Chapter 9

Symmetries

Symmetry turns out to be a very important concept in the analysis of music perception and its relationship to speech perception. There are five or maybe six identifiable symmetries of music perception. These are invariances under six corresponding types of transformation: pitch translation, octave translation, time translation, time scaling, amplitude scaling and (possibly) pitch reflection. Different symmetries apply to different aspects of music.

Some of the symmetries are **functional**, in that they correspond to required symmetries of perception. The other symmetries are **implementation** symmetries: they reflect the internal mechanics of speech and music perception.

9.1 Definition of Symmetry

As I developed my theory of music, based on the concept of perception of musicality as an aspect of speech perception, I came to realise that there are various symmetries of speech and music perception, and that these symmetries define very strong constraints on any theory that seeks to explain both the mechanics and purpose of music perception.

Symmetry is an everyday concept that we use when talking about shapes and patterns. For example, the human body has an approximate left-right symmetry. We recognise other types of symmetry in shapes such as rectangles, squares and circles, and we recognise repetitive types of symmetry such as found in wallpaper patterns.

Informally we can explain that a shape or pattern is symmetric if the shape or pattern is equal to itself when it is moved in some way. For example, a square is equal to itself if it is rotated 90 degrees. A circle is equal to itself if it is rotated any number of degrees. Both shapes are equal to themselves if they are picked up and turned over (in the case of a square it must be turned over around one of 4 axes that go through the centre, in the case of the circle we can turn it over around any axis that goes through the centre). The wallpaper is equal to itself if it is shifted by the distance between repetitions of the pattern (in the direction that the pattern repeats itself).

We can extend this informal intuition about what a symmetry is to give a more formal mathematical definition of symmetry:

> A **symmetry** is a set of **transformations** applied to a **structure**, such that the transformations preserve the properties of the structure.

Generally it is also presumed that the transformations must be **invertible**, i.e. for each transformation there is another transformation, called its **inverse**, which reverses its effect.[1]

Considering our left-right symmetry example, the transformation that preserves the structure is a reflection in the plane that divides the body down the middle, which swaps left and right. The left-right reflection is its own inverse.

We can formally define the other examples of symmetry already given, in terms of their corresponding sets of transformations:

- A circle has **circular symmetry**. The set of transformations consists of all possible rotations about the centre of the circle and all reflections about lines that go through the centre of the circle.
- The transformations defining the symmetry of a square are: all rotations that are multiples of 90 degrees, and all reflections about diagonals and about lines that join the midpoints of opposite sides.
- Considering an infinitely large wallpaper with a pattern (not itself symmetrical) that repeats every 10cm going up or down and every 10cm going left or right, the set of transformations for the wallpaper's symmetry consists of translations of the form $(n \times 10\text{cm}, m \times 10\text{cm})$ for arbitrary integers n and m.

All of these examples are **geometrical symmetries**. The sets of transformations are subsets of the full set of transformations that defines the symmetry of geometry itself. We can think of the symmetry of geometry as being represented by the transformations that preserve the properties of empty

[1] For most cases that we consider, if a transformation preserves the structure then the transformation has to be invertible. Non-invertible transformations can only preserve structure if there is not enough structure to require the transformation to preserve the distinction between different components of the structure.

space—in particular the property defined by the distance between any two points. If we restrict ourselves to a flat 2-dimensional geometry, this set of transformations consists of:

- All **translations**, i.e. shifting all of space a certain distance in a certain direction.
- All **rotations**, i.e. rotating all of space a certain angle (clockwise or anticlockwise) about a certain point.
- All **reflections**, i.e. reflecting all of space about a certain line.

It is entirely possible to define symmetries that have no direct geometrical interpretation. These are sometimes called **abstract symmetries**. For example, consider the structure consisting of addition on the real numbers.[2] This structure is preserved by any transformation that multiplies all the real numbers by one number c. For example, if we transform the real numbers by multiplying them all by 6.8 (so $c = 6.8$), then the operation of addition is preserved by this transformation, i.e. if $x + y = z$ then $6.8x + 6.8y = 6.8z$. If we extend the structure to include multiplication of numbers, then the symmetry no longer applies, because the structure of multiplication is not preserved by the transformation: it is not necessarily true that $x \times y = z$ implies $6.8x \times 6.8y = 6.8z$.

Symmetry has turned out to be a very powerful concept in mathematics. The **Erlanger Programm** was born out of recognition of the importance of symmetry. The "program" was created by the German mathematician Felix Klein, and emphasised the importance of studying the symmetries of mathematical structures.

9.1.1 Symmetries of Physics

Symmetry also matters in the study of the real world. The study of the most fundamental properties of reality is called **physics**, and it is in physics that symmetry plays the most important role.

The mathematics of physical symmetries is not an easy subject, and the more difficult parts of it do not have any direct bearing on understanding the symmetries of music, but there are enough similarities that it is worthwhile reviewing the role that symmetry plays in physics.

A dynamical physical system can be described by something called a **Lagrangian**. **Noether's theorem** says that for every symmetry of the Lagrangian, there is a corresponding **conservation law**. The symmetries of Lagrangians usually include the underlying symmetries of space and time, and these lead to standard conservation laws as follows:

[2] The **real numbers** are those numbers that can be expressed as either finite or infinite decimals, including both negative and positive numbers (and zero).

