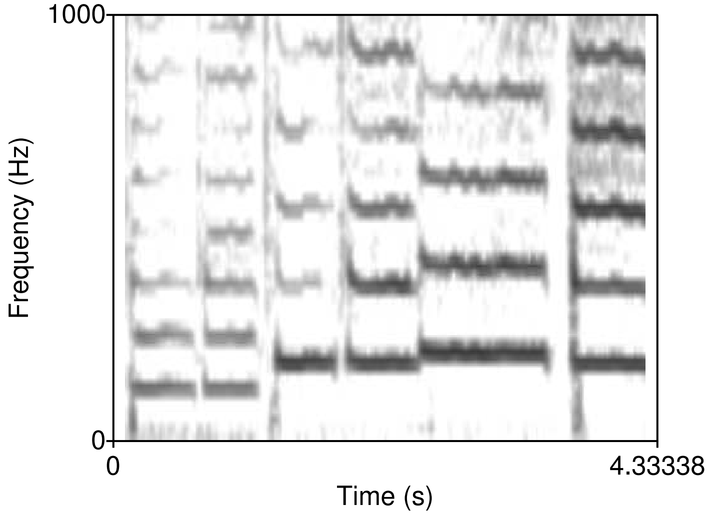

Figure 4.10. A spectrogram of the author singing "Twinkle Twinkle Little Star".

4.2.2 Intervals

An important component of music perception is the perception of **intervals** between notes. Perceived intervals correspond to *ratios* of frequencies. That is, the differences between two pairs of notes are considered equal if the ratios are equal. To give an example, the interval between two notes with frequencies 200Hz and 300Hz is perceived to be the same as the interval between 240Hz and 360Hz, since the ratio is 2 to 3 in both cases. Because intervals relate to ratios, it is often convenient to represent musical frequencies on a **logarithmic scale**.[5]

There are two types of interval that have special significance in music. Two notes whose frequencies differ by a power of 2 are psychologically perceived to have a similar quality. For example, a note at 250Hz would be perceived to have a similar quality to one at 500Hz, even though the 250Hz note is obviously a lower note than the 500Hz note. This ratio of 2 is normally referred to as an **octave** (the "oct" in "octave" means 8, and derives from the particulars of the scale used in Western music).

[5] A **logarithm** is a function f such that $f(x \times y) = f(x) + f(y)$. The **base** of a logarithm is the number b such that $f(b) = 1$. We will see that, in a musical context, the number of semitones in an interval is equal to the logarithm of the ratio of frequencies represented by the interval, where the base of the logarithm is $\sqrt[12]{2}$. A **logarithmic scale** is one that locates values according to their logarithms. (This is a non-musical meaning of the word "scale".)

The second type of musically important interval is any simple fractional ratio that is *not* a power of 2. Ratios that play a significant role in Western music include 3/2, 4/3, 5/4, 6/5 and 8/5. Two notes separated by such an interval do not sound similar in the way that notes separated by an octave sound similar, but the interval between them sounds subjectively "pleasant" (whether the notes are played simultaneously or one after the other). This phenomenon is known as **consonance** and the intervals are called **consonant intervals**.

As the reader may have already noticed, the ratios that define consonant intervals are the same as the ratios that exist between the harmonic components of a single (constant frequency) sound. For example, a musical note at 200Hz will have harmonics at 400Hz and 600Hz, and the ratio between these is 2:3, which corresponds to the harmonic interval that would exist between two notes with fundamental frequencies of 400Hz and 600Hz. It follows that two notes related by a consonant interval will have some identical harmonics: for example the 3rd harmonic of a 400Hz sound is 1200Hz which is identical to the 2nd harmonic of a 600Hz sound. However, harmonic intervals can be recognised even between notes that have no harmonics (i.e. pure sine waves), so the recognition of harmonic intervals is not necessarily dependent on recognising matching harmonics.

4.2.3 Scales

In most forms of music, including all popular and classical Western music, notes are taken from **scales**. A scale is a fixed set of pitch values (or **notes**) which are used to construct music.[6] Western Music has mostly adopted scales that are subsets of the **well-tempered chromatic scale**. The chromatic scale consists of all the black and white notes of the piano, as shown in Figure 4.11.

