WHAT IS MUSIC?

Solving a Scientific Mystery

The science of music started more than 2000 years ago, when Pythagoras made his observations about consonant intervals and ratios of string lengths.

But despite all the advances made in acoustics, psychology, neuroscience and evolutionary biology, scientists still have no idea *what music is*.

The theory in this book is the result of more than 20 years of research by the author. It explains in detail many of the familiar features of music: notes, scales, melody, harmony, chords, home chords, bass, rhythm and repetition.

It also explains the symmetries of music. These symmetries include invariances under pitch translation, octave translation, time translation, time scaling, amplitude scaling and pitch reflection.

Most importantly, the theory explains the emotional effects of music, and this explanation sits firmly within the framework of modern evolutionary theory. For the benefit of those not fully familiar with the concepts of theoretical biology, what this means is that the theory explains how our ability to respond to music helps us *have more grandchildren*.

Published by Philip Dorrell, 2005.

Revision Date: 22 March 2005

ISBN 1-4116-2117-4

The official website for this book is **http://whatismusic.info/**.
The author's personal website is **http://www.1729.com/**, and current contact details may be found at **http://www.1729.com/email.html**.

WHAT IS MUSIC?

Solving a Scientific Mystery

by Philip Dorrell

Dedicated to
Amanda and Natalie.

Contents

Acknowledgements

I would like to thank my wife Marcelina and my children Amanda and Natalie, for putting up with my efforts to write this book.

Thanks to my sister Jean who edited the book, and then, after I had done a substantial rewrite, edited the book a second time.

Also thanks to my Mum who read the book and made some useful suggestions, to Sean Broadley who made a remark about the musical quality of purely rhythmical music, and to Vasil Dimitrievski who told me about Macedonian dance music.

Any errors of style, grammar or content remain my own responsibility.

Chapter 1

Introduction

1.1 An Autobiographical History

1.1.1 The Facts of Life

In 1982 I was in the last year of a three year Bachelor of Science degree at the University of Waikato, New Zealand. I had lost interest in doing further study, but I did not really know what I wanted to do with my life. My degree was originally going to be a double major, but I had dropped out of physics, which left just mathematics as my major subject.

One of life's big problems, and one that (in 1982) I had no idea how to solve, is that of finding a satisfying career that enables one to be productive and happy—or at least not too unhappy—and pay the bills. And, if you can't solve that problem, then there is always Plan B, which is the *get-rich-quick scheme*.

Unfortunately, most get-rich-quick schemes don't work. Otherwise we'd all be rich, which, obviously, we aren't. To solve my career problem I needed more than just any old get-rich-quick scheme—I needed one that was truly original, and obviously different from all those schemes that didn't work. I had to find a way to exploit my own unique talents and knowledge.

As I was a nineteen year old university student about to graduate from my first degree, and I'd never held down a proper full-time job, I was somewhat lacking the experience of the "real" world that might be required to successfully operate a get-rich-quick scheme.

On the bright side, there were a certain number of things that I felt I knew and understood, which were not known or understood very well by most other people. I knew these things mostly because I had spent my childhood reading books about mathematics and science.

The "facts of life" that I had gleaned from studying mathematics and science were as follows:

- The universe operates according to laws which are very mathematical. We don't know what these laws actually are, but the laws that we currently use to describe the universe appear to be good limiting approximations to the actual laws that the universe operates under. For most purposes the difference between these approximations and the actual (but unknown) laws doesn't matter too much.

- Most people don't realise the full consequences of this, because they don't understand mathematics.

- Living organisms are part of the universe.

- Human beings are living organisms.

- The human mind is part of the human body.

- Therefore the human mind operates according to these same exact mathematical laws.

I discovered that most people believed that their own human nature was *not* the result of the operations of mathematical laws. The reasons they had for this belief might be that they felt they were too special to be subject to scientific laws (mathematical or otherwise), or they believed that they had a soul created by God (a soul almost by definition defies scientific explanation). To me, it seemed these people were paying too much attention to common sense and intuition, and not enough to our scientific understanding of the universe.

1.1.2 The Mathematics of the Universe

The mathematical nature of the universe was revealed to me (before I went to university) when I read books about the strange worlds of special relativity and general relativity.

Special relativity is something that contradicts common sense, but can be understood mathematically. I had read books that tried to explain special relativity in terms of people travelling on trains and signalling to each other with torches, but these books failed to make me feel that I understood what it was all about. Then I read *Electromagnetic Fields and Waves* by Lorrain and Corson (WH Freeman and Co, 1970), which had a section about special relativity. It described special relativity as the invariance of physical laws under the Lorentz transformation, and my eyes were opened. "Common sense" was replaced by abstract mathematical understanding.

I went on to read about general relativity. The first thing I learned was that books on general relativity explain special relativity better than books on special relativity. Or rather they simplify the mathematics, perhaps at the expense of divorcing the explanation even further from the common-sense

world view. Time becomes almost[1] just another dimension in a 4-dimensional space-time geometry.

I also learned that the theory of general relativity was the result of intelligent guesswork by Albert Einstein. He made certain assumptions about the comprehensibility of the universe, and then persisted with those assumptions for years, before finally discovering a satisfactory theory. At the time he formulated the theory (it was announced in a series of lectures he gave in 1915), there was only one piece of hard evidence in favour of it: an anomaly in the orbital precession of Mercury. The next item of evidence came in 1919, from measurements made during a solar eclipse of the deviation of starlight caused by the Sun's gravity, but these measurements were not so accurate as to confirm the theory very strongly, although they did have the effect of making Einstein instantly famous. Given this paucity of evidence, and the degree of speculation and mathematical intuition apparently involved in Einstein's attempts to find the best possible theory of gravity, it is amazing that the theory has since been confirmed by a range of different experiments and observations, and is now generally accepted by the scientific community as a correct description of both gravity and the large-scale structure of space and time in the universe.

I never persisted sufficiently to learn all the mathematics and theory of general relativity, but I understood enough to realise that here was a theory based on mathematics, which could only be developed by someone who knew the theory of special relativity, which itself could only be properly understood from a mathematical point of view. It followed that if you attempted to understand the universe, but you did not believe that the universe operated according to exact mathematical laws, then you were going to get hopelessly lost.

Later on, at university, I formally studied mathematics and science, which had the unfortunate effect of putting me off reading books on those subjects, so I expanded my horizons and read books about economics and psychology.

One thing I learned from studying economics was the connection between what people want and what you can do to get rich: you can get rich if you can find a new way to give people what they want and charge them for it.

1.2 The Science and Mathematics of Music

Towards the end of 1982, I devised a promising get-rich-quick scheme: compose and sell music. I wanted a way to make money with a minimum amount of effort. Songwriters sometimes make large sums of money from their compositions. The basic informational content of some of these compositions could

[1] "Almost", because the geometry is defined by a diagonal 4×4 tensor, where the time entry in this diagonal is -1 and the entries for the three spatial dimensions are each $+1$. This is the *only* difference between time and space in relativity (special or general).

easily be written on one page of notepaper—so it seemed like you didn't have to do too much work to compose one yourself.

My first attempt to compose music consisted of simply sitting down at a piano and trying to make something up. Unfortunately, I discovered, as many others have before and since, that it is very difficult to conceive new music that is any good. If you play something that sounds good, it always turns out to be part of something you already know.

But even if I lacked an innate talent for composition, I knew that there was a possibility of understanding music from a rational point of view. The mathematical simplicity of music implied that there might be some simple underlying mathematical theory that described what music was. If I could discover this theory, then I could use it to compose new music, and make my fortune.

The major constraint on any theory of music comes from biology and, in particular, from Charles Darwin's theory of **evolution by natural selection**. I knew that Darwin's theory was the explanation for the existence and origin of all living organisms, including myself and other human beings.

So the plan of action was straightforward:

- Analyse the mathematical structure of music as much as possible.
- From the mathematical structure of music, formulate mathematical theories about music.
- If that doesn't work, then take a biological approach, and develop theories about how music could arise from adaptive functionality in the human brain.
- Test predictions made by the theories.
- Try using the theories to compose new music (which is actually a special sort of prediction—you are predicting that the music you compose is going to be good).

1.3 A First Breakthrough: 2D/3D

Fast forward a few years, and I had what I thought was an exciting breakthrough. I analysed musical intervals as elements in a vector space, and discovered the 1D, 2D and 3D representations, as described in Chapter 5. This analysis showed why the **syntonic comma**[2] would always appear in any attempt to make a diatonic scale have only perfect consonant intervals between notes in the scale.

I discovered the natural mapping from the 3D representation to the 2D representation, which is analogous in an interesting way to the mapping from

[2]The **syntonic comma** is a ratio of 81/80, and gets discussed in full detail in Chapter 5.

3-dimensional space to a 2-dimensional visual image (e.g. on the retina of the eye). I knew that, by one means or another, the brain had the ability to process the visual mapping in both directions, i.e. going from 2D to 3D and from 3D to 2D.

Even better, I realised that a "non-loop" (or spiral) in musical 3D space maps onto a "loop" in musical 2D space, and these loops can plausibly be identified with simple chord sequences found in much popular music.

At the time it seemed that I had found the solution to the problem. But my attempts to flesh out all the details and develop a complete theory never progressed much further. I analysed many songs, attempting to assign 2D and 3D representations to the intervals that occurred in each song, but I was not able to find any rule for assignment that made the occurrence of a spiral-to-loop mapping depend on the musicality of the tune.

I also failed to complete the 2D/3D theory in a biological sense: even if we believe that neurons processing vision are somehow involved in processing music, why should the emotional and pleasurable effects of music occur? According to the 2D/3D theory, the looping logic of music is equivalent to the paradoxical logic of drawings by M.C. Escher, such as *Belvedere* (1958), *Ascending and Descending* (1960) and *Waterfall* (1961), where the paradox always depends on the fact that one position in a 2-dimensional drawing corresponds to an infinite number of positions in the 3-dimensional space represented by the drawing. Escher's drawings are interesting to look at, but they do not cause emotion and pleasure in the way that music does.

1.4 A Second Breakthrough: Super-Stimulus

Over a decade later, while idly thinking about the music problem, a simple idea occurred to me: many of the features of music are also features of *speech*, except that the corresponding musical features are regularised and discretised compared to those of speech. Perhaps the response to music is just a side-effect of the response to speech, and music is somehow contrived to maximise this response. To use a technical term, perhaps music is a **super-stimulus**.

From that one thought came all the rest of the theory outlined in this book. I do not (yet) have hard proof that the super-stimulus theory is correct, but it explains more things, and explains them better, than the 2D/3D theory did. I like to think it explains more things about music and explains them better than any other theory of music that has been published to date. The super-stimulus theory even provides a plausible explanation for its own incompleteness: that the principle of super-stimulus applies to some or all of the cortical maps that process speech, and not all of the relevant cortical maps have been properly identified and understood. The way that the theory works, a full explanation of all the causes of the musicality of a tune is only achieved when one understands the representation of meaning in *all* the relevant speech-related cortical maps in the listener's brain.

1.5 The Rest of This Book

1.5.1 Background Concepts

Chapter 2 lays down the problem. The main concepts required are that music is a biological problem—because people are living organisms—and that all biological problems must be solved within the framework of Darwin's theory of evolution by natural selection.

Chapter 3 reviews the assumptions that underlie most of the existing theories in the music science field. I give some references to specific papers and articles, and also summarise the different approaches used by music researchers in their attempts to solve the fundamental problem of what music is.

Chapter 4 reviews the basic theories of sound, hearing and music—as much as is needed for understanding the theory presented in this book. The required theory on sound and hearing is simple: sound consists of vibrations travelling through a medium, regular vibrations have a fundamental frequency, and arbitrary waveforms can be decomposed into sums of "pure" sine-wave tones, where the frequencies of the sine-wave tones are integral multiples of the fundamental frequency.

If you have learned to play a musical instrument, you will probably already know most of the required music theory.

Chapter 5 outlines very basic vector mathematics, which helps us to understand the relationships between consonant intervals on the well-tempered diatonic scale.

Section 5.3 introduces the **Harmonic Heptagon**. This diagram is useful when explaining the theory of home chords.

Chapter 6 gives some basic theory of how the brain works. This includes the brain and nervous system as an information processing system; what **neurons** are and how they are connected to each other; and the concepts of **cortical maps**, **binding** and **population encoding**.

Chapter 7 describes my older **2D/3D theory**, which relates 2D/3D relationships in music to 2D/3D relationships in visual processing. It may still have some relevance to a complete theory of music.

1.5.2 The Super-Stimulus Theory

Chapter 8 introduces the **super-stimulus theory**: that **musicality** is a perceived attribute of speech, and music is a **super-stimulus** for musicality. The difference between a super-stimulus and a normal stimulus is important to consider when analysing aspects of music. In particular, super-stimuli can have attributes that are never found in the corresponding normal stimuli.

One musical aspect that demonstrates this difference is **harmony**. Harmony is the simultaneous occurrence of multiple pitch values, but a listener to speech never attempts to listen to multiple speakers at the same time. The

normal stimulus corresponding to musical harmony turns out to be something somewhat different, and relates to the perception of consonant relationships between pitch values occurring at *different* times. The **harmonic cortical map** has the job of perceiving these relationships. It happens to operate in such a way that it can also perceive the same relationships between different pitch values occurring simultaneously, and in fact it responds more strongly to simultaneous pitch values.

Other attributes of music not found in speech are regularities of time and discontinuities of pitch. We must deduce that regular musical rhythms and discontinuous musical melodies are super-stimuli for parts of the brain that are designed to process *irregular* speech rhythms and *continuous* speech melodies.

Chapter 9 takes a slight diversion and considers the **symmetries** of music perception. These consist of transformations of musical data under which certain aspects of the perception of music are invariant. Six symmetries are identified: **pitch translation invariance**, **octave translation invariance**, **time scaling invariance**, **time translation invariance**, **amplitude scaling invariance** and **pitch reflection invariance**. All of these symmetries (except perhaps pitch reflection invariance) correspond to familiar features of music perception, but they are not normally understood as "symmetries". Considering them as symmetries forces us to ask particular questions, such as why do they exist, and how are they implemented? In particular, pitch translation invariance and time scaling invariance are non-trivial symmetries for the brain to implement, and therefore must serve some significant purpose.

The chapter on symmetries also compares musical symmetries to symmetries as studied in fundamental physics. The analogies between physical symmetries and musical symmetries presented in this book are strictly at an abstract level, mostly along the lines of "symmetries are more important than anyone originally realised in physics" and "symmetries are more important than anyone originally realised in the study of music". (So, for example, I do *not* attempt to apply Noether's theorem[3] to musical symmetries.)

Chapter 10 considers specific **cortical maps**—areas in the brain with specialised functionality—whose existence is implied by the various observed aspects of music. This consideration is guided by the concept of music being a super-stimulus, and the corollary that aspects of music are super-stimuli for specific aspects of speech perception. We will learn that each of these cortical maps processes a particular aspect of speech perception and a corresponding aspect of music perception.

Chapter 11 devotes itself to one particular symmetry—that of octave translation invariance. This invariance corresponds to the observation that notes separated by multiples of an octave have a similar subjective quality.

[3]Noether's theorem says that to every symmetry in a physical system there corresponds a conservation law. It is the most important theorem about symmetry in mathematical physics.

Existing terminology is that such notes are in the same **pitch class**. We find that octave translation invariance is not a required invariance of perception. Rather, it contributes to the efficiency of information processing related to pitch differences and, in particular, the implementation of compact "subtraction tables" required to calculate and compare the sizes of intervals between notes.

Chapter 12 discusses **calibration**. Pitch translation invariance—our ability to recognise the same melody played in different keys—implies an ability to perceive a 4-way relationship between pairs of notes separated by equal intervals. The question arises: *how is the perception of this relationship accurately calibrated?* Genetic predetermination seems implausible as an explanation, in which case there must be an explicit process of calibrating against some external standard, and this external standard turns out to be the intervals that exist between harmonic components of human voice sounds. The concept of calibration generalises to other aspects of music perception which are invariant under some symmetry—the time scaling invariance of rhythm perception being the other major example.

Chapter 13 is on the subject of repetition. Repetition is a feature of music not found in normal speech. We can distinguish between **free repetition**, where something is repeated an arbitrary number of times, and **non-free repetition**, where a phrase is repeated an exact number of times. How the brain models repetition is closely related to how it models sequential information (such as the sequence of notes in a melody).

Much can be deduced (or at least guessed) about music assuming only that there is such a thing as musicality, and that music is a super-stimulus for it. But eventually we have to develop a specific hypothesis about what musicality is: what it means, and how the brain perceives it. This happens in Chapter 14, where the hypothesis is developed that musicality corresponds to **constant activity patterns** (**CAP**) in cortical maps involved in speech perception. Perception of constant activity patterns in the listener's brain represents an attempt to detect corresponding patterns of activity in the brain of the speaker, and detection of constant activity patterns in the speaker's brain in turn indicates something important about the speaker's mental state. The final result of the perception of constant activity patterns is a validation of the listener's emotional response to the content of what the speaker is saying.

1.5.3 Questions, Review and the Future

Chapter 15 lists outstanding questions, and includes some suggestions for future research based on the assumptions and hypotheses of the theory developed in this book.

Chapter 16 is a summing up. It reviews the assumptions of the super-stimulus/CAP theory: which assumptions stand alone, and which depend on other assumptions.

Finally, Chapter 17 takes a look at the future—in particular a future where music is composed by an algorithm based on a proper theoretical understanding of what music is. There will be more and better music than ever before, most of it generated by music software running on home computers. There may even be too much good music, and some people ("music junkies") will give up work, play and everything else, and spend their whole life just listening to computer generated music.

Chapter 2

What is Music?

The problem with answering the question "What is music?" is understanding what would constitute a proper answer. Music arises from human behaviour, and the study of human behaviour is part of biology. So any question about music is a question about biology, and every question about biology requires an answer within the framework of Darwin's theory of evolution by natural selection.

2.1 Music is Something We Like

What is music? It's what comes out of the speakers when we play a CD on our stereo. It's what we hear on the radio. Music is singers singing and musicians playing. Music is a sound that we enjoy hearing.

Is this a proper answer to the question "What is music?"?

If I asked "What is a car?", you could answer by pointing at a large object moving up the street and saying "It's one of those." But this may not be a satisfactory answer. A full explanation of what a car is would mention petrol, internal combustion engines, brakes, suspension, transmission and other mechanical things that make a car go. And we don't just want to know what a car *is*; we also want to know what a car is *for*. An explanation of what a car is for would include the facts that there are people and other things (like shopping) inside cars and that the purpose of cars is to move people and things from one place to another.

By analogy, a good answer to the question "What is music?" will say something about the detailed mechanics of music: instruments, notes, scales, rhythm, tempo, chords, harmony, bass and melody. This matches up with the mechanical portion of our car explanation. It's harder to answer the

"What is it for?" part of the question. A simple answer is that music is enjoyable—it makes us "feel good". We could expand on this a bit and say that music creates emotions, or interacts with the emotions we already feel and, sometimes, it makes us want to dance.

2.2 The Biology of Feeling Good

The "feel good" explanation is worth something, but it isn't entirely satisfactory. Or, at least, it's not satisfactory if you're a professional theoretical biologist.

What does music have to do with biology? Music is something that people create and something that people respond to. People are living organisms, and biology is the study of living organisms.

We can compare music to eating. Eating is a well-known activity. People do it. Animals do it. We know what eating is: it is the ingestion of certain substances into our digestive systems. The ingested substances, or **food**, travel through the digestive system, where components of those substances are broken down and extracted by various means for use within the body. Leftover portions of the food get pushed out the other end.

We can explain eating at a psychological level: we eat when we feel hungry because it makes us feel good. Being "hungry" can be defined as a feeling of wanting to eat food. We can determine that we become hungry when we haven't eaten for a while,[1] and that we stay hungry (and slowly get hungrier) until we have eaten.

2.2.1 Having More Grandchildren

A professional biologist would explain the existence of hunger by saying that it is **adaptive** or, equivalently, that it is an **adaptation**.

A biologist calls something an adaptation if it contributes to having *more grandchildren*. Becoming hungry when we need to eat and eating when we are hungry contribute to having more grandchildren in the following ways:

- As children we need to eat food to grow up into adults.
- We need to eat to have the strength and energy to survive, to secure a mate, to do the mating itself, and then do all the work that comes afterwards, i.e. raise the children. In particular, we need to raise our children well enough that they can grow up and have children themselves.
- When a woman is pregnant, and also when she is breast feeding, she needs to "eat for two".

[1] There are other factors that influence hunger, such as whether it's the time of day at which we normally eat.

- We shouldn't eat when we already have enough food in us, because:
  - too much food at once will overload our digestive system,
  - once we have enough food in us, there are other more important things we should be doing instead of eating more food.

I refer to the need to contribute to having more *grandchildren*, rather than just children, to emphasise the importance of the continued cycle of birth, growth, development and reproduction. If something causes us to have more children, but has a negative effect on the ability of our children to raise their own children, to such an extent that it causes us to have fewer grandchildren, then that something is not an adaptation.

Strictly speaking, biologists think in terms of *long-term reproductive success*, i.e. having great-grandchildren and great-great-grandchildren, and so on forever. But, for our purposes, "grandchildren" is a close enough approximation. By the time most people get to having grandchildren, they no longer have the major responsibility to raise them, so whatever enabled their reproductive success to get that far will probably continue indefinitely anyway.

What made biologists think that everything had to be explained in terms of having more grandchildren? Most people would concede that if some species of organism does not have grandchildren, then pretty soon it is not going to exist at all. But does that mean that *every* purposeful behaviour of a living organism has to be explained in terms of long-term reproductive success?

2.2.2 Charles Darwin and His Theory

The most important discovery in the history of biology was Charles Darwin's theory of **evolution by natural selection**.

Even today, when his theory underpins all of modern biology, there are many people who refuse to believe that his theory is correct, or even that it could be correct. More than a hundred and forty years after Charles Darwin published his discovery, there is a whole industry of authors and pseudo-scientists "proving" that evolution does not occur, or that if it does occur then it is not occurring by natural selection.

This book is not aiming to change the minds of people who are skeptical about evolution. This is a science book, and it is based on a scientific point of view that the universe we live in appears to be *comprehensible* in the way that Albert Einstein remarked upon, and that furthermore it is reasonable to proceed on the basis that those bits of the universe that we do not yet comprehend will eventually turn out to be comprehensible.

The specific field of study concerned with understanding human behaviour according to Darwin's theory of evolution by natural selection is **evolutionary psychology**. The basic assumption of evolutionary psychology is that

our behaviour is determined in some manner and to some degree by our **genes**.

Genes are the information about how our bodies develop and operate. They are contained in molecules called **DNA**, which can be understood as long strings of text written in a language with a 4-letter molecular "alphabet". If you read molecular biology papers in scientific journals, you will see descriptions of genes written as strings containing the letters A, G, T and C. These are the first letters of the chemical names for the four molecular "letters" in the molecular alphabet: **adenine**, **guanine**, **thymine** and **cytosine**.

AGTTTCTAGGTCGTGAAACTGTTCAGGCTTAAGTTGCGGTA

Figure 2.1. A stretch of (single-stranded) DNA shown as a sequence of A, G, T and C.

For humans the strings of DNA are divided up into 23 pairs of **chromosomes**. Each chromosome is an unbroken stretch of DNA, usually tied up in complex spiral patterns (to keep it safe and out of harm's way when it is not being used). Every cell in your body has these 23 pairs of chromosomes, except for a few types of cell that don't need to reproduce themselves. (Also there are the **gametes** which are the intermediate stage between parent and child, and which have only one of each pair of chromosomes.) The chromosomes in each pair are similar to each other,[2] and we get one of each type of chromosome from each parent (via their gametes). For each pair of chromosomes, each of our parents supplies one chromosome from their own pair of chromosomes, or a mixture of both chromosomes in that pair. Darwin didn't know about DNA, and he didn't understand the mechanics of genetic shuffling and mixing that occurs when we have sex.[3]

When we reproduce, the central thing that reproduces is our DNA. For us, as multi-cellular organisms, this happens when we reproduce to create new organisms (i.e. babies), and also when the cells that make up our own bodies reproduce in order to make our tissues grow. Most of the time the DNA reproduces accurately, but bits of it can get changed or **mutated**. And when these mutations occur, they will on average be preserved, and the next time the DNA reproduces, the parts of the gene that were changed are no

[2]Exception: females have two X chromosomes, but males have one X chromosome and one Y chromosome per cell. Furthermore, one of the female X chromosomes is always rendered inactive within the cell.

[3]Gregor Mendel was the one who first learned about the genetics of sex. The science of genetics as we know it today began when Mendel did his experiments on sweet peas. Darwin's theory of genetics involved a theory of "blending", which didn't work very well. Unfortunately Mendel's work did not become widely known until some time after Darwin's death.

more likely to change the next time than any other part of the gene that was not changed.[4]

What happens to us if our DNA mutates? A lot of the time the answer is nothing, because much of the information in our DNA has little effect on how well our bodies work. In fact the notion of "gene" specifically refers to a portion of DNA which does affect some particular part of how our body develops or operates. Mostly this happens when a gene encodes the makeup of a particular type of molecule called a **protein**. There are many types of proteins that do many different things in our bodies. If DNA in one of your genes changes, then the protein encoded by the gene will change, and this could affect how the protein does whatever it does in your body. Ultimately, the changed protein could change your long-term reproductive success.[5] It might make it better, or it might make it worse (which is actually far more likely). If it makes it better, then you are going to have more grandchildren and great-grandchildren and so on. If it makes it worse, then you are going to have fewer grandchildren and great-grandchildren and so on than everyone else.

An important part of Darwin's theory is the idea that for every species there is some limit as to how many individuals of that species can ever exist at one time. Among other considerations, all life that we know of exists on planet Earth, and the Earth is finite in size. In practice, most species hit some limit long before they get to the point where their members occupy every square and cubic inch of the planet. As the more successful genetic variations form a constantly increasing proportion of the total population, the less successful genetic variations must eventually disappear altogether. When this happens, the species itself has undergone a permanent change. The removal of less successful variations is the **natural selection** and the resulting permanent change is the **evolution**.

Darwin realised that if the process of evolution went on for long enough, species could change into new species that were as different from their ancestors as different species are from each other. And if species sometimes split into separate populations, and those populations happened to evolve in different directions, then one species would turn into two or more species. Taking this idea to its logical conclusion, Darwin supposed that all life on Earth could have evolved from a single ancestral species:

> Therefore I should infer from analogy that probably all the organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed.[6]

[4]This is probably not 100% true, as some locations in the chromosome may be more susceptible to processes that cause mutation. It is more precise to state that the probability of mutation at any given location on the chromosome can be a function of location, but does not depend on whether the location in question has or has not recently suffered a mutation.

[5]A mutation will affect your descendants if it occurs in a **germ cell**, which is a cell from which the gametes (sperms or eggs) are descended.

The modern technical term for this hypothetical "one primordial form" is the **Universal Common Ancestor** (**UCA**).

Evolution by natural selection explains the characteristics of living organisms. Each living organism is the result of a long sequence of individual minor changes, and each minor change became fixed in the population because it resulted in increased reproductive success. There are a few caveats to this reasoning:

- Some changes may have resulted from genetic changes that had only a very marginal effect on reproductive success. There is a certain probability that some changes will become permanent even though they have no effect or even a slightly negative effect on reproductive success. This can happen particularly if a species is occasionally reduced to a very small population, or if a new species evolves from a very small sub-population of its ancestor species.[7]

- In some cases an observable aspect of a species' behaviour will be attributable to the effects of one or more evolved changes that occurred in the past, but this aspect may not currently contribute to reproductive success, even though the corresponding evolutionary changes did contribute to reproductive success at the time they occurred.

2.3 Explaining Purposeful Behaviour

Whether or not a particular aspect of human behaviour requires to be explained within the evolutionary framework is easier to decide if we restrict ourselves to consideration of **purposeful** behaviour.

Purpose can be defined as a type of reverse causality. Causality is something that flows forward in time. What *was* explains what *is*, and what *is* explains what *will be*. With explanations involving purpose it's the other way around: what *is* explains what *was*, and what *will be* explains what *is*.

A normal causal explanation might be applied to a soccer player kicking a ball that goes into goal: the ball with mass m was travelling at velocity v_1, when it made contact with the player's foot (via his boot) at position p_1, which caused it to change velocity to v_2, after which, according to the laws of physics, it travelled in a path that caused it to go into the goal. In the causal explanation, where and how the player kicked the ball determined the ball's path, which in turn determined the ball's final destination inside the goal.

In the purposeful or **teleological** explanation, the ball going into the goal explains the way that the player kicked the ball. That is, the result is treated as the explanation of the events that caused that result. "The player kicked

[6] *The Origin of Species* Charles Darwin 1859

[7] Motoo Kimura developed the **neutral theory of molecular evolution** which emphasises the importance of random (non-selective) processes in evolution.

the ball *so that* it would go into the goal." If the ball had initially been in a different location and travelling in a different direction, the player would have kicked it differently, but *he still would have kicked it in a way that would have caused it to go into the goal.*

Of course players don't always get the ball into goal, even if they try ("try" is a word whose meaning implicitly assumes purpose), but we still accept the explanation that goes backwards in time: the player kicked the ball the way he did because he was trying to get it into goal (and it nearly went in).

This distinction between causal explanations and teleological explanations goes all the way back to Aristotle: he used the term **efficient cause** to describe normal forward causality, and **final cause** to describe reverse teleological causality.[8]

Modern science only admits efficient causes. A very simple way of justifying this is to say that science only allows one explanation for any particular aspect of reality that requires explanation. If we have two explanations of the same phenomenon, either one explanation is not correct, or one of the explanations is redundant and could have been restated in terms of the other.

In the case of the soccer player kicking the ball into goal, we accept the correctness of both explanations: the ball went into the goal because of the way it was kicked, and the ball was kicked the way it was so that it could go into the goal. But these dual explanations only apply to purposeful phenomena. For all other phenomena only the efficient cause type of explanation ever applies. So we may assume that efficient causes are the more basic type of explanation, and we must look for a way to restate the final cause explanation in terms of efficient causes.

At which point we can directly apply Darwin's theory of evolution by natural selection. It is the cycle of reproduction and selection which converts efficient causes into final causes. Various soccer players try to kick the ball into the goal. The ones that get it in are seen as better players. The girls fall in love with the good soccer players, and they have lots of children. The children inherit the genes from their dads who were good soccer players, and some of these genes determine the behaviour that caused their dads to kick the ball into the goal. Maybe the genes give their owners stronger legs, or better coordination, or create a propensity to practice more, or give them a tendency to party less the night before an important match. Whatever the case, in the next generation of soccer players there is a higher proportion of those genes which make the players better at kicking balls into the goal.

This explanation does seem a little trite. The genes that contribute to players being able to kick accurately may be genes that have quite general effects, like being able to focus on achieving a result, or being able to develop coordinated action. The ancestors of a good soccer player may never actually have played soccer (or at least not professionally). They might have been

[8] Aristotle listed two other types of cause: **material** and **formal**, but we would tend to include them as parts of efficient and final causes respectively.

cricket players instead. Or perhaps the skills evolved to help them run away from lions and throw spears at edible prey animals.[9]

But the general idea holds good: natural selection converts a final cause explanation into an efficient cause explanation, protecting and preserving the unity of all scientific explanations.

It also means we can stop feeling guilty about using teleological explanations, *as long as they fit into the theory of evolution by natural selection.*[10]

Final causes can be chained together just like efficient causes. For example, a chain of efficient causes is: I was able to have many grandchildren because the girls liked me because I got rich because I kicked the ball into the goal because I had practiced a lot because I always arrived at practice on time. The corresponding chain of final causes is: I always arrived at soccer practice on time so that I could consistently kick the ball into the goal so that I could get rich from being paid well, so that all the girls would love me and I could choose the best one to marry so that I could have many grandchildren.

We can use Darwin's theory of evolution by natural selection to convert a final cause explanation into an efficient cause explanation, as long as the very last final cause in the chain of final causes is *lots of grandchildren.* If we end up with a final cause of something else, then our teleological explanation is not consistent with our otherwise consistent explanation of reality based on efficient causes.

2.3.1 Incorrect or Apparently Incorrect Sub-Goals

Where does music fit in to this theory of purpose and causality? Certainly we can identify purposeful causality in behaviours relating to music. "I worked at the shop so that I could save up money so that I could buy a fuzz box so that I could plug it into my guitar so that I could play 'Smoke on the Water'." But the chain of final causes seems to stop when we get to the music itself.

Many of the unsolved problems of evolutionary science involve the existence of final causes that appear not to have any explanation in terms of more grandchildren: the chain comes to a stop in a bad place. Any number of human behaviours seem to go directly against what is required for maximising long-term reproductive success, behaviours such as driving too fast,

[9]This is a reference to the **environment of evolutionary adaptedness** (**EEA**): the time when we lived in the jungle in hunter/gatherer tribes. The presumption is that not much evolution has happened between that time and the present day, so any evolutionary explanations must relate to those earlier circumstances as opposed to modern living conditions with cars, roads, supermarkets etc. The EEA (as an explanation for modern human behaviour) is discussed in more detail in Chapter 3.

[10]This is not a complete explanation of the existence of purpose in human (or animal) behaviour: in addition to natural selection, there are selective processes operating *within* the brain, which act to select those behaviours and behavioural strategies that (on average) help us to satisfy our biological goals. The physiological mechanisms that underlie these processes are themselves the result of evolution by natural selection, so there exists a two-level hierarchy of purposeful causality: natural selection has given rise to a purposeful system of internal selection which acts to select purposeful behaviours.

sky-diving, being generous, fighting for your country, eating too much fat (or just eating too much), eating sticky sweets that make your teeth go rotten, and drinking too much alcohol.

How can we explain the existence of these apparently **non-adaptive** purposeful behaviours? Plausible types of explanation include the following:

- The reproductive benefit is there, but just not so obvious to the untrained observer.
- The purposeful behaviour results from some more general purpose which benefits reproductive success on average.
- The behaviour used to benefit reproductive success, but times have changed and now it doesn't.

(The third explanation can be a special case of the second one: the behaviour used to benefit reproductive success, now it doesn't; in the future it may become beneficial again.) Another possible explanation is that the alleged behaviour isn't quite what it seems: for example, maybe generosity isn't quite as common as it appears to be, because people are always doing things to make themselves appear more generous than they really are.

Trying to explain non-adaptive purposes and purposeful behaviours is an ongoing activity in the world of evolutionary psychology, and some of the explanations that have been thought of are more convincing than others.

Here is a sample list of evolutionary explanations for some of the apparently non-adaptive human behaviours given above:

- Wanting to drive too fast used not to be non-adaptive, because there weren't any cars. The instincts that make drivers want to drive too fast had general benefits, encouraging our ancestors to learn how to move quickly and efficiently without crashing into anything.
- There weren't any opportunities to sky-dive in the distant past, on account of the non-existence of parachutes—so a desire to sky-dive would not have been non-adaptive.
- Dying for your tribe or country seems extremely non-adaptive, since dead people can't have children. But if society rewards warriors who risk their lives for the sake of the tribe, then it can be argued that the benefits going to those who risk their lives and survive more than make up for the losses suffered by those who risk their lives and get killed.
- Eating a lot of fat can be beneficial if there is a substantial risk of famine. The extra nutrients stored in the body of a fat person will help them to survive the hard times.

- In the past, most available sweet foods would have been either ripe fruit or honey. These are not quite as bad for your teeth as the boiled sweets and toffees that are available in large quantities in the modern supermarket. A desire to eat anything sweet is of particular advantage to children, as they need the extra energy to play, and play is important because it helps children develop their thinking and general life skills.

- Why people like to drink alcohol requires a different sort of explanation. Alcohol and other recreational drugs, legal or illegal, act directly on those parts of the brain that tell us if we have or have not achieved our goals. The most that evolutionary theory can tell us about drugs is that if a drug was widely available in the distant past, then humans should have evolved some resistance to that drug.

2.4 Proof of our Ignorance About Music

This issue of explaining non-adaptive purposes will come up when we investigate music. With music there is, however, a further complication: *we don't even know what music is.* Music is therefore a double mystery: we don't know the "what" and we don't know the "why". Maybe if we could solve the "what" that would help us answer the "why", or maybe if we could guess what the "why" is we could find out the "what".

There are a number of different ways I have found of demonstrating our ignorance of what music is, and each provides a useful insight into the nature of the problem:

- **Subjective and Objective.** The difference between knowing what something is subjectively and knowing what it is objectively.

- **The Martian Scientist.** Could we explain to a Martian scientist what music is?

- **The Incompleteness of Music Theory.** Here "music theory" refers to the kind of music theory that you learn when you learn to play music, and which will be presented in a basic form in Chapter 4. This music theory tells us something about the structure of music, but beyond a certain point it gives up.

- **Lack of Formula.** Despite common claims that some types of music are "written to a formula", there is no such formula, or if there is one, no one is telling us what it is.

- **The Economics of Music.** Those who compose good music get paid well, because making up good music is a hard problem. The very difficulty of the problem results from our ignorance about what music is.

2.4.1 Subjective and Objective

We know what we know about things in the world around us because information comes into our senses, and we process the information in our nervous systems and brains to create knowledge about those things. Sometimes we can convert this knowledge into symbolic natural language, i.e. by speaking or writing. Sometimes other people can relate our symbolic descriptions of things to their own experiences of the same things (or similar things).

If I see a sparrow, I can describe my observations of that sparrow to you. You can relate that description to memories of sparrows you have seen. If by some chance you have never seen a sparrow, I would first have to explain what a sparrow was, and you would have to relate that to your experience of seeing other types of bird. If you have never even seen a bird, then it becomes more difficult, and I would have to think more carefully about how to describe what a bird is to someone who has never seen one.

If I feel a pain in my leg, I can describe it to you, and you can relate that description to your own experiences of having pain in your legs. But we cannot feel the same pain. I cannot feel the pain in your leg, and you cannot feel the pain in my leg. It is almost impossible for one person to know exactly what pain another person is feeling. In fact we can argue that questions like "Is my pain the same as your pain?" are ultimately meaningless, as there is no meaningful way to make such comparisons.