- Symmetry under translation in space implies conservation of momentum.
- Symmetry under rotation in space implies conservation of angular momentum.
- Symmetry under translation in time implies conservation of energy.

There aren't any Lagrangians or Noetherian theorems in the theory of music, so I will not attempt to explain these concepts. But there is an illuminating parallel between symmetries as studied in physics and symmetries as we are going to study them in music:

> For every symmetry there is an important set of questions to ask.

In physics the main question is: *What is the conservation law that corresponds to this symmetry?* In studying music the questions derive from biological considerations: *What purpose does the symmetry have?* and *How is the symmetry achieved?*

Even though the analogy between physical symmetry and musical symmetry is fairly abstract, a number of specific concepts that arise when considering physical symmetries also apply to music:

- Symmetries in physics can be **global** or **local**. The examples given so far are all global because they are defined over the full structure being transformed. A **local symmetry** is one consisting of some type of transformation that can be defined pointwise, i.e. there is a transformation that can be specified separately over each location within the structure being transformed.[3] The choice of transformation at each point makes the set of transformations that define a local symmetry a very "large" set. An example of local symmetry does appear in our analysis of musical symmetries.
- Symmetries can be **partial**. This means that a symmetry only applies to part of a system. An example in physics is that of swapping protons and neutrons (these are fundamental particles that make up the nucleus of the atom). Swapping these two preserves aspects of the **strong force**, but does not preserve the **electromagnetic force**. The electromagnetic force is not preserved by the swap because the proton has electric charge, and the neutron doesn't. But in situations where the strong force dominates the evolution of a physical system, the symmetry between neutron and proton can be considered a full symmetry.
- Symmetries can be **approximate**. An approximate symmetry is one where the transformations only approximately preserve the structure

[3] Although the transformation can be different at each point, it is usually required to be a "smooth" function of position.

being transformed.[4] Approximate symmetries of Lagrangians give rise to approximate conservation laws. We will find that all musical symmetries are approximate to some degree. A special type of approximate symmetry is one that is exact for an arbitrarily small transformation. It will be the case for some musical symmetries that the symmetry is close to exact for small transformations but becomes less exact for larger transformations.[5] Unfortunately there does not seem to be any general term for this type of symmetry, so I will coin my own term and call such symmetries **limited symmetries**, to emphasise that the symmetry is exact or close to exact over a limited range.

- There are **broken symmetries** in physics. This has to do with situations where the set of potential evolutionary histories of a dynamical system has a certain symmetry, but any particular history must have less symmetry. The classical example of this is a circular pencil with an infinitely sharp point, balanced upright point downwards on an infinite flat surface. The system has circular symmetry as long as the pencil remains balanced upright. We know that the pencil is going to fall over. When it falls over it has to fall over in some particular direction. And when that happens, the system consisting of the pencil fallen down on the surface no longer has circular symmetry; in fact the symmetry is lost or "broken" the moment that the pencil starts to fall. (We will encounter an example of broken symmetry in music when we look at **pitch reflection invariance**.)

9.2 A Little More Mathematics

Before we consider the symmetries of music, I will define a few more mathematical ideas about symmetry.

9.2.1 Discrete and Continuous

Looking at the examples already given in Section 9.1, we can see variations in the number of transformations in a given symmetry. For example, in the left-right reflection example, there is only one transformation, i.e. reflection about the vertical line going through the centre. Actually every symmetry also includes the **identity transformation**—this is the transformation that does not change the structure—so we can say that the reflection symmetry contains two transformations. The symmetry of a square is defined by a set of eight transformations: four distinct rotations (including a rotation of

[4]There is some overlap between **partial** and **approximate**: a partial symmetry is approximate in those situations where the things it doesn't apply to have a "small" effect on the system or structure that the symmetry applies to.

[5]For this notion to be well defined, the set of transformations has to have some notion of size defined on it.

zero degrees which is the identity) and four distinct reflections. This set of transformations is **discrete**, because for any transformation there is no sequence of transformations distinct from it that get closer and closer to it. The set of transformations is also **finite**, because we can say how many there are, i.e. eight.

The set of transformations of our wallpaper is discrete, but it is not finite, because it includes the transformation $(n \times 10\text{cm}, m \times 10\text{cm})$ for arbitrary integers n and m, and the set of integers is infinite.

The set of transformations for circular symmetry is not discrete; rather it is **continuous**, because we can consider a rotation of x degrees for any real number x. Given any two different rotations, there will always exist another rotation that lies in between those two.

9.2.2 Generators

Our wallpaper symmetry example has an infinite set of transformations, but we can **generate** all these transformations from just two transformations, for example $(10\text{cm}, 0\text{cm})$ and $(0\text{cm}, 10\text{cm})$. These transformations from which all transformations in a set can be generated are called **generators**. We should note that the choice of generators is not necessarily unique, so being a generator is not a specific property of a particular transformation in the set. For the transformations that define the symmetry of a square, we can choose a rotation of 90 degrees and any reflection as a set of two generators for the full set of transformations.

The notion of generator can be extended to continuous symmetries in terms of **infinitesimal generators**. "Infinitesimal" can be understood to mean arbitrarily small. Thus we can define all rotations as being constructed as multiples of some very small rotation. For example, every possible rotation is approximately equal (with an error no greater than 0.0005 degrees) to a multiple of 0.001 degrees.