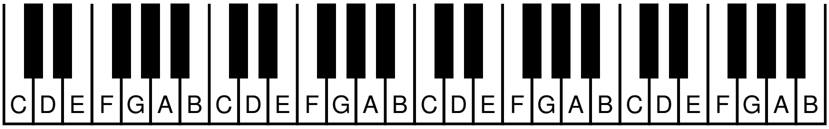

Figure 4.11. A musical keyboard.

Notes on the piano (and other keyboards) increase in frequency as you go from left to right. The interval between each note and the next is always the same, and is a ratio of $\sqrt[12]{2}$, which to ten decimal places (an accuracy that

[6]There are two subtly different usages of the word "note": to refer to a possible pitch value from a scale (e.g. "the note C sharp"), and (as defined earlier) to refer to a particular occurrence of a musical sound with that pitch value in an item of music ("the third note in this song").

far exceeds the capabilities of the human auditory system), is 1.0594630943. Each such interval is called a **semitone** (although this term can also be used to represent a similar sized interval on other scales). The expression "well-tempered" refers to the fact that all the semitones are the same ratio. The interval consisting of 12 semitones corresponds to a frequency ratio of exactly 2, which we have already defined as an **octave**. If we look at the piano or similar keyboard, we will see a pattern of 5 black notes (a group of 2 and a group of 3) and 7 white notes, which is repeated all the way along the keyboard. Each such pattern is 1 octave higher than the pattern to the left of it.

The notes within each pattern have standard names. The white note to the left of the group of 2 black notes is the note C. The names of the other white notes going upwards (i.e. to the right) are D, E, F, G, A and B. The black notes have names derived from their neighbouring white notes. The black note just to the left of a white note X is written $X\flat$ which reads "X flat", and the black note just to the right of a white note X is written $X\sharp$ which reads "X sharp". For example, the black note immediately to the right of C can be called either C♯ or D♭.

For the sake of standardisation, one particular note is tuned to a particular frequency. **Middle C** is the C that is found in the middle of a standard piano keyboard. The A above middle C is defined to have a frequency of 440Hz. This standardisation of frequency guarantees that everyone's musical instruments operate according to compatible tunings. A specific choice of frequency is *not* crucial to the effect that music generates, in fact we will see that one of the fundamental facts about music is that the absolute pitch of notes is relatively unimportant, and it is the intervals between notes, or their **relative pitches**, that matter.

The simplest forms of Western music are played on a subset of the chromatic scale called the **diatonic scale**. A simple example of this is the white notes on the piano, i.e. the notes C, D, E, F, G, A, B. Given the previous remark on independence of absolute pitch, we realise that what matters is the relative pitches of the notes. For example, taking C as a base note, the diatonic scale must include notes 0, 2, 4, 5, 7, 9 and 11 semitones above this base note. If we changed the base note to E, then the notes E, F♯, G♯, A, B, C♯ and D♯ would define what is effectively the same scale. Any music played on the notes C, D, E, F, G, A and B could be shifted to the corresponding notes of the scale with E as a base note, and it would sound much the same, or at least its musical quality would be almost identical. This shifting of music up or down the chromatic scale is known as **transposition**.

To emphasise the fact that music can be moved up and down by an interval that is not necessarily a whole number of semitones, I will talk about musical quality (or "musicality") being invariant under **pitch translation** (rather than saying it's invariant under transposition). This will be explained in detail in the chapter on symmetries.

In traditional musical language, scales are identified with a specific **home note**. For example, the white notes scale is usually either the **scale of C major** or the **scale of A minor** (also referred to as the **key of C major** and the **key of A minor**). Both of these scales contain the same set of notes, but the scale of C major has C as its home note, and the scale of A minor has A as its home note. The scale is regarded as both a set of notes and the home note. The home note is a note that the music usually starts with, and finally ends with. So, for example, if a tune in the key of C major is transposed 4 semitones higher, we would say that it has been transposed from the key of C major to the key of E major.