This problem seems related to questions like "Is my feeling of seeing red the same as your feeling of seeing red?". However, the colour of objects is something that can be specified in terms of physical theories about reflection and absorption of light. We know that human colour perception depends on reception of light by three specific types of colour receptor in the eye. In as much as two people have exactly the same colour receptors (which is mostly the case), there is some sense in which it can be said that they see the same red if they look at the same object under the same lighting conditions. Of course the internal processing of colour perceptions will still be different, because it is very unlikely that two people's brains are wired in exactly the same way.

If we doubt that I am seeing the same red as you are seeing, we can use a spectrograph to measure, for each frequency, the intensity of light falling onto the red surface and the intensity of light reflected off the surface. Then, for each frequency, the ratio between the intensity of light reflected off the surface and the intensity of light falling onto the surface gives us the absolute reflectance of the surface at that frequency. The values of all the ratios for all the frequencies of light define the colour of the surface. We can display these ratios as a function of frequency in a graph, or reduce them to a table of numbers. There is no real possibility of us disagreeing about what the numbers are. We can wonder if my experience of the number 3.567 is different from your experience of the number 3.567, but most of us are prepared to regard the meaning of "3.567" as completely independent of the person who

is reading the number.

This independence of observer is what we call **objective**. The opposite of **objective** is **subjective**. The meaning of the number 3.567 is objective. The pain in my leg is subjective.

Somewhere in between objective and subjective is **inter-subjective**. An inter-subjective perception is subjective, but we have some degree of confidence that my experience of it will be the same or at least similar to your experience of it. Most subjective phenomena are inter-subjective to *some* extent, in the sense that there is probably some person somewhere feeling something similar to what you are feeling now, and that person would understand what you were talking about if you described your feelings to them. Even pain is inter-subjective in this sense. Also it could be claimed that the difference between the objectivity of "seeing red" and the subjectivity of feeling pain is not so much that it is impossible to objectively describe what pain means, but just that our current understanding of the human mind and visual perception allows us to be more specific about what "seeing red" means.

2.4.2 The Martian Scientist

In Oliver Sacks' book *An Anthropologist on Mars: Seven Paradoxical Tales* (Vintage, 1996), the "Martian" is Temple Grandin, a well-known autistic, who has difficulty understanding the emotions and intentions of other people, and who has described herself (as quoted on p. 248 in Sacks' book) as feeling like "an anthropologist from Mars".

In general, the concept of the Martian Scientist is a good metaphor for the idea that there are things about ourselves that we are very familiar with, but which might be difficult to explain to an alien from outer space.

There is a presumption in this metaphor that there are at least some things that we could explain to an alien scientist. For example, it is presumed that it would not be too hard to introduce an alien scientist to our mathematical notations, so that we could talk about "3.567", and the alien scientist would know exactly what we were talking about. Similarly we would be confident that we could explain what a spectrograph was, and even explain the characteristics of colour receptors in the human eye, so that our alien friend could understand what we meant when we talked to him about the colour "red".

The concept of the Martian Scientist arises in discussions about consciousness. We all know subjectively what consciousness is, but as yet no one is able to explain what it is in an objective scientific sense. Could we explain consciousness to a Martian scientist? The problem is that a Martian scientist is quite likely to be conscious in exactly the same way that we are. Maybe it is not possible to be intelligent in a way that allows understanding and discussion of scientific concepts, unless one is conscious. So when we talk about consciousness with our friend from Mars, he could indicate that he knows what we are talking about. And yet we cannot say that this proves that either of us (human or Martian) has an objective understanding of what

consciousness is, because we may be doing nothing more than sharing our common *subjective* experiences of consciousness with each other.

Music is a bit different in this regard. Our ability to respond to music does not appear to play any essential role in our ability to comprehend the universe. Our perception of music depends in obvious ways on our systems for perceiving and processing sound. But being deaf does not in the least imply a lack of intelligence: quite plausibly our Martian scientist could be deaf. (Maybe the air on Mars is too thin for hearing to be of much use.) A deaf Martian scientist would not have any subjective understanding of what music is. This gives us a straightforward way to ask if we can find an objective description of music: could we explain what music was to a deaf non-musical Martian scientist?

Some people would explain music in terms of what they know about music, saying music is a sequence of sounds according to certain rules, which happens to have certain emotional effects on people. Given this explanation, and given an item of supposed music, the Martian could check if the supposed music satisfied the specified rules, and then check that it also had an effect on human listeners. But what we really want to know is whether the Martian scientist could learn to identify music, and in particular good music, when given only the music itself. In other words, could the Martian scientist *predict* the effect that an item of supposed music would have on human listeners? To use a term that I am going to use a lot throughout this book, would the Martian scientist be able to calculate the **musicality** of music?

2.4.3 The Incompleteness of Music Theory

It seems reasonable to assume that we could discuss mathematics with intelligent aliens. So if we could produce a description of music that was mathematical, then we could easily communicate that description to an alien scientist.

Much of music theory is mathematical. We will see details of this when basic music theory is introduced in Chapter 4. Notes have frequencies. Intervals between notes can be described as vectors and as certain fractional ratios between their frequencies. Notes and percussive sounds occur at certain times according to regular tempos. The relationships between fundamental and harmonic frequencies can be explained in terms of **Fourier analysis**, which is an important and non-trivial area of mathematics.

With all this existing mathematical music theory, we might wonder what the problem is. Can't we just tell our alien audience the mathematics of music theory, and then they will have an objective understanding of what music is? There are two main reasons why this might not be the case:

- Firstly, a mathematical description of music does not necessarily tell the aliens anything about what is going on inside the human brain when we listen to music.

- Secondly, our mathematical theory of music is not *complete*. Although music theory says quite a lot about the mathematical structure of music, it does not say enough to distinguish between really good music and mediocre music. Music theory fails to *predict* the musicality of supposed music.

These two problems are complementary: if we knew exactly what was going on inside the human brain when we listened to music, then this information could be translated into a procedure for calculating the musicality of music. The procedure for calculating musicality would be a simulation of the operation of those parts of the brain that play a role in perceiving music.

On the other hand, it may be possible to develop a complete mathematical description of music *without* developing any understanding about what happens inside the brain when we listen to music. But as you will see when you progress through this book, intelligent guesswork about what is happening inside the brain is the easiest way to make sense of the mathematical structure of music.

The incompleteness of music theory was my major motivation for performing the research which culminated in the development of the theories explained in this book.

Books that discuss music theory tend to skate around the issue of incompleteness. One good question to ask yourself, when reading a book (or paper) that discusses explanations of music, is what, if anything, the book says about why some music is better than other music. If an author ignores or denies the existence of musicality as something that a musical item can have more or less of, this makes it is easier for them to avoid confronting the question of what it is that determines musicality, and they can comfort themselves with discussions of "music", completely ignoring any comparison that can or should be made between "good" music and other music which is still recognisable as music, but not quite so good.

Even when a book does arrive at this issue, the author will admit (sometimes very implicitly), that they do not know what causes the difference between the good and the not so good, or they may just state categorically that this difference cannot be explained by "rules" (generally ignoring the possibility that they are talking about known rules, and that there might be other unknown rules that do explain the difference).

To approach a problem scientifically, we must not be afraid to confront our own ignorance. The more clearly we can state what we think we know, and what it is that we don't know, the more chance we have of finding some way to move forward. A precise statement of our ignorance about something can be an important first step in the development of a new theory, or in the design of an experiment likely to advance our understanding of the problem.

2.4.4 Musical Formulae

When people talk about music "written to a formula", they use this phrase in a derogatory sense, implying that some hack churns out musical items which are all very similar and just good enough to be marketable. The sophisticated listener is bored by this formulaic music, and hungers for musical creativity that comes from an inspired genius whose output could never be captured by anything as mundane as a formula.

No one ever says what the formula is. Or if they do, the formula suffers from the same incompleteness as music theory in general: the formula describes some aspect of the music, but it is not complete enough to generate the same creative output as the output of the person whose output the formula supposedly describes.

Now it is possible that someone somewhere *is* using a formula to generate music, and they are keeping it a secret. If you had a formula to generate music, you might want to keep it a secret too. You could use your formula to compose music which you could sell, but if everyone knew the formula then it would be too easy for anyone to make up good music, and the bottom would drop out of the market.

The type of formula I have just been talking about is a formula to generate music. In the world of mathematical computer science, they would call it an **algorithm** (rather than a "formula"). An algorithm is something that can be written down as a program written in some programming language, and executed on a computer. So we are talking about a computer program that can compose music, and not just any old music, but music that is as good as, or even better than, the best music as currently composed by professional composers and songwriters.

There is another type of algorithm which is relevant to the analysis of music, and that is an algorithm that calculates the quality or musicality of supposed music that is provided as input to the algorithm.

There is some degree of overlap between what these two types of algorithm achieve, but they are not the same thing. The **generative** algorithm produces music which has high musicality. The **predictive** algorithm accepts as input any music, or non-music, and tells what the musicality of that input is, and predicts its effect on the human listener.

If we had a predictive algorithm, then a naïve way to convert this to a generative algorithm would be to attempt an exhaustive search of all possible items of music, apply the predictive algorithm to each candidate, and output each item for which the predicted musicality was found to be high enough. This algorithm would work, but it might not be very efficient, because the set of possible musical items grows large very quickly as we consider items of greater and greater length, and only a very small proportion of all possible tunes might be at all musical.

Similarly, if we had a generative algorithm, there is no guarantee that this could be converted to an efficient predictive algorithm. Firstly, a particular

generative algorithm might not generate all possible strong pieces of music. Secondly, even if it did, the only way to use it as a predictive algorithm would be to run the algorithm and generate all possible items until one of them happened to be the same as the input data. If the algorithm terminated, you would know that your input data was musical. If it did not terminate, you would then know that the input data was not musical (but of course it takes an infinitely long time to determine that an algorithm does not terminate, unless you are able to provide a mathematical proof of non-termination).

In practice, we would assume that effective generative algorithms and effective predictive algorithms would both be based on a theoretical understanding of the human response to music, and that given information that could be used to formulate one type of algorithm, we could also formulate the other type of algorithm without undue difficulty.

There are algorithms for which conversion into a related type of algorithm is arbitrarily difficult and suffers from worst-case **complexity**.[11] The standard example is the **cryptographic hash algorithm**. This is an algorithm that produces a fixed length output—the **hash**—typically 128 or 160 bits long, which is derived from arbitrary sized input data, such as a computer data file. The algorithm is irreversible in the sense that it is very difficult to find an input value for a given hash value, unless you happen to already know an input value that generates that hash value. And if you have one input value that generates a hash value, it is equally difficult to discover a second distinct input value that generates the same hash value. In fact a cryptographic hash algorithm is considered broken if anyone ever discovers *any* pair of distinct input values that produce the same hash value.

However, cryptographic hash algorithms have been specially designed to be irreversible. In as much as music does not appear to be part of a biological digital security system, there is no particular reason to suppose that an algorithm for the evaluation of musicality could not be converted into an algorithm for generating music with a high level of musicality. In fact, based on the assumption that the human brain operates according to mathematically specified physical laws, we already have a method which in principle can generate high quality music: simulate the workings of the brains of those people who (at least occasionally) compose good quality music.

2.4.5 The Economics of Musical Composition

I have hinted that finding a musical "formula" would radically change the market for music. But what is the current state of the music composition economy? Who composes the really good music? How do they do it? How hard is it for them?

[11] **Complexity** is a computer science term meaning how much time and memory an algorithm uses when executed in a computer, often specified as a function of the size of the input data.

If existing well-known music theory was complete, then composing good quality music would be relatively easy because the theory would tell us how to do it. I would suggest that the existing economics of music implies that the composition of high quality popular music is far from easy:

- Some composers and songwriters write a lot of music, but others only ever write one or two very good items. This gives rise to the term "one hit wonder" (although this is used more typically of performers, who may or may not also be the composers of the music they perform).
- Some writers write a lot of good songs over a certain period, and then seem to dry up.
- The record industry churns out best-selling albums, many of which contain only one good song, with the rest being "album filler".
- You can get paid a decent amount for making up some good music. Generally nobody ever gets paid a whole lot for doing something that anybody could have done.

We can see that whatever knowledge it is that composers and songwriters have about music that allows them to write music, this knowledge does not exist in a form that enables them to generate arbitrary amounts of new high quality music. It is locked inside their brains as some type of intuitive understanding of music which, when combined with persistence and good luck, enables them to occasionally produce something great.

Trial and error may provide part of the explanation of how music is created: an experienced musician is familiar with many different musical patterns and structures, and combining this knowledge with their own subjective ability to evaluate music, they can generate possible new music, listen to it to see if it is any good, and remember the good stuff. Even when a new piece of music suddenly "comes" to a composer, this may have been the final result of an extended trial and error search that took place within the hidden mechanisms of their brain (a Freudian would say that their **subconscious brain** did all the work).

Although the inner workings of the brains of composers of great music is an interesting topic in its own right, it is not the major purpose of this book to explore the means by which people create new music. My primary focus is on what causes people to respond to the music that they listen to. I cannot rule out the possibility that learning more about musical composition might help us to better understand the listener's response to music, but in practice we will find more direct routes to solving the problem of why and how we respond to music.

The question of creation versus performance versus response cannot be completely ignored when considering the biological purpose of music. Some authors have suggested (and in some cases they just implicitly assume) that

the primary biological purpose of music has to do with creation and performance rather than response to music. I do briefly consider these possibilities, but I will show that there are reasons why hypotheses about the biological purpose of creating and performing music are both unnecessary and unconvincing.

Consideration of the economics of music leads to what I call the **luxury yacht test** (**LYT**) for a theory of music. It consists of the following steps:

- Discover a complete theory of music.
- The theory should specify an algorithm for calculating the musicality of music, possibly parameterised for variations in musical taste.
- Reverse this algorithm to create an algorithm for generating new good quality music.
- Sell the new music.
- Use the proceeds to purchase a luxury yacht.

So if you meet someone who claims to know the answer to the question "What is music?", ask them if they own a luxury yacht.

2.5 Universality

In the above discussion of musicality and predictive algorithms, I implicitly assumed that there existed some measure of musicality that was equal for all listeners. In practice there is a lot of commonality in musical taste, but the very fact that the phrase "musical taste" exists in the language tells us that musical preferences do vary from person to person.

It would be over-reacting to conclude that therefore an algorithmic and scientific theory of music cannot be discovered. People vary in how they react to strains of the flu, but that does not mean we cannot come to a scientific understanding of the influenza virus and its effect on people.

What it does mean is that we will have to **parameterise** our algorithms to take account of variations in musical taste. In other words, the algorithms will accept additional input data representing information about the musical taste of the listener. But, having said that, close enough is often good enough, and if a particular algorithm generates high quality music according to *your* tastes, then at least some of that music will also be considered high quality music according to *my* musical tastes. Suppose that I like only 1% of the music that you like, and we have an algorithm that generates new items of music that you like. To generate one item of music that I like, all I have to do is run the algorithm a hundred times. The 1% success rate (of this hypothetical algorithm) is far superior to the (very close to 0%) success rate of any currently known algorithm for generating music that I like.

The major factors likely to cause variations in musical taste are the following:

- Variations in exposure to music over one's lifetime.
- Variations in exposure to other sensory inputs that affect response to music (which could include language, non-verbal utterances, animal sounds and other natural sounds).
- Variations in personality type.
- Genetic variations in whatever it is in our brain that determines our response to music.
- Random/chaotic variations, i.e. points in the development of our bodies and brains where something could just as easily have developed one way as the other.

The most significant variations in musical exposure are where people belong to totally different cultures and each culture has its own distinct type of music. Not only are the tunes different, but the scales that the tunes live on are different (although usually there *are* scales, and those scales usually repeat every octave, but not always). The whole thing becomes relative: we like our music and not their music, and they like their music but not our music.

Cultural relativity spawns political correctness, and political correctness can discourage researchers from following lines of enquiry that they might otherwise follow. It might, for example, be deemed inappropriate to formulate a hypothesis that suggests (or assumes) that the music from one culture is "better" than the music from another culture.

The most politically incorrect candidate for a "best" type of music is probably Western music, as played on Western scales (i.e. the notes on a piano). Western music is coming to dominate over all other types of music, occasionally including ideas and forms from other cultures, but mostly just replacing them.[12] Is this because Western music is better than other music? Is it because Western countries are imperialistic and dominating? Is it all caused by capitalistic marketing machines?

One circumstance which reduces the accessibility of non-Western music to Western musicians is that most readily available musical instruments are tuned to Western scales, i.e. the well-tempered chromatic scale or some subset thereof. There may come a day when electronic keyboards routinely come

[12]The most substantial input into Western music from other cultures happened when American-African slaves and their freed descendants combined aspects of African music and Western music, giving rise to ragtime, jazz and blues. The African influence can probably be held responsible for most of what makes modern popular music different from older Western classical music. Despite this influence, Western popular music remains strongly tied to the diatonic scale and to underlying regular hierarchical tempo.

with options to select alternative tunings, and when that day comes the dominance of Western scales may be reduced somewhat, and alternative musics may be able to reclaim some of their lost ground.

Even ignoring the political questions, there are theoretical issues, like:

- Does a theory have to take account of all known types of music?
- Can I develop a theory that just applies to one musical culture?
- If my theory describes some aspect of music, does that aspect have to appear in all cultures, or in most cultures, or just in the biggest cultures?

There is the idea of **universality** current among those who study music (scientifically or otherwise), which is that theories about music have to apply equally to all known musical cultures. On one level it is a perfectly valid requirement, but if it is applied over-zealously then important sources of information about music can end up being ignored.

The concept of universality is being applied too strongly if it is used to reject any theory or hypothesis that cannot immediately be applied to all forms and genres of music from all musical cultures that have ever existed.

There is a useful analogy with the study of biology and the study of specific biological organisms. In studying biology we expect to find general principles that underlie the workings of all living species. At the same time, the biologist cannot simultaneously study all organisms at one time. He or she must necessarily concentrate their studies on one particular species, and indeed often just on one or a few members of that species. Eventually some of what is learned about particular species will turn out to generalise to theories that apply to many different species, or even to all species, but we cannot expect or require this generalisation to happen immediately every time we develop a new theory about something.

The criterion for accepting a scientific theory as being useful is not whether it unifies all knowledge in a domain, but rather that it unifies at least some set of distinct facts.

For example, it would be entirely possible and legitimate to develop a scientific theory about a single melody. Our observation of the melody could be regarded as a series of observations of individual musical notes. The occurrence of each note in the melody—its time, length and pitch—counts as one fact about the melody. The theory about the melody would be an explanation that described the notes in some way that was simpler and shorter than a full listing of the notes. Having found a theory about this one melody, we would hope that it could be generalised in some way to form a theory about other melodies, or even all melodies. But even if this is not immediately possible, the theory still has value if it can say something significant about just the one melody.

It follows that we should not feel guilty if we happen to develop theories of music that only apply to certain musical cultures, or to certain genres, or to the musical taste of one person (e.g. the person who developed the theory). The eventual aim of a theory of music is to be universal, and the theory I develop in this book certainly claims to be universal. But a theory about some aspect of music is not wrong or irrelevant just because it is not quite as universal as it could be.

2.5.1 Author's Declaration

Having justified the development of non-universal theories of music, it is perhaps now safe for me to declare my own musical tastes and preferences:

- Most of the music I listen to is the sort of thing you will hear on "Top of the Pops".
- Almost all the music I listen to is diatonic music with regular hierarchical tempo.
- I do not listen to, and do not enjoy, atonal music.
- I do not listen to classical music that much.
- I do not think that John Cage's infamous "4 minutes 33 seconds" is music.

The last example gets a mention in the introduction to *The Origins of Music* (see the next chapter for more discussion of the contents of this book and others), as part of the difficulty inherent in defining what music is, and it's not entirely clear if they are joking or not.

2.6 Scientific Theories

2.6.1 Testability and Falsifiability

The relationship between facts and theories is a large part of what science is about.

Consider a simple example: I throw a ball into the air in a certain direction. I take photos of its path with a camera that can take pictures rapidly at regular intervals. From the photos I record a series of positions at different times. The path and the recorded positions will look something like Figure 2.2.

I have a **theory** about the path of my ball. Writing t for time, x for horizontal position and y for height above some baseline, my theory can be written as a pair of equations that specify position as a function of time:

$$x = v_x t$$
$$y = v_y t - \frac{1}{2} g t^2$$

v_x represents initial horizontal velocity, v_y represents initial vertical velocity, and g represents acceleration due to gravity.

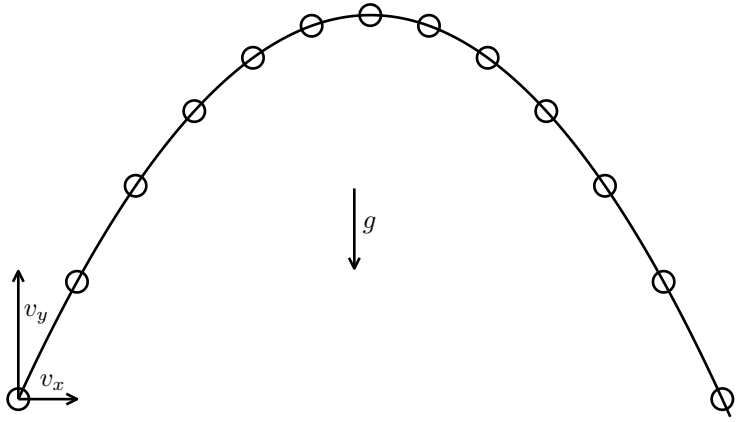

Figure 2.2. A ball thrown into the air with initial horizontal velocity v_x, vertical velocity v_y and downward acceleration g. The camera takes a photo of the ball's position at $t = 0$, $t = 1$, $t = 2$, etc.

The most important thing about the theory in relation to the facts is that the theory is specified using a fixed amount of information (i.e. those two equations), but it can explain a larger number of facts. In this case the number of facts that can be explained by the theory is virtually unlimited, because we can measure a large number of positions each time we throw the ball, and we can throw the ball any number of times, perhaps with different values of v_x and v_y each time.

Sometimes theories explain facts that can only be gleaned by observation, and the supply of facts may be more limited—a good example would be any theory that explains the positions of the planets, as we cannot easily throw new planets into space and observe them (although modern technology does allow one to fire small spaceships out into space). However, as long as the amount of information contained in the observations explained by our theory is larger than the amount of information contained in the specification of the theory, we can be confident that the theory is saying something useful about the world. We can be especially confident if the set of observations explained by the theory keeps on growing, without the theory itself requiring

any further improvement or adjustment.

There are a number of things that we can say about the ball example, which reflect on issues that arise generally when doing science:

- The theory can be related to more general theories. For example, the acceleration comes from gravity, and we can form a more general theory about gravity. The theory about gravity will tell us that g depends on height above the Earth, and that it has quite a different value if you happen to do the experiment standing on the moon.

- The theory is only approximately correct, in part because it makes various assumptions that are not quite true. Air resistance is ignored. It is assumed that the gravitational field is constant. (If we threw the ball hard enough to go into orbit, then the equation would turn out to be quite inaccurate.) Any effects due to the ball itself having a finite extent are ignored.

- The measurements of the ball's position will not be made with 100% accuracy. We will have to allow for this when verifying the theory against the data.

- We may not have any independent way of knowing the values of v_x and v_y, and they will have to be estimated from the data itself in each case. One consequence of this is that at least 3 data points have to be taken in order check the theory at all, since for any 2 data points there will be values of v_x and v_y that exactly match the data. If we don't know beforehand what g is, then its value also has to be calculated from the data, and at least 4 data points are required to be able to check anything. (We would, however, expect g to have the same value for different throws of the ball.)

- If we don't have a camera that can take pictures at regular intervals, it will be very difficult to do this experiment at all.

These issues all have to do with the concept of **testability**, or **falsifiability**. If we state a scientific theory, we expect it to make predictions about something; a theory that doesn't make any predictions that can be checked isn't really a theory. We then want to be able to compare the predictions with measurements and observations. If the predictions come out wrong, then the theory is **falsified**, i.e. proven wrong. We can never prove a theory true, but it becomes more convincing if it makes more and more predictions and never gets proven wrong.

This view is somewhat idealised—that a scientific theory is falsifiable by experimental observation and is rejected the moment it is contradicted by just one observation. Sometimes we have to be a bit forgiving of our theories, for various reasons:

- Sometimes a theory cannot be tested by any practical means, at least not when it is formulated, but it is testable *in principle*. Our theory about the thrown ball is difficult to test if we don't have the equipment for measuring its position at known times. Scientists sometimes deal with this difficulty by specifying **thought experiments**, i.e. experiments carried out only in their imaginations. If we don't have a camera that can shoot pictures at regular intervals, we can still imagine the existence of such a camera, and use this possibility to justify the testability of the theory about the position of a ball thrown into the air. Albert Einstein was famous for inventing thought experiments that tested certain aspects of quantum theory.[13]

- Sometimes the "facts" that disprove a theory turn out to be wrong.

- A theory may explain a whole lot of facts, and then fail on just one fact. Even if that one fact is quite reliable, and it disproves the theory, the theory is still telling us something about all the other facts that it does correctly predict. We know that the theory needs to be replaced with a better theory, but we don't throw away the old theory until we have found the new theory. In fact it becomes a requirement that any new theory should explain why the old theory works as well as it does. This sort of thing happened when special relativity "replaced" Newtonian physics,[14] and also when quantum mechanics replaced Newtonian physics (again).

2.6.2 Simplicity and Complexity

Science often progresses in a certain area because someone asks the right questions and does the right experiments. Real life phenomena can be very complicated, and theoretical descriptions of these phenomena must take into account many different factors. It is best if we can separate out the individual factors as much as possible.

In our thrown ball example, we remarked that air resistance was ignored. If we had tried throwing a piece of paper, or a feather, then it would have been impossible to ignore air resistance. We would not have been able to verify the theory contained in our simple equations. Now even an ordinary ball—like a tennis ball—might be affected by air resistance by a noticeable amount. If we had some idea that air resistance was a complicating factor, then we might guess that we could ignore it if the object being thrown was large and dense. Instead of throwing a tennis ball, we might choose to throw a

[13] Einstein was sure that the theory couldn't be correct, and the thought experiments (published in 1935 by Einstein and two other physicists, Boris Podolsky and Nathan Rosen) were intended to prove this—he believed that the results predicted by the theory were too strange to be possible. But when slightly altered versions of the thought experiments were carried out decades later, the results of the experiments confirmed the theory.

[14] But they still teach Newtonian physics in school.

solid iron ball. We would be rewarded by a very close fit to our mathematical equation, because the size and density of the solid iron ball would allow us to ignore air resistance.

By using a heavier ball, we have created a *simpler* phenomenon to study. If we didn't even know what the equation was going to be, we could have made observations on throwing the heavy ball, and looked for simple patterns in the data. For example, using the **method of differences**,[15] it would have been easy to discover the formula for height as a function of time.

In the case of music, we don't necessarily have a clear idea as to what all the complicating factors are, and whether they can be cleanly separated from each other. But there is one easy way we can avoid complexity, and that is to *study the simplest tunes possible*.

This means, given a choice between a symphony and a pop song, where the symphony has hundreds of bars, multiple motifs, several key changes and a whole orchestra of instruments, and the pop song has 12 bars, 3 chords, one melody, no key changes and can be performed by one guy singing while strumming a guitar, study the pop song first.

There is a tendency in musical academia to listen to "difficult" music, such as long complex symphonies, and strange contemporary music that ordinary folk don't listen to. If popular music is studied, this is done so apologetically.

But when we realise that music is a difficult scientific problem, and it has been studied for over 2000 years, and everyone is still clueless as to what music actually is, then no apology should be necessary. We should study the absolute simplest stuff possible. Even when studying pop music, we should simplify it as much as we can without rendering it unmusical. Is it just a melody line? Maybe, maybe not. Can we reduce the accompaniment to a simple chord sequence (like in a "Learn to Play Guitar" book)? Can we reduce the bass to just the root note of the chord? Can we leave out the rhythm accompaniment, or reduce it to a straightforward pattern of regular beats?

Another good example of scientific simplification is found in biology. Biologists have studied many different organisms, both complex and simple. But some of the most important discoveries in genetics and molecular biology have been made using the simplest possible organisms. The relationship between DNA and protein was discovered using viruses, which are usually just a small section of DNA wrapped in some protein. Other problems required self-contained organisms (viruses are always parasites), in which case bacteria were used as the object of study. And to study the mechanisms of development in multi-cellular organisms, a very simple multi-cellular organism was chosen: *Caenorhabditis elegans*, a 1mm soil nematode which not only has a

[15] Given a sequence of values, keep taking the differences of each element in the sequence and the next to get a new sequence, and repeat this procedure. If you arrive at a sequence of all zeros, you can reconstruct a polynomial which describes the original set of values, such that the degree of the polynomial is one less than the number of times the procedure was applied.

relatively small number of cells in its body, it contains an exact number of **somatic** cells as a fully developed adult—959. (**Somatic** cells are non-germ cells, i.e. those cells that are not destined to become ancestors of the cells in the organism's descendants.)

In all these cases, the biologists did not go around apologising for studying organisms that were too easy or too simple.

A more extreme example, where scientists can only solve the easiest version of the problem, is the dynamics of multi-body gravitational systems assuming Newtonian gravity: the interaction of *two* bodies in each other's gravitational fields is soluble with an **analytical** solution,[16] but solving for *three* bodies is too hard, except for certain special cases. Something similar is found when studying the quantum mechanics of the atom: the hydrogen atom with one nucleus and one electron is doable, the helium atom with one nucleus and two electrons is too hard, and scientists must resort to various approximations, or to brute force integration of the relevant equations on big computers.

If the calculations of the consequences of a theory cannot be calculated accurately (because we are not studying the simplest possible system described by the theory), then the predictions of the theory cannot easily be checked against the results of our observations. And if there is no simple equation that describes the behaviour of the system, there is much less chance that we will discover the theory describing the system just by analysing observations of its behaviour. This is demonstrated by the last example: significant discoveries about the quantum nature of the atom were made from observations of spectral lines of the hydrogen atom, which happen to exhibit certain simple regular patterns.[17] Similarly, Newton's discovery of universal gravity was helped by Kepler's discovery of the laws of planetary motion, which take a simple form because for each planet one can (to a first approximation) ignore the gravitational effect of all other bodies besides the Sun.

[16] An **analytical** solution is one that can be written down as a formula that you can work out on a basic scientific calculator, i.e. only containing algebraic operations, trigonometric and exponential functions, and their inverses.

[17] *Hydrogen: The Essential Element* by John S. Rigden (Harvard University Press, 2002) gives a very good account of how the simplicity of the hydrogen atom has contributed to the development of scientific knowledge.

Chapter 3

Existing Music Science

This is not the first book ever written about music science, and my theories aren't the first music theories either. This chapter summarises some of what has come before me.

Existing theories about music can be classified according to the assumptions that underlie them. The most common assumptions include: the Evolutionary Assumption (correct), the Music Assumption (incorrect), the Communication Hypothesis (incorrect), the Social Assumption (incorrect), the "In the Past" Assumption (incorrect), the Cultural Assumption (over-emphasised), the Cortical Plasticity Assumption (also over-emphasised), the Music-Language Assumption (correct but subject to misleading variations), and a few more technical assumptions about particular aspects of music (all of them probably incorrect). Although the Evolutionary Assumption is a good one to make, it has resulted in the development of many implausible evolutionary hypotheses about music.

3.1 Existing Literature

Each of the following five books is an edited collection of articles or papers written by different authors:

- *Handbook of Music Psychology* edited by Donald Hodges (Institute for Music Research 1996).
- *The Psychology of Music, 2nd Edition* edited by Diana Deutsch (Academic Press 1999).

- *The Origins of Music* edited by Nils Wallin, Björn Merker and Steven Brown (MIT Press 2000). These papers discuss different approaches to understanding the origins of music. Underlying most of them is the belief that we can understand more about music by understanding its origins.

- *Music and Emotion* edited by Patrik Juslin and John Sloboda (Oxford University Press 2001).

- *The Cognitive Neuroscience of Music* edited by Isabelle Peretz and Robert Zatorre (Oxford University Press 2003). This is the most recent music science book, although it is actually an expanded version of *The Biological Foundations of Music* (volume 930 of the Annals of the New York Academy of Sciences, June 2001).

For the purpose of quoting references, I will refer to these books as *Music Psych.*, *Psych. Music*, *Origins*, *Music & Emotion* and *Cog. Neuro. Music*. I am not going to attempt a full review of all the articles and papers—they are not light reading, and any attempts I make to clarify what I think they mean may not be all that helpful. If you are serious about learning all there is to know about music science, then you will probably want to read them yourself, and draw your own conclusions. In this chapter, I restrict myself to summarising existing work in music science as I understand it, and I give references where they seem relevant.

Some other books of interest include:

- *Emotion and Meaning in Music* Leonard B. Meyer (Univ. of Chicago Press 1956). Meyer, a professor of music, advances a theory of expectation, inhibition and completion, and discusses aspects of various musical items and excerpts in ways that match up with his theory.

- *Music and the Mind* Anthony Storr (Ballantine Books 1993). A partly philosophical, partly scientific book asking basic questions about the nature of music.

- *Music, the Brain and Ecstasy* Robert Jourdain (William Morrow 1997). A popularised introduction to music science.

For references to these books I will just quote the author's name.

3.2 The Origins of Music

Origins devotes itself to the origins of music, i.e. how and why did music come into existence? In practice this question is very closely related to the question of what music is now, and why it exists (now). In biology, the study of the present is inextricably linked to the study of the past. The current organism

is the result of a history of evolutionary steps consisting of mutation and recombination (i.e. sex), and natural selection acting on the resulting genetic variation. At each point in time natural selection acts on the species, and at each point in time—including the present—one can explain the purposes inherent in an organism's structure and behaviour in relation to the selective pressures acting at that point in time.

In the first chapter of *Origins*, the editors explain how the study of the evolution of music became "unfashionable" some time after 1940, and compare this to the famous 1866 ban by the Linguistic Society of Paris on discussion of the origins of language.

As the editors of *Origins* point out, discussion of the origins of *music* has never been specifically banned by anyone. But it has suffered from the same difficulties as discussion of the origins of language—scholars can endlessly speculate about origins (of music or language), and there is little reason to reject or accept one speculation over another, as the hard evidence required to do so is lost in the past. The speakers of pre-language and the players of pre-music are long since dead, and their language-like and music-like activities have not left any identifiable remains, at least not that have been discovered. (The musical fossil remains that have been discovered, as discussed in the next section, are of such a nature that their owners may have had musical capacities already equivalent to those of modern humans.)

There is one significant difference between discussing the origins of language and the origins of music: *we know what language is and what it is for*. We can guess what the major selective pressures on the human species were that determined the evolution of the human capacity for language: the need to send information and the need to receive information. (We could just say the need to communicate, but "communication" refers to an activity involving at least two entities, whereas natural selection must act primarily via the reproductive success of the individual.) When we discuss the origins of music, we are discussing the origins of something that we don't know what it is. Even if we do find out what the origin of music is, we may be left not knowing what music is for *now*.

Unfortunately the best guesses about the origin of music are just that: guesses—some plausible, others wild—but guesses just the same. And if the Music-Language Assumption is correct, and music is related to language, then we would expect the precursor of music to be related in an analogous way to the precursor of language. But, as the Paris ban implied, speculations about the precursor to language are also just wild guesses, and we are left with nothing very firm to hold on to.

3.3 The Archaeology of Music

The study of the archaeology of music consists almost entirely of the study of ancient musical instruments, and in particular the study of instruments

made from materials likely to fossilize (such as bone).

The most famous prehistoric musical artefact is the Divje bone "flute", as described in "New Perspectives on the Beginnings of Music: Archeological and Musicological Analysis of a Middle Paleolithic Bone 'Flute'", Drago Kunej and Ivan Turk (*Origins*). It was found in a cave in Divje, Slovenia in 1995. The dating of this fossil strongly suggests that it is a Neanderthal artefact: it was found in a deposit layer dated 50,000 BP (before present) to 43,000 BP, which was quite distinct from another layer dated 35,000 BP which was the most recent layer at the site containing **Aurignacian** artefacts. (Aurignacian culture is a European stone age culture going back to 40,000 BP at the very earliest, and is strongly associated with "modern" humans, with a degree of innovation in art and tool manufacture that contrasts somewhat with that of the **Mousterian** Neanderthal culture.)

Given that it is now believed that modern humans are not all that closely related to Neanderthals, the Divje flute appears to push the origin of music a long way back in time: the common ancestor of Neanderthals and modern humans could have lived as long ago as 400,000 BP.

Much depends, however, on this one piece of evidence. One major uncertainty is that the object may not be a flute. The artefact is a broken piece of a cave bear thigh bone, with two holes in a line, and signs of two other holes on each of the broken ends, and another hole underneath. There may have been some other reason why the artefact's creator decided to drill holes in a bone. But given that it is difficult to think of any other practical purpose for a bone with holes in it, one would be forced to attribute some symbolic significance to it, and there is very little evidence that Neanderthals created artefacts with symbolic meaning (the evidence that does exist is ambiguous and controversial, and contrasts with overwhelming evidence of symbolic artefacts created by modern humans who lived in Europe at the same time as the Neanderthals).

A second uncertainty is that the holes might not have been the result of human activity, the most plausible alternative being that some carnivore bit down on the bones. However, the number of holes and partial holes, and the regularity of their placement, is just a bit too much coincidence for this explanation to be believable. (The paper by Kunej and Turk contains a detailed analysis of the nature of the holes and different cutting processes that could have created them, with the conclusion that the holes were most probably the result of deliberate human manufacture, and very probably not the result of a large carnivore biting on the bone.)