9.2.3 Stronger and Weaker Symmetries

As the reader may already have noticed, the set of transformations for one symmetry can be a subset of the transformations for another symmetry. We call a symmetry with more transformations in it a **stronger** symmetry, and one with fewer transformations a **weaker** symmetry. For example, the set of eight transformations defining the symmetry of a square is a subset of the transformations that define the symmetry of a circle that has the same centre. The circular symmetry is **stronger** than the square symmetry. The set of transformations for circular symmetry is in turn a subset of the set of transformations for the symmetry of empty 2-dimensional space.

A general rule is that more structure implies weaker symmetry, unless the additional structure is symmetrical with regard to the existing symmetry.[6] For example:

- We start with the symmetry of empty space, which is defined by the set of transformations consisting of all possible translations, rotations and reflections.
- We add a single point to the space. The set of transformations that preserve this new structure is reduced to those transformations that do not change the position of the point: this reduced set consists of rotations about the point and reflections about the lines that go through the point.
- We add a circle whose centre is the same point. As it happens this does not alter the symmetry of the system, because the circle is symmetrical under the same rotational and reflective transformations.
- We add a square whose centre is the same as the centre of the circle. The set of transformations is now reduced to those eight transformations of the discrete square symmetry.
- To the square we add arrows to each side that point clockwise: now the structure has only discrete rotational symmetry, and the set of transformations consists of rotations of 0 degrees, 90 degrees, 180 degrees and 270 degrees clockwise.
- We add one point to the structure distinct from our first point: now the system has no symmetry at all, and the set of transformations consists only of the identity transformation.

9.3 Musical Symmetries

So what are the musical symmetries? There are five symmetries that can be readily identified, and a possible sixth symmetry whose existence is not so obvious. They can be categorised according to the sets of transformations that define them:

- **Pitch Translation:** adding a certain interval to each note.
- **Octave Translation:** adding a multiple of an octave to a note.
- **Time Scaling:** playing music slower or faster.

[6] Actually, components added to an asymmetrical shape can make it *more* symmetrical. We can avoid this difficulty by requiring that added structure always be *labelled*. So if I have a structure consisting of 3 points of a square, which has only reflective symmetry, and add the missing point, but with a unique label, e.g. “A”, then the new structure is a square with one point specially labelled, and it still has only reflective symmetry.

- **Time Translation:** playing music earlier or later.
- **Amplitude Scaling:** playing music more quietly or more loudly.
- **Pitch Reflection:** (this is the possible symmetry) reflecting pitch about a particular pivot note.

For each symmetry we want to answer the following questions:

- What does the symmetry apply to? Some symmetries apply to a piece of music as a whole; others apply to different portions of a piece of music. Some symmetries only preserve some aspects of music.
- Does the symmetry apply to speech? It is a consequence of the musicality perception hypothesis that every symmetry of music perception must be a symmetry of speech perception.
- Does the symmetry serve a functional requirement of perception? Is there a requirement that our perception of speech be invariant under the transformations of the symmetry? Or does the symmetry exist because of internal implementation details of the perceptual process?
- If the symmetry is a functional requirement, how much effort and machinery is devoted to achieving that symmetry?
- If the symmetry exists for implementation reasons, what does it achieve?
- If the symmetry is limited (and they all are to some degree), how limited is it?

We will not be able to answer all of these questions straight away because some of the answers will only become apparent when we investigate the nature of cortical maps that respond to the various observable aspects of music.

9.3.1 Pitch Translation Invariance

Pitch translation is the transformation where a fixed interval is added to all the notes in a piece of music. In musical terminology this corresponds to **transposition** into a different key. However, the translation interval does not have to be an exact number of semitones. The basic observation is that translating a piece of music does not alter the musical quality of that music in any significant way. This **pitch translation invariance** is so strong that we do not normally regard a piece of music transposed into a different key as being a different piece of music.

Furthermore, when we listen to music, we cannot normally tell what key it is in. Some people do have what is known as **absolute pitch**. A person with absolute pitch can identify a note that is played to them without any context, e.g. a single piano note. However, even listeners with absolute pitch do not regard the musical quality of music as being different if it is played in a different key.

Absolute pitch is sufficiently uncommon that we are amazed when someone demonstrates the ability to identify the pitch of musical notes. And yet we are perhaps being amazed by the wrong thing. If we consider the point at which musical sounds enter our nervous system, i.e. the hair cells in the organ of Corti, the set of hair cells stimulated when a tune is played in one key is *completely different* from the set of hair cells stimulated when the tune is played in a different key. Yet by the time this information is processed through all those processing layers in the brain that process music, the resulting processed information is *exactly the same* in both cases.

The sheer perfection of the computational process that achieves pitch translation invariance suggests that there must be some important reason for it. And it suggests there may be a significant amount of brain machinery devoted to achieving it.

The word "translation" in "pitch translation" implies that musical intervals can be combined by addition. However, musical scales are logarithmic in nature, or to put it another way, an interval is actually a *ratio* between frequencies. Pitch "translation" is really a frequency *scaling*, where "scaling" refers to multiplication by a constant. But the notion of intervals being things you add together is so predominant that I will continue to use the term "pitch translation" to refer to the corresponding invariance.