For the purposes of this book, I want to refer to a scale as a set of notes, without specifying any particular note as a special note. Determination of a home note is deferred to a separate stage of analysis. So I will define the **white notes scale** to be the scale consisting of the notes C, D, E, F, G, A, B. The term "diatonic scale" in effect describes any scale that is a transposition of the white notes scale. In many places I discuss properties of the diatonic scale, but when I want to give concrete examples with specific notes, I use the white notes scale.[7]

A **tone** is defined to be 2 semitones. You will notice that the intervals between consecutive notes on the diatonic scale are all either 1 semitone or 1 tone.

4.2.4 Consonant Intervals

An interval of 12 semitones is exactly equal to an interval that corresponds to a ratio of 2. I have already said that intervals equal to simple ratios, i.e. so-called "consonant intervals", are important to music. But how do powers of $\sqrt[12]{2}$ fit into this picture? It can be mathematically proven that no integral power of $\sqrt[12]{2}$ other than exact multiples of 12 can ever be an exact fraction.

What happens in practice is that some of the intervals are close enough to consonant intervals to be recognised as such by those parts of our brain that respond to consonant intervals, and they are close enough to make music played on the well-tempered scale sound musical. It is also possible to define scales where the intervals are exactly consonant. However, there are difficulties in trying to do this, and I do an analysis of these difficulties in Chapter 5 when discussing vector representations of musical intervals.

The following table shows all the exact well-tempered intervals and the corresponding approximate consonant intervals, up to and including an octave, which can be found between notes on the chromatic scale:

[7] There is a musical terminology **do**, **re**, **mi**, **fa**, **sol**, **la**, **ti**, (made famous in a song sung by Julie Andrews) which can be used to refer to positions in the diatonic scale without assuming any absolute location, but this notation is both clumsier and less familiar to most readers.

Semitones	Ratio	Consonant Ratio	Fraction	Note
0	1.0	1.0	1	C
1	1.05946309			$C\sharp(D\flat)$
2	1.12246205			D
3	1.18920712	1.2	6/5	$D\sharp(E\flat)$
4	1.25992105	1.25	5/4	E
5	1.33483985	1.33333333	4/3	F
6	1.41421356			$F\sharp(G\flat)$
7	1.49830708	1.5	3/2	G
8	1.58740105	1.6	8/5	$G\sharp(A\flat)$
9	1.68179283	1.66666666	5/3	A
10	1.78179744			$A\sharp(B\flat)$
11	1.88774863			B
12	2.0	2.0	2	C

The right hand "Note" column shows the notes such that the interval from C to that note is the interval whose details are shown on that row. For example, the interval from C to E is 4 semitones.

There are some standard names used for different sized intervals. Four that I will often refer to in this book are:

- an **octave** = 12 semitones = a ratio of 2,
- a **perfect fifth** = 7 semitones ≈ a ratio of 3/2,
- a **major third** = 4 semitones ≈ a ratio of 5/4, and
- a **minor third** = 3 semitones ≈ a ratio of 6/5.

4.2.5 Harmony and Chords

Harmony is where different notes are played simultaneously.

Harmony can often be described in terms of **chords**. A chord is a specific group of notes played together, either simultaneously, or one after the other, or some combination of these. Typically the notes in a chord are related to each other by consonant intervals.

The most common chords found in both popular and classical Western music are the **major chords** and **minor chords**. Each chord has a **root note**. A major chord contains the root note and the notes 4 semitones and 7 semitones higher than the root note. A minor chord contains the root note and the notes 3 semitones and 7 semitones higher. So, for example, C major consists of C, E and G, and C minor consists of C, E♭ and G.

The musical quality of a chord is—at least to a first approximation—unaffected by notes within that chord being moved up or down by an octave.

The following list shows some of the ways that the chord of C major can be played (with notes listed from left to right on the keyboard):

- C, E, G
- G, C, E
- C, C (an octave higher), G, C, E

However, having said this, there is a tendency to play some of the notes at certain positions. For example, with the chord C major, the lowest note played would normally be C, and one would not play the note E too close to this lowest C. In general the root note of the chord is the one that is played lowest. In practice this rule is usually satisfied by the existence of a separate **bass line** (see section on bass below) which includes the root notes of the chords.