The next oldest known fossil flute is one found in a cave at Geissenklösterle, Germany, dated to 30,000 BP–37000 BP (found by a team from the University of Tübingen).[1] This is associated with the Aurignacian culture, and thus reflects the capabilities and musical preferences of prehistoric modern

[1] **http://www.uni-tuebingen.de/uni/qvo/pm/pm2004/pm824.html** (University of Tübingen press release)

humans, not necessarily much different from those of modern humans living today.

3.4 Common Assumptions

Although there are many different theories of music, and many different approaches that have been taken by those trying to understand music, a relatively small number of basic assumptions underlie most of these theories.

3.4.1 The Evolutionary Assumption

One assumption that I do not dispute is the requirement that music must be explained within the framework of evolution by natural selection.

It's one thing to suppose that music evolved by natural selection as a result of satisfying some biological purpose. It's another thing to determine what that purpose is. Possibilities that have been considered by music scientists include the following:

- Young men sing to attract young women. In *The descent of man, and Selection in relation to sex* (1871), Charles Darwin considered the possibility that music had evolved as a result of sexual selection. **Sexual selection** is where a female has to choose a male according to the same preferences as other females, otherwise her own sons will not have the genes required to make them attractive to the next generation of females. In this way sexual selection can create and maintain preferences that do not serve any other useful purpose, or which may even be counterproductive, like the peacock's tail, which just gets in the way. In "Evolution of Human Music through Sexual Selection" (*Origins*), Geoffrey Miller reviews evidence for and against sexual selection as an explanation for music, his conclusion being that the hypothesis is at least plausible.

- Young women sing to attract young men. Sexual selection does operate in both directions: a male must choose a female mate according to the same preferences as other males, otherwise his daughters will not have the genes required to make them attractive to the next generation of males. Men are generally less choosy about who they have sex with, which implies that sexual selection will not influence male choice as much as it does female choice. But men are reasonably choosy about who they form long-term relationships with, and we do observe that men are apparently more obsessed with physical attractiveness than women are (although it is debatable as to what proportion of the attributes that determine physical attractiveness are the result of sexual selection). So if sexual selection can plausibly explain the musical abilities of males, it can just as plausibly explain the musical abilities of females.

- It's easier to remember something if you sing it as lyrics in a song. See "How Music Fixed 'Nonsense' into Significant Formulas: On Rhythm, Repetition and Meaning" (Bruce Richman, *Origins*) and "Synchronous Chorusing and Human Origins" (Björn Merker, *Origins*).

- Performing music as part of a group improves one's membership within the group—the "social bonding" theory. See "A Neurobiological Role of Music in Social Bonding" (Walter Freeman, *Origins*).

One difficulty with all of these theories is that they allow for music to be completely arbitrary, and therefore say nothing about why music is like it is.

A recent review of evolutionary theories is found in "Is Music an Evolutionary Adaptation?" (David Huron, *Cog. Neuro. Music*). See also "Human Musicality" (Donald Hodges, *Music Psych.*).

Some evolutionary theories of music are stated in terms of what music evolved *from*. Music evolved from something else, where the something else had or has a discernible purpose, and somehow this something else evolved into music. Unfortunately A cannot evolve into B unless B itself has some purpose. Otherwise there is nothing to drive the evolution required. To put it another way, the fact that A might have been a precursor of B does nothing to explain why B exists. It's like explaining what wings are good for by saying that they evolved from legs, and that legs serve the purpose of getting the animal from one place to another by walking or running: we still don't know what the wings are good for.

A list of things that music might have evolved from includes:

- Mothers making communicative noises and gestures to their babies, and babies to their mothers. See "Antecedents of the Temporal Arts in Early Mother-Infant Interaction" (Ellen Dissanayake, *Origins*).

- Language, or specific aspects of language, such as the rhythm and melody of language.

- Alternatively, language evolved from music, and music just carried on existing as well. See "The 'Musilanguage' Model of Music" (Steven Brown, *Origins*), which lists various models of language/music evolution.

The language-related evolutionary explanations are a subset of those explanations subject to the Music-Language Assumption (see below).

3.4.2 The Music Assumption

Perhaps the most dominant and yet unjustified assumption in the field of music science is the assumption that it is *music* that must be explained. Within the framework of evolutionary theory, this translates into an assumption that

music has a biological purpose—that music somehow contributes to reproductive success. Many of those studying the evolutionary theory of music seem to make this assumption implicitly, without even considering the alternative: that the human tendencies that cause people to compose, perform and/or appreciate music can serve some biological purpose, but music itself does not serve any such purpose, rather music is just a side-effect of those tendencies.

On the other hand, sometimes it is recognised that music does not appear to serve any useful purpose, but this is presented as a fatal difficulty within the evolutionary framework.

Musical activity can be divided roughly into three activities:

- Composing
- Performing
- Listening

For each of these activities we can suppose that there exists a corresponding tendency to engage in that activity. My theory not only rejects the Music Assumption, it also supposes that only the tendency to *listen* to music requires biological explanation, because the other activities, i.e. composition and performance, are ultimately motivated by the desire to listen to music. Composers compose and performers perform in order to satisfy their own desire to listen to good music, and to satisfy the desire of their audience to listen to good music.

3.4.3 The Communication Hypothesis

The Communication Hypothesis depends on the Music Assumption—that music must be explained—and states that the explanation for music is that it is a form of communication. The problem is to determine what it is that is being communicated. Given the observed effects of music on listeners, we might suppose that one or more of the following is being communicated:

- Emotional quality
- Dance! (as a command)
- Feel good! (as a command)

There are several major objections to this hypothesis:

- The amount of information inherent in a piece of music far exceeds what is necessary to impart information on any of these topics. “Dance” and “Feel good” are just simple commands, and there are not that many distinct emotional qualities in the world that are worth communicating. Yet music has a level of complexity, even in the simplest of tunes, which seems out of proportion to what is required to communicate any of these items of information.

- Composing music is not easy to do. How can you musically communicate anything if you don't know how to compose music? At best you can make use of the repertoire available to the culture you live in. Compare this to language: we all know a "repertoire" of words and syntax, but we do not rely on a "repertoire" of sentences, rather we freely compose our own sentences as the need arises.

- It does not feel subjectively that we perform music to communicate. We perform to entertain (ourselves or others), or because the occasion demands it. When we do want to communicate, we generally speak, and this is often supplemented by other forms of communication, such as facial expression, body language, and non-linguistic vocalisations such as laughing and crying. But we do not sing.

The first part of *Origins* consists of articles about animal calls and songs and their relationship to human language and music. Given that almost all animal calls are believed to be some type of communication, it would follow that if human music evolved from non-human animal calls, then music must also be a type of communication.

Patrik Juslin in "Communicating Emotion in Music Performance: A Review and Theoretical Framework" (*Music & Emotion*) presents a theory of how music communicates the emotions of the performer to the listeners.

3.4.4 The Social Assumption

The Social Assumption is the assumption that music plays some crucial role in creating and maintaining human society. It is true that people gather together to make music, and to listen to music, and to respond in other ways such as dancing. And people often sing songs or make music that reflects membership in their society or religion.

But none of these observations are really evidence that music exists *for the purpose* of maintaining social connections or increasing social bonding. People listen to music together, but they also drink alcohol together. One would hardly say that the purpose of alcohol is to increase social bonding. In Western society our use of alcohol and other recreational drugs is fairly informal (and even legally prohibited in some cases). In other societies particular drugs may play a central role in the formal rituals of those societies. But we would still not say that the *purpose* of mind-altering drugs is to facilitate social bonding. Rather we would say that the drugs have effects on their users which lead to them being chosen as a component of social rituals. Similarly for the use of alcohol at a party. And similarly for the performance and appreciation of music, whether in a formal ritual or at an informal party—it is the effects of music that encourage its use in those situations.

Humans are very social animals—almost anything they do can be done

socially.[2] So just because an activity occurs in social situations, that is no reason to suppose that the activity in question serves a social purpose.

This reasoning applies even where the performance of music *requires* group activity, like a choir singing in harmony, or a band playing different instruments. It typically requires group activity to make a house. But it is not the purpose of house-building to bond society together—the purpose of building a house is to make a house that someone can live in.

3.4.5 The "In the Past" Assumption

Reference to the past is a general strategy for solving hard problems about evolutionary human biology: the thing to be explained doesn't serve any useful purpose now, but it was very useful in the past when we were all hunter gatherers living in small tribes. The technical name for this past life that explains everything about us is the **environment of evolutionary adaptedness**[3] (**EEA**). Now it is true that there was a time when all of our ancestors lived in this environment, and currently many of us don't live in such an environment. *Some* evolutionary problems can be solved by comparing the past with the present. A good example is the set of desires that cause us to eat more of certain foods than are good for us. In the EEA these foods were not freely available, and when they were occasionally available, the short-term benefits of eating them outweighed the long-term costs. Most people were going to die early anyway, and malnutrition presented a much greater immediate threat than cancer, diabetes and circulatory disease.

But EEA-based explanations must be used with caution, and here is a list of problems that can arise:

- Some EEA-based explanations make further suppositions about the nature of human culture in the EEA. But the big thing about human culture is that it *varies*. Culture is a manner of creating and passing on variation that operates somewhat independently of genetic evolution, and also considerably faster. Any evolutionary explanation that assumes some particular and peculiar characteristics of primitive human culture is ignoring this intrinsic tendency towards variation.

- There are still people living today in circumstances that approximate the EEA. That is, they live in small tribes and feed themselves by hunting wild animals and gathering wild plant foods. If you were invoking the EEA, hoping that your theory could not be tested against a real live stone age hunter-gatherer culture (and found wanting), you could be out of luck.

- Even if a theory of musical behaviour depends on characteristics of life in an environment and culture that no longer exists, the human

[2] Although there are *some* activities that we mostly prefer to do in private.

[3] A term invented by John Bowlby, the psychiatrist who developed Attachment Theory.

musical tendencies that the theory is trying to explain do still exist. Any theory must be consistent with our *current* experience of those tendencies. If, for example, music was used by males to flirt with females in the past, are modern day males observed to flirt with females by singing to them? Do they show even a tendency to behave in this way? (There are indeed circumstances where young men are observed to sing or perform music to females in the hope of creating or enhancing romantic interest, but there is no real evidence that this is an *instinctive* behaviour. Rather it appears to result from a conscious plan based on a conscious understanding of the likely effects of such performance.)

3.4.6 The Music-Language Assumption

At its most general, the Music-Language Assumption states that music and language have some relationship to each other. It is an assumption that I agree with, and if you read on you will see that my theory of music quite explicitly relates the perception of music to the perception of language.

There are, however, many different ways that music and language can be related. There are also many different choices to make as to which aspects of language relate to which aspects of music, and why. For example, some authors relate musical harmony to linguistic syntax—an analogy not included in my theory.[4]

Papers that relate music to language include "Comparison between Language and Music" (Mireille Besson and Daniele Schön, *Cog. Neuro. Music*), "Toward an Evolutionary Theory of Music and Language" (Jean Molino, *Origins*) and "The 'Musilanguage' Model of Music Evolution" (Steven Brown, *Origins*).

Poetry is one phenomenon whose characteristics place it in the gap that lies between music and language, and some authors consider the relationship between poetry and music, for example, Fred Lerdahl in "The Sounds of Poetry Viewed as Music" (*Cog. Neuro. Music*).

3.4.7 The Cultural Assumption

Music is a cultural phenomenon, and people respond primarily to music from their own culture. Some conclude from this that the evolution of music is subject only to laws of cultural evolution, and that it is not appropriate or relevant to explain music in terms of genetic evolution by natural selection.

[4] A **syntax** is formally defined as a set of rules for accepting a sequence of symbols. Thus a syntax of English would be a mathematical description of what constituted a grammatically correct English sentence. Although the syntaxes of natural human languages have so far defied complete formal description, there are approximate descriptions that are convincingly close, and good enough to enable computers and people to chat on some level (usually bounded by the limitations on the computer's ability to handle semantics rather than by its inability to deal with syntax).

It is true that culture strongly affects the musical behaviour and the musical tastes of individuals. But the existence of human culture does not remove the need to explain human behaviour in a biological evolutionary framework. Human culture exists because there are human tendencies to copy attitudes, preferences and behaviours from other people. These tendencies to copy are themselves necessarily determined by our genes, and are subject to natural selection just like any other genetically determined aspect of human nature.

Human culture is not a simple fixed attribute of human behaviour. There are many possible variations in the way that information is copied from one person to another. You can pay more or less attention to the attitudes and behaviours of other people, according to any number of relevant criteria: whether or not another person is admirable in some way, whether they are successful, whether they belong to your family group, whether they are the same gender as yourself.

Different *kinds* of information can be copied in different ways. There are almost certainly special mechanisms that exist for learning and reproducing natural language. At the same time, many behaviours are not substantially determined by cultural transmission, behaviours such as running, walking, eating and breathing (the basic mechanics of these behaviours are not culturally determined, although culture may still affect some peripheral aspects of them).

There also exist specific "anti-culture" mechanisms, which have the effect of negating or reversing culturally determined attitudes. In particular there is teenage rebellion, where at a certain age the individual goes out of their way to behave in ways consistent with their peers but inconsistent with the mores of their parents and the larger society they live in.

And, as a final complication, different individuals have varying tendencies to copy or not copy the attitudes and behaviours of others. Some people have a strong tendency to "fit in", even where this conflicts with common sense. Others live in a world of their own, yet may still make a useful and unique contribution to the society they live in, perhaps as a result of their individualism.

It is very likely that separate genes affect each of these different mechanisms and aspects of the transmission of culture. So we can't just say "Music is determined by culture, so forget about the biology". We still have to ask what the cultural mechanisms are that cause music to propagate from one generation to the next, and perhaps change along the way, and what the biological purpose is of those cultural mechanisms (i.e. what the forces of natural selection are that act on the genes that affect those mechanisms).

3.4.8 The Cortical Plasticity Assumption

I investigate **cortical plasticity** in more detail in Chapter 10. Cortical plasticity refers to the brain's ability to rewire itself to process whatever type of information it needs to or wants to process. In the context of music science,

the concept allows us to believe that the brain rewires itself however much is necessary to process the patterns and structures of music. The problem with this belief in flexibility is that it distracts us from an opposite possibility: that aspects of music evoke a response in cortical maps which already exist for some other purpose, and these cortical maps exist independently of any exposure to music.

The Cortical Plasticity Assumption is related to the Cultural Assumption, in that it is generally assumed that a person's brain adapts to the music of their culture by means of cortical plasticity.

In "Musical Predispositions in Infancy" (*Cog. Neuro. Music*), Sandra Trehub reports on studies of the musical capabilities of infants. The results show that many aspects of music perception are already found in infants, even though they are so young that their previous exposure to music must be very limited. The conclusion is that we come into this world to some extent already "wired" for music perception.

3.4.9 The Simultaneous Pitch Assumption

Compared to the assumptions I have discussed so far, the Simultaneous Pitch Assumption is quite a technical assumption. It is assumed that, to understand the basis of musical harmony, we must understand how the brain processes perception of simultaneous notes with pitch values related (or not related) to each other by consonant intervals.

This may seem almost common sense, since harmony is by definition the performance of different notes simultaneously in music. However, this assumption is a subtle corollary of the Music Assumption—the assumption that we must explain music, as opposed to explaining human musical tendencies.

Harmony is one aspect of music where this assumption makes a large difference. One form of harmony is **chords**: groups of notes related by consonant intervals. It is an empirical fact that the listener to music can perceive chords as groups of notes played simultaneously, but can also perceive chords as groups of notes played sequentially. It may be that the response to sequential notes is what actually matters and requires explanation in an evolutionary framework, and that the response to simultaneous notes is an accidental side-effect of the ability to respond to notes of a chord sequentially.

An example of research into harmony and the perception of consonance and dissonance is "Neurobiology of Harmony Perception" (Mark Tramo, Peter Cariani, Bertrund Delgutte & Louis Braida *Cog. Neuro. Music*).

Tramo *et al.* conclude from their research that consonance and dissonance of simultaneous tones are encoded in the form of **interspike interval**[5] (**ISI**) distributions as measured in the auditory nerve of a cat (there is no claim

[5]The **interspike intervals** are intervals between action potentials. Calculating the distribution of intervals is equivalent to calculating the **autocorrelation** function of the signal, and doing so extracts periodic features from the signal.

that cats perceive music, but it is reasonable to presume that this aspect of auditory perception is not too different from what occurs in humans).

This encoding would be an example of **temporal coding**, i.e. encoding of information in the precise *timings* of neural activity. The paper does not make any suggestions as to how such an encoding might be translated into other forms of encoding, such as position within a cortical map. However, it seems likely that temporally encoded information must eventually be re-encoded into a positional form if it is to be integrated and processed with all the other information that the brain processes.

Tramo *et al.*'s research is part of a long history of attempting to determine neurophysiological correlates of the subjective perception of consonance and dissonance, which includes the work of scientists such as Hermann von Helmholtz, Carl Stumpf, and R. Plomp and J.M. Levelt (the last two developed the **critical band theory** of consonance). Although consonance and dissonance appear to be major aspects of music, there are difficulties that arise in interpreting these attempts to understand the perception of consonance and dissonance:

- Most experiments in this field involve asking subjects to judge the consonant/"pleasant"/"non-rough" quality of pairs of tones, which are usually played simultaneously. But our knowledge of the relationship between subjectively perceived consonance and musicality is very limited: we observe that dissonant chords tend to "resolve" into consonant chords, and that's about it. So even if we determine that neurophysiological phenomenon X is perfectly correlated with the perception of consonance and dissonance, we still don't know what, if anything, phenomenon X has to do with musicality.

- As already mentioned above, harmonic relationships matter both between simultaneous tones and sequential tones. The ISI distribution measured by Tramo *et al.* is quite explicitly dependent on the simultaneity of the tones: the distribution is a function only of the *current* tone or tones being perceived. An observation readily made by anyone who has played music with different types of accompaniment (including no accompaniment at all) is that very often the difference between simultaneous and sequential has only a minor effect on how the harmonic relationships between notes contribute to the musicality of the music. In many cases the harmonic relationships are already found *in the melody* (which is sequential), and playing an explicit accompaniment at most helps to emphasise those relationships.

3.4.10 Other Musical Aspect Assumptions

The Simultaneous Pitch Assumption is just one of a group of technical assumptions that derive from the Music Assumption. A brief description of some of these other assumptions is:

- **Scale Assumption:** that there is some part of the brain that responds to musical scales, and the purpose of this part of the brain is to perceive musical scales. A common follow-on conclusion is that scales exist so that the brain can *categorise* pitch values, similarly to how it categorises other continuums into discrete values, as happens with vowel sounds and colours. For example, see "Intervals, Scales and Tuning" (Edward Burns, *Psych. Music*).

- **Regular Beat Assumption:** that the occurrence of regular beat in music relates to the importance of regular beats from some other source or sources. One popular candidate for this is the human heart, either the person's own heart, or their mother's heart which they heard before they were born. In either case it is not clear why hearing a regular beat under particular circumstances should result in the development of our appreciation of the complex rhythms of music. Nor is it clear why there should be a major perceptual system devoted to listening to heart beats: the infant in the womb cannot do much in response to its mother's heart beats, and even when we do hear our own hearts beating, we do not normally act on the information in any significant way. Our bodies have other ways of providing and processing information relevant to the functioning of the heart (like wanting to rest when we get tired from doing too much exercise).

- **Hierarchical Segmentation Assumption:** I originally made this assumption myself, that, to understand music, we must understand how the brain processes hierarchically organised data, because music has a hierarchical structure. In particular musical time has a hierarchical structure.

  Musical time is hierarchical in the sense that a tune consists of bars, which—assuming for instance typical 4/4 time—sub-divide into half bars and then into counts and then half counts and finally quarter counts. Often the hierarchy of grouping also proceeds in the opposite direction: bars are grouped into groups of bars and even into groups of groups, in a way that matches the phrasal structure of the melody. A natural mathematical representation of this hierarchical division is a discrete N-dimensional space, where N is the number of hierarchical levels. Unfortunately, cortical maps in the brain are only 2-dimensional (with the 3rd physical dimension being too small to represent information values), so there is no "natural" way to represent this N-dimensional space in the brain.

When I developed a full understanding of the **regular beat cortical map** (see Chapter 10) and how it processes information about rhythm and tempo, I found that the hierarchical nature of musical time is a consequence of the constraint that musical rhythm should contain multiple regular beats, so there is no need to make specific assumptions about the existence and perception of hierarchy just to explain this feature of musical time.

The regular beat cortical map may not account for musical hierarchy that exists on a time scale greater than bar lengths, and large scale hierarchy may result from constraints determined by other aspects of musicality. One such aspect is repetition: components of music within an observable hierarchy are often repetitions or partial repetitions of previous components of the same music.

A Generative Theory of Tonal Music by Fred Lerdahl and Ray Jackendoff (MIT Press 1983) describes a formal system for analysing music into strict hierarchies.

3.5 Questions That Have to be Answered

Perhaps the biggest problem with most theories of music is that they fail to confront *all* the questions that can be asked about music.

There are many things that we know about music—most of these become obvious to anyone who learns to perform music. A complete theory of music must explain all of these things that we know about music, not just some of them. The theory must explain why music is what it is, and why it isn't what it isn't.

One point of view is that many aspects of music are culturally determined, and for any such aspect one can specify "culture" as being the reason for that aspect's existence. A corollary of this view is that only those features observed across all or most cultures need to be explained.

I have already discussed this issue in the previous chapter, in the section on Universality. In developing my own theory of music I have decided to take what might be called the strong approach, and I assume that in the first instance a theory of music should be capable of explaining all observed features of music, whether or not those features are found across all cultures, as long as it can be established that the features contribute substantially to the musicality of music for a substantial number of listeners. This implies that you cannot dismiss a feature of music from the scope of a general theory just because there are some listeners who do not respond to that feature or to music containing that feature.

Even if we don't accept this strong approach, and instead settle for a weaker approach of only requiring explanation for those features that are universal, or at least found across a large proportion of all musical cultures,

there are still many questions that need to be answered.

The questions in this first list relate to universal or near universal aspects of music:

- What selective pressures have resulted in the human capacity to respond to music?
- Why do melodies consist of notes with constant pitch values taken from scales, where a scale consists of a finite set of possible pitch values?
- Why are notes sometimes "bent" (breaking the rule about constant pitch values stated in the previous question)?
- Why do scales usually repeat every octave?
- Why are notes separated by multiples of an octave perceived as having a similar quality? (And this is not true for other consonant intervals.)
- Why do scales usually contain 5 to 7 notes per octave?
- Why are scales usually uneven?
- Why does melody mostly go up and down the scale one step at a time?
- Why is the musical quality of music invariant under transposition into a different key?
- Why do consonant intervals play such an important role in music?
- Why is musical beat usually completely regular?
- Why is musical beat sometimes *not* completely regular (e.g. irregular bar lengths found even in popular music, and polyrhythm found in some types of non-Western music)?
- Why is musical time consistently divided up into intervals by factors of 2 (mostly) or 3 (sometimes)?
- How are we able to recognise the same rhythm played at different tempos?
- Why does music have an emotional effect? Why does it sometimes cause goosebumps or shivers down the spine?
- Why do we enjoy music?
- Why do we like some music more than we like other music?
- Which parts of the brain respond to music, and do different parts respond to different aspects of music?

- Do the parts of the brain that respond to music serve some other purpose, or have they been specifically recruited as a result of exposure to music?

The next list consists of questions that relate more specifically to popular forms of Western music, but I would still expect a complete theory of music to answer them:

- Why does the well-tempered diatonic scale work as well as it does?
- Why do chords change mostly at the beginning of a bar?
- Why do the more strongly emphasised notes in the melody usually correspond to notes in the current chord?
- Why are there home chords, and why are they almost always either C major or A minor (on the white notes scale)?
- Why is the final home chord often preceded by a dominant 7th chord, i.e. G7 precedes a final C major, or E7 precedes a final A minor?
- Why is there a bass line which generally starts with the root note of the chord when there is a new chord?
- What determines the minimum number of chords found in popular tunes: very rarely less than 3, and usually at least 4?
- Why are syncopated melodies so common in modern popular music?
- Why do listeners prefer music containing singing?
- Why do song lyrics almost always rhyme (although sometimes the rhymes are weak)?
- Why do melodies contain repeated components, or components that repeat some but not all aspects of the music (e.g. rhythm only)?
- Why do certain instrumental timbres work better with certain genres of music? (A good example of this is the over-driven electric guitar, which appears to be entirely responsible for the previously unknown genre of heavy metal, elements of which are contained in much of modern popular music.)
- Why do we like to watch groups of people dancing synchronously in time to music (but not the synchronous motion of anything else)?
- What are the constraints, as yet undetermined, which make it non-trivial to compose original commercial quality music, even if one knows all the "rules" of musical composition?

(Some of these questions contain technical musical terms that some readers may not be familiar with. These will be explained as necessary in the next chapter on "Sound and Music".)

3.6 Approaches to Studying Music

When no one has any idea what the answer is, there aren't any rules about what is the correct way to attack the question, and as a consequence there are many different approaches that music scientists (and philosophers and theorists) have taken in their efforts to solve the basic mystery of music.

Here is a list of research and analysis methods that I am aware of:

- Cognitive and perceptual experimentation that attempts to discern the processes involved in music perception and related types of perception including language cognition. This experimentation may be combined with the use of brain imaging techniques that measure the intensity and location of neural activity in the brain while a subject performs certain cognitive tasks.
- Comparison of human music to various kinds of animal "song".
- Comparison of music to language.
- Studying the development of musical competence in the growing child. (At a given point in time, some aspects of music perception may be well developed and others may not be—so studying development can help to analyse music perception into its components.)
- Studying the archaeology of music, in particular fossil musical instruments such as the Divje bone flute.
- Formulation of hypotheses about how music contributes to reproductive success.
- Analysis of individual musical items, attempting to explain the subjective effects of the music being analysed. Most such analysis is done within the discipline of traditional music theory, which unfortunately tends to be somewhat unscientific: the "theories" are not formulated as proper scientific theories, and the theorists do not treat the study of music as a sub-discipline of biology.
- Statistical analysis of either individual items (small or large) or of collections of different musical items.

  Hit Song Science (*http://www.hitsongscience.com/*) is a commercial service that claims to be able to distinguish hits from non-hits based on a statistical analysis of a large historical database of hit music.
- Mathematical modelling of music perception. Many such models are based on neural networks (which are in effect mathematical models of networks of neurons in the brain). For example, in "Tonal Cognition" (*Cog. Neuro. Music*), Carol Krumhansl and Petri Toiviainen describe a neural network model that perceives key changes.

- General philosophical discussions of music and any aspects of the human condition assumed to be relevant to an understanding of music—in particular human emotion. Unfortunately such philosophical discussions suffer the same problems as traditional music theoretic analysis: they are usually not very scientific.

- Investigation into the differences between the brains of musicians and non-musicians. Learning to play music well enough to make a living from it causes significant and observable changes in the brain. For example, see *Cog. Neuro. Music*, "The Brain of Musicians" (Gottfried Schlaug), "Representation Cortex in Musicians" (Christo Pantev, A. Engelien, V. Candia and T. Elbert) and "The Brain that Makes Music and is Changed by it" (Alvaro Pascual-Leone).

  Of course it is likely that reorganisation of the brain occurs with many types of specialist; for example, the way that mathematics is represented in the brains of mathematicians may be different to how it is represented in the brains of non-mathematicians. And the representation of information about driving in a racing car driver's brain may be different to the representation of the same information in the brain of an ordinary driver. Thus the reorganisation of cortical maps in the brains of musicians is interesting, but it may tell us more about the consequences of becoming a specialist in something than it tells us about what music is.

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.

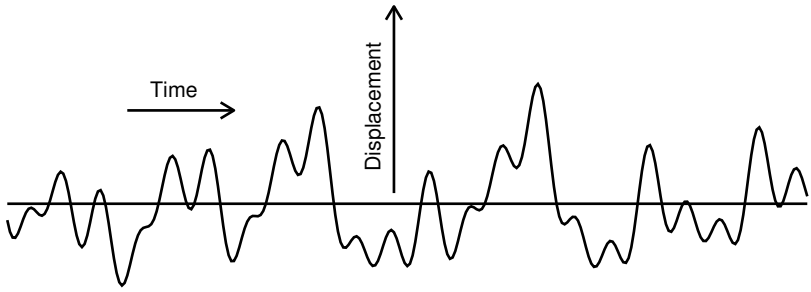

Figure 4.1. A graph of sound waves passing a fixed point, showing displacement as a function of time.

> **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

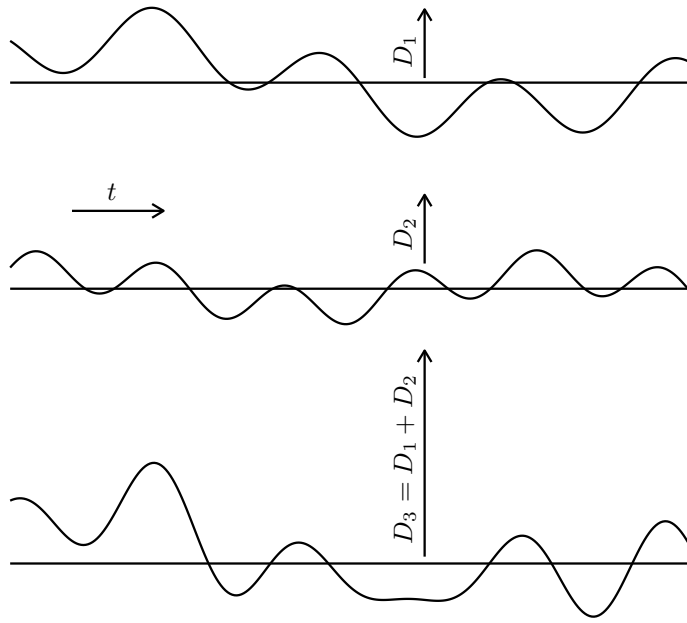

Figure 4.2. Linear superposition. Displacement D is a function of time t. $D_1 + D_2 = D_3$ for each time t. D_1 as a function of time is the displacement at a given point caused by one sound, D_2 is the displacement at the same point caused by another sound, and D_3 is the total displacement caused 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.

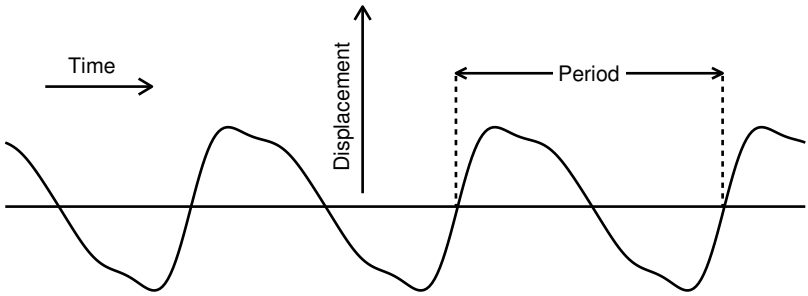

Figure 4.3. Sound consisting of a regular repetitive vibration.

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]

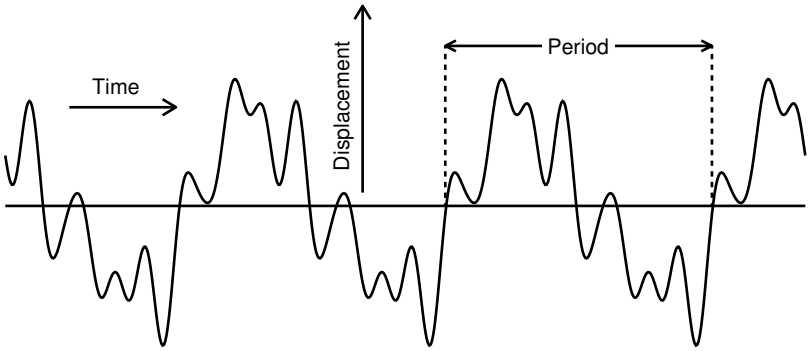

Figure 4.4. Sound consisting of a regular repetitive vibration but with little ups and downs within the main vibration.

[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.

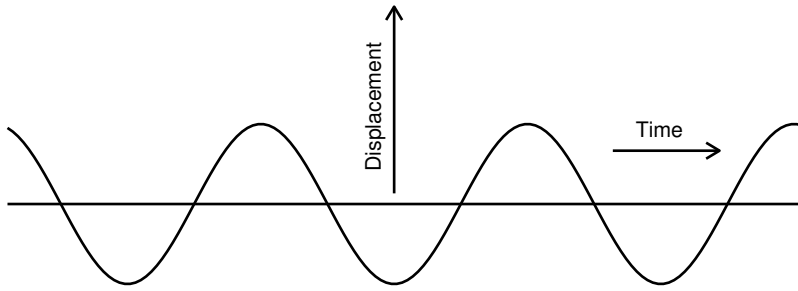

Figure 4.5. Sine wave vibration.

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).

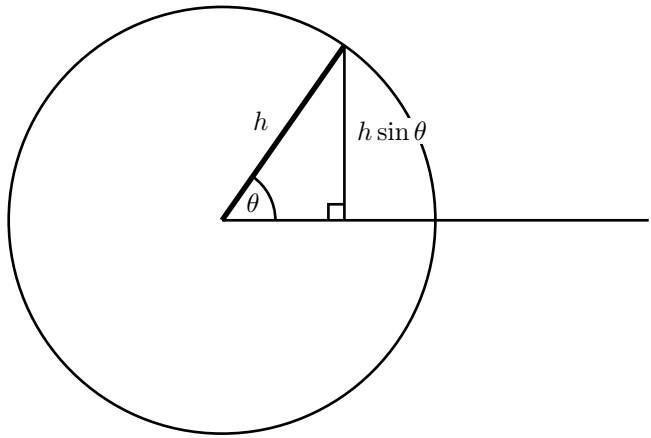

Figure 4.6. Definition of the **sine** function: $\sin\theta$ is the length of the side opposite the angle θ divided by the length of the hypotenuse h ("sin" is the abbreviation for "sine" used in mathematical equations and formulae).

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.

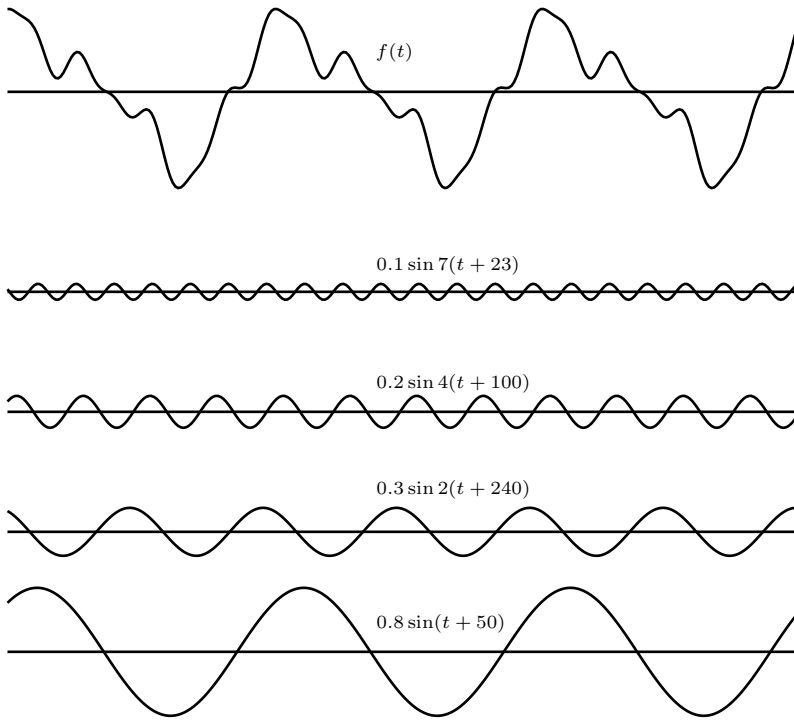

Figure 4.7. The periodic function f can be decomposed into the sum of four sine wave functions: $f(t) = 0.8 \sin(t+50) + 0.3 \sin 2(t+240) + 0.2 \sin 4(t+100) + 0.1 \sin 7(t + 23)$. (Here t is assumed to be measured in degrees.)

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/**.

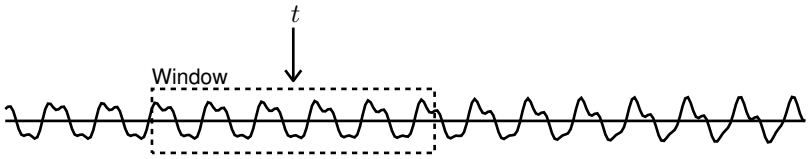

Figure 4.8. Vibration analysed inside a sliding "window". A window size is chosen such that the pattern of vibration is approximately constant within the window. Frequency analysis at each time t is based on analysis of vibration within the window centred on that time.

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

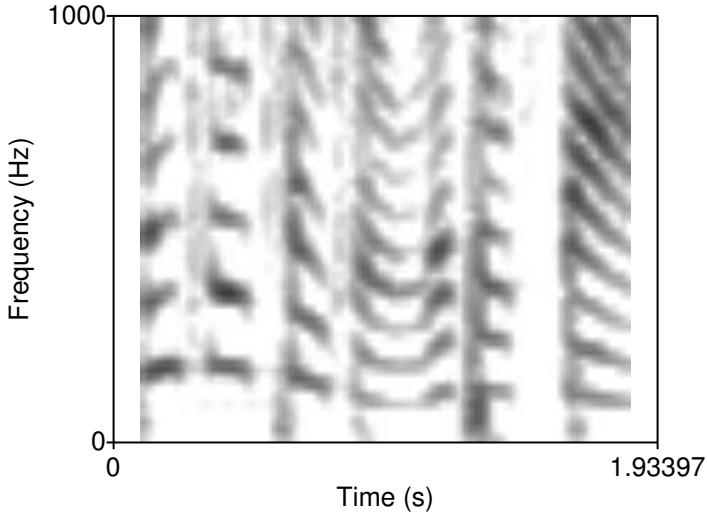

Figure 4.9. A spectrogram of the author saying "Twinkle Twinkle Little Star".