Is pitch translation invariance a functional requirement? The answer comes from considering speech melody. Is our perception of speech melody pitch translation invariant? We know that different speakers speak in different pitch ranges. Pitch translation invariance means that these different speakers can speak the *same* speech melody, by translating the speech melody into a range that is comfortable for them.

There are limits to pitch translation invariance. If music is translated too low or too high, we will not be able to hear it. Even before it gets translated that far, it will start to lose its musical quality. There are limits to the variation in pitch range that occurs in human speakers, and this would explain why pitch translation invariance in perception of speech is limited.

Although variations in speaker pitch range explain the need for pitch translation invariance, they don't explain why it has to be a translation on the log frequency scale. What we deem to be the "same" speech melody depends on the nature of pitch translation invariance. Conceivably, some other form of translation could have been used to define a correspondence between speech melodies in different pitch ranges. For example, addition of frequencies could have been used (instead of multiplication by a scale factor). Part of the answer may have to do with the relationships between speakers with different pitch ranges. Many of the differences between speakers depend on difference in size: a child is smaller than an adult. In as much as the vocal apparatus of a child is a scaled down version of an adult's, the same vocal operations will result in corresponding speech melodies translated in accordance with a scaling of frequency values by the scale factor that exists between the sizes of the child and the adult.

But even this explanation doesn't explain why the translation between different pitch ranges has to be as precisely equal to a frequency scaling as it actually is. We will discover a good explanation for this precision when we consider the **calibration theory** in Chapter 12. We will also discover that many of the mechanisms of pitch translation invariance have to do with consonant ratios—frequency ratios that are simple fractions. These ratios are intrinsically pitch translation invariant, so the significance of consonant ratios explains both how pitch translation invariance is achieved, and also why it exists as a precise frequency scaling.

Pitch translation invariance is a global symmetry: the translation must be applied to a whole piece of music. If we translate only portions of a piece of music, or translate only some notes, we will certainly break the tune (we are assuming translation by an arbitrary interval—we will see that another form of invariance exists if the translation interval is an octave, or a multiple thereof).

Pitch translation invariance necessarily applies to all aspects of music that concern pitch and differences between pitch. These aspects include melody, scales, harmony, chords, bass, home notes and home chords.

9.3.2 Octave Translation Invariance

Octave translation invariance refers to the sameness of the quality of musical notes that differ by one or more octaves. (As is the case for pitch translation invariance, octave translation invariance is really a *scaling* invariance, i.e. the invariance applies when frequency is multiplied or divided by a power of 2, but I will continue to use the terminology of "translation".)

There appear to be two main aspects of octave translation invariance in music:

- Musical scales repeat every octave.
- Notes within chords and in bass accompaniments can be translated up or down by octaves, without significantly altering their musical quality or effect. (This is the example of a *local* symmetry that I mentioned above, because the transformations defining the symmetry are translations which can be applied to individual notes. Compare this to pitch translation invariance, which is a *global* symmetry because the transformations defining the symmetry are translations which must be applied to all of a musical item at once.)

The repetition of scales every octave is not just specific to Western music—it appears in many different musical cultures. This suggests that something quite basic—and hard-wired into the brain—is going on here.

The statement that notes within chords and bass can be translated individually must be subject to a caveat. As stated in the music theory of chords, there are some constraints on where notes in a chord are usually placed, and if you translate the notes by too many octaves then those constraints will be broken. Also if the chords and bass line are constructed in such a way as to create their own separate melodies, then this will constrain the notes not to be moved from their location within those melodies.

These issues should not cause us to disregard the local nature of octave invariance as applied to chords. If a tune is played simply, as just melody and chords, then much of the musical quality of the tune is revealed by this simple mode of performance. When music is played this way it *is* possible to shift individual bass notes, individual chords and individual notes within chords up or down by an octave, without having any significant affect on the quality of the music.

One aspect of music that is definitely *not* octave translation invariant in a local sense is the identity of individual notes of the melody. When a major component of a melody is freely repeated, such as a verse or a chorus, we may be able to translate an occurrence of such a component by an octave, without breaking the musical effect. But if we translate individual notes up and down by octaves, we will certainly ruin the melody. As I have already noted, the most common note that follows any particular note in a melody is either the same note, the note above it or the note below it. Adding random multiples of an octave to notes breaks this pattern.

Octave translation invariance does not seem to serve any functional requirement. There are no significant octave relationships within individual speech melodies, nor between speech melodies of different speakers. When we investigate cortical maps that represent and process information about musical pitch and musical intervals, we will find that octave translation invariance enables the efficient implementation of calculations relating to the pitch translation invariant characteristics of music and speech. In particular it facilitates the efficient implementation of "subtraction tables" that calculate the interval between two pitch values by subtracting the first value from the second value.

9.3.3 Octave Translation and Pitch Translation

Octave translation invariance and pitch translation invariance are the only example (from the set of musical symmetries discussed here) of a pair of weaker and stronger symmetries, i.e. pitch translation invariance is stronger than octave translation invariance, and the set of transformations representing octave translation is a subset of the set of transformations for pitch translation. Pitch translation invariance means being able to add any interval to musical notes; octave translation invariance means being able to add any interval that is a multiple of one octave.