The next most common chords (after the major and minor chords) are certain 4-note chords derived from the major or minor chords by adding an extra note:

- **Seventh** or **dominant seventh**: 0, 4, 7 and 10 semitones above the root note, e.g. G7 = G, B, D and F.
- **Major seventh**: 0, 4, 7 and 11 semitones above the root note, e.g. C major 7 = C, E, G and B.
- **Minor seventh**: 0, 3, 7 and 10 semitones above the root note, e.g. A minor 7 = A, C, E and G.

The five types of chord described so far account for a large proportion of the chords that appear in traditional and modern popular music. Other less commonly used chord types include **suspended chords**, such as CDG and CFG, where the D and the F represent “suspended” versions of the E in C major. Such chords often **resolve** (see next section on home chords and dominant 7ths for more about resolution) to their unsuspended relations.

Chords with 5 or more notes have a softer feel, and occur more often in jazz music. Even music with more than the average number of 4-note chords (most popular music has more 3-note chords than 4-note chords) has a similar softer feel.

Sometimes 2-note chords appear, in particular 2 notes separated by a perfect fifth, e.g. CG, which is like a C chord that doesn’t know if it’s a major chord or a minor chord. This type of chord has a harder feel.

4.2.6 Home Chords and Dominant Sevenths

Scales that have home notes also have **home chords**. The root note of the home chord is the home note, and usually the notes of the home chord are all notes on the scale. So if the home note of the white notes scale is C, then the home chord will be C major, i.e. C, E and G. If the home note on the white notes scale is A, then the home chord is A minor, i.e. A, C and E.

The dominant seventh chord has a strong tendency to be followed by a chord, either major or minor, that has a root note 5 semitones higher (or 7 lower). We say that the following chord **resolves** the preceding dominant 7th, and there is some feeling of satisfying a tension created by the dominant 7th. Typically the dominant 7th appears just before a corresponding home chord, as the second last chord of the song or music. For example, in the key of C major, the second last chord will be a G7, i.e. G, B, D and F, which will resolve to a C major, i.e. C, E and G. Similarly, in A minor, the second last chord will be E7, i.e. E, G♯, B and D, which will resolve to A minor, i.e. A, C and E.

The G♯ in E7 is not contained in the scale of the key of A minor. It may, however, still occur within a tune to match the occurrence of the E7 chord. Usually G♯ appears where we might otherwise expect G to occur. If G does not occur at all in the tune, then we can consider the scale as being changed to one consisting of A, B, C, D, E, F and G♯. This scale is called the **harmonic minor scale**. It has an interval of 3 semitones between the F and the G♯. We can "fix" this over-sized interval by moving the F up to F♯, to give the scale A, B, C, D, E, F♯, G♯, which is known as the **melodic minor scale**.

4.3 Musical Time

The second major aspect of music, after pitch, is time. Music consists essentially of notes and other sounds, such as percussion, played at certain times.

Musical time is divided up in a very regular way. The simplest way to explain this is to consider a hypothetical tune:

- The time it takes to play the tune is divided up into **bars**. Each bar has the same duration, which might be, say, 2 seconds. The tune consists of 16 bars. The structure of the tune might consist of 4 identifiable **phrases**, with each phrase corresponding to 4 bars.[8]

- The tune has a **time signature** of 4/4. The first "4" tells us that the duration of each bar is divided up into 4 **beats** (or **counts**[9]). The second "4" in the signature specifies the length of the note. So a tune

[8] A phrase 4 bars long might, however, not be neatly contained inside 4 bars—it might (for instance) consist of the last note of one bar, three whole bars and then the first 3 notes of a fifth bar.

[9] I sometimes prefer the word "count" to "beat" in this context, because I use "beat" in a more generic sense when talking about "regular beats", which may or may not correspond to the "beats" in "n beats per bar".

with signature 4/4 has 4 **quarter** notes (also known as **crotchets**) per bar. A quarter note is one quarter the length of a "whole" note (or **breve**), but there is no fixed definition of the length of a whole note. Therefore the choice of note length for a time signature is somewhat arbitrary, and partly a matter of convention. A tune with a 4/2 time signature is probably intended to be played more slowly than one with a 4/4 time signature. The fraction representing the time signature is usually chosen to be not too much greater than 1 and not less than 1/2. Typical signatures include 4/4, 6/8, 3/4, 2/2, 12/8 and 9/8.