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.

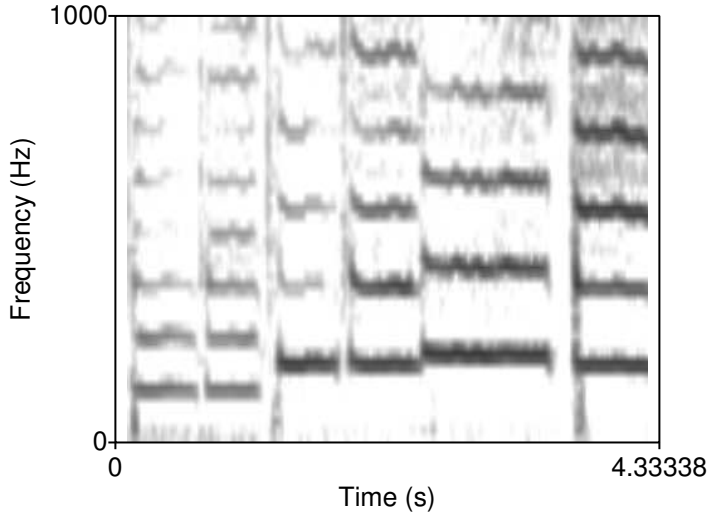

Figure 4.10. A spectrogram of the author singing "Twinkle Twinkle Little Star".

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.

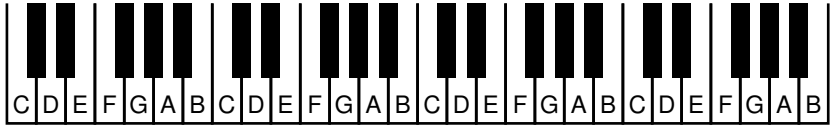

Figure 4.11. A musical keyboard.

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$\sharp$ or D$\flat$.

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$\sharp$, G$\sharp$, A, B, C$\sharp$ and D$\sharp$ 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 $\approx$ a ratio of 3/2,
- a **major third** = 4 semitones $\approx$ a ratio of 5/4, and
- a **minor third** = 3 semitones $\approx$ 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$\flat$ 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?

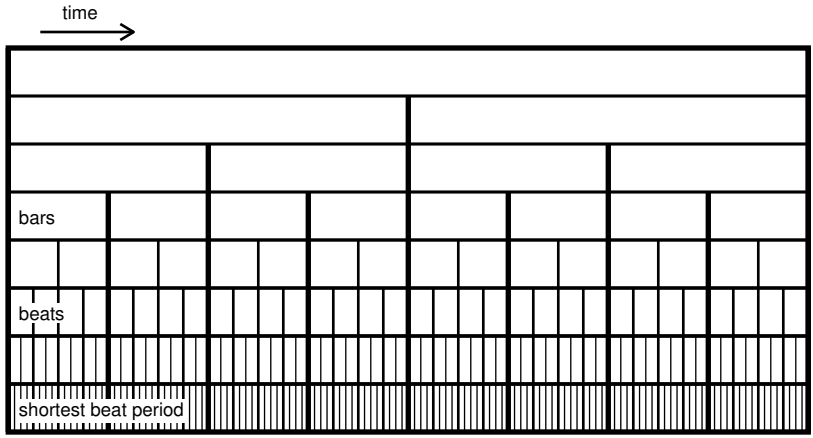

Figure 4.12. Hierarchical division of musical time. This example shows eight bars. The time signature is 4/4, i.e. 4 beats to a bar, and the tune contains notes that are 2 or 4 notes to a beat, so that the shortest beat period is 1/16 of the length of the bar.

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**.

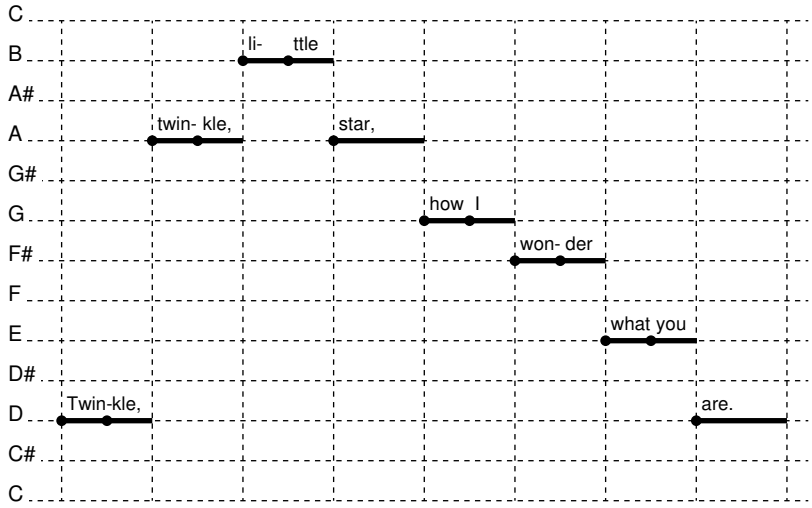

Figure 4.13. The first phrase of a well-known melody, drawn as a graph of log frequency versus time. The vertical lines show the times of the beginnings of the bars. The time signature is 2/4, and the melody is shown here as played in the key of D major.

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.

Chapter 5

Vector Analysis of Musical Intervals

The intervals between musical notes can be regarded as **vectors** in a **vector space**. Intervals in the diatonic scale have natural 1-, 2- and 3-dimensional vector representations, and there are also natural mappings from 2 to 1, 3 to 1 and 3 to 2 dimensions. The **kernel** of the natural 3D to 2D mapping is generated by the **syntonic comma** which equals 81/80. The **Harmonic Heptagon** provides a compact visualisation of all the consonant relationships between notes in the diatonic scale, and a trip once around the heptagon corresponds to one syntonic comma.

5.1 Three Different Vector Representations

When I first started trying to understand all the relationships between notes on the diatonic scale, there seemed to be almost too many different ways to describe the intervals between pairs of notes.

(The analysis in this chapter applies to the diatonic scale; however, for the sake of concreteness, all examples are based on the white notes scale, i.e. C, D, E, F, G, A, B.)

Firstly, an interval between two notes can be described as the logarithm of the frequency ratio between the notes. On the well-tempered scale all intervals are integral powers of $\sqrt[12]{2}$, so this is equivalent to a simple count of semitones. For example, the interval from a C to the next higher G is 7 semitones (i.e. the ratio $2^{7/12}$).

The representation as a count of semitones can describe all possible in-

tervals, but it does not take into account the structure of the diatonic scale. Some steps from one note to the next are tones, and some are semitones. So a second formulation is to count the number of tones and semitones in an interval separately. For example, the interval from a C to the next higher G would be 3 tones and 1 semitone.

So far, we have two ways to describe intervals between notes, yet neither of them says anything about the most important feature of musical intervals, which is that chords, harmony and the repeating structure of the scale are all based on intervals that correspond to simple fractional ratios of frequencies. For example, the interval from C to the next higher G corresponds to a ratio of approximately 3/2.

These simple fractional ratios form the basis of a third representation. The third representation is different from the other two, because it only applies to some intervals—only the consonant intervals have obvious representations as fractional ratios. Any ratio assigned to other intervals is somewhat arbitrary, and there is no best way of making such an assignment to all intervals.

To analyse the relationships between these three representations of musical intervals—semitones, semitones plus tones, and fractional ratios—we need a common framework for specifying them. Luckily there is a ready-made mathematical structure that we can use: each of the three representations defines a **vector space**.

5.1.1 What is a Vector Space?

Vectors are mathematical objects with **magnitude** and **direction**. Vectors can also be formulated in terms of **components**. The component formulation will turn out to be more useful for the current analysis. Also the vector spaces that we will define are all **finite dimensional**, which makes everything a lot easier.

We can define a **finite dimensional vector space** V as follows:

- The vector space has some number n of dimensions. We say that V is n-dimensional, or nD for short. (The only values for n that we use in this chapter are 1, 2 and 3.)

- A vector belonging to an n-dimensional vector space V has n components. Each component is a number.[1] We can write the components as a comma-separated list in brackets; for example, $(2, 3)$ is an example of a vector belonging to a 2-dimensional vector space, and $(-1, 0, 5)$ is an example of a vector belonging to a 3-dimensional vector space.

- Two vectors from the same space V are equal if and only if all their corresponding components are equal. (We can say that equality is defined

[1] Generally a **real** number, although most of the components of the vectors we are dealing with will be integers.

componentwise.) For example, $(3, 2, 1) = (3, 2, 1)$ but $(3, 2, -1) \neq (3, 1, -1)$ because the second components 2 and 1 are not equal.

- Vectors from a vector space V can be added together by adding their corresponding components. For example, $(1, 0, -2) + (3, 3, 4) = (1 + 3, 0 + 3, -2 + 4) = (4, 3, 2)$, and $(3, 4) + (1, -3) = (4, 1)$, as in Figure 5.1. (Thus addition is componentwise.)

- A vector from a vector space can be multiplied by a number, often called a **scalar** to distinguish it from a vector, by multiplying each of its components by the number. This is called **scalar multiplication**. The scalar is normally written on the left of the multiplication. For example, $4 \times (1, 0, -2) = (4 \times 1, 4 \times 0, 4 \times -2) = (4, 0, -8)$, and $3 \times (2, 1) = (6, 3)$ (see Figure 5.2).[2] (Scalar multiplication is also componentwise.) Note that the definition of scalar multiplication is consistent with the definition of addition, in that, for example, $2\mathbf{x} = \mathbf{x} + \mathbf{x}$ for any vector $\mathbf{x}$ belonging to a vector space V.

We also want to define an n-dimensional **point space**. Just like a vector in a vector space, a **point** in an n-dimensional point space can be written as a list of n components, where each component is a number. The important difference between a point space and a vector space is that they have different operations defined on them. There is no way to add points to each other or to multiply points by a scalar. We can, however, add a point to a vector to get another point, which we do by adding corresponding coordinates (exactly as for vector addition—see Figure 5.3).

In any vector space there is a well-defined zero vector $\mathbf{0}$ which is the vector whose components are all zero. For every vector $\mathbf{x}$, $\mathbf{x} + \mathbf{0} = \mathbf{x}$, and for every point p, $p + \mathbf{0} = p$. In a point space there will be a point called the **origin**, which has all components 0, but if we are choosing coordinates for a space, it is somewhat arbitrary which point in the space we choose to be the origin.[3]

In our musical point spaces, the points in each point space will represent musical notes, and the vectors will represent intervals between pairs of musical notes. We will generally choose the note middle C to be the origin in the coordinate systems of our point spaces. Note that some of our point spaces

[2] In this book I use three different notations for multiplication. For example, to multiply 2 by x we can write $2x$ or $2 \cdot x$ or $2 \times x$. Sometimes in mathematics different multiplicative notations are used for different types of multiplication, but here we are not defining more than one kind of multiplication for any pair of mathematical objects that can be multiplied together. The first notation is the most compact, but it cannot be used to multiply numbers ($3 \times 2 \neq 32$); the "dot" notation is the next most compact, and is OK as long as there is no danger of confusing the dot with a decimal point; and the traditional "$\times$" is the least compact but most explicit notation. ("$\cdot$" and "$\times$" are also standard notations for different ways of multiplying vectors together, but there is no multiplication of vectors by other vectors in this book, so we can get away with using them to represent numerical and scalar multiplication.)

[3] To put it another way, there is no operation that we are allowed to define on the point space that can actually tell us whether or not a point is the origin.

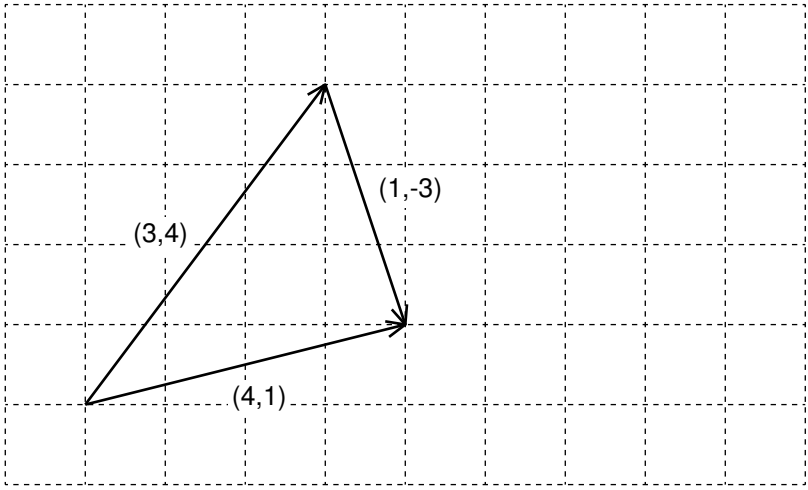

Figure 5.1. Vector addition: $(3, 4) + (1, -3) = (4, 1)$.

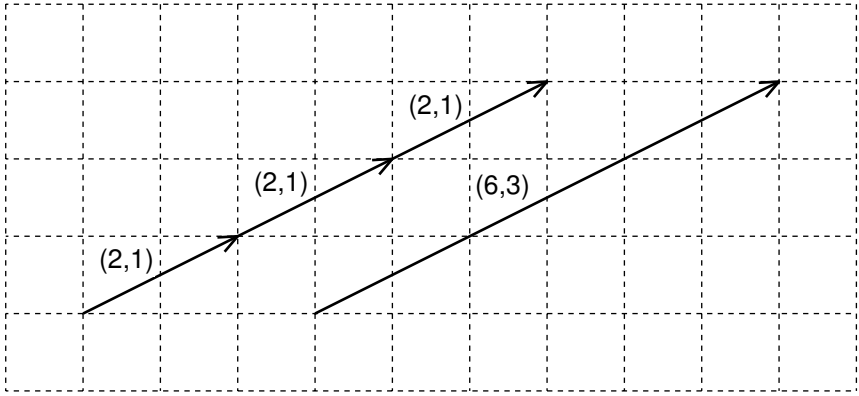

Figure 5.2. Vector scalar multiplication: $3 \times (2, 1) = (6, 3)$.

will have multiple points representing each note, in which case it will be more precise to say that we will choose an origin such that the origin represents middle C (and other points in the point space may also represent middle C).

Having done the theory, we can see what vector spaces and corresponding point spaces we get from our three representations of musical intervals.

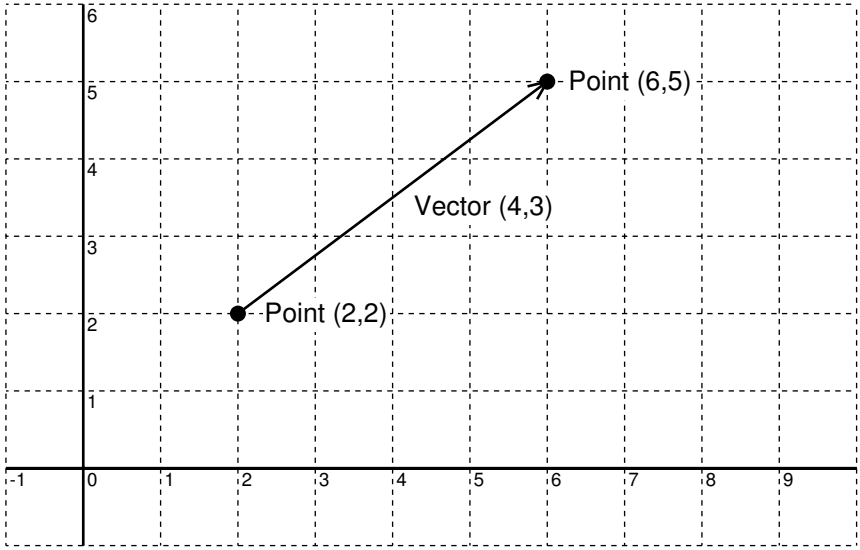

Figure 5.3. Addition of a vector to a point: point (2,2) + vector (4,3) = point (6,5).

5.1.2 1D Semitones Representation

The semitones representation is the simplest. Imagine all the notes of the chromatic scale laid out evenly along a straight line. An interval from one note to another is represented by a vector containing one component, which is the number of semitones it takes to get from the start note to the end note (positive if we are going up, negative if we are going down). For example, middle C is the point (0), the G above it is (7), and the interval from C to G is represented by the vector (7), which means 7 semitones.

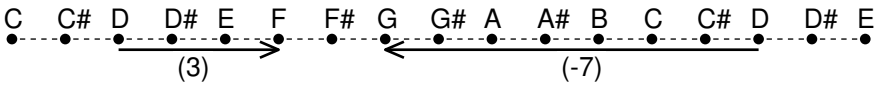

Figure 5.4. 1D semitones representation. Two vectors are shown: a 3 semitones interval (a minor third) going from D up to F, and a −7 semitones interval (a "negative" perfect fifth) going down from D to G.

5.1.3 2D Tones/Semitones Representation

For this representation we take the number of tones in an interval to be the first component, and the number of semitones to be the second component. So for our interval C → G, the vector is $(3, 1)$, which means 3 tones + 1 semitone. We can imagine the point space as a slightly irregular stairway: for each tone step we travel one unit to the right, for each semitone step we travel one unit upwards.

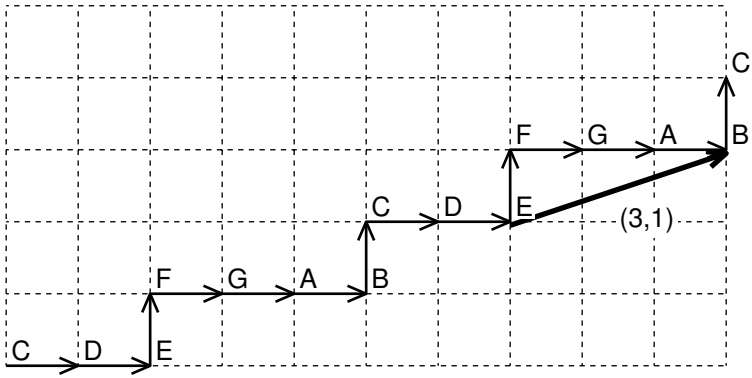

Figure 5.5. 2D tones/semitones representation. The perfect fifth interval from E to B is represented by the vector $(3, 1)$ = 3 tones + 1 semitone.

5.1.4 3D Consonant Interval Representation

When we combine consonant intervals, we have to *multiply* the corresponding fractions. For example, the equation

minor 3rd + major 3rd = perfect fifth

is represented by $6/5 \times 5/4 = 3/2$.

The way to convert multiplication of fractions to addition of vectors is to create vectors where each component is the power of a prime number that occurs in the fraction (as a factor of the numerator or the denominator). As it happens, we assume that the only primes occurring in the relevant fractions are 2, 3 and 5. The three different primes are where our three dimensions come from. For example, $6/5 = 2^1 \times 3^1 \times 5^{-1}$, so we can represent it by the vector $(1, 1, -1)$. Similarly $5/4 = 2^{-2} \times 3^0 \times 5^1$ is represented by $(-2, 0, 1)$, and $3/2 = 2^{-1} \times 3^1 \times 5^0$ is represented by $(-1, 1, 0)$. We can check that the vector addition gives the right answer: $(1, 1, -1) + (-2, 0, 1) = (1 - 2, 1 + 0, -1 + 1) = (-1, 1, 0)$. This corresponds to the multiplication $6/5 \times 5/4 = (2^1 \times 3^1 \times 5^{-1}) \times (2^{-2} \times 3^0 \times 5^1) = (2^1 \times 2^{-2}) \times (3^1 \times 3^0) \times (5^{-1} \times 5^1) = 2^{1+-2} \times 3^{1+0} \times 5^{-1+1} = 2^{-1} \times 3^1 \times 5^0 = 3/2$.

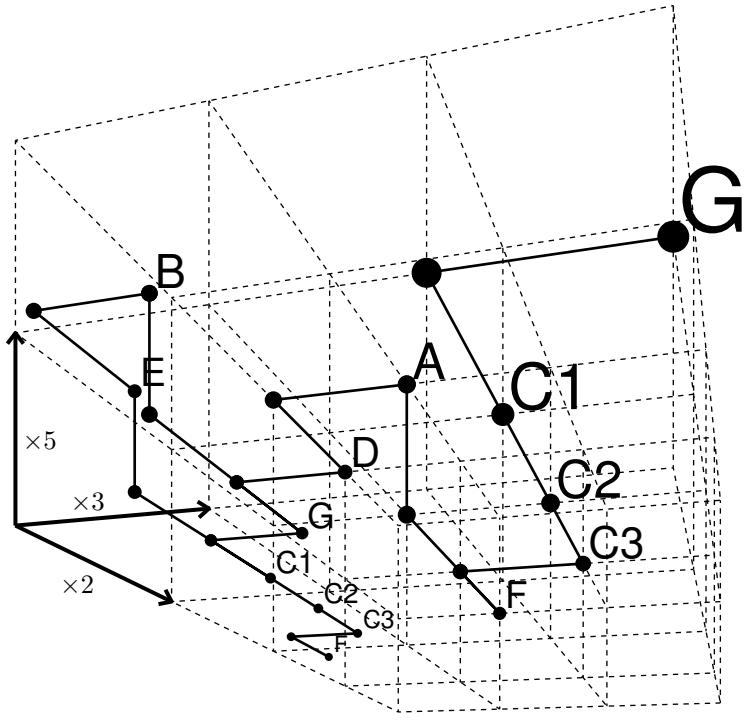

Figure 5.6. 3D musical point space. The C's are labelled according to which octave they are in. The repetition of the pattern is shown for F, C1, C2, C3 and G.

Figure 5.6 attempts to show notes in the 3D point space. Unfortunately 3D visualisation is difficult to do on 2-dimensional paper. One way we can simplify the 3D representation is by removing the least important dimension. The best dimension to remove is the $\times 2$ dimension, i.e. the octaves, because the 3D representation represents harmonic relationships between notes, and as we will see, the brain represents harmonic relationships modulo octaves anyway.

Figure 5.7 shows the notes on the scale in this "flattened" 3D representation. Octaves have been flattened to zero, so the $\times 3$ unit vector is equivalent to a perfect fifth ($\times 3/2$) and the $\times 5$ unit vector is equivalent to a major third ($\times 5/4$).

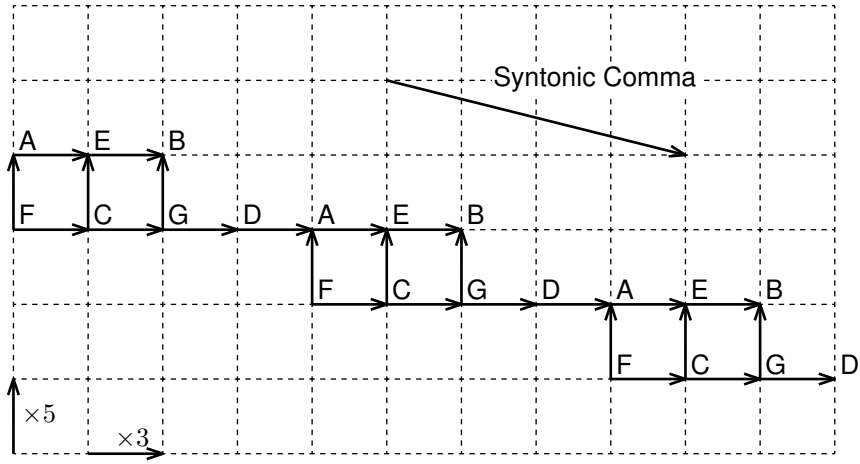

Figure 5.7. "Flattened" 3D musical point space. It is much easier to see the repeated representations of notes in this representation. The period of repetition is the **syntonic comma**, as shown in the diagram, which is discussed in more detail later in this chapter.

5.2 Bases and Linear Mappings

An important fact about an n-dimensional vector space is that we can **generate** all the elements of the vector space from a suitably chosen **basis** (plural **bases**) of n vectors. The required generation is by means of addition and scalar multiplication. For example, in a 2-dimensional vector space, we can choose $(1, 0)$ and $(0, 1)$ as a basis. Then for any vector (x, y), we can generate it as $x(1, 0) + y(0, 1)$. (Calculating: $x(1, 0) + y(0, 1) = (x \times 1, x \times 0) + (y \times 0, y \times 1) = (x \times 1 + y \times 0, x \times 0 + y \times 1) = (x, y)$). Similarly, for a 3-dimensional space we can choose $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ as a basis.

These examples of vector bases are formed by selecting n vectors e_i, $i = 1..n$ where the jth component of e_i is 1 if $i = j$ and 0 otherwise. But there are many other bases that we can choose. In fact, almost any randomly chosen set of n vectors will form a basis for an n-dimensional vector space. The only requirement is that the n vectors must be **independent**, which means that none of the basis vectors can be generated from any or all of the other basis vectors.

Vectors spaces are a type of **mathematical structure**. The structure consists of the vectors in the vector space and the operations of addition and scalar multiplication. Give mathematicians a structure and they will always ask: what **mappings** are there from one space to another that preserve

(or "respect") the structure? A **mapping** (or **function**) is some rule that specifies for each input value from one set called the **domain** a single output value from another set called the **codomain**.

What exactly do mathematicians mean by "preserving" a structure? In the case of a mapping from a vector space to another vector space, what this means for a mapping f is that if we add two input values and apply f to the result, it's the same as if we applied f to the input values first, and then added those two output values together. Writing this is as an equation we get:

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$$

Similarly the mapping must respect scalar multiplication: if we multiply the input value by a scalar and apply f, we get the same result as if we applied f first and then multiplied by the scalar. As an equation this is:

$$f(a \cdot \mathbf{x}) = a \cdot f(\mathbf{x})$$

A mapping with these properties is a **linear mapping**. If we have a mapping from a point space to another point space which generates a corresponding linear mapping between the associated vector spaces, then the point space mapping is called an **affine mapping**. If we already have a linear mapping f between vector spaces associated with a pair of point spaces, then for any pair of points p_i in the input space and p_o in the output space, we can define an affine function g such that $g(p_i) = p_o$, and for any vector $\mathbf{v}_i$ in the input vector space, $g(p_i + \mathbf{v}_i) = g(p_i) + f(\mathbf{v}_i)$. In other words, we can extend the mapping of vector spaces to a corresponding affine mapping of point spaces that is consistent with the vector mapping, and which maps the chosen point in the input space to the chosen point in the output space.

In the case of our musical point and vector spaces, we are only going to be interested in affine mappings that consistently map points representing notes to points representing exactly the same notes. Similarly for vector mappings and intervals, we are interested in vector mappings that map vectors representing intervals to vectors representing the *same* intervals. These mappings are **natural mappings**, where "natural" can be interpreted to mean the most obvious or meaningful.

5.2.1 2D to 1D Natural Mapping

How do we naturally map from the 2D tone/semitones representation of intervals to the 1D semitones representation? This is an easy one: we want

to reduce an interval described as x tones and y semitones to the form of z semitones. We know that 2 semitones = 1 tone, so the answer is $z = 2x + y$.

As mathematicians are never happy with anything that is *too* easy, given a question that was easy to answer, they try to think of a related question that might be a bit harder. A good question for this mapping is: can we *reverse* the mapping? In other words, given the output value, can we determine an input value, and will this input value be unique?

In the first instance it seems obvious that there are many possible input values for each output value. For example, $(2, 0)$, $(1, 2)$ and $(0, 4)$ all map to (4). In other words, an interval of 4 semitones could be 2 tones + 0 semitones, or 1 tone + 2 semitones, or 0 tones + 4 semitones. But, only one of these input vectors represents an actual interval between two notes on the diatonic scale, i.e. $(2, 0)$. So the question that follows is: is there a unique input value for each possible output value if we restrict input values to those vectors that correspond to intervals between notes on the diatonic scale? To give an exhaustive answer to this question, the following tables list all non-negative intervals on the diatonic scale which are not greater than an octave, showing them as tone/semitone vectors and as semitone vectors, grouped by equality of the semitone vector:

Interval	Tone/Semitone	Semitone
C → C	(0, 0)	(0)
D → D	(0, 0)	(0)
E → E	(0, 0)	(0)
F → F	(0, 0)	(0)
G → G	(0, 0)	(0)
A → A	(0, 0)	(0)
B → B	(0, 0)	(0)

Interval	Tone/Semitone	Semitone
E → F	(0, 1)	(1)
B → C	(0, 1)	(1)

Interval	Tone/Semitone	Semitone
C → D	(1, 0)	(2)
D → E	(1, 0)	(2)
F → G	(1, 0)	(2)
G → A	(1, 0)	(2)
A → B	(1, 0)	(2)

Interval	Tone/Semitone	Semitone
D → F	(1, 1)	(3)
E → G	(1, 1)	(3)
A → C	(1, 1)	(3)
B → D	(1, 1)	(3)

Interval	Tone/Semitone	Semitone
C → E	(2, 0)	(4)
F → A	(2, 0)	(4)
G → B	(2, 0)	(4)

Interval	Tone/Semitone	Semitone
C → F	(2, 1)	(5)
D → G	(2, 1)	(5)
E → A	(2, 1)	(5)
G → C	(2, 1)	(5)
A → D	(2, 1)	(5)
B → E	(2, 1)	(5)

Interval	Tone/Semitone	Semitone
F → B	(3, 0)	(6)
B → F	(2, 2)	(6)

Interval	Tone/Semitone	Semitone
C → G	(3, 1)	(7)
D → A	(3, 1)	(7)
E → B	(3, 1)	(7)
F → C	(3, 1)	(7)
G → D	(3, 1)	(7)
A → E	(3, 1)	(7)

Interval	Tone/Semitone	Semitone
E → C	(3, 2)	(8)
A → F	(3, 2)	(8)
B → G	(3, 2)	(8)

Interval	Tone/Semitone	Semitone
C → A	(4, 1)	(9)
D → B	(4, 1)	(9)
F → D	(4, 1)	(9)
G → E	(4, 1)	(9)

Interval	Tone/Semitone	Semitone
D → C	(4, 2)	(10)
E → D	(4, 2)	(10)
G → F	(4, 2)	(10)
A → G	(4, 2)	(10)
B → A	(4, 2)	(10)

Interval	Tone/Semitone	Semitone
C → B	(5, 1)	(11)
F → E	(5, 1)	(11)

Interval	Tone/Semitone	Semitone
C → C	(5, 2)	(12)
D → D	(5, 2)	(12)
E → E	(5, 2)	(12)
F → F	(5, 2)	(12)
G → G	(5, 2)	(12)
A → A	(5, 2)	(12)
B → B	(5, 2)	(12)

And the answer to our question, as to whether we can reverse the mapping from the tone/semitone representation to the semitone representation, is that the mapping can be inverted uniquely for all intervals except 6 semitones. There are two possible input values—$(3, 0)$ and $(2, 2)$—which map to an interval of 6 semitones. The tables above only list interval sizes from 0 semitones to 12 semitones. However, an octave is always represented by $(5, 2)$, and all other possible intervals can be created (from one of the intervals listed) by adding a whole number of octaves to either the start note or the end note. So a full answer is: the mapping can be inverted uniquely for all intervals, except $12n + 6$ semitones (for any integer n). It perhaps should be noted that this "almost" inverse mapping is not at all linear; for example, $f^{-1}((4)) = (2, 0)$ but $f^{-1}((8)) = (3, 2) \neq 2 \times (2, 0)$.[4]

This answer will turn out to help us when we come to define a natural 3D to 2D mapping. We will want to know that the 2D to 1D mapping can be uniquely reversed for all consonant intervals, which is the case, because the only interval for which it cannot be uniquely reversed is 6 semitones, and 6 semitones is not a consonant interval.

Finally, the natural 2D to 1D mapping can be represented by a **matrix**:

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Each row of the matrix represents a component position in the input value, and each column of the matrix represents a component position in the output value. Each number represents the contribution that the corresponding input component makes to the corresponding output component. This matrix has 2 rows and 1 column because our mapping is from a 2D space to a 1D space.

[4]Or rather the reverse mapping is only linear if we restrict it to head-to-tail addition of vectors corresponding to pairs of intervals on the diatonic scale.

5.2.2 3D to 1D Natural Mapping

How do we naturally map from the 3D fractional ratio representation of intervals to the 1D semitones representation?

We determined the 2D to 1D mapping by calculating the value we expected to get for basis vectors $(1, 0)$ and $(0, 1)$. We can do the same thing to construct a 3D to 1D map. $(1, 0, 0)$ represents a ratio of 2, which equals 12 semitones. $(0, 1, 0)$ represents a ratio of 3—more precisely an *approximate* ratio of 3—which equals 19 semitones. And $(0, 0, 1)$ represents a ratio of approximately 5, which equals 28 semitones. So our corresponding matrix is:

$$\begin{pmatrix} 12 \\ 19 \\ 28 \end{pmatrix}$$

5.2.3 3D to 2D Natural Mapping

We can continue to use the same pattern to construct a natural 3D to 2D mapping: specify suitable values for the basis vectors. In keeping with our need to interpret intervals as intervals between actual notes on the scale, we want to choose 2D vectors that represent such intervals. Our chosen basis vectors in the 3D space represent consonant intervals, and we have already seen that all consonant intervals have unique representations as tone/semitone vectors. 12 semitones (an octave) equals 5 tones plus 2 semitones, 19 semitones equals 8 tones plus 3 semitones and 28 semitones equals 12 tones plus 4 semitones. We can write these numbers out as another matrix:

$$\begin{pmatrix} 5 & 2 \\ 8 & 3 \\ 12 & 4 \end{pmatrix}$$

The question remains as to whether this mapping gives correct answers for all consonant intervals. Why is this important? We have proven that the mapping works for our basis vectors. But does this proof automatically extend to *all* consonant intervals? One way to be sure is to check all the possibilities (for consonant intervals less than an octave):

$$\begin{aligned} &3 \text{ semitones} \approx 6/5 = 2^1 \times 3^1 \times 5^{-1} : \\ &\quad f(1, 1, -1) = (1 \cdot \mathbf{5} + 1 \cdot \mathbf{8} - 1 \cdot \mathbf{12}, 1 \cdot \mathbf{2} + 1 \cdot \mathbf{3} - 1 \cdot \mathbf{4}) \\ &\qquad\qquad = (5 + 8 - 12, 2 + 3 - 4) \\ &\qquad\qquad = (1, 1) \end{aligned}$$

$$
\begin{aligned}
&4 \text{ semitones} \approx 5/4 = 2^{-2} \times 3^0 \times 5^1 : \\
&\quad f(-2,0,1) = (-2 \cdot \mathbf{5} + 0 \cdot \mathbf{8} + 1 \cdot \mathbf{12}, -2 \cdot \mathbf{2} + 0 \cdot \mathbf{3} + 1 \cdot \mathbf{4}) \\
&\qquad\quad = (-10 + 0 + 12, -4 + 0 + 4) \\
&\qquad\quad = (2,0) \\
&5 \text{ semitones} \approx 4/3 = 2^2 \times 3^{-1} \times 5^0 : \\
&\quad f(2,-1,0) = (2 \cdot \mathbf{5} - 1 \cdot \mathbf{8} + 0 \cdot \mathbf{12}, 2 \cdot \mathbf{2} - 1 \cdot \mathbf{3} + 0 \cdot \mathbf{4}) \\
&\qquad\quad = (10 - 8 + 0, 4 - 3 + 0) \\
&\qquad\quad = (2,1) \\
&7 \text{ semitones} \approx 3/2 = 2^{-1} \times 3^1 \times 5^0 : \\
&\quad f(-1,1,0) = (-1 \cdot \mathbf{5} + 1 \cdot \mathbf{8} + 0 \cdot \mathbf{12}, -1 \cdot \mathbf{2} + 1 \cdot \mathbf{3} + 0 \cdot \mathbf{4}) \\
&\qquad\quad = (-5 + 8 + 0, -2 + 3 + 0) \\
&\qquad\quad = (3,1) \\
&8 \text{ semitones} \approx 8/5 = 2^3 \times 3^0 \times 5^{-1} : \\
&\quad f(3,0,-1) = (3 \cdot \mathbf{5} + 0 \cdot \mathbf{8} - 1 \cdot \mathbf{12}, 3 \cdot \mathbf{2} + 0 \cdot \mathbf{3} - 1 \cdot \mathbf{4}) \\
&\qquad\quad = (15 + 0 - 12, 6 + 0 + -4) \\
&\qquad\quad = (3,2) \\
&9 \text{ semitones} \approx 5/3 = 2^0 \times 3^{-1} \times 5^1 : \\
&\quad f(0,-1,1) = (0 \cdot \mathbf{5} - 1 \cdot \mathbf{8} + 1 \cdot \mathbf{12}, 0 \cdot \mathbf{2} - 1 \cdot \mathbf{3} + 1 \cdot \mathbf{4}) \\
&\qquad\quad = (0 - 8 + 12, 0 - 3 + 4) \\
&\qquad\quad = (4,1)
\end{aligned}
$$

These all give the right answer. Alternatively, we could have realised that all other consonant intervals can be constructed from our basis vectors by doing **head-to-tail**[5] addition of vectors between points on the scale, so the answers would have had to come out right anyway (because we can simultaneously perform the head-to-tail additions in the 3D space and 2D space, so the answers always have to match).

5.2.4 Images and Kernels

If we tell a mathematician that we have a linear mapping from one vector space to another, there are two questions that he or she is likely to ask us about our mapping:

- What is the **image** of the mapping?

[5] "Head-to-tail" refers to adding vectors by directly adding them represented as displacements from one point to another. For example, if $p_1 = p_0 + \mathbf{x}_{01}$, $p_2 = p_1 + \mathbf{x}_{12}$ and $p_2 = p_0 + \mathbf{x}_{02}$, it follows that $\mathbf{x}_{02} = \mathbf{x}_{01} + \mathbf{x}_{12}$. p_1 is both the "head" of $\mathbf{x}_{01}$ and the "tail" of $\mathbf{x}_{12}$. See also Figure 5.1.

- What is the **kernel** of the mapping?

The image of a mapping is the set of all output values for the mapping (or function). If a point y is in the image of a function f, then there must be some value x in the domain such that $f(x) = y$.