This strong/weak relationship is relevant to understanding the relationship between these two symmetries in the roles they play in music and speech perception. Pitch translation invariance is a functional requirement and octave translation invariance is an implementation requirement. Any computation that starts with absolute pitch values and results in a pitch translation invariant output will still be pitch translation invariant if the input values are first reduced to a value modulo octaves. We will find that every aspect of music/speech perception that is octave translation invariant will be a component of pitch perception that is pitch translation invariant.

9.3.4 Time Scaling Invariance

If a piece of music is played faster or slower, then we can recognise it as being the same piece of music. The quality of the music is not preserved quite as strongly as in the case of pitch translation invariance; indeed most tunes have a preferred tempo that maximises the effect of the music, and the music is correspondingly weakened if we play it at a different tempo. But the fact that we can recognise music independently of its tempo suggests that there is some aspect of the perception of music that is preserved under time scaling. As is the case with pitch translation invariance, achieving this invariance is more non-trivial than we realise.

When we look at how tempo is represented in cortical maps, we will see that neurons stimulated by a rhythm played at a fast tempo are completely different to those neurons stimulated by the same rhythm played more slowly. To achieve time scaling invariance, the brain has to perform a calculation such that its final result is a pattern of neural activity in a neural map which is the *same* for either the slower version or the faster version of the same rhythm.

There is a plausible functional purpose for time scaling invariance: some people talk faster than other people, and the same person can talk at different speeds on different occasions. There are perfectly good reasons for wanting to talk at different speeds: sometimes it matters more to say what you have to say as quickly as possible, other times it matters more to speak slowly so that your audience can easily understand what you are saying. In as much as the rhythms of the language being spoken assist in the comprehension of language, it is important that the same rhythms can be recognised at different tempos.

9.3.5 Time Translation Invariance

Of all the symmetries listed, this is the one that seems most trivial. If I play a tune now, the musical quality is the same as if I play it in 5 minutes time. It probably comes closer than any other symmetry to being exact. Even if I play a tune in 5 years time, the musical quality will not be much different. If I wait long enough, like 200 years, then my existing audience will all be dead, and therefore unable to perceive the musical quality of music. So time translation invariance does have some limits.

Time translation invariance satisfies an obvious functional requirement: if I use a certain speech melody now, or in 5 minutes time, it should be perceived identically by my listeners.

Even though time translation invariance seems trivial, it is not necessarily completely trivial to implement. If we tried to design a computer system to perceive patterns of sound occurring in time, very likely we would use an externally defined timing framework (i.e. a *clock*) to record the times at which events occurred. In order to recognise the repeated occurrence of the same pattern, we would have to find some way to realign our frame of reference relative to the sets of observed events themselves.

One simple way of doing this is to define the first note of the melody as being at time zero. But this will not produce an entirely satisfactory result. If I add just one extra note to the beginning of the melody, all my notes will have their times offset by the duration of that extra note. If we try to compare notes in these different occurrences of a melody by comparing their values and their times, then the melody with the extra note will be completely different to the original melody, because all the notes will be labelled by different times.

This does not correspond to our own experience of melody recognition: we do not have any difficulty recognising a melody that has had an extra note added to the beginning.

A better theory of how time translation invariance is achieved is given in Chapter 13, which is about repetition. The basic relationship between repetition and time translation is that a repetition of a component of music corresponds to a translation of the first occurrence of the component to the time (or times) at which the component occurs again.

9.3.6 Amplitude Scaling Invariance

This invariance seems almost as trivial as time translation invariance.[7] If we turn the volume up or down, it is still the same music. There are limits, but these correspond to obvious extremes. If we turn the volume down too far then we can't hear anything; if we turn it up too much then the perceived sound becomes distorted (and eventually physical damage occurs to the sensory cells in the ear).

Amplitude scaling invariance satisfies an obvious functional requirement in speech perception: some people talk more loudly than others. Also if someone is farther away, then they are going to sound quieter, but it's still the same speech.

[7]It is not, however, necessarily trivial to implement. Within the organ of Corti, a louder sound activates a larger population of hair cells. Non-trivial computation is therefore required (within the auditory cortex) to recognise similarity between louder and softer versions of the same sound.

Amplitude scaling does not affect our perception of the *quality* of music, but it does affect our enjoyment of music. If we like a particular item of music, then we like it turned up louder, and we experience that the emotional effect is more intense if it is played more loudly. A significant portion of the money spent by consumers (and performers) on music is spent on equipment whose sole purpose is to make the music *louder*. One of the consequences of the correlation between loudness and musical enjoyment is that deafness caused by music being too loud is the major health risk associated with listening to music.

9.3.7 Pitch Reflection Invariance

This is the most obscure musical symmetry; in fact I am not completely certain that it exists. But its existence is plausible. The diatonic scale has a reflective symmetry. If we consider the white notes, it can be reflected in the note D. This symmetry can also be seen in the Harmonic Heptagon. It is not the symmetry of this scale that makes the case for pitch reflection invariance; rather it is that, given the symmetry of the scale, a certain property of the scale is also invariant under reflection. This property is the **home chord**. In fact the home chord of tunes played in the white notes scale is always either C major or A minor.[8]

C major and A minor are reflections of each other around the point of symmetry D. When we look at home chords in detail, we will consider what forces exist (in our perception of music) that cause the home chord to be one or the other of these two chords. It seems at least possible that these forces involve interactions between notes separated by consonant intervals, and that the force from note X to note Y is the same as the force from note Y to note X. It is this symmetry of forces between notes that gives rise to the symmetry between C major and A minor as home chords, given the symmetry of the scale itself.