- Each of the 4 beats in a bar has an implicit intensity: beat 1 is the strongest, beat 3 is the next strongest and beats 2 and 4 are the weakest and similar to each other. We can regard each beat as corresponding to the portion of time that starts at that beat and finishes at the next beat.

- The time within beats can be further divided up into smaller portions. In most cases, durations are divided up into 2 equal sub-durations, and the beat at the half-way point is always weaker than any beat at the beginning of a duration the same length as the duration being divided. Our hypothetical tune might contain durations of 1/2 and 1/4 the main note length. The smallest duration of time such that all notes in a tune can be placed on regular beats separated by that duration defines the finest division of time that occurs within that tune, and in most cases is equal to the shortest note length occurring in the tune. There appears to be no standard term for this duration, so for the purposes of this book I will call it the **shortest beat period**.

To sum up the division of time in this hypothetical tune: there are 4 groups of 4 bars, each bar has 4 beats, and each beat can be divided into 4 sub-beats. Where things occur in groups of 4, there is a tendency for these 4's to be actually pairs of 2. As a result the division of time can be written as:

$$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \text{ sub-beats}$$

(Figure 4.12 shows the division of time in 8 of these bars—if I tried to fit 16 bars into the width of the page, the divisions corresponding to the shortest beat period would be too fine to print properly.)

The number 2 strongly dominates division of time within music. But the number 3 does make occasional appearances. That famous dance the **waltz** requires music that is 3 beats to the bar. Music can also be found that is 6, 9 or 12 beats to the bar, in which the beats on the bars are interpreted as 2 groups of 3 beats, 3 groups of 3 beats or 2 groups of 2 groups of 3 beats respectively. Divisions of time within a beat are almost always in powers of 2, but sometimes music contains **triplets**, which are groups of 3 notes within a duration that normally contains 1 or 2 notes.

There are a very few tunes where the beats are grouped into groups of 5 or 7.[10] In these cases the groups of beats may be grouped into uneven halves, e.g. alternating 2 and 3 beat bars, or alternating 3 and 4 beat bars.

The division of time into smaller and smaller pieces, step by step, where each step is either a factor of 2 or 3, forms a sequence. The notions of "bar" and "beat" represent two particular positions within this sequence. Are these positions truly special, or are they assigned arbitrarily?

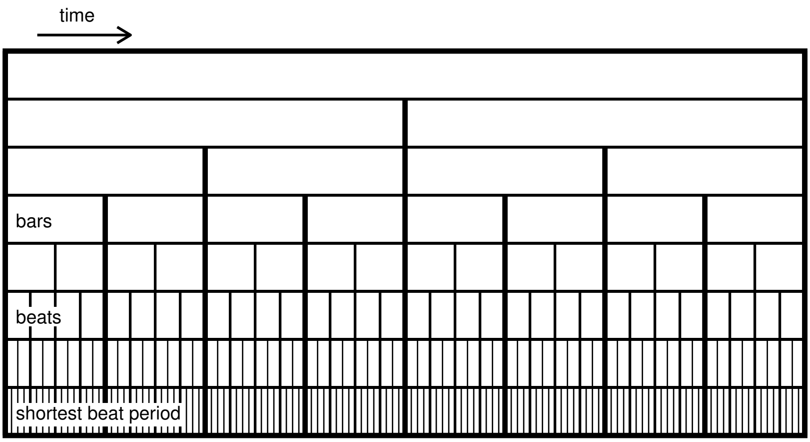

Figure 4.12. Hierarchical division of musical time. This example shows eight bars. The time signature is 4/4, i.e. 4 beats to a bar, and the tune contains notes that are 2 or 4 notes to a beat, so that the shortest beat period is 1/16 of the length of the bar.

Firstly, with respect to bar length, if you had a tune of 16 bars with 4 beats to the bar, with each bar 2 seconds long, could you claim that actually it was 8 bars with 8 beats to the bar and each bar 4 seconds long? I've already mentioned that the notes within a bar have different strengths according to their position. A general criterion for bar size is that there is no variation in beat strength from one bar to the next. If bars for a tune are in pairs, where the first bar in each pair has a stronger beat than the second one, then we have chosen the wrong bar length, and what we notated as pairs of bars should be joined together into single bars of twice the length.