The set of output values for our natural 3D to 2D mapping is indeed the whole of the 2D tone/semitone vector space. To show this is true it is enough to choose a basis for the output space, and show that all the basis vectors are in the image.

For example, we can choose $(-3, 2, 0)$ as an input value that maps to $(1, 0)$, i.e. 1 tone, and $(4, -1, -1)$ as an input value that maps to $(0, 1)$, i.e. 1 semitone, as checked by the following calculations:

$$\begin{aligned} f(-3,2,0) &= (-3 \cdot \mathbf{5} + 2 \cdot \mathbf{8} + 0 \cdot \mathbf{12}, -3 \cdot \mathbf{2} + 2 \cdot \mathbf{3} + 0 \cdot \mathbf{4}) \\ &= (-15 + 16 + 0, -6 + 6 + 0) \\ &= (1, 0) \end{aligned}$$

$$\begin{aligned} f(4,-1,-1) &= (4 \cdot \mathbf{5} - 1 \cdot \mathbf{8} - 1 \cdot \mathbf{12}, 4 \cdot \mathbf{2} - 1 \cdot \mathbf{3} - 1 \cdot \mathbf{4}) \\ &= (20 - 8 - 12, 8 - 3 - 4) \\ &= (0, 1) \end{aligned}$$

(These choices correspond to asserting that a tone is equal to a ratio of $9/8$ and a semitone is equal to $16/15$, but these ratios are not the only possible choices. Two alternative choices are tone $= 10/9$ and semitone $= 27/25$. We will analyse the mathematics behind this non-uniqueness shortly.)

Having found input vectors that map to all the basis vectors, we can use them to construct input vectors that map to any output vector according to the construction of that output vector from the basis vectors. For any vector (n, m) in the 2D space, the vector $n(-3, 2, 0) + m(4, -1, -1) = (-3n + 4m, 2n - m, -m)$ is mapped by f onto (n, m).

We can conclude that our 3D to 2D mapping is an **onto** mapping, which means that its image is the whole of the codomain.

We can find an input value for any chosen output value, but this choice of input value is not unique. Uniqueness of input values for output values is exactly what the concept of **kernel** is about. The domain of our natural mapping has 3 dimensions, and the image has 2 dimensions. The theory of vector spaces tells us that we can do some simple arithmetic on these dimensions:

$$3 - 2 = 1$$

to conclude that our mapping has a kernel of dimension 1. So what is the kernel? It is the **subspace**[6] of the domain which is reduced to zero by the mapping. It's the 1 dimension that we "lose" as we go from 3 dimensions to 2 dimensions. The kernel is a measure of the non-uniqueness of input values for a given output value. If two input values differ by an element of the kernel, then they will map to the same output value. Conversely, if two input values map to the same output value, their difference must belong to the kernel.

The kernel of our natural 3D to 2D mapping has 1 dimension, so we must be able to find a basis for this subspace consisting of a single unit vector such that all elements of the kernel are multiples of that unit vector. If the unit vector is (x, y, z), then x, y and z must satisfy the following equations:

$$\begin{array}{rcrcrcl} 5x & + & 8y & + & 12z & = & 0 \\ 2x & + & 3y & + & 4z & = & 0 \end{array}$$

One solution is: $x = -4, y = 4, z = -1$, i.e. the vector $(-4, 4, -1)$ maps to zero. And to check:

$$\begin{aligned} f(-4,4,-1) &= (-4 \cdot \mathbf{5} + 4 \cdot \mathbf{8} - 1 \cdot \mathbf{12}, -4 \cdot \mathbf{2} + 4 \cdot \mathbf{3} - 1 \cdot \mathbf{4}) \\ &= (-20 + 32 - 12, -8 + 12 - 4) \\ &= (0, 0) \end{aligned}$$

The most general solution is $x = -4t$, $y = 4t$ and $z = -t$ for arbitrary t. So if $\mathbf{y}$ is an output value, and $\mathbf{x}$ is an input value such that $f(\mathbf{x}) = \mathbf{y}$, then $f(\mathbf{x} + t(-4, 4, 1)) = \mathbf{y}$ for any number t. However, if we restrict ourselves to 3D vectors that can be constructed by adding together 3D representations of intervals between notes on the diatonic scale, the components of $\mathbf{x}+t(-4, 4, 1)$ will always be integers, and if the components of $\mathbf{x}$ are all integers, the only way this can happen is if t is an integer. (Because if t is not an integer, then the last component of $t(-4, 4, 1) = (-4t, 4t, t)$ will not be an integer.)

$(-4, 4, -1)$ corresponds to the fractional ratio $2^{-4} \times 3^4 \times 5^{-1} = 81/80$. The zero vector $(0, 0, 0)$, which also maps to zero, corresponds to $2^0 \times 3^0 \times 4^0 = 1$. So the statement that $(-4, 4, -1)$ generates the kernel of the natural 3D to 2D mapping in effect tells us that 81/80 represents an interval of 0 semitones, and that in some sense 81/80 is the same as 1. Now 81/80 could perhaps be considered close to 1, but it's definitely not equal to 1. In fact if we listen to two notes whose frequencies have a ratio of 80 to 81, we will be able to tell the difference, as they will separated by about 22% of a semitone.

[6] A **subspace** of a vector space is a vector space consisting of a subset of the vectors in the original vector space with the same operations of addition and scalar multiplication defined. Any sum of two vectors in the subspace must also be in the subspace, and any scalar multiple of a vector in the subspace must be in the subspace.

The value of 81/80 is an example of what is called a **comma**; in particular it is called the **syntonic comma**, **Ptolemaic comma** or **comma of Didymus**. In general a comma is something that happens when we define musical scales according to rules that require one note (or more than one note) to be in two places at once. The comma is the interval between the two different places where the note wants to be.

We can summarise all this analysis of the relationships between consonant intervals and intervals on the diatonic scale as follows:

> The kernel of the natural linear mapping from the 3D representation of music intervals, as natural ratios based on powers of 2, 3 and 5, to the 2D representation of musical intervals, as sums of tones and semitones, is the vector space generated by the vector representing the ratio of the syntonic comma, which is equal to 81/80.

5.2.5 Visualising the Syntonic Comma

There are many sequences of intervals (between notes on the diatonic scale) that we can follow to realise the syntonic comma, and "prove" that $81 = 80$. One example is (with all steps going up in frequency):

$$\begin{array}{ccccccccccc} C & \to & G & \to & D & \to & A & \to & C \\ & 3/2 & \times & 3/2 & \times & 3/2 & \times & 6/5 & \end{array}$$

versus

$$\begin{array}{ccccc} C & \to & C & \to & C \\ & 2 & \times & 2 & \end{array}$$

By equating these two paths, which both start at a C and arrive at a C two octaves higher, we get $162/40 \approx 4$, which reduces to $81/80 \approx 1$.

But this is not the only path. For example, we could replace the first path with:

$$\begin{array}{ccccccccccc} C & \to & E & \to & B & \to & D & \to & F & \to & C \\ & 5/4 & \times & 3/2 & \times & 6/5 & \times & 6/5 & \times & 3/2 & \end{array}$$

Reconciling with the second path above this tells us $1620/400 \approx 4$, again reducing to $81/80 \approx 1$.

To find the full set of such paths that can tell us $81/80 \approx 1$, we need to start by categorising all pairs of notes whose frequencies are related to each other by simple fractional ratios.

We have already noted that all consonant ratios between notes on the diatonic scale come from intervals of 0, 3, 4, 5, 7, 8, 9 semitones, or from one of those values with a multiple of 12 semitones (i.e. 1 octave) added on. We also noted that the consonances of x semitones and $12 - x$ semitones are directly related. For example, A → C is 3 semitones which is the approximate

ratio 6/5, and C → A is $12 - 3 = 9$ semitones which is the approximate ratio $2 \div 6/5 = 5/3$. Independently of the operations of adding octaves to intervals and subtracting intervals from octaves, there are only 3 distinct consonant intervals. We can choose one representative from each of the pairs 3 and 9, 4 and 8, and 5 and 7.

It will turn out to be simpler to choose 3 (minor third), 4 (major third) and 7 (perfect fifth). The reasons for this are as follows:

- Every interval of 3 or 4 semitones is 2 steps on the diatonic scale, and conversely, every interval between notes separated by 2 steps on the diatonic scale is either 3 or 4 semitones.

- Every interval of 7 semitones is 4 steps on the diatonic scale, which in all cases divides up into parts of 2 steps plus 2 steps, where one part is 3 semitones and the other is 4 semitones.

- Every interval between a pair of notes on the diatonic scale separated by 4 steps is either an interval of 7 semitones that divides up into intervals of 3 and 4 semitones as just stated, or it is the dissonant interval of 6 semitones that divides up into two portions of 3 semitones each.

It follows that we can visualise all consonant intervals and the relationships between them by stepping along the white notes scale two notes at a time, i.e. C, E, G, B, D, F, A, C. Every consonant interval is represented by either one step on this "double-stepped scale" (i.e. 2 steps on the white note scale), or by two steps (i.e. 4 steps on the white note scale). And for every case where three notes are related pairwise to each other by consonant intervals, the three notes will be found arranged in consecutive order on the double-stepped scale.

The double-stepped scale and the relationships between notes are shown in the following diagram:

$$\begin{array}{ccccccccccccccc}
C & \Rightarrow & E & \rightarrow & G & \Rightarrow & B & \rightarrow & D & \rightarrow & F & \Rightarrow & A & \rightarrow & C \\
C & & \longrightarrow & & G & & \longrightarrow & & D & & \longrightarrow & & A & \rightarrow & C \\
C & \Rightarrow & E & & \longrightarrow & & B & & & & F & & \longrightarrow & & C
\end{array}$$

(Key: $\rightarrow$ = minor third = 6/5,
$\Rightarrow$ = major third = 5/4,
$\longrightarrow$ = perfect fifth = 3/2)

All possible paths from C to C that "travel along" the syntonic comma can be stepped along these lines. Starting on the left, step a minor third, a major third or a perfect fifth to the right each time, switching lines as necessary. Notice that there is no arrow from B to F on the third line—this is the dissonant 6 semitones interval.

5.3 The Harmonic Heptagon

If we take the double-stepped scale, and close it up into a loop, we get a diagram that I call the **Harmonic Heptagon**,[7] as shown in Figure 5.8.

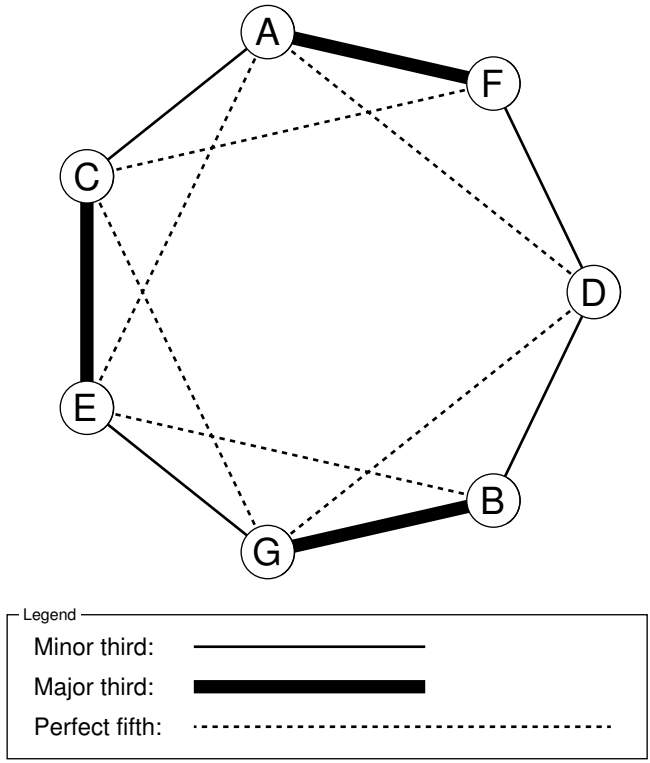

Figure 5.8. The Harmonic Heptagon

It is very easy to explain the syntonic comma in terms of the Harmonic Heptagon:

> Every time we travel once around the Harmonic Heptagon, we travel a distance in 3D space corresponding to the syntonic comma. And every time we travel a distance in 3D space corresponding to the syntonic comma, we must have travelled once around the Harmonic Heptagon.

[7] Those familiar with music theory might know about the **Circle of Fifths**. The Harmonic Heptagon looks a bit similar, but it is a different diagram. The Circle of Fifths contains all the notes (both black and white), and it only shows connections representing a perfect fifth (7 semitones $\approx 3/2$). The Harmonic Heptagon contains only the white notes, but it shows connections for all consonant intervals between those notes.

There are a few other things we might note in passing about this heptagon:

- Every note belongs to one minor third interval and one major third interval, except for D, which belongs to two minor third intervals.
- Every note belongs to two perfect fifth intervals, except for B and F which both belong to one perfect fifth interval and one dissonant 6 semitones interval.
- Every major and minor chord that can be played on the white notes consists of 3 consecutive notes going around the heptagon. These are: C major (CEG), E minor (EGB), G major (GBD), D minor (DFA), F major (FAC) and A minor (ACE). The only such triad that is not a chord is (BDF) which is a dissonant combination of notes, although this group of notes does appear as part of the G7 chord (GBDF).
- Other common chords form sequences of 4 consecutive notes on the heptagon, including G7 (GBDF), Dmin7 (DFAC), Fmaj7 (FACE), Amin7 (ACEG), Cmaj7 (CEGB) and Emin7 (EGBD).
- There is a reflective symmetry between D,F,A,C and D,B,G,E.
- Alternative home chords C major (CEG) and A minor (ACE) are opposite the D, and are symmetrical with respect to the symmetry mentioned in the previous point.

The significance of some of these points will be explained when I develop aspects of the theory in later chapters.

Chapter 6

The Brain

The human brain is an information processing system, which can be analysed in terms of input, output, calculation and storage. At this level of abstraction the brain is like a computer. The smallest information processing components in the brain are the **neurons**. Each individual neuron can be considered to be an information processing system, with its own input, output, calculation and storage.

In between the whole brain considered as an information processing system, and individual neurons considered as information processing systems, it is possible to some extent to identify subsystems in the brain (variously known as **maps**, **functional maps** or **modules**), consisting of groups of neurons that perform a particular information processing function.

The fundamental problem of brain research is to determine how and where meaning is represented in the brain.

6.1 An Information Processing System

To understand the brain it is easiest to see it as being part of the **nervous system**. Taken as a whole, the nervous system and brain constitute a very sophisticated **information processing system**.

The functions of any information processing system can be divided roughly into four components:

- Input of information from external sources.

- Output of information to external destinations.
- Calculation: using available information to create new information.
- Storage of information, so that it can be retrieved and used again at some later time. Some information processing systems do not have any storage. Such systems can be described as **stateless**, because they do not have any **state** that represents information stored in the system. Other information processing systems have a very limited amount of state.[1]

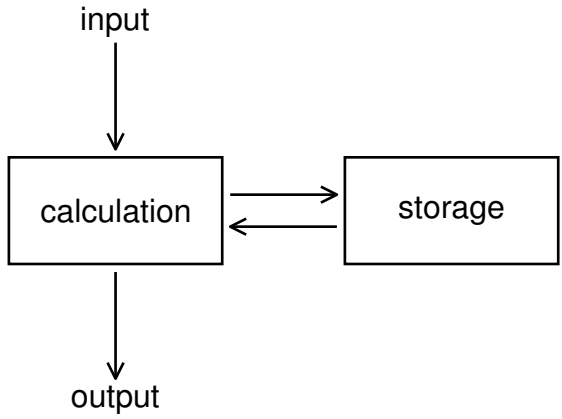

Figure 6.1. Basic components of an information processing system.

6.1.1 Analogy with Computers

Is the brain like an electronic computer? The best answer to this question is "yes" at an abstract level, but "no" when you look at the particulars. An electronic computer is certainly an information processing system. We can identify aspects of its functionality according to the list above. For example, considering the personal computer I am using to write this book, its information inputs include the keyboard, my Internet connection (when receiving data), the mouse, the microphone, and the scanner. Its information outputs

[1] Any type of calculation other than simply passing the input to the output requires temporary storage of the current state of the calculation, so the concepts of calculation and storage cannot be completely separated from each other. The notion of a **Turing Machine**, devised by Alan Turing to describe the capabilities of any physically plausible information processing system, divides storage into a **state machine** allowing a finite number of states, and an infinitely long **tape** which is read, written and moved along according to the operation of the state machine.

include the monitor, my Internet connection (when sending data), the speakers, and the printer. The computer performs calculations on the information it has available to it to create new information, and it can store information, either temporarily in main memory, or more permanently on the hard disk.

There are some ways that the fine details of the nervous system look like a computer: individual components are connected to each other through connections that look a bit like wires, and electrical signals of a sort travel through these connections. But we will also see many ways that the human brain is not at all like a computer. Computer components and connections between components are almost always driven by regular clock signals, whereas no such thing exists in the brain. Computers are very fragile: a failure in even one tiny hardware component can render the whole system unusable. Brains tend to be more robust than that.

Arguing about whether we can prove that the brain is or isn't like a computer is not a useful end in itself, but the analogies between the two are often illuminating. Some information processing tasks can be better performed by electronic computers, and others are better performed by the brain. Understanding the reasons for these differences in performance can help us understand why certain things in the brain happen the way they do.

6.2 The Neuron

The fundamental information processing component of the brain and nervous system appears to be the **neuron**, which is a particular type of cell[2] found in the brain and nervous system. I say "appears to be" because there is enough mystery and uncertainty about how the brain works that some scientists believe there must be more to it than just neurons and the connections between them.

Informally people often talk about "brain cells" as being the cells in our brain that do the thinking, but neurons are not the only brain cells. Other types of cell found in the brain include the **glial cells**, which are in fact more numerous than neurons. The evidence is that glial cells play a supporting role, which includes controlling ionic concentrations around neurons and recycling neurotransmitters released from synapses.

The **neuron doctrine** says that neurons are the fundamental information processing components of the brain and nervous system, and that the flow of information through the nervous system occurs via the physical connections between neurons. This is a "doctrine" in the sense of a useful working as-

[2] **Cells** are the basic components of all living things. Some living things, like germs, consist of only one cell. Other organisms (including us) are **multi-cellular**. Almost all cells are created by one cell splitting up into two cells (the main exception being that sometimes cells merge, like the sperm and the egg at the moment of conception). Different body tissues are formed from conglomerations of different types of cells.

sumption.[3] It is a working assumption accepted by most but not all working neuroscientists. (**Neuroscience** is the study of brains and nervous systems.)

So what do neurons look like, and how are they connected to each other? A neuron consists of a **soma**, which is its central cell body, and an **axon** and **dendrites**. The axon and dendrites are thin branching tubes that form tree-like structures coming out of the soma.

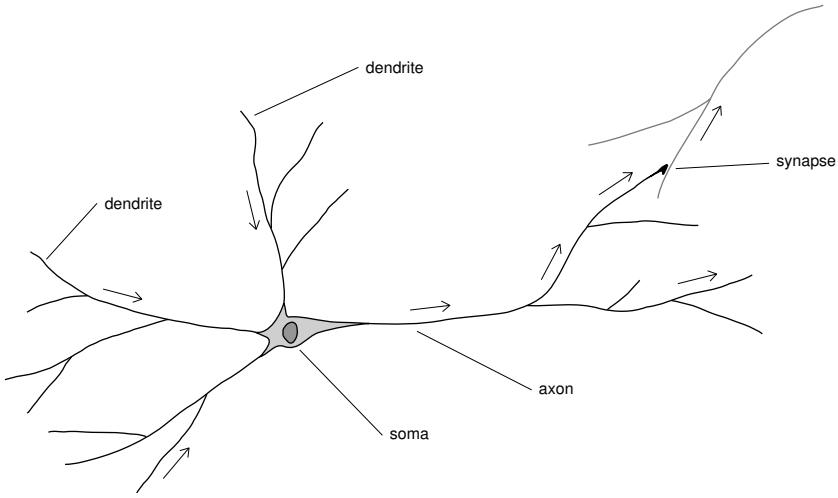

Figure 6.2. A simplified picture of a neuron, including a synaptic connection to another neuron. The arrows show the direction of the flow of information through the neurons.

The primary activity of a neuron is the generation and propagation of **action potentials** that start from the soma and propagate along the axon. The action potential is the signal that neurons use to communicate with each other. It is a type of electrical signal, but it is not a current flowing through a wire as in a computer: it is a complex transfer of sodium and potassium ions between the outside and inside of the axon. The ion transfer becomes self-propagating once initiated from the **axon hillock**, which is the point of the axon where it starts on the soma.

The branches of the axon are called **axon collaterals**. These axon branches have well defined end-points called **terminal boutons**. The boutons form connections to other neurons (and occasionally back to the same neuron). These connections are known as **synapses**.

[3]Other famous scientific doctrines include the **cell doctrine**, which says that living organisms are completely constructed from cells, and the **central dogma of molecular biology**, which says that DNA encodes for RNA which encodes for protein. These doctrines have turned out to have various caveats and exceptions, but they nevertheless continue to provide the major framework for understanding the phenomena that they describe.

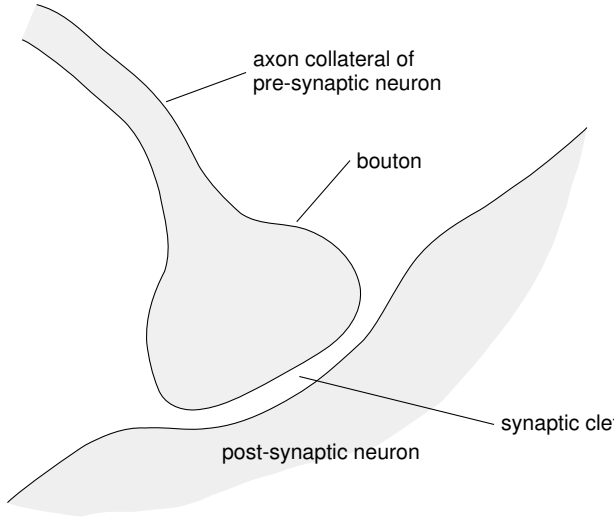

Figure 6.3. A neural synapse.

When referring to a particular synapse, the neuron that sends information into the synapse is the **pre-synaptic** neuron, and the target neuron that receives information from the synapse is the **post-synaptic** neuron.

In effect the synapse is a point of attachment, but there is actually a gap between the bouton and the post-synaptic neuron which is called the **synaptic cleft**. When an action potential arrives at a bouton, it is not transmitted as such to the post-synaptic neuron. Rather certain chemicals called **neurotransmitters** are released across the synaptic cleft. Different synapses release different types of neurotransmitter. Common neurotransmitters include **glutamate**, **GABA** (**gamma-aminobutyric acid**), **norepinephrine**, **dopamine** and **serotonin**.

On the other side of the synaptic cleft, on the post-synaptic neuron, are the **receptors**, which receive the neurotransmitter. Different neurotransmitters have different effects on the neuron that they target, but the intention is the same in all cases: the release of neurotransmitters at a synapse affects the tendency of the target neuron to propagate an action potential on its axon.

A neuron is said to **fire** when an action potential is propagated. Action potentials are an all-or-nothing affair: once one starts it continues until it reaches the terminal boutons of the axon. An action potential propagates in a manner somewhat analogous to the burning of a fuse, in that there is an advancing front of activity (which consists of an exchange of sodium and potassium ions), such that the activity at one point initiates activity at neighbouring points that have not yet been activated. A major difference between

action potentials and burning fuses is that the axon is not permanently used up by the action potential: there is a gradual recharging process that makes it ready to propagate a new action potential on the next occasion.

Some neurotransmitters transmitted via a synapse make the target neuron more likely to fire; these are called **excitatory** neurotransmitters. Other **inhibitory** neurotransmitters make the target neuron less likely to fire. The terms "excitatory" and "inhibitory" are used to describe both the neurotransmitters and the synapses that transmit them.[4] The effects of neurotransmitters also depend on the locations of synapses: synapses on the central soma have a more immediate effect than synapses on distant branches of the dendrites, and some synapses on the dendrites act only to cancel out the effects of synapses further away from the soma on the same dendrite. Another thing that alters the tendency of a neuron to fire is how long it was since the last time it fired. As already stated, there is a recharging system, and the more time this has had to act since a previous firing, the more readily the neuron will fire again.

There is considerable complexity in the workings of each neuron, and scientists do not yet understand everything that goes on in individual neurons. As well as neurotransmitters transmitted across synaptic clefts, there are other neurotransmitters that leak somewhat into the surrounding medium, and act as broadcast messages that can be delivered to multiple neurons. There is also so-called **retrograde transmission** of **nitric oxide** (**NO**) from the post-synaptic neuron back to the pre-synaptic neuron that activated it. Some type of retrograde transmission of information is needed if neurons are to provide feedback about the value of information received to the neurons that sent them the information—this may be the function that NO transmission performs.

The description of a neuron I have given here illustrates the basic concept of the neuron as an information processing component. In fact we can readily identify three out of the four information processing functions:

1. The inputs of the neuron are the neurotransmitters received by receptors on its dendrites.

2. The outputs of the neuron are the action potentials propagated along its axon.

3. The calculation performed by the neuron is determined by the effect that input signals have on its tendency to fire.

How a neuron stores information is not so obvious. In the first instance, information is stored temporarily according to the neuron's firing state: whether or not it is currently firing, or if it is not firing, how much excitation would

[4] And the neurons, because, as it happens, many neurons primarily release one particular neurotransmitter across the synapses that they form with other neurons.

be required to make it fire. Secondly, information may be stored by changes to the long-term state of the neuron, which will mostly consist of:

- Changes in the strength of synaptic connections, i.e. how much effect an incoming signal has on the target neuron.
- Growth and formation of new connections between neurons, and the disappearance of existing connections.

6.2.1 Comparison to Computer Components

Circuits in computer components such as CPUs (central processing units) primarily process information in the form of currents flowing through wires, or voltages between pairs of points. In either case, there are generally only two states: either there is a current or voltage, or there isn't. Mathematically, these two states can be understood to represent the numbers 1 (for "on") and 0 (for "off"). Most computer circuits are driven by a regular clock signal. Thus the value of a current or voltage is determined for each interval between two clock ticks.

The smallest units of functionality within computer circuits are **logic gates** and **flip-flops**. These normally have only one or two inputs and one or two outputs. Logic gates have output values that are an immediate function of their input values.[5] Flip-flops have their output values in each clock cycle determined by their input values in the previous cycle.[6] For example, a logic gate with two inputs and one output might determine its output value according to the following logic table:

Input 1	Input 2	Output
0	0	0
1	0	1
0	1	0
1	1	0

To state this table in a sentence, the output is on (i.e. equal to 1) only if input 1 is on (i.e. equal to 1) and input 2 is off (i.e. equal to 0). If we want to use our neural terminology, we could say that input 1 is an excitatory input and input 2 is an inhibitory input. Also we note that the inhibitory effect of input 2 overrides the excitatory effect of input 1. So we can see some resemblance between the operation of a neuron and the operation of a logic gate in a computer circuit. We can even identify inputs as being either inhibitory or excitatory.

[5] There is necessarily some delay, and circuits must be designed so that any accumulated delays do not extend from the beginning of a clock cycle into the next clock cycle.

[6] Most types of flip-flop have their output as an implicit input, so that when a clock tick occurs, the values of the other inputs determine whether or not the current output "flip-flops" to the opposite value.

But a major difference with a neuron is that it is not controlled by clock cycles. The input signals and output signals in neural circuits are discrete events that can happen at any time. We will see that this has implications for understanding and comparing the **representation of meaning** in computers and in the brain. A set of electronic components in a computer can have one set of meanings for one clock cycle, and then have a completely different set of meanings in the next clock cycle. The lack of such a precise and global control of time periods in the brain means that the meanings represented by signals in neurons must be fairly independent of time (at least in the short term: processes of learning and cortical plasticity can cause meaning to change in the longer term).[7]

6.2.2 How Many Connections?

Another big difference between computers and brains is the number of connections between components. Neurons don't have one or two input and one or two output connections; they have *thousands* of connections to other neurons. The average is about 10,000 inputs and 10,000 outputs. Some neurons have more than 100,000 connections to other neurons.

There are about 100,000,000,000 (one hundred thousand million) neurons in the human brain. You can do the arithmetic, and see that this means there are about 1,000,000,000,000,000 synapses. (This number is so big that it has no common name, so we can just call it a thousand million million.) In some ways it might be more realistic to compare synapses (rather than whole neurons) to the individual components that occur in computer circuits.

We can compare the human brain to a personal computer, comparing numbers of components, numbers of connections and speed of operation:[8]

- 1,000,000,000,000,000 synapses in the brain compares to 100,000,000 transistors in a modern CPU, maybe 8,000,000,000 bits stored in RAM, and 1,000,000,000,000 bits stored on a typical hard disk.
- Individual components in computers do things much faster than anything in the brain: a 2GHz CPU is performing 2,000,000,000 operations per second. Very few neurons fire more than 1000 times a second, and most fire less than 100 times a second.
- Computers are terrible at making full and continuous use of their circuitry: your computer's RAM might have 8,000,000,000 bits, and operate at 500MHz, but you will be lucky if more than 128 bits of that memory are in use at any one time.

[7]There can be meaning in the actual timings of action potentials—this is **temporal coding** which is explained later in this chapter. The concept of temporal coding is distinct from the concept of the meaning of action potentials changing over time.

[8]The figures given are all very approximate, and the values for computers change as computer technology develops.

A lot of design effort has gone into making the CPU do at least a few things in parallel, but it only contains a small fraction of the overall number of components in the system.[9] Neurons in your brain do not have to wait for some central authority to tell them to do something—each neuron reacts directly at all times to the inputs of the neurons immediately connected to it.

- A computer's hard disk retains information even when the power is turned off. Compared to RAM, hard disks are usually larger but slower, and the processing bottleneck is even more extreme: a typical hard disk might store 1,000,000,000,000 bits of information, stored on several **platters**, with two heads per platter, each head capable of transferring just *one* bit at a time at a rate of perhaps 100,000,000 bits per second.

These differences are revealed in the different abilities of human and computer information processing systems: all the different parts of your brain can operate simultaneously to calculate the relevant consequences of information made available to it, whereas a computer has to work its way through all the potential deductions and conclusions sequentially. On the other hand, if there is a need to multiply a million numbers together—and be sure of getting exactly the right answer—the computer is going to finish the job a whole lot quicker.

6.3 Modularity in the Brain

We can identify the four information processing components—input, output, calculation and storage—for the brain and nervous system as a whole:

- Information is input from sensory cells, also called **receptors**. There are sensory cells that supply the input for the traditional five senses, and also for various internal senses:
  - Sight: neurons in the retina that respond to light. There are four types of retinal receptors: three colour-sensitive types and one "black and white" receptor type for low light conditions.
  - Hearing: cells that receive sound information. These are the hair cells in the **organ of Corti**, which have already been mentioned in the previous chapter.
  - Taste: receptors in the tongue for sweetness, saltiness, sourness, bitterness and "umami".
  - Smell: olfactory receptors in the nose.

[9] This problem of not being able to use more than a small portion of the computer's circuitry at any one time is called the **Von Neumann bottleneck**, named after John von Neumann, a famous physicist, mathematician and computer scientist.

  - Touch: various receptors in the skin that detect pressure and temperature.
  - Internal senses include receptors for balance, position and motion of various parts of your body, and other receptors that provide information about the internal state of bodily organs.

- The major output of information from the nervous system is via the **lower motor neurons**. Each motor neuron activates a single muscle fibre. There are two types of lower motor neuron: **alpha motor neurons** activate **extrafusal** muscle fibres which do the real work, and **gamma motor neurons** activate **intrafusal** muscle fibres which play a role in managing feedback to the nervous system about the contracted state of muscles. Other outputs occur via the **autonomic nervous system** which controls such things as heart rate, blood pressure, digestion and the release of various hormones.

- The brain stores information: this is what we call "learning" and "memory".

- The brain calculates: this includes all the processes of perception, where raw sensory information is translated into knowledge and understanding of things in the external world and within ourselves, and the processes of decision-making, which eventually result in us making or controlling muscle movements required to carry out those decisions.

We can analyse the neuron as an information processing system, and we can analyse the whole brain as an information processing system. In both cases we can identify the four components of input, output, calculation and storage. Are there any in-between levels of organisation and functionality that we can analyse?

When we look at an electronic computer system, we can see that it consists of various circuit boards plugged together, and each circuit board consists of integrated chips and other electronic components that have been soldered onto the board and connected by etched connections on the board. There is a lot of modularity in how computer systems are constructed. This partly has to do with the economics of design and manufacture: it is easier to design systems constructed from general purpose components that have already been designed, and it is easier to make profits from manufacturing general purpose components because they can be used in many different systems.

The "economics" of the design and manufacture of the human brain and nervous system is a bit different from that of electronic computers. The "design" has resulted from an accumulation of incremental mutations over millions of years of evolution. The "manufacture" is the process of conception, growth and development. These processes of natural design and manufacture may result in a form of biological modularity, but it is not clear if it is a

form of modularity that it going to help us analyse the brain into functional components.

When man-made information processing artefacts are made from components, the components are generally manufactured separately, and then attached to each other by various means to make the final product. It is often easy to pull such an artefact apart into its separate components, especially if we are armed with a screwdriver, or perhaps with a soldering iron that lets us remove components from a circuit board. If a component is general purpose, then it will have a well-defined functionality independent of its role in that particular artefact, and it will be easy to understand that functionality by analysing the design of that individual component.

The "components" of the body of a living organism have to grow and develop **in-place**, i.e. connected as they are to all the other components of the body. And they are also constrained to *evolve* in-place. For example, in all the history of the evolution of lungs and hearts, at no point were the lungs and hearts ever disconnected from each other. Because there is no "assembly" stage in its manufacture, it is not so easy to disassemble the components of a living organism. The boundaries between biological components are not always as sharply defined as in a man-made artefact.

These differences between man-made and biologically-made are most acute when looking at the brain. In a modern computer, the component with the most connections to other components is the CPU, and yet the number of pins on even the latest CPU is no more than a few hundred. Each of these pins has a specific function that is determined at the time the CPU is designed, somewhat independently of the design of any particular computer system that is going to include that CPU.

The design of the human brain (and that of other animals) favours as many connections as possible between components, in as much as components can be identified at all. The functionality of connections between different brain areas is partly genetically pre-programmed and partly determined by the processes of growth, development and learning. The larger scale components of the human brain are not plug-in modules as such; rather they are different areas of functional localisation. For example, the colour-processing component of the brain is an area that contains neurons whose firing is a function of perceived colour, such that processing of colour appears to depend strongly on the presence of that area. And there will be millions of connections between that component and other components that provide its inputs and process its outputs.

This high level of interconnectedness implies that it is not going to be so easy to analyse the brain as an information processing system by breaking it up into a moderate number of smaller information processing components.

6.3.1 The Representation of Meaning

The analysis of signals and components in any information processing system should ultimately result in an understanding of how *meaning is represented in that system.*

Here is a very simple example: a thermostat, as shown in Figure 6.4. In this particular example, the thermostat consists of several components, to make the flow of information more explicit:

- A thermometer, which measures temperature and outputs a signal representing the current temperature.
- A target temperature unit (presumably set by the user), which outputs a value representing that temperature.
- A "comparison" unit, which receives as input the values output from the thermometer and the target temperature unit, and which outputs a signal if the measured temperature from the thermometer is less than the target temperature.
- A relay, which receives the signal from the comparison unit, and switches on when it receives a signal, and switches off when it receives no signal. The relay switch controls a heating circuit which includes a power source and a heating element.

The aim of our analysis is to understand the meaning of the signal travelling from the comparison unit to the relay. In fact there are two meanings, one from the point of view of the comparison unit, and one from the point of view of the relay:

1. Coming out of the comparison unit, the signal means "the temperature is too cold".
2. Going into the relay, the signal means "turn the heater on".

Given these two meanings, we can also assign a meaning to the connection between the comparison unit and the relay:

> "If the temperature is too cold, turn the heater on."

That was an exhaustive analysis of the meaning of just *one* signal travelling through a connection between two active components in a very simple information processing system. We would like to do a similar analysis for every neuron and every synapse in the brain and nervous system. Given the way that the brain works, there are two types of question to ask:

1. For each neuron, what does it mean when it fires?

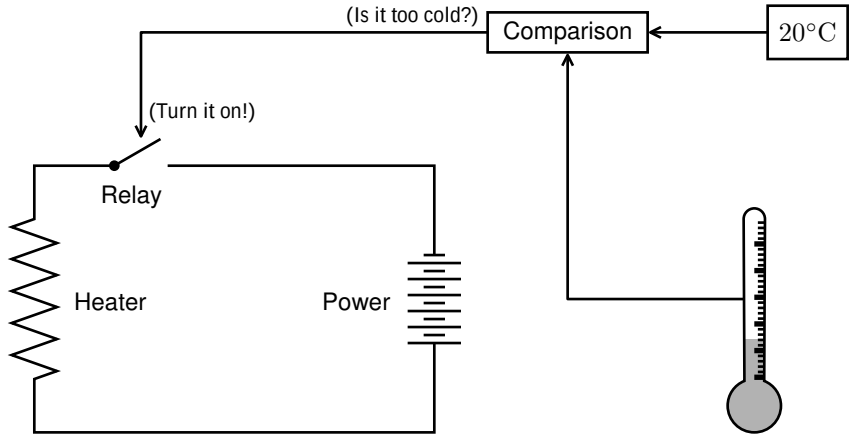

Figure 6.4. Analysis of meaning in a thermostat. The temperature measured in the thermometer is compared to the target temperature of 20°C. If the temperature is too low, a signal is sent to the heating circuit causing the heating circuit to switch on. We can give two interpretations of the signal going from the comparison unit to the relay: coming out of the comparison unit it means "the temperature is too cold", and going into the relay it means "turn the heater on".

2. For each synapse, what is the meaning of the connection between the pre-synaptic neuron and the post-synaptic neuron?

As mentioned above, neurons are also affected by neurotransmitters that are transmitted in a more non-specific manner, and by retrograde messengers like NO. So we can also ask about the meaning of those signals.