If pitch reflection invariance is a genuine musical symmetry, then it comes into the category of implementation symmetries. The home chord does not appear to directly represent any information about speech as such; rather it is the result of a process that attempts to define a frame of reference for categorising the notes (or frequencies) in a melody in a way that is pitch translation invariant. It is also a broken symmetry (in the same sense this term is used in physics), because one of the two possible home chords must be chosen for each particular tune.

[8] Musical theory does allow for other home chords, but if you survey modern popular music it's almost always one of these two chords.

9.4 Invariant Characterisations

A major question to be asked about functional symmetries in perception is: how is each invariant perception represented in the brain?

It is useful to consider the same problem stated for a simple mathematical example. We can compare idealised solutions to this type of problem to the more pragmatic solutions that the brain might actually use.

Consider the set of finite sequences of numbers. An example might be (3,5,6,−7,4.5). We will allow our numbers to be negative or positive (or zero) and to have fractional components. A natural definition of equality for this set is to define two sequences to be equal if they have the same number of elements and if the corresponding elements in each position are equal. So (4,5,6) is equal to (4,5,6), but it is not equal to (4,5) because (4,5) has only two elements, and it is not equal to (4,7,6), because the elements in at least one position (i.e. the second) are not equal. And it is not equal to (4,6,5), because order matters.

Next we want to define a symmetry represented by a set of transformations. The set of transformations consists of all those transformations that add a constant value c to each element of a sequence, where c is any number.

Then we define what is called a **quotient set**. Elements of the original set are elements of the quotient set, but they have a different rule of equality: elements of the quotient set X and Y are considered equal if there exists a transformation (from the set of transformations defining the symmetry) that transforms X into Y, i.e. if there exists some number c which can be added to each element of X to give Y. This relationship between elements that we want to consider equal is called an **equivalence relation**[9] on the original set.

To give an example, consider two sequences X and Y, where X =(4,5,−6) and Y =(5.5,6.5,−4.5). X is equivalent to Y because we can transform X into Y by adding 1.5 to each element of X. But X =(4,5,−6) is not equivalent to Z =(5,6,−9), because there is no number that we can add to all the elements of X to get Z: we would have to add 1 to the first two elements but −3 to the last element.

This mathematical model could be considered a simplified model of pitch translation invariance as it applies to music—considering notes of a melody to be a simple sequence of values, and ignoring considerations of tempo and rhythm. The numbers in the sequence correspond to pitch values (e.g. as positions in a semitone point space), and the equality of a sequence to a translation of that sequence by a fixed value corresponds to the musical identity of a tune to a version of itself transposed into a different key.

The important question is: how can we represent distinct members of the quotient set? If we represent (4,5,−6) as (4,5,−6) and (5.5,6.5,−4.5) as (5.5,

[9] In general an equivalence relationship must have the following properties: $x = x$ (**reflexive property**), $x = y$ implies $y = x$ (**symmetric property**), $x = y$ and $y = z$ implies $x = z$ (**transitive property**).

6.5,−4.5), then we are using *different* representations for what are meant to be the *same* elements.

This is not necessarily a bad thing, but it complicates the definition of operations on the quotient set. If a formula specifies a calculation applied to a non-unique representation, then we have to verify that the calculation applied to different representatives of the same element always gives the same result.

Mathematicians have struggled with the issue of how to define unique representations for elements of a quotient set. One neat but somewhat tricky approach is to identify each element of the quotient set with the **equivalence class** of members of the original set that correspond to it. So the representative of the element (4,5,−6) is the set of elements of the set of sequences which are equivalent to (4,5,−6). This is a neat trick, because the equivalence class of (4,5,−6) contains the same members as the equivalence class of (5.5,6.5,−4.5), and sets are equal if they have the same members. But there is no easy way to write this equivalence class down, because it contains an infinite number of elements. (We could just write "the equivalence class of (4,5,−6)" and "the equivalence class of (5.5,6.5,−4.5)", but this gets us back to where we started, because we will have different ways of writing the same equivalence class.)

There is another way out of this quandary. It doesn't work for all examples of quotient sets, but it works fine for the one we are considering. What we need to do is find a well-defined procedure for choosing a **canonical representative** for each equivalence class of the original set. This canonical representative will be the representative of each equivalence class. As long as the members of the original set can be written down somehow, then the canonical representatives can be written down, and we have unique written representatives for elements of our quotient set.

In the example we are considering, there are various rules we could use to choose the canonical representative for each equivalence class. One is to choose the representative whose first element is 0. So the canonical representative of (4,5,−6) would be (0,1,−10). Another possibility is to choose the representative such that the total of the elements in the sequence is 0. The representative of (4,5,−6) would then be (3,4,−7).

Canonical representatives are not the only means of defining unique representatives for equivalence classes. In our current example, we could represent each sequence by the sequence of *differences* between consecutive elements in the sequence. So the representation of (4,5,−6) would be (1,−11). The sequence of differences always has one less element than the original sequence, so it is not a member of the equivalence class. The sequence of differences represents the equivalence class because:

- if we add a constant value to the sequence then this doesn't affect the differences, and,

- if two sequences have the same sequence of differences, then there must be a constant difference between all their corresponding elements.