Secondly, with respect to beat length, what is the difference between 4 quarter notes per bar and 8 eighth notes per bar? The distinction between these two possibilities seems somewhat more arbitrary, as a tune with a time signature of 4 quarter notes per bar can still contain eighth notes and sixteenth notes.

[10] At least there are very few such tunes in modern Western popular music. The traditional folk music of some cultures makes heavy use of "complex" time signatures with 5, 7, 9, 11 or even 13 beats to the bar. For example, 7/8 is a common time signature in Macedonian folk dances.

To give an example of how convention determines the assignment, 6 beats to the bar *always* represents 2 groups of 3 beats. If you have a tune that has 3 groups of 2 beats in each bar, this always has to be notated as 3 beats to the bar (notating each group of 2 beats as if it was 1 beat).

4.3.1 Tempo

Given a division of musical time into bars and beats, the **tempo** refers to the number of beats per unit of time. Normally the unit of time is minutes, so tempo is given as beats per minute. Often tempo varies gradually during the performance of a musical item. There can be some leeway as to what tempo a given piece of music is played in, but at the same time there is usually an optimal tempo at which the music should be played.

4.4 Melody

Having described pitch and musical time, we can now explain what a melody is:

> A **melody** is a sequence of notes played in musical time.

Usually the notes of a melody are all played using the same musical instrument (where the notion of "instrument" includes the human voice). The notes do not overlap with each other, and in general the end of one note coincides exactly with the start of the next note.[11] However, it is also possible for a melody to contain **rests**, which consist of silent periods that occur in between groups of notes (or **phrases**) in the melody.

The sounds produced by musical instruments used to play melodies are usually sounds that satisfy the requirements of frequency analysis into harmonics. That is, they consist of regular vibrations at a fixed frequency, with the shape of vibration either constant or varying slowly in a manner typical for that instrument. These sounds will therefore have identifiable harmonic components. Examples of instruments that satisfy these criteria include the human voice; string instruments like the violin, guitar and piano; and wind instruments like the flute, clarinet and trumpet. Electronic instruments allow an almost unlimited range of sounds; but when they are used to play melodies, the timbres are usually either imitations or variations of the sounds produced by traditional instruments, or they are artificial sounds that still satisfy the criteria of being regular vibrations with identifiable harmonic components.

The notes of a melody generally come from a particular scale. Most melodies in popular Western music exist on the diatonic scale. However, some melodies contain **accidental notes** that consist of additional notes from the chromatic scale temporarily included in the tune. In many cases the inclusion

[11] The technical term for this is **legato**.

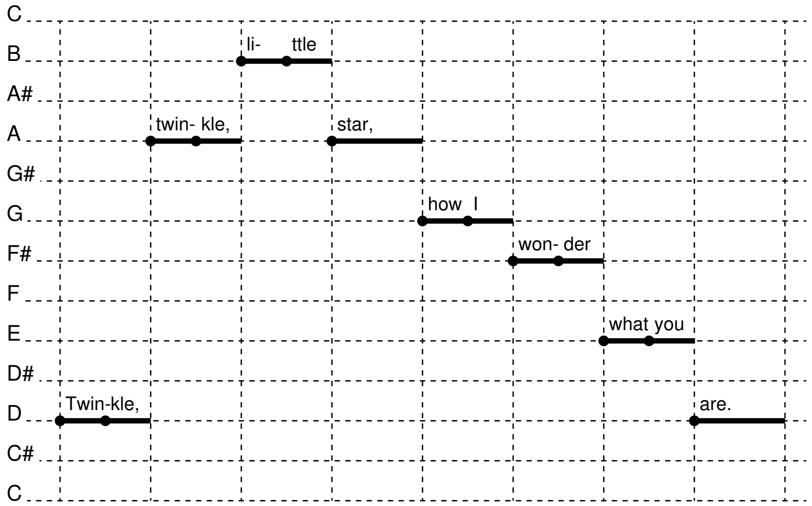

Figure 4.13. The first phrase of a well-known melody, drawn as a graph of log frequency versus time. The vertical lines show the times of the beginnings of the bars. The time signature is 2/4, and the melody is shown here as played in the key of D major.

of an accidental amounts to a temporary change of key. For example, the only difference between the scales of C major and F major is that C major contains B but not B♭, and F major contains B♭ but not B. A tune that starts and ends in C major may have some portions in the middle where B♭ occurs but B doesn't, and this can be interpreted as a temporary change into the key of F. The important point is that the tune has not migrated to the chromatic scale (where all notes are allowed); rather it has shifted from one diatonic scale to another. Such changes of key are called **modulations**. In classical music multiple modulations can occur within longer pieces of music, and one of the historical reasons that the well-tempered scale was adopted over alternatives was to make such changes of key viable without losing the consonant quality of musical intervals.