The first two types of question cover a lot of ground in themselves. In fact the first question is actually 100,000,000,000 questions, one for each neuron, and the second question is 1,000,000,000,000,000,000 questions, one for each synapse. That's a lot of questions!

And it could get worse. It could be that those are not even the right questions to ask. It may be that we cannot hope to understand the representation of meaning in the brain just by learning the meaning of each neuron and of each synaptic connection between neurons.

It may be that the firing of one neuron has a meaning highly dependent on the firing of other neurons. The meaning may also depend on the relationships between the times at which those other neurons fire, and there may be a complex dependence between the meaning of a neuron firing and the immediate past history of that neuron's own firings.

And when we look at the meaning of a synapse, it may not be sufficient to consider it as just a relationship between the meaning of the pre-synaptic

neuron and the meaning of the post-synaptic neuron. I mentioned earlier that the effects of synapses on the post-synaptic neuron can depend on the relationship between the positions of synapses on the dendritic tree. So we have to take into account the locations of synapses on the dendritic tree of the post-synaptic (target) neuron, and separately assign a meaning to the state of each portion of the dendritic tree, and relate the meaning of each synaptic connection to the states of the dendrite on each side of that synapse.

6.3.2 Temporal Coding

The question of how easily we can specify the meaning of a neuron's firing independently of the context of its previous firings relates to the theory of **temporal coding**. Temporal coding refers to the idea that information is encoded in the precise sequence of timings of action potentials in a neuron. It raises the bar on the difficulty of analysing the effects of all the connections between neurons, because for each synapse we must take into account the relationship between the firing times of the pre-synaptic neuron and the firing times of the post-synaptic neuron.

There is one particular type of temporal coding which does occur in the ear, the auditory nerve and auditory processing areas in the brain, which we might call **direct temporal coding**—"direct" because there is a direct relationship between the times of neural firings and the times of the events encoded by those neural firings. (In practice it's just called "temporal coding", as the possibility of temporal coding which represents information symbolically *without* any direct physical relationship to the original information is somewhat more hypothetical.)

This form of temporal coding starts in the ear, in the organ of Corti, where neurons responding to frequencies from 20Hz to about 4000Hz are **phase-locked**. This means that the firings of neurons in a group representing the same (or similar) frequency are locked in phase with the frequency of the original sound. In fact, for the lower range of frequencies, from 20Hz up to about 200Hz, information about frequency is *only* encoded temporally, as location on the basilar membrane does not distinguish between different frequencies in this range. For higher frequencies the frequency being represented is actually faster than the rate at which neurons can fire. This difficulty is solved by having multiple neurons represent the high frequency signal, according to what is known as the **volley principle**, whereby different subsets of neurons within a group of neurons fire signals for each frequency cycle.

We can understand that lower frequencies have to be temporally encoded, because the mechanics of the ear do not allow them to be positionally encoded. Assuming that "higher-level" processing requires positional encoding at some point, we would presume that temporal encoding gets converted to positional encoding somewhere in the auditory cortex, although it appears that current scientific understanding of this part of the brain is not sufficient to say with any certainty where or how (or even if) this actually occurs.

But if frequencies from 200Hz to 4000Hz are positionally encoded, why do they also need to be temporally encoded? A general answer is that the brain will represent information in as many different ways as possible that can help it to process that information. Temporal coding disappears above 4000Hz, because it is not worth the effort for the brain to maintain the quantity and accuracy of neural firings required to represent sounds at those frequencies temporally. A more specific answer is that the perception of the direction of lower frequency sounds depends on the perception of differences between times at which events are perceived in the left ear and the right ear. Temporal coding of sounds at these frequencies must be maintained at least as far as an area called the **superior olive**, where there are neurons that can compare the times of signals received from both ears. (And for higher frequency sounds, the brain uses relative intensities between left and right to determine direction—a secondary problem for determining direction from temporally coded high frequency sounds is that of knowing exactly which individual vibration perceived by the left ear corresponds to which individual vibration perceived by the right ear.)

One might suppose that the perception of music depends on temporal coding. In particular the harmonic relationships between frequencies related by simple integer ratios would give rise to corresponding relationships between neural firings in response to those frequencies. For example, if neuron A is responding to a frequency at 100Hz by firing 100 times a second, and neuron B is responding to a frequency at 200Hz by firing 200 times a second, then there will be exactly 2 firings of B for every 1 firing of A.

Despite this possibility, the theory of music perception developed in this book mostly ignores temporal coding, and indeed defines a general principle of musicality which is entirely a function of unchanging (or not very often changing) *spatial* patterns of activity in cortical maps that process musical information.

Apart from direct temporal encoding of sounds by phase-locked neurons, there are other basic types of temporal encoding that have been recognised as occurring in the brain. The first is simply that many neurons encode information entirely in terms of *frequency* of firing, i.e. frequent firing means that there is a lot of something, and less frequent firing means there is less of it.

The second type of temporal encoding gives meaning to the phase relationships between the firing of different neurons, and one theory supposes that different neurons fire in **synchrony** with each other (i.e. the same phase and frequency), if they are *referring* to information about the same entity. This theory is an attempt to solve the **binding problem** (discussed in more detail in the next section).

There is necessarily some conflict between different types of temporal coding. For example, neurons phase-locked to different frequencies cannot represent a relationship by firing in synchrony, because synchrony would require

them to match frequency. And the volley principle mentioned above can be seen as a way to allow phase-locking of a group of neurons representing frequency to coexist with frequency encoding within individual neurons of information about intensity.

6.3.3 Localisation and Functional Maps

Despite the possibility that the massive interconnectedness of neurons makes it impossible to understand how meaning is represented as neural activity in the brain, there are some grounds for optimism that naïve **reductionist** theories may be sufficient for us to understand how the brain works. In particular we hope to *reduce* the problem of understanding how the brain works to the simpler problem of understanding how individual neurons or groups of functionally similar neurons represent and process information:

- The relationship between meaning and neural activity is straightforward enough when we look at the periphery: we can directly describe the meaning of sensory cells in terms of the sensory input that they respond to. For example, the meaning of a retinal receptor firing is that a particle of light has landed on that receptor, most likely of a frequency which the receptor is sensitive to. Similarly, we can understand the relationship between meaning and activity for motor neurons: the meaning of a signal propagated along the axon of a motor neuron is "contract this muscle fibre".

- The history of medical neurology consists mainly of a list of ailments of the mind associated with damage to specific areas of the brain. For example, damage to specific areas of the brain results in specific deficiencies in language: damage to one area reduces fluency, damage to another area reduces comprehension.[10] Damage to areas relating to visual perception results in specific deficiencies in vision, such as inability to perceive motion, or inability to perceive colour. The associations between functional deficiencies and localised damage suggest very strongly that specific functionalities and representations of specific meanings are implemented in specific areas of the brain.

- Work on animals has shown that many neurons can be identified whose firing is a function of specific perceptions that the animal must be making in response to its environment. For example, by recording activity of individual neurons, scientists can do more than observe that one area processes colour—they can determine that each neuron in that area is maximally activated by a colour that is particular to that neuron.

[10] It is possible to be fluent without having comprehension. In such cases of **fluent aphasia**, patients speak quickly and easily, but the content of their speech tends towards meaningless nonsense.

A **cortical map** is an area of the **cerebral cortex** where neurons are specialised to perform some particular type of processing, and where there is some identifiable relationship between the position of a neuron in the map and its meaning. The cerebral cortex is the area of the brain which is most developed in mammals and in particular in humans, and it carries out most higher functions. The "map" concept can also apply to areas of the brain outside the cortex, and a general term is **functional map**, or sometimes just **map**. The cortex does, however, have a specific structure that is somewhat flat. The human cortex can be spread out to cover an area of about 0.2 square metres, it is approximately 2mm thick, and it contains 4 to 6 layers of neurons (the layers are fairly fuzzy—one cannot necessarily assign each neuron to a precise layer numbered from 1 to 6). So a cortical map is physically not unlike a real map on a sheet of paper.

In principle the physical position of a neuron has no particular meaning, because what matters is how neurons are connected to each other. However, meanings of signals from some types of sensory cells are necessarily position-dependent: the position of a retinal receptor relates to its position in the image projected onto the retina, the position of a receptor for touch is its actual position on the body, and the position of a receptor for sound in the organ of Corti is a function of frequency. Furthermore, these positional relationships are often preserved by the way that connections are formed travelling from one part of the brain to another. We may presume that the development of the nervous system and brain has evolved in a manner that uses these positional relationships to organise the brain in a way that enables effective processing and use of information from sensory sources.

When scientists look at the **auditory cortex**, which is that part of the cortex that processes sound information, they find many maps that are **tonotopic**, which means that one dimension of the map is correlated with frequency of harmonic components of perceived sound, or with pitch of perceived sound (which more or less corresponds to the frequency of its lowest harmonic). In later chapters, as I present my theory of music, we will have reason to speculate on the existence and purpose of a number of distinct tonotopic cortical maps, and on the relationship between perceived musicality and the patterns of neural activity in those maps.

6.4 Separation and Binding

One general theme that has emerged as scientists have analysed functional localisation in the cortex and elsewhere in the brain is that of *separate processing of different aspects of perception.*

The most studied area of perception is that of vision. Experimenters have used monkeys and other animals to investigate the relationship between brain activity and perceptual functions. Although experimentation on animals is an ongoing ethical controversy, you can get away with inserting probes into

monkeys that you couldn't insert into the brains of human subjects.[11] As it happens, there isn't a whole lot of difference between the visual capabilities of monkeys and those of ourselves, and most of our visual capabilities have evolved from the need to be able to climb and jump through the treetops without falling off and getting hurt.[12]

What scientists have found is that different areas of the visual cortex are specialised for different aspects of visual processing. For example, there are areas that specialise in perception of shape, and others that specialise in perception of motion, and yet others that perceive colour. There are about 30 distinct visual processing areas that have been identified in the monkey and/or human brain.[13]

6.4.1 Colour Perception

To give a specific example, there is a colour perception cortical map that encodes information about the colour of an object. The definitive book on this subject is Semir Zeki's *A Vision of the Brain*. This book is specifically about perception of colour, but its underlying themes are cortical mapping and functional localisation.

The colour of an object is quite distinct from the colour of light reflected from that object. The colour of light reflected from an object is a function of both the colour of light falling on the object, and the colour of the object itself. From an information processing point of view, the information about the colour of light is the input, and the information about the colour of the object is the output. One reason it took scientists a while to discover the difference between these two notions of colour is that our colour perception system is so good that we can reliably identify reflective colour of objects under quite extreme variations of lighting conditions. It is so good that we don't realise how good it is: we just take it for granted that we see the colours of objects.

A similar situation occurs with **pitch translation invariance**—our *inability* to perceive absolute pitch when we listen to music (which is analysed in detail in Chapter 9 on symmetries). We take it for granted that a tune sounds much the same if we transpose it into a different key, but actually there must exist a large amount of sophisticated machinery in the brain to convert the raw incoming information into the desired invariant perception.

[11] There are occasions where, for the purposes of planning brain surgery, it is necessary to map the functionality of a patient's brain by means of electrode stimulation, so as to discover which portions are acceptable to remove, and which parts should be left alone. Such exploration can provide useful scientific data about the localisation of function in the human brain.

[12] Even though it has been millions of years since our ancestors ceased to be full-time tree dwellers.

[13] In *The Astonishing Hypothesis* (page 149), neuroscientist Francis Crick mentions 20 visual maps and 7 partly visual maps, and suggests that at least one of the visual maps will turn out to be several distinct maps.

And once we realise how much machinery there is performing this task, we will be led to ask ourselves what the purpose of this calculation is, because it must be something important if so many resources are devoted to it. In the case of colour perception, it is important to determine the actual colours of objects, both to identify them reliably, and to determine their properties. To give a simple example, if we are looking for ripe fruit on a fruit tree, we select which fruit to pick and eat based on the colour of the fruit.

6.4.2 The Binding Problem

Returning to the issue of separation, what we find with visual processing is that the earlier stages of visual processing encode information in maps that encode for all aspects of vision, including both position and colour. But as information proceeds to so-called "higher" processing areas, the cortical maps separate out the different aspects of that information. Thus the colour map encodes for colour, *almost without regard to position*, and other maps encode for position *without encoding any information about colour*. This seems a little paradoxical: surely in any scene we see different objects of different colours in different positions, so how does the brain properly track the connections between these aspects of colour and position?

The paradox would perhaps not exist if the brain only processed information about one thing at a time: that one thing would have a certain colour and a certain position, its colour would be encoded by the colour cortical map, its position would be encoded by the position cortical map, and that would be that.

But there are many situations where we perceive (and respond to) multiple characteristics of multiple objects. So neuroscientists are stuck with the problem of how (and where) we actually "see" a scene in which all the aspects of vision, including position, motion and colour, are correctly combined in different parts of the scene. This problem is known as the **binding problem**, referring to the need to "bind together" the different aspects of perception that have been separated.

To give a specific example, we might see a red ball in one position going up and a green ball in another position going down. There will be neurons active in two positions in the cortical map representing positional information, and neurons active in two positions in the cortical map representing motion (one group for "up" and another for "down"), and neurons active in two positions in the cortical map representing colour (one group for "red" and another for group for "green"). How do we know that actually the red ball is going up and the green ball is going down, and not vice versa?

The problem is not just one of *how* binding occurs, but also *where*. The changing retinal images encode information that will be used to calculate all the different aspects of visual perception such as colour and motion. As this information is processed, the different aspects are processed separately in different areas, and there does not seem to be any area where they are *joined*

back together. In as much as our conscious visual perception (or "seeing") must combine these different aspects, it is apparently distributed in some mysterious manner across different parts of the brain.

Ultimately our high-level perceptions must be made accessible to those parts of the brain that think about the world and make decisions about what to do. For example, if we are playing a game with different balls, and we know that the red ball is the one we need to catch, we need to be able to move appropriately towards the red ball, in response to its perceived position and direction of movement.

Some scientists have felt the binding problem to be so difficult that they have been motivated to provide rather esoteric explanations of how the brain does the binding. For example, quantum mechanical correlations have been invoked to explain the mystery binding. This hypothesis has been advanced by Roger Penrose (a theoretical physicist) and Stuart Hameroff (a professor of anaesthesiology). Most scientists find this combination of quantum mechanics and neuroscience somewhat implausible and perhaps unnecessary. It doesn't help that the quantum components of Penrose's theory depend on as yet undiscovered theories of quantum gravity.

There are two possible solutions to the binding problem that are both simpler and less esoteric than quantum consciousness:

- The first is that different aspects of information are never completely separated: for example, cortical maps encoding for colour still weakly encode for positional information. This weak encoding may be sufficient to enable re-assembly of information in some manner.

- Second is the theory of **synchronous firing**. This says that neurons whose firing is associated with the same object are bound together by firing synchronously (i.e. all at the same time and in phase with each other). So the neurons representing the direction "up" will fire synchronously with the neurons representing the colour "red", and the neurons representing the direction "down" will fire synchronously with the neurons representing the colour "green". The presumption is that information is generally encoded by a neuron as a rate of firing, without regard to particular timing, and that there is therefore the freedom to choose specific firing times in relationship to firing times of other neurons, in order to specify binding. The phase of neural firing can be changed without altering the overall firing rate, and therefore without altering the information value encoded by that neuron. There is an intrinsic plausibility to this theory: if two neurons A and B have inputs to C, and if the activation of C is stronger when its inputs come in repeatedly at almost exactly the same time, then neuron C will be more strongly activated if its inputs A and B are synchronous. Thus C will be activated more strongly by A and B if A and B are referring to the same object, and if this happens then C will also be referring to

that object. Synchronised neural firing is observed experimentally to occur, and there is some evidence that it occurs in relation to aspects of a stimulus that either are or need to be bound together.

The concept of separation of aspects would appear relevant to the development of a theory of music perception. For example, following the analogy of how the visual system separates processing of different aspects of vision such as location, colour and motion, we might reasonably expect that the auditory system separates the processing of pitch relationships and temporal relationships. And we would expect that the results of these separated aspects of processing are combined back together again to provide the final conscious percept.

It follows therefore that we should consider the binding problem when analysing how the human brain processes music. On the other hand, whatever solution the binding problem has, it is probably going to be the same solution for all different types of perception, whether visual or aural or anything else. So when the theory requires me to state that certain perceptions are bound together, I am quite happy to state that I don't know for sure how the binding happens, but I know that binding has to happen somehow, and the same "somehow" is how it happens in the case of music.[14]

6.5 Population Encoding

There is another complication in the representation of meaning in cortical maps. As a simplification, we could consider a cortical map which was effectively a one-dimensional map, and which responded to one numerical aspect of a stimulus, for example the frequency of a sound.

The encoded value comes from a continuous range of values: it could be any real number between 20 and 20000 (representing frequency in Hz). But the set of neurons in the cortical map is finite. If we assign a particular frequency to each neuron, then only a discrete number of frequencies can be represented by the map. Some ad-hoc mechanism would be required to deal with the in-between frequencies; for example, we could round to the nearest value that had a representation.

There are a number of reasons why such a simple representation of meaning will not be satisfactory:

- If we consider sensory neurons, it is very unlikely that a neuron is going to have a sharp cutoff in what it responds to, in such a way that there is no overlap between what different neurons in a map respond to.

[14] Although there is the difficulty, as previously mentioned, that if different auditory neurons are phase-locked, then whether or not they can or do fire in synchrony is dependent on the relationships between the frequencies that they are firing at.

- If a particular neuron gets damaged or lost, the values it represents will cease to be represented at all.
- If only one signal appears, or only one signal appears within a certain portion of the cortical map, then only one out of all the neurons in that portion will be active, which seems to be a waste of information processing capacity.

Population encoding is a manner in which neurons in a cortical map encode numerical values. Very simply, we can say that for each neuron, and for each possible signal value, the rate of firing of the neuron is a function of the encoded value. Each neuron has its **encoding function**.

This method of encoding would be equivalent to the first method of encoding that we described, *if* the encoding function for each neuron was equal to a maximum value for all the values in the range that the neuron represented, and a minimum (or zero) value for all values outside that range (as in Figure 6.5). But what happens in practice with population encoding is that the encoding function still has a peak value, i.e. an encoded value that results in a maximum firing rate, but this encoding function falls away smoothly as the encoded value moves away from this peak value (as in Figure 6.6).

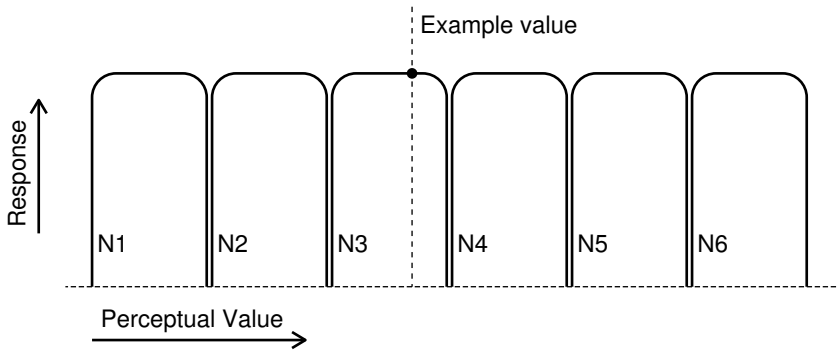

Figure 6.5. Neural response without population encoding. The encoding functions for a perceptual variable are shown for 6 neurons in a hypothetical cortical map. Each neuron has a maximum response to values in the range of values it represents, and the ranges represented by different neurons are all disjoint from each other. An example value is shown, such that only neuron N3 responds to it.

Thus, for any encoded value, the neurons whose peak values are nearest to that value will fire most strongly, and neurons with peak values further away from the encoded value will fire less strongly, or not at all.

Given the observed firing rate of neurons responding to a single encoded value, it is relatively straightforward to determine what the encoded value is.

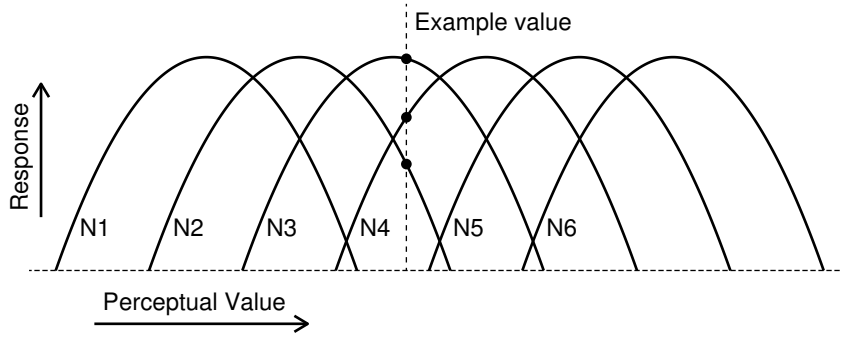

Figure 6.6. Neural response with population encoding. The encoding functions for a perceptual variable are shown for 6 neurons in a cortical map. For each neuron there is a value to which it gives a maximum response, but there is overlap between the ranges of values that different neurons respond to. Neuron N3 responds most strongly to the example value shown, but neurons N4 and N2 also show a response. N3, N4 and N2 constitute the "population" of neurons responding to that perceived value.

Thus the firing of all those neurons accurately represents the encoded value. Population encoding can quite accurately represent encoded values that are in between the peak values of the neurons in the map. For example, to determine the value represented by the firing of a group of neurons, take the average of the neurons' peak values, weighted by their firing rate. (A more accurate procedure is to calculate a **maximum likelihood value**, which is the value for which the current pattern of neural firing would be most probable.)

One problem with population encoding is that if the encoding functions are too broad, then it will be difficult to distinguish two distinct values from one value equal to the average of those distinct values. There is a simple musical example that illustrates this phenomenon: when we hear people singing in chorus. As long as the singers are singing in tune on average, we will hear the singing as being perfectly in tune, even if the individual singers are all slightly off.

In some cases the distinction between one signal and two signals close together may be made by the above-mentioned mechanism of synchronous firing. That is, neurons responding to one signal will fire synchronously with each other, and neurons responding to a second signal will fire synchronously with each other, but not synchronously with those representing the first signal.

It is also possible that the breadth of the encoding functions is itself adjustable by some means, so that neurons in a cortical map can choose broad or narrow encoding depending on which is the most useful in the current circumstance.

Population encoding is pretty much a universal property of cortical maps. So whenever I make a statement like "Cortical map X encodes for values Y and Z", this can be correctly interpreted as "The neurons in cortical map X fire at a rate that is a function of the closeness of their peak values of Y and Z to the observed values of Y and Z".

It's also worth noting that population encoding bears very little resemblance to how numerical values are normally represented in electronic computers. In computers we do not use a linear sequence of components to represent numerical values according to position. Generally we pick a base N (almost always 2), and then write the number as a sequence of digits, where each digit is an integer in the range 0 to $N - 1$. So, to represent 1000 possible values, we would need 10 components (i.e. 10 digits in base 2), and to represent 1,000,000 possible values we would need just twice as many components, i.e. 20. To represent 1000 possible values in a cortical map, the brain would need 1000 neurons, although with population encoding this could be reduced by some fixed factor—for instance 2—to 500 neurons, given the ability of population encoding to represent the "in between" values. To encode 1,000,000 possible values would still require 1000 times as many neurons as required to represent 1000 possible values, i.e. 500,000. This contrast between the efficiency of digital and analogue representations appears in the theory of **octave translation invariance** in Chapter 11.

Chapter 7

2D/3D Theory of Music

This chapter describes my older 2D/3D theory of music, which was formulated in response to observations about the vector representations of musical intervals and the various mappings between them.

Firstly we look at some more vector and point space mappings: a 2D to 1D vector mapping which maps both tones and semitones to "steps", and the visual 3D to 2D point space mapping which maps the 3D world to 2D (retinal) images. Then I discuss the major concept in the 2D/3D theory, which is the suggestive analogy between the musical 3D to 2D mapping and the visual 3D to 2D mapping.

7.1 More Vector Space Mappings

7.1.1 Another Mapping from 2D to 1D

We've looked at the "natural" mapping from 2 dimensions to 1 dimension, i.e. the one that maps tones and semitones to semitones. But there is another mapping from the 2-dimensional space to a 1-dimensional space that could be considered relevant to understanding music perception. This is the mapping that maps both a tone and a semitone to a step. The "step" represents a step on the diatonic scale that one takes as one goes from one note to the next note on the scale. We cannot consider the target space of this mapping to be the same as the 1-dimensional semitone space, so perhaps we can call it the **1-dimensional step space**. This mapping "forgets" the difference between a tone and a semitone, in the sense that looking at an output vector consisting

of n steps, we cannot tell which of those n steps in the input vector were semitones and which were tones. It is represented by the following matrix:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

So why might this mapping be important for understanding music perception? There are many tunes where a first phrase consists of some sequence of notes played in a certain rhythm, and then a second phrase consists of the same sequence of notes *transposed along the diatonic scale*, played in the same rhythm. This transposition is different from the normal sort of transposition, which refers to an exact translation such as when a key change occurs. The exact pattern of intervals in the second phrase will be different from that in the first phrase, because some tones will change to semitones, and vice versa. To give a simple example, the first phrase might be CDEDEE, and the second phrase could be DEFEFF, which is transposed one "step" up the scale.

But if we apply the forgetful 2D to 1D mapping that we have just described, then the mapped version of the second phrase is an exact translation of the mapped version of the first phrase.

This seems a promising notion. But if it really forms an aspect of music perception, there would have to be some cortical map that performs this mapping. If we assume that the cortical maps that process music already exist to serve some other purpose, then it is unlikely that such a cortical map exists, because there is no other reason why the brain would want to process information about musical intervals in this way; in particular scales do not occur outside music, and speech melodies do not have a structure which can be factored into independent dimensions of tone and semitone. In Chapter 10, the **melodic contour cortical map** is introduced. This map ignores the difference between tones and semitones in many cases, not because there is a 2D to 1D mapping, but rather because it processes pitch information with a reduced level of precision.

7.1.2 Another Perceptual 3D to 2D Mapping

The world we live in is 3-dimensional. We make representations of parts of the world in pictures and photographs which are 2-dimensional. The images on the retinas of our eyes are 2-dimensional, and our brain reconstructs a model of the 3-dimensional world from the information in these two 2-dimensional images. The correspondence between a 3D scene and its 2D picture can be described as a mapping from a 3D point space to a 2D point space. By considering vectors defined by pairs of points in the 3D and 2D spaces, we can define a corresponding mapping from a 3D vector space to a 2D vector space. As already mentioned in Chapter 5, a mapping between point spaces that defines a corresponding well-defined linear mapping between vector spaces is called an **affine mapping**. The mapping between a 3D scene and a 2D

picture is *not* an affine mapping. This has to do with the fact that things far away are smaller on the picture than things that are close. The technical name for such a mapping is a **projective mapping**.

However, if we consider a very small portion of the 3D scene ("small" in the sense of being a small volume of limited diameter), which is a large distance from the point of view that defines the picture ("large" compared to the size of the "small" portion), then the mapping is *approximately* affine, and there is a corresponding approximately linear mapping of displacement vectors.[1]

Furthermore, the human brain necessarily has an ability to process the correspondence between 3D scenes and 2D pictures of those scenes. This ability underlies our ability to perceive 3D from the 2D information provided by our retinas.

The first assumption of the 2D/3D theory of music is that there is a significant analogy between the two different 3D to 2D mappings:

- the musical 3D to 2D natural mapping which maps from the 3D representation of musical intervals to the 2D tone/semitone representation of musical intervals, and,

- the visual 3D to 2D natural mapping which maps from arbitrarily small displacement vectors in an arbitrarily small portion of a 3D scene to their images in a 2D picture (with the point of view not too close to said portion).

Translated into the language of neurons and cortical maps, this analogy suggests two possible hypotheses about the relationship between the two types of 3D/2D mapping:

1. There is a cortical map somewhere in the brain that processes the relationship between 2D and 3D in the brain, and this cortical map also processes the relationship between 2D and 3D in music, or,

2. there is a set of neurons somewhere in the brain with an intrinsic ability to process 2D/3D relationships. Most of them are recruited to process the relationship between 3D objects and 2D images, but some of them get recruited to the task of processing 2D/3D relationships in music.

The problem with the first hypothesis is that we would then expect listening to music to *feel like* visual perception of the real 3-dimensional world. We would expect this because that is the generally observed fact about cortical

[1] If a point space mapping is *not* affine, not only will the corresponding vector mapping not be linear, it won't even be well-defined (the mapped value of a vector will vary depending on which two points are used to define it). But if we assume that the point space mapping over a small enough portion of the point space is sufficiently close to affine, then the corresponding vector space mapping will be correspondingly close enough to being well-defined.

maps: two different experiences or perceptions or emotions feel the same if and only if the same neurons are active in both cases.

The second hypothesis is an attempted work-around to this problem, i.e., the same sort of neurons process visual 2D/3D and musical 2D/3D, but there is no actual overlap in which neurons are active in each case, and that is why music does not feel like visual perception of 3D space.

As stated so far, the 2D/3D theory provides an explanation for the diatonic scale, and it explains the relevance of harmonic relationships between notes in the scale, but it does not explain any other features of music.

7.2 The Looping Theory

The second assumption of the 2D/3D theory is based on two observations:

1. Music tends to go around in circles. Tunes start on a home note and a home chord (prototypically the note C and the chord C major which consists of the notes C, E and G), travel a path visiting other notes and chords, and finally return to the home note and the home chord.

2. The Harmonic Heptagon (see end of Chapter 5) defines a cyclic path around the diatonic scale.

So maybe the 3D representation of notes, as defined by the 3D representation of the intervals between different notes, travels once around the Harmonic Heptagon as it travels from the initial home note and chord to the final home note and chord. This implies that the final home note is displaced from the initial home note by the 3D vector $(-4, 4, -1)$ which represents the syntonic comma of 81/80 (or by $(4, -4, 1)$ representing 80/81, depending on which way we go around the loop). In 3 dimensions the tune travels along something like a spiral, and the 2-dimensional picture is seen from a point of view such that the spiral looks like a closed circle. To close the gap corresponding to the syntonic comma, the point of view has to be one such that points separated by a multiple of the $(-4, 4, -1)$ vector are in the same line of sight, and thus occupy the same position in the 2D image.

The looping theory adds some extra constraint into the 2D/3D theory. Furthermore, we can relate common chord sequences to a trip around the Harmonic Heptagon. For example, a common chord sequence is C major, F major, G7, C major. To make the theory work there has to be some method of determining where each chord would be placed on the Harmonic Heptagon relative to previous chords that have already occurred in the tune. The tune starts with C major (CEG). Next is F major (FAC). It seems reasonable to regard the F major as being connected to the C major via the shared note C. Moving on to G7 (GBDF), it seems reasonable again to connect it to F major by the shared note F. And then the G7 will be connected to the final C major by the shared note G, which completes a full circle going clockwise around

the heptagon. In 3D space, the final C major chord is located in a position displaced from the initial C major chord by the vector $(4, -4, 1)$.

7.3 Outlook for the 2D/3D Theory

Unfortunately my development of the 2D/3D theory has not made any further progress. And I have now developed the newer **super-stimulus theory**, which has a much better foundation in biological theory, and is able to explain many aspects of music in plausible and convincing detail. But given uncertainty about some parts of the super-stimulus theory, and the incompleteness of that theory, I can't rule out the possibility that the older 2D/3D theory has some relevance to a final and complete explanation of music.

The concept of the Harmonic Heptagon does turn out to be important for developing certain aspects of the super-stimulus theory, in particular the theory of home notes and home chords. And the 1D/2D/3D vector theory of intervals gives a complete picture of all the relationships between intervals described as tones plus semitones and intervals described in terms of simple fractional ratios (if those ratios are considered not to have any prime factors in the numerators and denominators other than 2, 3 and 5). So the analysis of intervals as vectors was a useful analysis to do, even if the full 2D/3D theory turns out to be incorrect.

I will finish this section with a list of unresolved issues around the 2D/3D theory:

- The analysis of chordal movement around the heptagon doesn't say anything about melody. We have to find a way to relate the notes of the melody to the notes of the harmony within the framework of the theory.

- One can attempt to place or locate notes of the melody in 3D space in the same sort of way that I described chords being located. This requires us to define rules as to which harmonic intervals between which notes are to be used to locate notes relative to each other. The desired result is that the final home note is located at a position in 3D space displaced from the position of the initial home note by the syntonic comma. Presumably the displacement calculated by calculating the locations of notes in the melody should be consistent with the rules for calculating the locations of chords, particularly if the chords are implied by the melody.

- Consecutive chords do not always share notes, so shared notes cannot always be used as a basis for determining where to locate chords relative to each other in 3D space. They can also share more than one note, in some cases giving rise to two different choices of relative location.

- The theory doesn't say much about time and rhythm. The best it can do is suppose that the times that notes occur play a role in the rules that determine which relationships between which pairs of notes determine relative locations in 3D space. A bigger difficulty is that there is some degree of musicality in music that consists only of rhythmical percussion—something that a theory based on frequency ratios cannot possibly explain. (The super-stimulus theory does better here, as it can explain the musicality of music that has no melody or harmony at all.)

- The 2D/3D theory depends too much on specific features of the well-tempered diatonic scale, in particular that the steps are all one of two sizes.

- The theory assumes that ratios involving 7 (or higher prime numbers) are musically unimportant. For example, adding 7 would increase the number of dimensions from 3 to 4. This is less of an issue with the super-stimulus theory. The construction of the Harmonic Heptagon is based on powers of 2, 3 and 5; and the super-stimulus theory does make use of the Harmonic Heptagon to analyse some aspects of Western diatonic music. But the super-stimulus theory does not *depend* on the existence of this heptagon to explain *all* music—it only makes use of the heptagon to explain relevant properties of music based on the scale that the Harmonic Heptagon is constructed from.

- As already mentioned in the introduction, the 2D/3D theory is analogous to the paradoxical drawings of M.C. Escher, which exploit the ambiguity in 3D space of the location of points represented on a 2D image. But looking at an Escher drawing does not "feel like" listening to music, whereas one might expect it to do so if the same paradox applied to perceptions processed by the same cortical maps in each case.

Chapter 8

The Perception of Musicality

The starting point of my newer theory of music, which is the main theory developed in this book, is the assumption that music itself has no purpose, but that our *response* to music has a purpose. Music is a **super-stimulus** for the perception of **musicality**, and musicality is a perceived attribute of *speech*.

All of the cortical maps that respond to music are actually cortical maps whose purpose is to respond to speech. Every aspect of music corresponds in some way to an aspect of speech perception, although the nature of this correspondence may not always be obvious.

8.1 Where is the Purpose?

We have already asked the question: What is the biological purpose of music? But biological purpose only exists within the structure and behaviour of living organisms. Music is not a living organism; the living organism in this case is the human being who enjoys music. Music only exists as a result of human behaviour in relation to music. The most significant human behaviours that relate to music are:

- Composing music
- Performing music
- Listening to music

- Dancing to music

In which of these resides the true biological purpose of music? Is it in composition, where the composer creates music in order to express their emotions, or communicate in some way with the listener? Or is it the performance of music, where the performer expresses and communicates? Is there something special about the group performance of music, which bonds the members of society together as they perform together? Is it dancing, another group activity, with bonding and good exercise thrown in as an extra? Is it just the listening that does something useful, letting us understand the emotions of the composer and performer?

8.2 That Which is Like Music

Perhaps we are assuming too much when we list the choices above. Maybe the biological purpose associated with music is not actually about the music. Is this possible? Could the biology of music actually be about something else?

A simple question follows from this line of thought:

> What is the thing that is most like music which is *not* music?

We can try to answer this question by looking at different aspects of music. These include at least the following:

- Melody
- Scales
- Rhythm
- Harmony
- Chords
- Home notes and home chords
- Bass
- Instrumental timbres with harmonic frequencies that are integral multiples of the fundamental frequency
- Repetitions: exact and partial, free and non-free
- Rhyme

As it happens, there exists one very important human behaviour that has at least three of these aspects, and that important behaviour is *speech*. The three aspects that speech has in common with music are melody, rhythm and timbres with harmonic frequencies that are integral multiples of the fundamental frequency.

Speech has melody, because the pitch of the voiced portions of speech goes up and down as the speaker speaks. This "speech melody" can include—depending on the language—**lexical pitch**, **pitch accent** and **intonation**. Languages that use lexical pitch, where each individual word has its own little melody, are called **tone languages**; one well-known example is Cantonese. Pitch accent is where the accents of words are partly defined by changes in pitch; typically a rise in pitch represents an accent. Intonation is where a sentence or phrase has an overall melodic shape that says something about the meaning, intention or emotion associated with that sentence or phrase. (Classifications of languages into those that do or do not contain pitch accent and/or intonation are not absolute. Some languages indicate accent almost entirely by pitch, such as Swedish; in other languages it forms part of the indication of accent, as in English. And there is often variability in the occurrence of pitch accent across different dialects of one language. Intonation occurs in many languages, more so in those languages that do not have lexical pitch.)

Speech also has rhythm. We can define rhythm as the patterns of timings of syllables in words. Each language has its own typical patterns of rhythm. This rhythm plays a significant role in the perception of language: in particular it helps to predict the locations of syllable boundaries. If you have ever tried to write speech recognition software, you'll know that syllable boundaries are more difficult to spot than one might suppose. And it's rather hard to identify the content of syllables if you are not even too sure of where they start and end. One of the ways that our brains can solve this problem is to use the observed times of previous syllable boundaries to predict the likely times of future syllable boundaries. The known properties of a language's rhythms are what make this prediction possible. It has been established that babies become sensitive to the rhythms of the language spoken around them at quite an early age.

And finally, the spoken human voice has an instrumental timbre with harmonic frequencies that are integral multiples of the fundamental frequency. (To be precise, it has these qualities when it is uttering **voiced** sounds, i.e. vowels and voiced consonants.)