There is a certain aesthetic to this representation: it doesn't involve trying to choose a special representative of the equivalence class; instead we just perform some operation whose result is unaffected by the transformations defining the equivalence relationship. And the operation preserves enough information about its input value to retain the distinction between different equivalence classes. Further on we discuss what happens if not enough information is preserved by the operation that generates the representation—in that case we are left with an **incomplete representation**, i.e. one that gives the same representative for all members of an equivalence class, but which does not always distinguish different equivalence classes.

9.4.1 Application to Biology

This theory of equivalence relationships and quotient sets is a gross simplification of the concept of symmetries in biological perception. Elements of a quotient set are either equal to each other or they are not. But biological perception also has requirements of *similarity*. To give a basic example, we don't consider two melodies completely different if they only vary by one note.

This requirement affects what constitutes a plausible theory about the representations of perceptions invariant under some symmetry. In particular, if two objects are perceived as similar, then the internal representations of the perceptions of those objects should be correspondingly similar.

We can attempt to apply this to our current example of sequences as a model of pitch translation invariant perception of melody.

We could consider the first example of a representation of a melody that is invariant under a constant pitch translation, where we choose a representative such that the first element (i.e. note) of the sequence is 0. For example, the sequences (2,1,4,3,1,1,1,2,3) and (3,2,5,4,2,2,2,3,4) belong to the equivalence class whose canonical representative is (0,−1,2,1,−1,−1,−1,0,1). What happens if we add an extra note to the beginning of the sequence? If the extra note is not the same as the original first note, then all corresponding elements of the canonical representative will be different. For example, we might add 1 to the start of (2,1,4,3,1,1,1,2,3) and 2 to the start of (3,2,5,4,2,2,2,3,4) to give (1,2,1,4,3,1,1,1,2,3) and (2,3,2,5,4,2,2,2,3,4), and the canonical representative becomes (0,1,0,3,2,0,0,0,1,2). The similarity between the representatives (0,−1,2,1,−1,−1,−1,0,1) and (0,1,0,3,2,0,0,0,1,2) is obscured by the fact that the elements of the latter have been translated 1 up from their values in the former.

But our common experience is that we can easily recognise the similarity between a melody and the same melody with an extra initial note added to the start.

If we consider the representation where the total value of notes in the sequence is zero, which is the same as saying that the average value is zero, then adding an extra note will change all the components of the representative, but only by a small amount. For example, the sequence (2,1,4,3,1,1,1,2, 3) has an average value of 2, so its representative is (0,−1,2,1,−1,−1,−1,0,1). If we add 1 to the start, i.e. (1,2,1,4,3,1,1,1,2,3), the average value is now 1.9, so the representative becomes (−0.9,0.1,−0.9,2.1,1.1,−0.9,−0.9,−0.9,0.1,1.1). The old representative and the "tail" of the new representative (i.e. all those elements after the additional first element) are different, but only by 0.1.

The representation as a sequence of differences seems more promising: if we add an extra note to the start of a tune, the representative will be changed by the addition of one difference to the beginning. For example, the representative of (2,1,4,3,1,1,1,2,3) and (3,2,5,4,2,2,2,3,4) is (−1,3,−1,−2,0,0, 1,1). Adding 1 and 2 (respectively) to the start of these sequences results in a representative (1,−1,3,−1,−2,0,0,1,1). The old representative and the tail of the new representative are now identical.

We can characterise these different representations in terms of how they are affected by a change at a certain point in a sequence being represented, and in particular by how soon a representation "forgets" the effect of a change that occurs at the beginning of the input sequence. The representation with the first element set to zero never forgets the effect of a change at the start of a sequence. The representation with the average (or total) set to zero also never forgets, but the effect of the change is diluted in proportion to the overall length of the sequence. The differences representation, on the other hand, forgets the effect of a change almost straight away.

There are, however, other types of change to a sequence that affect the differences representation in ways that are inconsistent with our experience of how we perceive melodies and changes to them:

- If one note in the middle of the tune is changed, two consecutive differences will change in the representative. For example, changing (2,1,4, 3,1,1,1,2,3) to (2,1,4,3,2,1,1,2,3) changes the representative from (−1,3, −1,−2,0,0,1,1) to (−1,3,−1,−1,−1,0,1,1).

- If all the notes past a certain point are increased by a constant value (like a sudden change of key), one difference will change. Changing (2,1, 4,3,1,1,1,2,3) to (2,1,4,3,2,2,2,3,4) changes the representative from (−1, 3,−1,−2,0,0,1,1) to (−1,3,−1,−1,0,0,0,1,1).

For each of these two changes, the corresponding change to the differences representative consists of a change to a small number of values in one part of the differences sequence, but subjectively we experience these changes differently: a changed note is a changed note, but a change in key feels like a permanent change to the state of the melody, and the effect of this change in state is felt until our perception of the melody "settles in" to the new key.

So even the sequence-of-differences representation does not fully match up with our subjective experience of pitch translation invariant perception of melody. In some sense it is *too* forgetful, and the other representations we considered are not forgetful enough. But consideration of these possibilities has given us a flavour of what we want to look for when considering how the brain represents musical information.

9.4.2 Frames of Reference

Attempting to find a canonical representative for a tune can be described as trying to find a "frame of reference", to use a term from physics. A physicist analysing the motion of an object moving at a constant velocity through space will try to choose a frame of reference that simplifies the analysis, for example one where the object is not moving at all.