Usually more notes of the melody occur on strong beats than on weaker beats, and in general if a note occurs on a weaker beat then there will also be at least one note on the stronger beat either immediately preceding or immediately following the weaker beat. If this doesn't happen (so that consecutive notes occur on weaker beats, and no notes occur on the strong beats in between the weaker beats), then you have **syncopation**. When a rhythm is syncopated, the weaker beats are often performed with a degree of emphasis that the omitted stronger beats would have had (if they hadn't been omitted). Syncopation is used heavily in modern popular music.

4.5 Accompaniments

4.5.1 Harmonic Accompaniment

In popular music the most important accompaniment to the melody is usually the **harmonic accompaniment**, which in its most basic form consists of a **chord sequence** or **chord progression**. The durations of chords that accompany a melody are generally longer than the durations of individual notes, and are usually a whole number of bars for each chord. For example, a tune that has 12 bars might have 4 bars of C major, 4 bars of F major, 2 bars of G7 and 2 bars of C major. The **chord change**, where a new chord starts, almost always occurs at the beginning of a bar. However, syncopation of chords is not unknown and, for example, occurs frequently in **salsa** music, which is a strongly syncopated genre of music.

The notes of the chords usually relate to the notes of the melody. In particular, most of the time the notes of the melody that fall on the strongest beat are also notes contained within the chord. For example, the notes in a bar might be **C**, D, **E**, D. The notes C and E have the strongest beats, and so would be expected to occur in the chord for that bar, which might (for instance) be C major. It is much less likely that the chord for such a bar would be (for instance) B major, because B major doesn't have any of the notes C, D or E in it.

It is also very common for the intervals of a chord to appear in the portion of the melody that it accompanies. For each note in a melody we can ask "What is the next note?", and most often it will be one of three possibilities:

1. The same note again.
2. A note above or below that note on the scale.
3. A note separated from that note by a consonant interval, such that both notes occur in the accompanying chord.

The main exception occurring outside these possibilities is when an interval crosses a chord change. In which case there may or may not be a relationship between the interval between the notes and the intervals within or between the old chord and the new chord.

There is a sense in which the melody implies its chords, and the actual chords can be regarded as supplementing implied chords which arise from our perception of the unaccompanied melody. In particular, if you made up a new melody, say by humming it to yourself, without the help of any musical instrument, and you conveyed your new melody to an experienced musician, it is likely that they would be able to easily determine an appropriate harmonic accompaniment for it.

Usually the notes of the chords are notes from the scale that contains the melody, but accidentals do occur in chords, and in fact may occur more often in chords than they do in the melody. However, the more notes a chord has that are not in a scale, the less likely that chord is to appear in a melody on that scale.

4.5.2 Rhythmic Accompaniment

Chords are not the only component of music that accompanies melody in a musical performance. There are also **rhythmic accompaniments**. These accompaniments are usually played on **percussion instruments**, which are instruments that do not produce sounds with well-defined harmonics. Either the sounds are **noise**, which contains a continuous frequency range rather than discrete frequency components; or, if a percussive sound can be analysed into discrete harmonics, the frequencies of the harmonics are not multiples of the fundamental frequency.

Rhythm is also often suggested by the manner in which the chordal accompaniment is played, and by the **bass line**.

4.5.3 Bass

Bass notes are the lowest notes in a tune. In modern popular music there is almost always a well defined **bass line**. The primary purpose of this bass line is to provide the root notes of chords. For example, if a bar starts with the chord C major, the bass line will most likely start with the note C. This bass note seems to reinforce the feeling of the chord. Bass lines can also serve to reinforce the rhythm of the tune, and in some cases the bass line forms a melody of its own.