This is not by any means a complete match between the aspects of speech and the aspects of music, at least not with regard to those aspects that are obvious to us. But it does appear that speech is closer to music than anything else is.

We could consider poetry as another candidate for something which is "like music but which is not music". Poetry has a regular rhythm, and it

has rhyme. Poetry can be seen as something that lies in between ordinary speech and music in its characteristics. However, in the general scheme of things, poetry is less significant than either speech or music. By any simple economic measure, the amount of time, money and effort that the general population puts into the production and consumption of poetry is far less than what they put into the production and consumption of either speech or music. And poetry does not have any obvious biological purpose any more than music does. So, for the moment, we will ignore poetry. And later we will see that the theory we develop enables us to formulate a more abstract definition of music, such that poetry turns out to be a minor form of music in itself.

Having "matched" music with speech, we should mention the obvious discrepancies, both in what we matched and in what we did not match:

- A melody can be defined as pitch which is a function of time. In speech melody, pitch is generally a continuously varying function of time, whereas in musical melodies pitch is constant for the duration of a note, and then jumps suddenly to the pitch of the next note. Also the pitch values in a musical melody take on only a finite set of values corresponding to the notes in a scale. There are no scales in speech.

- Speech does not have harmony or chords, unless perhaps we consider several people talking at once. And if there are no chords, then there are definitely no home chords.

- The rhythm of speech is not regular in the way that musical rhythm is regular. There is no regular hierarchical division of time as there is in music.

- Speech does not have a bass accompaniment.

- Exact repetitions of phrases or sentences do not normally occur in speech.

- Ordinary speech does not rhyme.

The only match that appears without any caveat or discrepancy is the match between the spoken human voice and the human voice as a musical instrument; in fact the human voice is the most popular musical instrument, as most popular music is in the form of song.

There are also many aspects of speech that have no obvious equivalent in music, including vocabulary, syntax and semantics (although some authors have tried to draw analogies with these aspects of language).

One day I was thinking along these lines, of how music resembles speech and that in some ways music seems almost like a parody of speech, and I had an idea. My flash of inspiration was this: there is a somewhat limited match between speech and music, but maybe the real match isn't between speech

and music *per se*, maybe the real match is between the *perception* of speech and the *perception* of music. Maybe there is a match between the cortical maps that respond to speech and those that respond to music.

Now the main result of perceiving music appears to be the perception of the **musicality**[1] of music, which causes a pleasurable emotional effect. The main result of perceiving speech is understanding the semantics of what is being said. But that is not the only result of perceiving speech. There are other things that we perceive, like the identity of the speaker, and the emotion of the speaker, and clues as to whether the speaker is being honest with us. So if speech perception includes these various extra perceptions, maybe *the perception of musicality is yet another aspect of speech perception.*

In other words, musicality is an attribute of speech, which provides the listener with some significant information about speech. (Musicality is assumed to be significant, whether or not we think we know what that significance actually is. But we will make hypotheses about that as well—later on.)

But if musicality is an attribute of speech, what then is music? Music is a *contrived stimulus* or **super-stimulus**, which is contrived so as to have a high level of musicality.

This idea of a super-stimulus is well known in the field of biology. Ethologists are fond of taking apart the ways some animals respond to their environment, and discovering super-stimuli that create a more extreme response in the animal than the normal stimuli.

Some well-known super-stimuli have been discovered by scientists studying the feeding of baby birds by their parents. In fact super-stimuli have been discovered for both parent and chick behaviour: artificial parents that the chicks prefer to beg food from, and artificial chicks that the parents prefer to feed. The super-stimuli are often just over-simplified models of the parent or child bird, generally with exaggerated versions of markings that have been observed to play a role in the begging or feeding reflexes in question.

An example is given by Professor Vilayanur Ramachandran in his 2003 Reith Lecture (on BBC Radio).[2] He refers to work done by the ethologist Niko Tinbergen on herring gulls. The chick's begging reflex is tied to the colour pattern on its parent's beak, which is yellow with a red spot. The chicks will beg from a beak that has been separated from its owner, and they will beg even more enthusiastically from a long yellow stick with three red stripes on it. This coloured stick is the super-stimulus.

Ultimately the real beak and the super-stimulus beak must create a response in the same cortical (or maybe non-cortical) map in the baby gull's brain. The nature of the super-stimulus tells us something about what that particular cortical map responds to.

[1] The word "musicality" has at least two common meanings: firstly describing how musical a *person* is, and secondly describing how musical some *music* is. It is the second meaning of the word that is used throughout this book.

[2] Also published as a book: *The Emerging Mind* (Profile Books 2003).

The purpose of Professor Ramachandran's lecture was to explain his theory of human art based on various principles that could be derived from our scientific understanding of how the brain works—and one of those principles was that of the super-stimulus.

If music is a super-stimulus, then it is certainly a more complex super-stimulus than a yellow stick with three red stripes. Although popular music may be simpler than classical music, and some popular music is very simple compared even to other popular music, there is still a minimum level of complexity. There are also many distinct items of music: all those that have ever been composed, and probably an even larger number that have not yet been composed. And music has various different aspects: melody, rhythm, harmony, bass, home chords etc.

These complexities of music suggest, although they do not absolutely prove, that the perception for which music is a super-stimulus is one that has a moderately level of complexity in itself. The human cortical maps involved in the perception of musicality probably perform information processing much more complex than that which is performed by the cortical map in the herring gull chick's brain that responds to the yellow and red pattern of its parent's beak.

8.3 Corollaries to the Hypothesis

Firstly I will restate the hypothesis as developed so far:

- The perception of musicality is an aspect of speech perception.
- The perceived musicality of speech represents useful information about the speech being listened to. The benefits of perceiving this information have provided the selective pressure that has driven the evolution of the ability to perceive musicality.
- Music is a contrived super-stimulus, contrived so as to have a high degree of musicality.

This hypothesis explains at least some of the properties of music in relation to speech, and it explains why music is not exactly like speech. And it fits plausibly into the framework of Darwin's theory of evolution by natural selection.

But at the same time it gives rise to a whole range of new questions:

- What is musicality? Or more specifically, what information does musicality provide about speech? And how does that relate to the emotional effect of music?
- Is musicality a one-dimensional attribute of music? That is, does it reduce to a simple "this music has low musicality", and "that music has high musicality"?

- Why is it not subjectively obvious to us that musicality is a perceived feature of speech?
- There are claims in the scientific literature that there is **double dissociation** (a term that I will explain when I attempt to answer this question) between music perception and speech perception: does the evidence supporting this claim contradict the musicality hypothesis?
- For the aspects of melody and rhythm, how do we explain the differences between speech melody and musical melody, and between speech rhythm and musical rhythm?
- What about other attributes of music that appear not to exist in normal speech at all, i.e. scales, harmony, home notes, home chords, bass and rhyme?
- What does the hypothesis tell us about the cortical maps that respond to music and speech? Can the aspects of music help us understand the nature of the cortical maps involved in speech perception?

These are not all the questions that need to be answered. An investigation of musical symmetries and invariances raises more questions. Symmetries turn out to be so important that I devote a whole chapter to listing and describing the full set of symmetries of music perception.

8.3.1 What is Musicality?

What is musicality? Or more specifically, what information does musicality provide about speech? And how does that relate to the emotional effect of music?

The first thing to say is that the musicality hypothesis, i.e. the hypothesis that perception of musicality is an aspect of speech perception, was enough to lead me on a long path of successful investigation into the mechanics of music perception. I found out a lot of things about music, *without even knowing what musicality was or what information it represented*. In the end, I was able to come up with a plausible answer to this question: musicality represents an estimate of a certain aspect of mental state of the speaker, corresponding roughly to **conscious arousal**. But I remain more confident of the general musicality hypothesis than of my more specific theory as to what musicality actually is.

The theory of **constant activity patterns** (or **CAP**) is fully explained in Chapter 14. It tells us that musicality is caused by the occurrence of activity patterns in neural maps that remain constant, and that the means by which the listener's brain calculates musicality represents an attempt to detect the occurrence of similar activity patterns in corresponding neural maps in the *speaker's* brain, which provides information about the speaker's

level of conscious arousal, which in turn influences the listener's perception of the speech content and in particular influences their emotional reaction to that content.

Before we can get to the details of the CAP theory, we need to investigate and understand musical symmetries and cortical maps, so for the moment it is best to defer consideration of the meaning of musicality, and just continue under the assumptions that (1) there is such a thing as musicality and (2) the perception of it is something that matters.

8.3.2 The Dimensionality of Musicality

Is musicality a one-dimensional attribute of music? That is, does it reduce to a simple "this music has low musicality", and "that music has high musicality"?

On the one hand, the musical quality of music does seem to be multi-dimensional, in that different types of music evoke different emotions and different feelings. On the other hand, the musicality-arousal hypothesis suggests that there is one primary dimension to musicality, which determines the music's ability to evoke an emotional response. Furthermore, the musicality-arousal hypothesis states that an emotional response is only *supported* by the musicality, and if any specific emotion is evoked, it must be evoked by some other aspect of the music.

One of the most often asked questions about music and emotion is: why do minor chords sound "sad" and major chords sound "happy"? The theory in this book, unfortunately, does not have much to say about this issue. But if musicality is a one-dimensional attribute, then this implies that the *quality* of emotion evoked by a tune—if there is a definite quality of emotion evoked by that tune—is independent of whatever it is that determines if a tune is very musical or not very musical. If a very musical tune has mainly minor chords it will evoke a strong feeling of sadness, and if a very musical tune has mainly major chords it will evoke a strong feeling of happiness. If a tune is not very musical, then it will evoke an emotion in accordance with the chord type, but the emotion evoked will be weaker on account of the tune's lower level of musicality.

8.3.3 Subjective Awareness of Musicality

Why is it not subjectively obvious to us that musicality is a perceived feature of speech?

According to the theory, music is a *super-stimulus*, meaning that the effect generated by music is much stronger than for normal speech. We are aware of the emotional effect of music, but we are not internally aware of the processes that generate that emotional effect. When musicality affects the perception of normal speech, it probably has a mild reinforcing effect on our emotional

response to the content of the speech being perceived. The musicality of normal speech affects our emotional responses, in a subtle way, and it always has done, but without us realising it. The question "Why are we not consciously aware of this effect?" is a little bit like the question "How come we don't notice the centrifugal force caused by the Earth's rotation?". The answer to the latter question is that the gravity we feel is the gravity that would be there if the Earth wasn't rotating, minus the effect of the centrifugal force. We are not aware of the presence of the centrifugal force, because we have never experienced what it would be like to stand on an Earth that was not spinning. Similarly, we might become consciously aware of the effect that the perception of musicality has on our perception of speech if it was suddenly disabled in some way.

Another way of looking at this is to compare machine-generated speech to human speech. All man-made speech machines to date are not capable of accurately simulating normal speech, in as much as they do not sound completely natural to a human listener. The unnaturalness of artificial speech corresponds to various aspects missing from it. One of those missing aspects is the musicality of the speech, since no provision is made for musicality in the algorithms that generate the speech.

8.3.4 Double Dissociation

There are claims in the scientific literature that there is double dissociation between music perception and speech perception: does the evidence supporting this claim contradict the musicality hypothesis? (Some examples are given in *The Cognitive Neuroscience of Music*, John Brust "Music and the Neurologist: A Historical Perspective", and Isabelle Peretz "Brain Specialization For Music: New Evidence from Congenital Amusia".)

Double dissociation of two components of cognition A and B refers to finding subjects with disabilities, where one subject has A but not B, and another subject has B but not A. Double dissociation of music and speech refers to subjects who can perceive musical qualities of music but cannot understand speech, and subjects who can understand speech but cannot perceive the musical qualities of music. **Amusia** is a general term for loss of musical ability, and **aphasia** refers to loss of language perception or ability.

I can answer this question better when I consider the musicality-arousal hypothesis in more detail, but the following points can be noted:

- If our theory claims that music perception is a *subset* of speech perception, then only amusia without aphasia provides any challenge to the theory—aphasia without amusia is entirely possible since an aphasia may apply to an aspect of speech which is not an aspect of music.
- You cannot dissociate an aspect of music perception from speech perception if you don't know what music perception is perception of—you

cannot dissociate two components of perception if at least one of them is unknown in function and purpose.

- Perception of musicality may only be relevant in some circumstances, i.e. when the speaker is aroused and the semantics of what they are saying has emotional consequences for the listener. (An alternative hypothesis is that the *emotional* effects of musicality only apply in some circumstances, but there may be other effects of musicality on our perception of speech that apply more generally, in which case amusia would affect those aspects of speech perception in all cases.)

Thus, in all likelihood, the subjects with comprehension of speech but lack of music perception *also lack those aspects of speech perception that depend on their ability to perceive musicality.* Since the scientific observers studying these subjects do not know what role the perception of musicality plays in speech perception, they will fail to observe that their subjects lack those aspects of speech perception.

8.3.5 Differences in Melody and Rhythm

For the aspects of melody and rhythm, how do we explain the differences between speech melody and musical melody, and between speech rhythm and musical rhythm?

There are two steps that connect "normal" stimuli to super-stimuli:

1. The requirements for perception of normal stimuli determine the structure and operation of the cortical maps that perform that perception.

2. The structure and operation of the cortical maps that perform a given perceptual task determine the nature of the super-stimuli for that perception.

A common consequence is that super-stimuli are qualitatively different to the corresponding normal stimuli.

We can consider this explanation even in relation to individual neurons. For example, perceptual neurons in the brain that encode for colour of light (not colour of objects) will have the strongest response to pure spectral colours. Such colours hardly ever occur in nature.[3] But the proper purpose of these neurons is not to perceive pure spectral colours: it is to perceive and distinguish all the other colours that occur naturally.

Hair cells in the organ of Corti respond maximally to pure sine tones at a particular frequency. Again such tones hardly ever occur in nature, and it

[3]Even rainbows are blurred. In my own personal experience, the nearest one gets to pure spectral colours in nature is when seeing small flashes of spectral colour from sunlight refracted through individual drops of dew in the grass.

is not the purpose of these hair cells to respond to pure tones; rather their purpose is to perceive sounds in general.

There is an interesting analogy between pure spectral colours and pure tones: both are pure regular sine wave vibrations at a particular frequency. It is quite commonly the case that super-stimuli for a perceptual sub-system are more regular in form than the ordinary stimuli that the perceptual sub-system is designed to respond to. This can explain some of the differences between music and speech. For example, the rhythms of music are much more regular than the corresponding rhythms of speech, but the purpose of neurons that respond strongly to regular rhythm may actually be the perception of *irregular* rhythms.

Similarly, we might speculate about the regular patterns of frequency that occur in musical melody versus the patterns that occur in speech melody. We will find, however, that properly solving the problem of musical scales versus smoothly varying speech melody requires a more in-depth understanding of the cortical maps underlying the perception of melody.

8.3.6 Attributes Apparently Absent in Speech

What about other attributes of music that appear not to exist in normal speech at all, i.e. harmony, home notes, home chords, bass and rhyme?

This question will—for each attribute—have a similar answer to the previous question. But for these aspects it is less obvious what the corresponding attributes of speech are. In the end we will find it easier to backtrack along the line of implication that goes:

$$\text{Perception of Ordinary Stimulus} \Rightarrow \text{Cortical Map} \Rightarrow \text{Super-Stimulus}$$

In other words, we will look at the aspects of music (the super-stimulus), and we will make an intelligent guess as to what types of cortical maps respond to those aspects, and then we will make further intelligent guesses as to how those cortical maps fit into the requirements of speech perception.

For a simple example, we can look at **harmony**. Harmony is multiple notes played together. We might suppose that this has something to do with multiple speakers speaking at the same time. But our brains do not normally attempt to comprehend more than one speaker at a time; indeed it is hard enough to fully perceive all relevant aspects of the speech of just one speaker. This difficulty can be solved if we distinguish the super-stimulus that activates a cortical map from the normal stimuli that are intended to activate it. In the case of harmony, we can suppose that there exists a cortical map that responds strongly to multiple pitch values separated by consonant intervals, but at the same time suppose that the purpose of that cortical map is to respond to just one pitch value at a time. One clue comes from observing how chords can be played: we can play all the notes of a chord at once,

but we also get the effect of the chord by playing the notes of the chord in sequence. So we have proof that the cortical map that responds to multiple simultaneous pitch values can also respond to the same pitch values occurring sequentially.

Of course the actual purpose of this "harmonic/chordal" cortical map will not be to respond to distinct notes of *constant* pitch value separated by consonant intervals, because speech melody does not have this form. We will eventually develop a hypothesis that the purpose of this map is to calculate the durations between points in speech melody that differ by consonant intervals. These calculated durations give partial information about the "shape" of a melody (and in a way that happens to be **pitch translation invariant**, more of which in the next chapter).

8.3.7 Implications for Cortical Maps

What does the musicality perception hypothesis tell us about the cortical maps that respond to music and speech? Can the aspects of music help us understand the nature of the cortical maps involved in speech perception?

These questions have already been partly answered by my answers to the other questions: investigating the aspects of music will enable us to guess the nature of cortical maps that respond to those aspects, and then we will make further guesses as to the purpose of those maps in perceiving normal speech.

The musicality hypothesis will ultimately (see Chapter 14) allow us to place a stronger constraint on the relationship between responses of cortical maps to music and their responses to speech, which is:

> If the activity patterns of a cortical map contribute to determining musicality, then the primary purpose of that cortical map is *not* the perception of musicality, and in fact must be related to some other aspect of speech perception.

Roughly speaking, we can explain this by saying that musicality is a *secondary* aspect of a cortical map's activity, so there must be some other reason that the cortical map exists in the first place.

8.4 Explaining Musical Behaviours

I started this chapter with questions about the purpose of musical behaviours, but then went on to suggest that the real purpose lies within the *perception* of musicality. For the sake of completeness, we should verify that all musical behaviours can be explained this way.

One issue that comes up, when scientists investigate different musical cultures, is that in the more "traditional" (i.e. small tribal) cultures, there is a much greater degree of participation in musical activity. Almost everyone

participates in structured musical performances. There may even be a greater involvement in the composition of music. If we compare this to modern Western culture, where many people's participation in music is to switch on the radio and listen, then Western culture seems to be the odd one out. Maybe any hypothesis about music should be based on the creation and performance of music, rather than the consumption of it, since the majority of cultures give greater emphasis to those aspects of music.

One can argue, however, that the differences between music creation and consumption in small tribes and in the modern Western world have to do with economics, technology and numbers. In the modern world we have easy and fairly cheap access to the best possible music, created and performed by experts who work full time on nothing else. A very small number of composers and performers can do all the work needed to make the music, and everyone else can just listen and enjoy. In a small tribe, if you want to hear good music, you are probably going to have to perform it yourself. So it can plausibly be argued that in both cases—traditional tribal society and modern technological civilisation—the driving force is the desire to *listen* to music.

Even if there are many people in Western society who occasionally perform to the best of their ability (singing not quite in tune in the shower, or singing "Happy Birthday" at a birthday party), most people do not find such performances very satisfying unless they can achieve something that also gives enjoyment to themselves as a listener. (Counter-argument: maybe Western culture—with its emphasis on not even trying to do something unless you can do it really well—artificially discourages the practice required to become musically competent, and in other cultures musical competence may be more commonplace.)

The modern musician may hope to make lots of money, or meet lots of attractive groupies, or just get kudos for being a great entertainer, but achieving all these things rests on their ability to please their audience, which rests on the audience's desire to listen to good music, i.e. music which has a high level of musicality.

Another type of behaviour that has been suggested as showing the purpose of music is behaviour as part of a group, i.e. performing or listening to music in a group.[4] We know, however, that musical performance and listening to music can both be enjoyed on a purely solitary basis, particularly as modern technology allows the listener to enjoy music without immediate involvement by any other person, and the performer can also perform without anyone else performing with them and without any listeners (other than themselves) directly listening to them. It is not clear that music has a social purpose any more than other activities such as eating, drinking or going for a walk in the countryside. Each of these activities occurs socially, but it is not necessarily the *purpose* of any of them to promote social bonding.

[4] As already mentioned in Section 3.4.1 (page 49).

8.4.1 Dance

There is also a relationship between dance and music: given the right type of music many people will enjoy dancing to it, and most dancing is accompanied by some kind of music. Perhaps the real purpose of music is to facilitate or encourage dancing. This of course raises the question as to what the biological purpose of dancing is. Dance can be good exercise, and it can be a social activity, and it also plays a role in sexual/romantic interactions between the sexes. But it is not at all clear why dance should be necessary to further any of these aims—they can all be achieved quite satisfactorily without it.

People also enjoy *watching* dance, and at least part of the reason that people dance is for the effect that it has on those watching. The musical theory that I develop in this book readily explains the multiple aspects of music—melody and rhythm etc.—and relates these aspects to the perception of an individual speaker speaking to the listener. Perception of a speaker speaking is more than just listening to the sounds made; it also involves watching the posture and movements of the speaker. It is entirely possible therefore that dance is a visual super-stimulus relating to the visual perception of movements made by a speaker. We can even incorporate dances involving multiple dancers into this theory, in the same way we incorporated harmony: the cortical map that responds strongly to watching multiple dancers dancing in synchrony has as its primary function the perception of the motion of *one* person.[5]

Including dance in the theory of musicality explains another fact about dance that we take for granted without realising it: there is no such thing as *non-human dancing*. We can make objects, pictures and animals move around to the music. This can look amusing or mildly interesting, but it lacks the emotional effect of watching human dancers dance.

One musical aspect of dance is its obvious relationship to rhythm: whether we dance or watch others dance, we prefer the rhythm of the dance movements to match the rhythm of the music.

Another possible musical aspect of dance, which I can only confirm from personal observation, is an apparent **stepped constancy of motion**. At any given point in time, we will have a general perception of how fast a dancer is moving their body. In some forms of modern dancing, one can observe a smoothness of motion with a subjectively constant speed, sometimes with sudden changes from one speed to another, where these changes are synchronised to the rhythm of the dance and the music. This can almost be interpreted as a form of legato, analogous to the legato of melody. In the melodic legato it is the pitch that steps from one constant value to another; in the dancing "legato" it is the subjectively perceived speed of motion that steps from one value to the next.

[5] Although the effect of dancer multiplicity may be more analogous to the "chorus" effect (which occurs when we hear multiple singers singing in unison, i.e. all singing the melody), than it is to the effect of harmony.

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.

Chapter 10

Musical Cortical Maps

Having found that the nature of music is determined by the response characteristics of cortical maps involved in speech perception, we can make intelligent guesses as to what some of these cortical maps are, based on observations about the structure of music.

Hypothetical maps "discovered" this way include the **regular beat cortical map**, the **harmonic (chord) cortical map**, the **bass cortical map**, the **scale cortical map**, the **home chord cortical map**, the **note duration cortical map** and the **melodic contour cortical map**.

We also discover a similarity between patterns of neural activity in pitch-related cortical maps and those in time-related cortical maps. This similarity is a hint of something deeper going on: a hint as to what musicality actually represents.

10.1 Cortical Plasticity

Cortical plasticity is a term that refers to the ability of areas in the brain to take on different functions. Cortical plasticity is related to a concept of **competitive recruitment**, where the processing functionality required "recruits" an area of neurons to perform that function. There is an ongoing battle between competing functionalities to recruit the most neurons, and the competition is decided by some measure of how important the different functionalities are.

Some well-known examples of cortical plasticity are:

- If a part of the body is lost, for example an arm, the cortical neurons that responded to sensation in that part of the body will be recruited by sensations in other parts of the body, for example a part of the face. The result is that when you touch a certain part of the person's face, they may report that you are touching their arm.[1]

- If a particular sensory input is lost or suppressed, then neurons that processed that sensory input may be recruited by other similar (but unsuppressed) inputs. This type of plasticity can depend strongly on age.

  For example, there are areas in the brain that process inputs from both eyes. A common problem is that some children have a "crooked" eye which fails to align with their direction of sight, and as a result that eye (the **weak eye**) fails to provide useful information to the visual processing areas in the brain, and neurons in the visual areas preferentially develop connections to the other eye (the **strong eye**).[2]

  A crooked eye can often be fixed by appropriate surgery, and the result is that the newly straightened eye is capable of re-forming connections to the visual processing areas, as long as the eye is fixed before the end of a **critical period**. After the critical period (which starts at about age 5 and ends about age 10), the cortical plasticity of these areas is lost, and it is no longer possible for the previously weak eye to recruit the neural connections required to make use of the useful information now coming in from it. (Recovery can be assisted by patching the strong eye for a period of time, which forces the child to make use of their weak eye and the information coming from it. This prevents the vicious circle where they only look at things with their strong eye because they can only "see" with that eye, and they only see with that eye because they only look at things with that eye.)

  It is also observed that temporary suppression of visual input from one eye *after* the critical period does not result in loss of neural connections from that eye.

There is a strong economic flavour to the concept of plasticity. We might imagine a town where all the bakers were killed by some disaster. The demand for bread would motivate some other people, perhaps the cake makers, to move into the business of bread making. And other food manufacturers might take up some of the resulting slack in the cake cooking business, and so on.

Similarly, if for some reason people stopped eating bread, then the bakers would have to consider a change of career, perhaps into the cake business, which in turn would put some of the cake makers out of business (due to increased competition), and so on.

[1] *The Emerging Mind* Vilayanur Ramachandran, page 14.

[2] The medical term for crooked eye is **strabismus**, and the loss of functionality in the weak eye is called **amblyopia**.

If the concept of plasticity was taken to a logical extreme, it would imply that any area of the brain could perform any function. But it is observed in practice that certain functions are always performed in certain areas. Processing of visual information always takes place in areas at the rear of the brain (in the **occipital cortex**). Processing of sound information always takes place in certain areas on the sides of the brain (within the **temporal lobes**). So cortical plasticity does not represent complete freedom to relocate functionality anywhere, but it does represent freedom to relocate functionality to some extent. This constrained relocation can be incorporated into the economic analogy: in principle bakers can bake bread anywhere, but in practice their preferred location is a function of the location of their supplies, like flour and cooking fuel, and the location of the customers who come in to buy their bread, and the locations of buildings that happen to have built-in ovens and gas connections.

We can consider cortical plasticity as a means of allocating resources to a fixed set of information processing functionalities, where the brain is in some way prepared to develop the ability to perform those information processing tasks. But we can also consider it a means of explaining how a brain can allocate resources to information processing tasks that did not exist in the previous evolutionary history of the owner of that brain.

A good example is the set of cortical maps that support the human ability to read. Most people are able to learn to read without too much difficulty, even if none of their ancestors have ever had the opportunity or the need to do any reading. And it seems reasonable to suppose that reading will end up having its own cortical maps devoted to the specific information processing tasks that make up reading, i.e. deciding where to direct the eyes, recognising shapes of letters, recognising sequences of letters as words, translating letters and words into sounds (maybe), and passing this information through to the those parts of the brain that process speech and language. (Although it is likely that some aspects of reading will be implemented in cortical maps that also implement similar aspects of other tasks.)

We invoke cortical plasticity to explain how the relevant areas of a person's brain are recruited in order to perform the information processing tasks related to reading. And we assume that if the person had not learnt to read, then those areas would have been devoted to other information processing tasks.

This relates to the issue of "hard-wiring" and "soft-wiring". Considering neural circuits as our "wires", the question is, for any given circuit, how predetermined is the nature of the information that is represented by activity in that circuit? Just how much plasticity is there? We have already considered the issue of representation of meaning, and now we are asking about how the representation of meaning *develops*, and how flexibly it can change.

The degree of plasticity is going to vary between different parts of the brain and nervous system, and between different functionalities. As we have already

seen, the meaning of activity in a motor neuron connected to a muscle fibre is "contract that muscle fibre", and there isn't really any way that its meaning can change. Similarly for sensory neurons: the meaning of activity in a heat-detecting neuron in the skin at a certain position is "this position in the skin is hot", and there is no way that its meaning can change. In contrast, somewhere in the mysterious inner workings of the human brain there are regions that give us the ability to learn new ways of thinking and understanding, and the ability to develop skills that relate to circumstances that may be considerably different from anything in our past evolutionary history, such as reading. Some other activities that seem to involve information processing somewhat different to anything that our hunter-gatherer ancestors would have done include:

- Playing chess.
- Playing the piano.
- Driving a car.
- Doing mathematics.

The inner regions of the brain must be sufficiently plastic to be able to provide these new types of functionality. Given that our understanding of the organisation and operation of these regions is very limited, it is difficult to know with any certainty what sort of plasticity occurs, and how plastic the corresponding brain areas are. But we might expect that there is a continuum of plasticity, ranging from the hard-wired sensory and motor neurons, to the more flexibly soft-wired inner regions.

If one brain area A receives most of its inputs from another area B whose neurons are mostly hard-wired, the variation in meaning of neurons in area A will be limited by the nature of the connections from region B (where the neurons have relatively fixed meanings).

A similar limitation will occur if a brain area C sends most of its outputs to another area D whose neurons are mostly hard-wired. The general pattern suggested by this reasoning is that those functional maps closest to the external world, (i.e. sensory and motor maps) are the most hard-wired, and the maps connected to those maps are somewhat less hard-wired, and the next layer of maps can be even less hard-wired, and so on.

However, even when there are multiple processing layers, the inner layers may still be quite hard-wired, especially if their function is largely predetermined by specific requirements that have evolved under natural selection over a long period. These areas will have neurons specialised to process the particular types of information they are intended to process. For example, the areas that process specific types of visual information like colour, motion and depth will have evolved to perform the processing of those types of information.

Even an area of the brain that performs a function like recognising faces is likely to have its function predetermined, because of the importance of this task. But the mapping of neurons to individual faces must necessarily be soft-wired, because the set of faces that any individual has to recognise is going to be different for each individual.

10.1.1 Plasticity and Theories of Music

So what does all this talk of cortical plasticity lead to?

Having distinguished different degrees of plasticity, we can ask a basic question about the cortical maps that respond to music:

How plastic are those maps?

Because music is so strange and unlike anything else, it is easy to fall into the assumption that the cortical maps that respond to it are developed from scratch in response to the patterns that occur in music. These cortical maps are assumed to develop in parts of the brain that have a high degree of plasticity.

This describes my thinking when I developed the 2D/3D theory of music. I assumed that the patterns of music determined the development of cortical maps that responded to those patterns. It seemed unlikely that there were pre-existing cortical maps to process things like scales and chords and hierarchical systems of rhythm and tempo.

And it was not implausible to me that if musical intervals had natural 3D representations and 2D representations related by a linear 3D to 2D projection, then the requirements of music perception could recruit neurons in the brain already designed to process that type of projection, regardless of the fact that those neurons were "designed" (by evolution) for visual processing and not auditory processing.

Since (under this assumption) the cortical maps that processed music were not pre-existing maps designed for that purpose, I supposed that musicality was some very generic property of music that translated somehow into a generic property of the response of those musical cortical maps to the music. The 2D/3D theory seemed to supply a plausible candidate in the form of the "80 = 81" paradox (corresponding to the syntonic comma). Somehow, I supposed, the paradox of the syntonic comma gave rise to pleasure and emotionality. I will have to admit, however, that I was never able to plausibly provide any details of that "somehow".

As it happens, the super-stimulus theory of music, based on musicality perception as part of speech perception, also explains musicality as resulting from a generic property of neural responses. But the new theory is quite specific as to the why and how of this property. And, what is more interesting, the new theory does not depend on cortical plasticity to explain how we perceive those features of music that appear to only exist in music.

10.2 Musicality in Cortical Maps

If musicality is a perceived attribute of speech, then it follows that all of the cortical maps that respond to music *are also cortical maps that respond to speech.* Since speech is a significant component of human behaviour, and listening to speech is a significant component of human perception and cognition, it is very likely that many of the cortical maps that perceive and process universal aspects of speech are substantially predetermined in both their location and their functionality.

Thus each identifiable aspect of music is a super-stimulus for one of these predetermined cortical maps that plays a role in speech perception. But, as already discussed in Chapter 8, the corresponding speech aspect may lack some of the qualitative features of the musical aspect. To give an example: a cortical map might respond to multiple notes in music, but that does not mean that its purpose is to respond to simultaneous speech from multiple speakers.

With these considerations in mind, we can proceed to the next major steps in the analysis:

- For each musical aspect, make an intelligent guess as to what type of cortical map would respond to that aspect.

- And, having made such a guess, try to discover a plausible purpose that such a cortical map would have in the perception of speech.

When I started making these guesses about cortical maps, it took me a while to realise the importance of symmetry, and initially I did not take it into account. Perhaps this was a good thing: some of my hypothesised cortical maps seemed to represent information in ways that were unnecessarily obscure and indirect, but then later I realised that these representations made sense given the requirement for perception of speech melody that was invariant under both pitch translation and time scaling. It was a good thing because one hypothesis—that music is a super-stimulus for musicality which is an aspect of speech perception—had caused me to suppose the existence of cortical maps for calculating representations of speech which happened to be invariant under certain symmetries, and this prediction coincided with the implications of a second hypothesis, i.e. that speech and music perception include non-trivial mechanisms to achieve invariance under these symmetries, where the second hypothesis had been made for reasons independent of those for making the first hypothesis. When you are formulating speculative scientific theories based on a limited supply of hard facts and evidence, it's always comforting to discover that different and independent lines of thinking succeed in arriving at the same destination.

10.3 The Regular Beat Cortical Map

Underlying the human perception of rhythm is a basic response to a regular beat, where a regular beat is defined as some constant sound repeated at an exact and constant interval.

It is not too hard to imagine a neuron within a cortical map that specifically responds to such a beat. The following is one possible arrangement of inputs and outputs for a neuron that is intended to respond to a regular beat with a period of 500 milliseconds:

- A direct input of the current sound.
- An input of the current sound delayed by 500ms.
- An input of the neuron's own output delayed by 500ms.

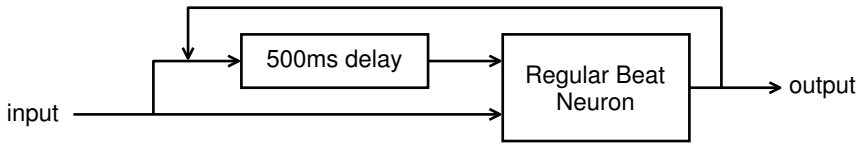

Figure 10.1. A regular beat neuron. The neuron is only activated if both inputs are active: this happens when an input is followed by an input delayed by 500ms, or when an input occurs 500ms after the neuron's own output.

The first two inputs activate the neuron when two beats occur separated by an interval of 500ms. The third input reinforces the neuron's firing when the beat occurs regularly. (Since the second and third inputs both require a 500ms delay, it is possible that they are combined *before* entering the delay, so that only one delay unit is required instead of two. This is how the delayed inputs are shown in Figure 10.1.)

Our perception of a regular beat is able to "jump over" missing beats. For example, if we program a drum machine to emit beats at a particular regular interval, we can detect the regularity of the beat, and are sensitive to any errors or changes in the timing. If we further program the machine to randomly omit a certain portion of the beats, we can still detect changes or errors in the underlying regular beat. This implies that we maintain an internal beat that helps us to fill in the missing beats.

If we are listening to a regular beat with a period of 500 milliseconds, and then one beat is omitted, there will be a gap of 1000 milliseconds. Can the regular beat neuron fill in the missing beat somehow? It is the third input in the list above that enables the neuron to maintain its own internal beat: if an external beat is omitted, the neuron will still have responded to the input

from the previous beat, and at the time of the omitted beat, the delayed input from the neuron's own output will stand in for that omitted beat.

One problem with this simple model is that if a regular beat neuron can fire in response to its own delayed input, then once it starts firing it will continue firing forever, once every beat period, whether or not there is any further input signal.

A solution to this problem is to be able to generate an output signal of varying intensity. That way there can be an output signal in response to a delayed previous output signal, whether or not there is a current input signal, but at the same time the output signal will be stronger if there is both a current input signal and a delayed output signal (delayed by the correct period). If no further input signals occur, then the output signal will repeat by itself, but will gradually fade away.

This solution almost works. But individual neurons usually represent different intensities of signal by how often they fire, and unfortunately the regular beat neurons are constrained to fire on the beat.

One way to solve that problem is to represent an output signal of varying intensity by a group of neurons, such that intensity is represented by the percentage of neurons in the group that fire at one time.

Thus a given beat period is represented by a group of neurons, all of whose outputs feed into each other as delayed inputs, and all of which also take the current input signal as a direct input. The probability of each neuron in the group firing is then a function of how much input signal there is, and how many delayed output signals are received from other neurons in the group. (This idea is similar to the **volley principle** that applies to phase-locked neurons representing a frequency of sound. However, given that the frequency of regular beats is much lower than the frequency that neurons can fire at, it may not be absolutely necessary for neurons representing beats to fire at the exact millisecond the beat occurs: it may, for example, be sufficient to represent a beat by a short burst of firings, in which case the number of firings in the burst represents the perceived intensity of the beat.)

To properly recognise a regular beat of a given period, the regular beat neurons have to do more than be excited by a delayed output signal and a direct input signal—they have to be inhibited by an output signal delayed by the wrong amount (i.e. by a period which is not a multiple of the beat period). This inhibition means that a regular beat neuron can only respond to one phase of a regular beat at any particular time, because if it responds to two different phases of beat, one phase will inhibit the response to the other phase.

One way to achieve the required inhibition is to have an additional inhibitory connection from the output of the regular beat neuron which inhibits the neuron's response to new inputs that occur too soon after the output activation, i.e. less than 500ms afterwards (see Figure 10.2). In effect the neuron chooses its phase when it fires, and the additional inhibitory connection has

the effect of making it insensitive to inputs with the wrong phase.

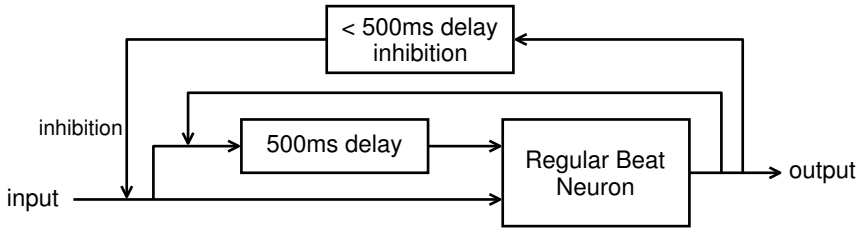

Figure 10.2. Adding inhibition derived from output signal of regular beat neuron to suppress response to out-of-phase inputs.