Setting the first note to zero and setting the average note value to zero can be seen as simple strategies for finding this frame of reference. Now the musical aspects of scales and home chords do seem to act like a frame of reference. Given that all the notes from the melody are from a certain scale, and given that the scale has an uneven structure (although repeated every octave), it is possible to identify certain notes in the scale as being "special", and use those notes to define the frame of reference for choosing a canonical representative of the melody. For example, for a tune played entirely on a diatonic scale, we would transpose it until it was in the key of C major, and that would be our canonical representative.

There is only one thing seriously wrong with this theory of scales and home notes as choosing a frame of reference: there are no identifiable scales or home chords in speech melody, and the super-stimulus theory implies that the biological purpose of pitch translation invariant melodic representations is to represent speech melody, not musical melody. So, for example, it would not make any sense to specify that the canonical representative of a speech melody could be determined by transposing it into the key of C major.

Having said that, we will find that the purpose of those cortical maps that respond to scales and home chords is to provide representations of melody that are pitch translation invariant, and these maps provide invariant representations for both speech melody and musical melody.

9.4.3 Complete and Incomplete Representations

If we asked a mathematician to find a representation for members of a quotient set, they would look for a representation that loses the distinction between members of an equivalence class, but which does not lose any other distinctions, i.e. between members of different equivalence classes.

But biological pragmatism often finds solutions to problems that may not offer mathematical perfection. Losing exactly the right amount of distinction is like sitting exactly on a very fine line. Perhaps in real life we sit on one side or the other of the line, or even straddle it in some way.

We do know that the pitch translation invariance of melody perception is close to perfect for moderate translations, so we can be sure that the distinction between elements within an equivalence class is definitely lost. But it is possible that the brain uses representations that are **incomplete**, in the sense that they do not completely represent *all* the information about an equivalence class of melodies (because they have lost *more* information than would be lost by a complete representation).

One consequence of incompleteness is that there would exist distinct melodies having the same representation, and therefore perceived as being the same melody. This might seem strange, given that many people can easily learn to reproduce a melody fairly exactly (if they can sing in tune), but we must remember that a musical melody is a discrete thing, compared to a speech melody which is continuous, and the set of discrete musical melodies is a much smaller set than the set of melodies in general. So we may be able to reliably distinguish different musical melodies, but not necessarily be able to distinguish different non-musical melodies, even where there might be a significant difference that we would spot if we happened to view a visual representation of the melody as a function of frequency against time. It is also possible that the brain uses multiple incomplete representations, which, when taken together, form a fairly complete representation of a member of the quotient set. More research may need to be done on the brain's ability to identify and distinguish non-musical melodies.

Recall the representation of a sequence of notes as a sequence of differences (i.e. intervals). This can be defined in terms of the operation of calculating the differences between consecutive note values. There are many other calculations that we could define to act on the original note information. Some of these will produce results that are pitch translation invariant, and some of those pitch translation invariant representations will be incomplete. Some examples of incomplete pitch translation invariant representations include the following:

- Whether or not each note is greater than, less than or equal to the previous note. So, for example, (2,1,4,3,1,1,1,2,3) and (3,2,5,4,2,2,2,3, 4) would be represented by $(<,>,<,<,=,=,>,>)$. The representation is incomplete in the sense that there are sequences which have the same representative but are not equivalent to each other. For instance (2,0,5, 2,0,0,0,3,4) has the same representative as the first two sequences, even though it is not equivalent to them.

- For each note, the number of steps since the last note (if any) that was harmonically related to the current note (but not the same). For example, supposing that intervals of 3 or 4 semitones are harmonic and intervals of 1 or 2 semitones are not harmonic, the representative of

(2,1,4,3,1,1,1,2,3) and (3,2,5,4,2,2,2,3,4) is (?,?,1,?,2,3,4,?,?), where "?" means no other note harmonically related to that note has previously occurred.

- For each note, the number of times the same note has occurred previously. The representative of (2,1,4,3,1,1,1,2,3) and (3,2,5,4,2,2,2,3,4) would be (0,0,0,0,1,2,3,1,1).
- For each note, the number of times a note harmonically related to that note has occurred previously. Again, for the same example, the representative would be (0,0,1,0,1,1,1,0,0).

It is not too hard to see that each of these functions defines a result that is pitch translation invariant. None of these four functions is a complete representation, because for each function there are distinct melodies not related to each other by pitch translation for which the function gives the same result.

We could define a number of incomplete representations invariant under some symmetry, and then combine them into a complete (or almost complete) representation. But why bother? Why not just calculate a single simple complete representation?

A possible answer to this question has to do with the biological pragmatism that I mentioned earlier. It may not matter that a representation is perfectly complete. And the cost to calculate a perfectly complete representation may be exorbitant. We must also consider the constraints of evolution: a simple calculation of a complete representation may be feasible, but there may be no way that it could have evolved from less complete (and less invariant) representations.[10]

I will adopt the terminology of **invariant characterisations** of musical structures, preferring "characterisation" over "representation", because "characterisation" is a term that emphasises both the biological purpose of such representations and their potential incompleteness.

[10]This is similar to the explanation of why there are no wheels in nature: there is no way that something which is not a wheel could have evolved continuously and gradually into something that is a wheel, while all the time being useful to its owner.