4.6 Other Aspects of Music

4.6.1 Repetition

Music is often quite repetitive. There are several identifiable kinds of repetition:

- Repetition of rhythmic accompaniment within each bar.
- **Free** repetition of an overall tune, or major components of it.
- **Non-free** repetition of components of a tune within a tune.
- Occurrence of components within a tune which are not identical, but which are identical in some aspects. This is **partial repetition**.

The difference between free and non-free repetition is how many times you are allowed to do it. For example, in the nursery rhyme "Ring a Ring o' Rosies", you can sing any number of verses, so this is free repetition. But within one verse, the melodic phrase for "pocket full of posies" is an exact repetition of "Ring a ring o' rosies" (except, of course, for the words). You have to repeat this phrase exactly twice: doing it just once or doing it three times does not work. The repetition is non-free.

Partial repetition is where phrases are not identical, but some aspects may be identical, for example their rhythm, and/or the up and down pattern that they follow. For example, "Humpty Dumpty sat on the wall" is followed by "Humpty Dumpty had a great fall". There is an exact repetition of melody and rhythm in the "Humpty Dumpty" parts, but after that the melodies of the phrases are different, although the rhythm is still exactly the same.

There isn't much existing musical terminology to describe repetition in music, and the terms "free", "non-free", "exact" and "partial" are ones I have made up.

4.6.2 Songs, Lyrics and Poetry

In most modern popular music, the instrument carrying the melody is the human voice.[12] And the singers don't just use their voice to make the notes: the melody is sung with **lyrics**, which are the words of the song. There is usually some interaction between the emotional effect of the music and the emotional effect of the lyrics. There also generally needs to be some consistency between the rhythm of the melody and the rhythm of the lyrics. Usually one syllable of lyric maps to one note of melody, but syllables are sometimes broken up into multiple notes.

One of the most specific and peculiar features of lyrics is **rhyme**. Rhyme is where the last one or more syllables of the words at the ends of different phrases sound the same. The matching portions of words that rhyme must include an accented syllable. And the match must not just be caused by the words actually *being* the same.

Rhyme is a very persistent feature of song: popular song without rhyme is rarer than popular music that isn't sung. There is some tolerance for **weak rhymes**: these are rhymes that are not exact. In a weak rhyme either the vowels are the same but the consonants are only similar, or vowels are altered to create a rhyme that would not exist using normal spoken pronunciation. But, in general, the vowels have to sound the same; and the more natural the match between vowels is and the more the consonants also sound the same, the better the rhyme is.

Rhyme isn't just found in music—it's also found in **poetry**, or at least in the more traditional kind of poetry that rhymes and scans. **Scanning** refers to poetry having a regular rhythm that is consistent with the lexical accents[13] of the words. The regular rhythm of poetry is another feature that it shares with music. These similarities between poetry and music will lead

[12] Additional voices may also sing some or all of the harmony.

[13] **Lexical** means it is an intrinsic property defined on a per-word basis, i.e. each word knows which syllable or syllables within it are accented most strongly.

us to the suspicion, which will be revisited, that rhyming scanning poetry is actually a weak form of music.

Another musical art form that lies in between song and ordinary speech is **rap**. The main feature of rap is that the music has a spoken component, and this spoken component has rhythm, but it does not have melody. Any melody is carried by other instruments, or by accompanying singers. Rhyme is preferred in rap just as much as it is in song and poetry. The rhythm of rap exists in musical time, as for sung melody, but it is not required to be regular as is the case for poetry.

4.6.3 Dance

There is a strong association between dance and music. People like to dance to music, and people like to watch other people dance to music.

Some features of dance particularly relevant to the analysis of music carried out in this book are the following:

- Movement that is visibly rhythmical.
- Short-term constancy (or smoothness) of perceived speed of motion.
- Synchronised motion of multiple dancers.

The super-stimulus theory has radical implications for the association between dance and music: it suggests that dance is more than something *associated* with music, that dance actually *is* music. This will follow from the general nature of the final theory developed in Chapter 14, where musicality is defined as a secondary feature of many different aspects of speech perception. In particular musicality appears not just in the aspects of speech perception related to the perception of sound—it also appears in the visual aspects of speech perception.