The neuron in the example given above is one that responds most strongly to a beat with period 500ms. To explain our perception of all the components of musical rhythm, from bar length down to fractions of a note, and our perception of music that plays at different tempos, we will have to suppose the existence of an array of neurons, all maximally responsive to different beat periods.

And, as for all cortical maps, we must take population encoding into account: there cannot be one neuron for each exact beat period, since the set of possible periods forms a continuum. As in other cases of population encoding, the response of a neuron peaks for a particular perceived value, but the response is still strong for values close to but not equal to the neuron's value of maximal response. Thus a particular beat will be represented by the firings of neurons with peak responses near to that beat period, even though none of those neurons has a peak period exactly equal to the beat period in question.

It is interesting to speculate about the mechanism that implements the required delay function, and this might involve a circuit of several neurons specialised for creating fixed delay periods between input and output signals. However, for our current purposes it is sufficient to suppose that a delay *can* be achieved somehow, without worrying too much about exactly *how* it is achieved.

The period with maximal response defines one dimension of the regular beat cortical map. It is possible that the second dimension relates to the timbre of the beat sound. Perception of a regular beat is affected by the similarity or not of the sounds in the beat; for this reason drummers have drum kits with many different drum sounds, and a given accompaniment will contain several percussive timbres, each defining a regular or semi-regular beat at a particular tempo.

Music has a hierarchical structure of regular beats. What does this imply about the pattern of activity in the regular beat cortical map when a person listens to music? There will be a series of active zones in the cortical map, with

each zone corresponding to one period in the beat hierarchy. For example, with music in 4/4 time and containing sixteenth notes, there will be an active zone for each of the following periods:

- 1 bar
- 1/2 bar = 2 counts
- 1 count = a quarter note = a crotchet
- 1/2 count = an eighth note = a quaver
- 1/4 count = a sixteenth note = a semi-quaver

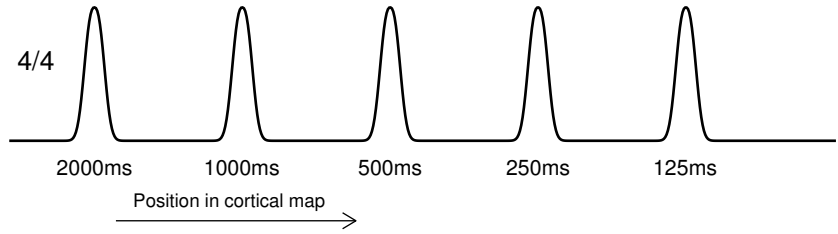

Figure 10.3. Response of regular beat neurons to 4/4 time. The graph shows response of neurons in the map to the beat periods of 1 bar length (2000ms), 1/2 bar, 1/4 bar = 1 count, 1/2 count and 1/4 count.

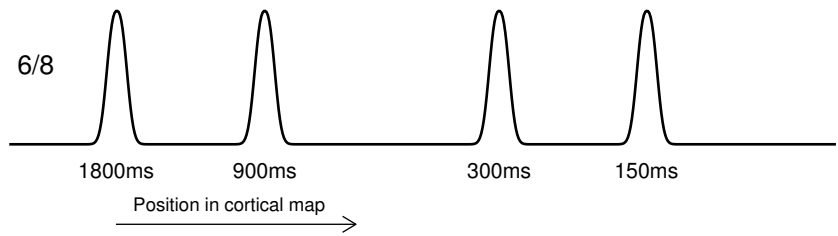

Figure 10.4. Response of regular beat neurons to 6/8 time. The graph shows response of neurons in the map to the beat periods of 1 bar length (1800ms), 1/2 bar, 1/6 bar = 1 count, and 1/2 count.

If we imagine the beat periods to be arranged on a logarithmic scale, then these zones will form a regular pattern, as in Figure 10.3. Not all time signatures, however, are based on powers of 2. There are time signatures where the beat hierarchy contains one or two factors of 3, and in these cases the spacing between zones will not be completely even, as in Figure 10.4.

What will the response of this cortical map be to the rhythms of speech? Since the rhythms of speech are not regular like those of music, there will not be a set of fixed active and inactive zones. Figure 10.5 shows a typical response that might occur to speech rhythm. The peaks are much less pronounced than in the case of the response to musical rhythm. Also the response pattern will change over time, whereas the response to musical rhythm remains constant (except in as much as the tempo gradually changes).

There is one slight simplification that I have made in the diagrams showing response to musical rhythm: all the peaks are the same height. In practice we would expect the peaks to be different heights, depending on how much the rhythm of a piece of music emphasises the different periods in the beat hierarchy. There is, however, a further complication that counteracts this variation, which is that of **saturation**, where any very high peaks get trimmed down to a size that reflects the dynamic range of the neurons in the cortical map. I explain this in more detail when I describe the scale cortical map in Section 10.6.

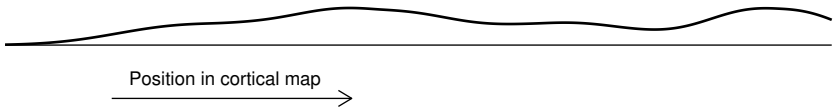

Figure 10.5. Response of regular beat cortical map to an irregular speech rhythm.

10.3.1 Symmetries of Regular Beat Perception

What are the symmetries of this cortical map? It is certainly time translation invariant, because the frame of reference used to define the response of neurons in the map is created by the immediate past activity of each neuron and other neurons close to it in the map—there is no global frame of reference.

This cortical map is *not* time scaling invariant. Slower and faster versions of the same rhythm will activate different neurons in the regular beat cortical map.

But, if we imagine the map laid out on a logarithmic scale (i.e. neurons separated by the same ratio of peak response beat period are separated by the same physical distance), then the effect of speeding up a rhythm by a certain factor will be to shift the activity pattern of regular beat neurons by a corresponding distance along the cortical map. This will happen for both regular musical rhythms and irregular speech rhythms.

So to achieve time scaling invariance, what is required is a second layer of processing that relates activity of each neuron in the regular beat map to activity in other neurons in the map that are fixed distances from that

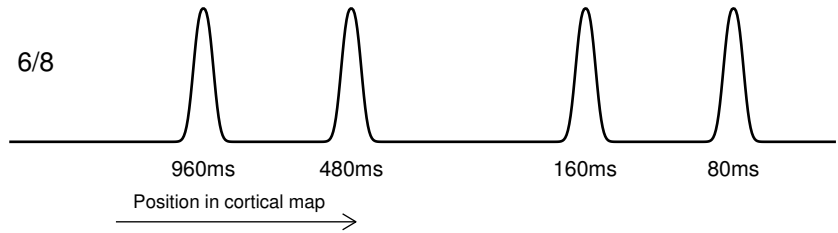

Figure 10.6. Response of regular beat neurons to 6/8 time but with faster tempo. The pattern of responses is the same as in Figure 10.4 except that it has been translated a fixed distance to the right.

neuron. There are various ways this might be done, but our strongest clue will come when we look at the brain's response to pitch information, because the strategies applied to achieve pitch translation invariance can also be applied to achieve time scaling invariance.

10.3.2 Unification

Scientists always like a good **unified theory**. Newton managed to unify falling apples and orbiting planets. Einstein unified time and space and he tried to unify gravity and electromagnetism. Modern physicists have unified two out of the four basic forces,[3] and consider it almost axiomatic that some day they should be able to unify everything.

The most satisfying aspect of my theory of music is that it achieves a convincing unification of time scaling invariance of rhythm perception and pitch translation invariance of melody perception, and does this even though the implied analogies between melody and rhythm are not immediately obvious.

The details of this unification will become apparent as we look at more cortical maps that respond to music.

10.4 The Harmonic Cortical Map

What type of cortical map would respond to chords? From the properties that chords have, we might suppose the following:

- Neurons in the map respond to notes as a function of their pitch.
- The response of the neurons to pitch is octave translation invariant.

[3]The four forces are the strong force, the weak force, the electromagnetic force (itself previously identified as a unification of the electric and magnetic forces) and the gravitational force. The weak and electromagnetic forces have been unified into an "electro-weak" force.

- The response of the neurons to a given pitch value is a function of whether that pitch value is harmonically related to other pitch values that the cortical map is already responding to, and whether those pitch values came from sounds with the same timbre as the sound with the new pitch value.
- Once neurons in the map respond to a given pitch value, they continue to respond to it, until activity in the map is reset (i.e. setting all neurons inactive) in some way.
- Neural activity in the cortical map is reset by a strong beat and by occurrence of a relevant low-pitch value.

These properties of the harmonic cortical map are all reasonable guesses that follow from the observed properties of chords:

- A chord consists of a particular set of pitch values.
- The occurrence of notes in chords is octave translation invariant.
- The notes occurring in a chord tend to be related to each other by consonant intervals, and are usually all played on the same instrument.
- A chord exists within a piece of music for a certain duration, even if not all notes of the chord are being played simultaneously for all of that duration.
- New chords generally start at the beginning of a bar.
- Chords are accompanied by a bass line where the dominant bass note for each bar corresponds to the root note of the chord.

What could be the purpose of this cortical map? We have already noted that, even though chords can be played as simultaneous notes, it is very unlikely that the purpose of the harmonic cortical map is to respond to simultaneous speech from different speakers. It is more likely that its purpose is to identify harmonic relationships between frequencies occurring at different times in the speech melody of *one* speaker.

Why does the harmonic cortical map have a reset function? We can say that specific notes **enter** the map when the neurons that represent those notes become (and remain) active. For example, suppose the map is initially empty (i.e. all neurons are inactive), and we play the notes of the chord C major in sequence. First we play the note C, so the neurons for the note C become active, and C has entered the map. Then we play the note G. Because G is harmonically related to C, it also enters the map, and the neurons for the note G become active. Finally we play the note E. This is harmonically related to both of the notes already in the map, so it also enters. Now we have the map in a state where neurons for the notes C, E and G are all active. If any other

notes try to enter the map, they will not do so easily because they will not be harmonically related to all the existing notes in the map. Of course any re-occurrences of C, E and G will continue to reactivate the corresponding neurons.

(There are additional factors that determine which notes count as part of the chord and which do not—in particular the notes of a chord are typically played on one instrument as notes with a common timbre, which has the effect of grouping those notes together, and presumably this affects the extent to which the harmonic cortical map recognises the mutual relationships between those notes and ignores relationships among notes in the melody which are not in the current chord. Without this grouping effect, chords containing groups of notes not so strongly related to each other by consonant intervals might not be recognised as chords at all. Continuity of timbre across different chords also helps to define which notes are perceived as being part of the chord, so that, for example, notes from the melody but not in the chord do not incorrectly enter the harmonic map.)

As a result of the mutual reinforcement between neurons representing C, E and G, and the inhibition of neurons representing pitch values for other notes, the map will become stabilised into a pattern of activity involving those neurons representing C, E and G. If we assume that information is to be derived from the changing patterns of activity in the harmonic cortical map, then no more information is going to be derived once the pattern of activity becomes stabilised. An easy solution to this problem is to start again: clear the map of all activity, allowing a new set of notes to enter the map.

What event triggers the reset? Empirically we observe that a reset tends to happen when there is a strong beat at the beginning of a long beat period (i.e. the beginning of a bar). Probably it doesn't matter too much exactly what causes a reset,[4] as long as it is an event that happens occasionally, and as long as it is an event that is defined in a manner which makes it invariant under all the relevant symmetries. A beat-based trigger is unaffected by all pitch-related symmetry transformations, and it is time translation invariant and time scaling invariant.

I have described the operation of the harmonic cortical map in terms of how it responds to notes and chords in music. In speech there are no notes held at constant frequency, so the pattern of discrete notes entering the map will not occur with speech melody. What will happen is that, at any point in the speech melody, the current pitch will activate corresponding neurons in the harmonic map, but in a way that depends on which neurons are already active in the map, and on the harmonic relationships between the pitch values of the neurons already active and the current pitch value.

[4]When I say it doesn't matter too much, I mean that evolution could perhaps have chosen some other criterion for resetting, and the map would serve its purpose just as well.

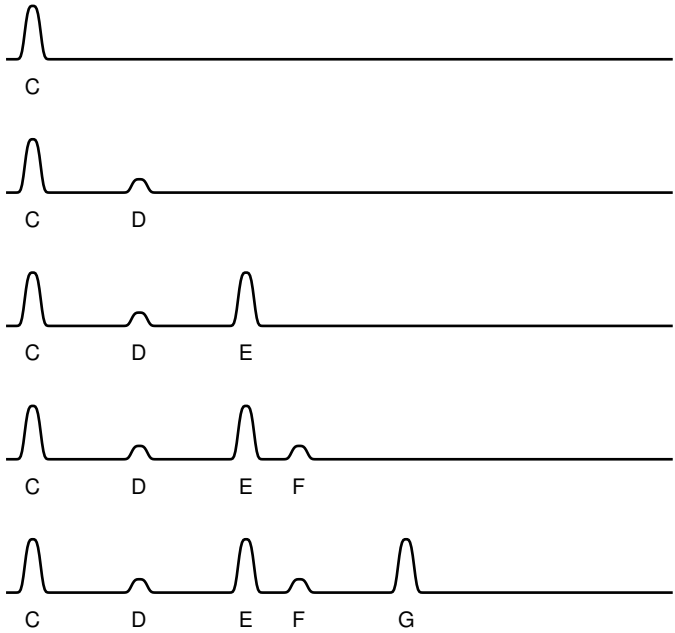

Figure 10.7. Notes entering the harmonic map. First the note C enters the map. No other note is active yet, so it enters unopposed. Then D tries to enter. Because C is already active in the map, and D is not harmonically related to C, the activation of D is suppressed. Next E enters. E is harmonically related to C but not to D. Since C is active in the map, and D is not very active, the consonant relationship with C causes E to be activated in the map. When F tries to enter the map, the dissonance between E and F suppresses the activation of F, even though F is harmonically related to C. Finally G enters the map and becomes activated because it is harmonically related to both the notes already active.

The rhythm of speech will affect the activity in the map in such a way that stronger beats of longer period will cause a reset of activity in the map back to zero (or back to a lower level). To explain the effect of bass, we must suppose that a lower frequency pitch value X activates corresponding neurons in a related bass cortical map, and these neurons in turn predispose activation in the harmonic cortical map of neurons representing pitch values that correspond to low harmonics (1st, 2nd, 3rd, 4th, 5th and 6th) of the pitch value X.

10.4.1 Active Zones

We have already observed the occurrence of active and inactive zones in the activation of the regular beat cortical map by music. We see something similar happening with the harmonic cortical map, but with the difference that the pattern of activity changes suddenly each time the map resets and responds to a new chord. And, as is the case for the regular beat map, the pattern of active and inactive zones in the harmonic cortical map does not occur in response to speech melody—the response to speech consists of changing patterns of activity with a continuous range of activity levels found across the map.

10.4.2 Octave Translation Invariant Representations

We have seen that the harmonic cortical map represents pitch in an octave translation invariant manner. This means that if a C occurs, then no matter which C it is, the same neurons will be activated. To avoid saying "octave translation invariant representation", I will use the simpler terminology that the map represents pitch values **modulo octaves**. "Modulo" is a mathematical term meaning "ignoring multiples of". For example, two numbers are equal **modulo 10** if they have the same last digit. Two musical notes are equal modulo octaves if the interval between them is a whole number of octaves.

10.4.3 Intensity Invariance

The patterns of activity in the harmonic cortical map are not pitch translation invariant. But, if we measure the *intensity* of activation of the currently entering pitch, then this intensity (as a function of time) *is* pitch translation invariant. In effect the degree of activation of a given note in the map constitutes a pitch translation invariant encoding of that note, derived from its relationship to the occurrence of other notes in the melody.

Invariance of intensity can also be found in the regular beat cortical map. In this case we are looking for time scaling invariance. We have noted that the effect of time scaling is to cause the pattern of activity to be translated within the cortical map. The intensity of the response to a given beat reflects the current dominance of the beat period whose corresponding neurons are activated by that beat. In as much as the cortical map consistently relates this dominance to intensity of activation (for different beat periods), the sequence of intensity values will be time scaling invariant.

The analogy to the harmonic cortical map is quite strong—the regular beat cortical map has time scaling invariant intensity and the harmonic cortical map has pitch translation invariant intensity. In the harmonic cortical map, translating the pitch will correspondingly translate the pattern of activity in the map. Because the map represents notes modulo octaves, there will

Figure 10.8. An encoding of the melody CCDEDEFGFEDDC, where each note is encoded according to its degree of activation in the harmonic cortical map as in Figure 10.7. Because the activations are based on the relationships between notes, the sequence of intensities is a pitch translation invariant characterisation of the melody. It is not a *complete* characterisation: in the current example there is no distinction between the encodings of C, E and G which all have the same level of activation in the harmonic cortical map when it is responding to the chord of C major, and there is no distinction between the non-chord notes D and F, which have the same level of activation as each other.

be a wrap-around with this translation, i.e. values translated off one end of the map will reappear in corresponding positions at the other end of the map. But as long as the map consistently represents the strengths of activation and the effects of consonant relationships for different pitch values, the intensity as a function of time will be pitch translation invariant.

10.5 The Bass Cortical Map

Bass notes are tightly coupled to chords. The major function of a bass note appears to be to emphasise the identity of the current chord which has that bass note as its root note. So we can guess the existence of a corresponding cortical map with the following properties:

- It responds most strongly to notes of lower pitch.
- It affects the entry of notes into the chord map, in particular favouring the entry of notes that are equal (modulo octaves) to the bass note, and of notes that are equal (modulo octaves) to other low harmonics of the bass note.

Because a new bass note representing the root note of the chord starts at the same time as the chord, the bass can be understood as helping to trigger the reset of the harmonic cortical map, to deactivate the old chord and start activating a new chord.

The bass cortical map sits in between being octave translation invariant and not being octave translation invariant. The effect of the input is not necessarily octave translation invariant, as the bass cortical map responds most strongly to the lowest note. If I add an octave or two to the lowest note, then it is not going to be the lowest note anymore. But the effect that the output of the bass cortical map has on the harmonic cortical map *is* octave

translation invariant. For example, if C is identified as a bass note, then this reinforces any chord with a root note C, and it does not matter which octave the bass C note was in.

As is the case for other aspects of music, we must remember that music is a super-stimulus. Bass notes in music are often played much lower than the notes in the melody. The only constraint that seems to exist on how low bass notes can go is that we be able to hear them. Speech melody does not include extra notes so much lower than the main speech melody. The response of the bass cortical map to speech melody is such that the response to lower pitch values is greater than the response to higher pitch values. The consequence of this response function is that the super-stimulus for the map consists of notes with very low pitch values.

10.6 The Scale Cortical Map

Scales are a major component of almost all music. This includes not just modern Western music and traditional Western music, but the music of most cultures. As has already been explained in the chapter on music theory, a scale is a set of notes from which a melody is constructed, and scales are normally octave translation invariant.

There are two basic difficulties we encounter when trying to understand how scales relate to perception of speech melody:

- There are no notes in speech melody.
- There are no scales in speech melody (if there aren't any notes, then there cannot exist a scale from which notes are chosen).

By contrast:

- Speech melody consists of pitch (or frequency) as a smoothly varying function of time.
- The only jumps in pitch occur when there is a gap in the voiced sound.

But, as is the case for all other aspects of music, the dissimilarity between the musical aspect and the speech aspect is not a fatal obstacle in our plan to relate the musical aspect to the perception of speech. We must try to discover a cortical map that responds to the musical aspect, and then see if the same cortical map serves a useful purpose in the perception of speech.

If we regard the scale as being a property of a melody, then we can ask what is required to perceive this property. The scale is determined according to the occurrence of notes in the melody, but independently of the times at which they occur. This suggests a cortical map that responds to the occurrence of pitch values, and continues responding indefinitely, so that it builds up a picture of the full set of pitch values that have occurred in the

tune. In a continuous speech melody, each pitch value only occurs for an infinitesimal period of time, so the "set" of pitch values is better thought of as a continuous density function. Another consideration is that we would not expect the neural response to last too long, since the details of any melody (speech or musical) eventually become irrelevant (and the cortical map needs to clear itself so that it can process new melodies). The response will have to "fade away" at some finite rate.

So we have a cortical map with the following characteristics:

- Neurons in the scale map have a response to pitch modulo octaves. (This immediately explains why scales are octave translation invariant.)
- Neurons are activated by pitch, in inverse proportion to the speed at which the pitch in the melody is changing.
- The activation of neurons by incoming pitch values decays slowly. We might expect the rate of decay to correspond to the time scale of a spoken sentence.

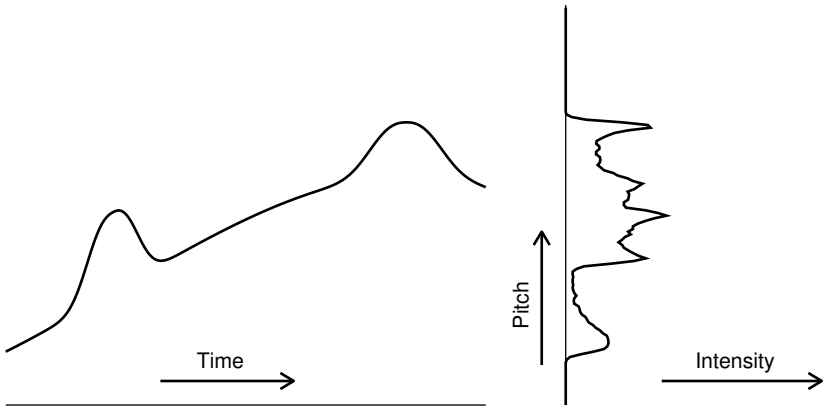

Figure 10.9. Response of the scale cortical map to a smooth melodic contour. The graph on the left is an arbitrary melodic contour as a function of pitch against time. On the right is the graph of intensity of activation in the scale cortical map as a function of pitch. The graph of intensity against pitch is rotated and reflected so that it shares a common pitch axis with the first graph. The intensity is greater for those pitch values where the contour is increasing or decreasing more slowly, and for those pitch values that occur more than once in the melody.

As in other descriptions of cortical maps, the statement that neurons respond to particular pitch values is subject to the caveat of population encoding: i.e. each individual neuron has a response that is a function of how

close the incoming pitch value is to the pitch value that the neuron has a peak response to.

So what will be the response of this scale map to music which is composed from notes on a scale?

- Because musical notes do not move up or down, but remain constant for their duration until they change to a new note, the activation of the neurons for those pitch values will be at a maximum rate.
- Because pitch values between the notes on the scale do not occur, the neurons for those in-between pitch values will not be activated.

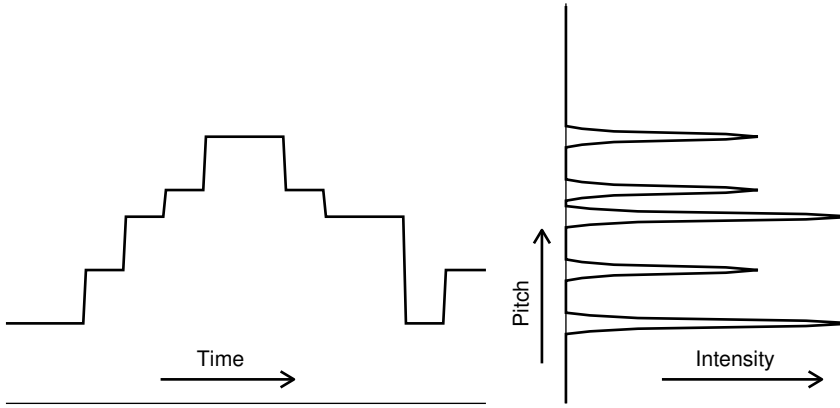

Figure 10.10. Response of scale map to musical melody. The same algorithm is used to calculate the response of the cortical map to the melody as in Figure 10.9. However, because the musical melody consists of notes held at fixed pitches selected from a finite set of values, the response of the cortical map consists of high activity of some neurons and very low or zero activity of other neurons.

The result will be a series of active zones separated by inactive zones. We have already seen this pattern with the regular beat cortical map and the harmonic cortical map. Seeing the same pattern occur in maps that relate to qualitatively different aspects of music strongly suggests to us that there is something deeper going on here. We may not be so far from a full answer to the question "What is music?". But for the moment I will carry on with the analysis of individual maps.

The active zones in the scale cortical map are very likely to be **saturated**. The response to pitch values is inversely proportional to the rate of change (of pitch): in effect the cortical map is measuring how much each pitch value has occurred as integrated over time. The constant notes of music are unnatural, and it is therefore likely that the degree of activation will go outside the

dynamic range of the cortical map, since the cortical map was not designed (by natural selection) to deal with the contrived extreme patterns of musical melody. What happens when a neuron attempts to encode a numerical value that goes outside its normal dynamic range? The most that could be expected of a neuron in this situation is that it fire at its maximum possible rate.

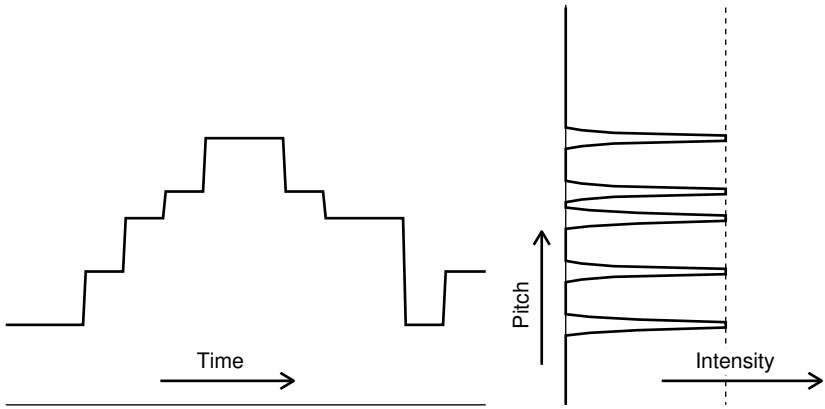

Figure 10.11. The effect of saturation on the response of the scale cortical map to music. The dashed line on the right shows the maximum intensity represented in the map (corresponding to the maximum rate of neural firing). Saturation occurs when the intensity function is capped by this maximum value. Saturation does not occur with smooth speech melodic contours because the measured intensities are lower.

It is likely that similar phenomena of **saturation** will be found in the regular beat cortical map and in the harmonic cortical map.

We can also consider the invariances of intensity for the output from the scale map. Intensity will be pitch translation invariant, for similar reasons to those that cause invariance of intensity in other maps. But because activation for a given pitch value depends on the rate of change in pitch, the intensity of the output will not be time scaling invariant for speech melody (for musical melody it will be invariant due to the saturation, but that is not relevant to how the cortical map provides invariant perception for the case of speech). Only the *relative* intensities will be time scaling invariant; i.e. there will need to be further processing that provides a final output value invariant under the operation of scaling the intensity of the scale map's output by a constant factor.

10.7 The Home Chord Cortical Map

Most simple tunes in Western music have a home note and a home chord. The tune starts with the home note and chord, moves on to other notes and chords, and eventually returns back to the home note and chord.[5]

The home chord for a tune always contains notes from the scale that the tune exists in, but different tunes on the same scale can have different home chords. However, for a given scale, most tunes on that scale have a home chord taken from a very limited set of choices. In particular, on the diatonic scale, the home chord is almost always one of two choices. On the white notes scale, these choices are A minor or C major.[6] We can postulate that the tendency to have one of these two chords as the home chord is an intrinsic property of the scale itself. This makes it much easier to speculate about the forces that determine which chord becomes the home chord, because a scale is a much simpler thing than a piece of music.

Given that the home chord of a tune played on the white notes scale is either A minor or C major, what determines which of these two it is? A very simple rule appears to work in all cases that I know of: whichever of the two chords occurs first.

The set of possible home chords for a scale is obviously a pitch translation invariant function of that scale. To put it another way, the choice of notes from the scale to make a home chord must depend entirely on the relationships that the notes in the scale have with each other. We can also mention that since the home chord and the scale are both octave translation invariant, the processes that determine the home chord must also be octave translation invariant.

An important determinant of the home chord of a scale must be the very unevenness of the scale. The diatonic scale is invariant under an octave translation (or multiple thereof), and *it is not invariant under any other translation.* If, for example, it consisted of two identical halves, so that it was invariant under a translation of half an octave, then the set of possible home chords would have to be invariant under the same translation (of half an octave).

We could assume, as a first approximation, that the choice of home chord is determined by the relationships between notes considered pairwise: it is always simpler to connect items of information two at a time. Relationships between pairs of notes can be identified by two main criteria:

1. How close the notes are to each other, i.e. proximity relationships.

2. If the notes are related by harmonic intervals.

[5] However, both the home note and home chord may occur in the middle of the tune, separately or together, without the tune being finished at that point.

[6] And on some of the variants of the white notes scale, where G is replaced by G♯ and optionally F is replaced by F♯, there is really only one choice, which is A minor.

Which of these two criteria has the most influence on the choice of home chord?

If we look at the white notes scale, the strongest (i.e. most common) choice of home chord is C major. If we look at the immediate environment of the note C, the three steps below it (going upwards from G to C) are tone, tone, semitone, and the two steps above it are tone, tone. The only other note that has this environment is the note F. But F does not occur as a home note for melodies in the white notes scale. It seems that we can therefore rule out proximity relationships as a major contributor to determining the choice of home note or home chord.

The second possibility is to consider harmonic relationships. We have already found a concise way to present and view all harmonic relationships between notes on the white notes scale: the **Harmonic Heptagon**. When we look at the location of the two possible home chords on this diagram (as in Figure 10.12), the following aspects are very suggestive:

- The notes of the two possible home chords exist on one side of the diagram, opposite the location of the note D.
- The Harmonic Heptagon has a reflective symmetry, and A minor and C major are mirror images of each other under this symmetry.
- The note D, as well as being the centre of symmetry, and opposite the home chords, is surrounded by the notes B and F, which are the two notes that have fewer harmonic relationships between themselves and other notes. The interval between B and F is not a consonant interval, whereas all other intervals between notes two steps apart on the heptagon are a perfect fifth.

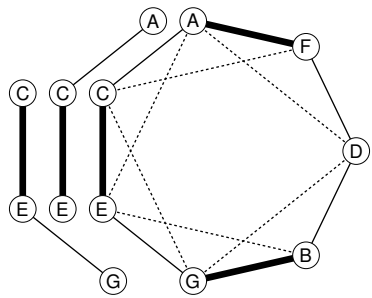

Figure 10.12. The preferred home chords (for the white notes scale) C major (CEG) and A minor (ACE) and their positions in the Harmonic Heptagon.

These observations suggest the existence of a home chord detecting cortical map with the following properties:

- Neurons in the map represent pitch modulo octaves.

- Neurons representing one pitch value will reinforce the activity of neurons representing another pitch value if the interval between the two pitch values is consonant.

- The level of reinforcement (as a function of interval size) is the same between two notes in each direction—this underlies the reflective symmetry.

- Notes that are not harmonically related to each other mutually inhibit one another in the map.

This results in the observed home chords of C major or A minor as follows:

- The notes B, F and D are weak in the map (i.e. the neurons representing those notes are only weakly activated) because the missing consonant interval between B and F weakens the mutual reinforcement between those notes and other notes.

- The mutual inhibition between neurons representing notes not harmonically related to each other means that only one of A or G can be in the home chord. Thus, given that the home chord tends to contain the notes A, C, E and G, it will be either A, C and E or C, E and G. This is the broken symmetry referred to earlier—"broken" in the sense that the set of possible home chord notes A, C, E and G has full reflective symmetry, but one of two non-symmetric subsets of this set must be chosen to be the actual home chord.

- Given that the tune is being played on the white notes scale, the home chord map can stabilise in one of two states: activation of neurons representing A minor, or activation of neurons representing C major. The competition for the two states is won by whichever chord has its neurons activated first, i.e. which of the two chords is played first.

As already mentioned, the choice of home chord is very much a function of the scale, and not of any details of the tune, except for determining the choice between the two possible chords. In particular, the strength of activation of neurons for a given note in the home chord map is independent of how many times that particular note occurs, which suggests that the inputs to the home chord map are filtered through some cortical map that responds to the occurrence of notes, while ignoring how many times or for how long those notes occur. But this describes the scale cortical map whose existence we have already hypothesised. We conclude that the home chord map probably receives its input from the output of the scale map.

10.7.1 Why Reflective Symmetry?

If there really does exist a musical symmetry of pitch reflection, it appears to be derived from a symmetry of mutual reinforcement between neurons representing notes as a function of the interval between them being consonant, or not, as the case may be. That is, the reinforcement (or inhibition) of activity in neurons representing note X by activity in neurons representing note Y is equal to the reinforcement (or inhibition) of activity in neurons representing note Y by activity in neurons representing note X.

Is there any particular reason why this symmetry needs to exist? We can consider mutual reinforcement as giving rise to an iterative voting system. Notes vote for and against each other, and the more votes a note receives in its favour, the more its own votes count for. Asymmetrical reinforcement could give rise to instabilities in this iterative process, and these instabilities would result in artefacts of changing activation independent of the incoming data.

For example, instead of the Harmonic Heptagon, we might have a scale that was approximately even, with all notes separated by the same interval. No note would occupy a special position relative to the others, and on average each note would be equally reinforced by all the other notes. But if reinforcement was stronger in one direction around the circle than the other, e.g. reinforcement of notes separated by a certain interval was stronger going clockwise than anticlockwise, then we would get "waves" of activation travelling clockwise around the circle. As a result the patterns of activity in the cortical map would fail to stabilise, or would take an undesirably long time to stabilise, and the travelling "waves" would be sensitive to minor variations in the input data. These properties conflict with the requirement that a frame of reference should be stable and insensitive to small changes in the data.

10.7.2 Alternative Theory: The Dominant 7th

An alternative theory of the preferred home chord is based on the observation that the white notes scale contains the dominant 7th (GBDF) and the chord it normally resolves to, C major (CEG), which is the home chord. This theory possibly better accounts for the choice of A minor as home chord for the harmonic minor scale (A, B, C, D, E, F, G♯) and the melodic minor scale (A, B, C, D, E, F♯, G♯), since these scales contain both the dominant 7th (EG♯BD) and the chord A minor (ACE).

In effect this alternative theory states that the preferred home chord is determined by more than just pairwise relationships. The full story may be a mixture of this theory and the theory based on pairwise consonant relationships. Note that A minor can be the home chord for tunes on the white notes scale that do not have the chord (EG♯BD) appear at all, so the dominant 7th is not essential to the "hominess" of the home chord.

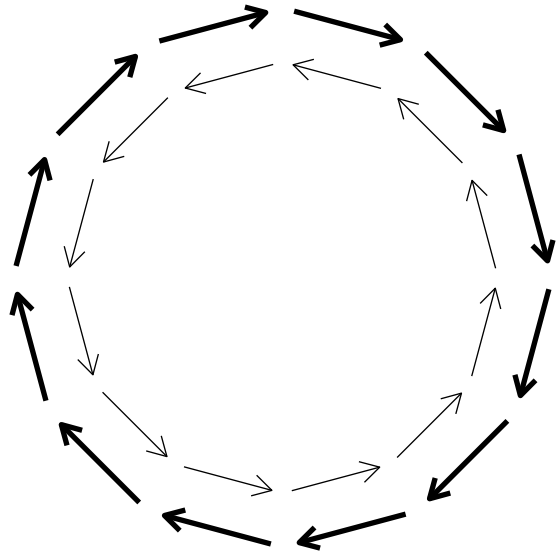

Figure 10.13. A pattern of mutual reinforcement which is pitch translation invariant but not pitch reflection invariant (the cycle represents one octave modulo octaves). Reinforcement is stronger in one direction than in the opposite direction. This will cause waves of activation to travel in the dominant direction.

10.7.3 The Evolution of Cortical Maps

The home chord cortical map is somewhat similar to the harmonic cortical map. Both maps represent pitch modulo octaves, both have an activation of neurons that persists after the occurrence of the relevant pitch, and both have mutual reinforcement between consonantly related notes and mutual inhibition between notes not consonantly related.

Here is a list of differences:

- The home chord map does not have any reset function. Thus the home chord of a simple tune played on a constant scale remains constant for the whole tune.

- The mutual inhibition between notes not consonantly related to each other is much stronger in the home chord map. It is not possible to have four notes in a home chord, whereas in general four note chords occur quite commonly in music, and five note chords are not unknown.

When we see groups of cortical maps that are not exactly the same, but somewhat similar in their properties, then this has an obvious evolutionary interpretation: at one point in the history of our species there only existed

one map, but then this map evolved into multiple copies of itself, and each copy evolved useful variations in the activation characteristics of its neurons.

This is not the only example of musical cortical maps that may be relatives of each other in evolutionary terms. We will see this next when we investigate the perception of note length.

10.8 The Note Duration Cortical Map

Remember that our first model of the regular beat cortical map contained neurons with three inputs:

- Current input signal
- Current input signal delayed by beat period
- Current output delayed by beat period (feedback input)

It was the last input which gave these neurons the ability to detect ongoing regular beats, in a manner robust to omission of occasional beats.

If we omit the feedback input, then what we are left with is a neuron that responds to a particular duration of note, if we regard the duration of a note as being delimited by a beat at the beginning of the note and a beat at the end of the note (which is usually also the start of the next note).

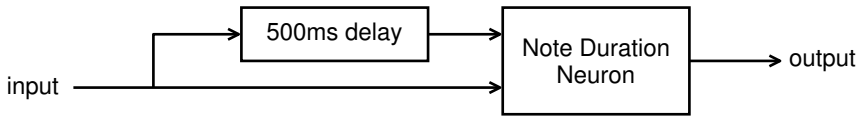

Figure 10.14. A neuron that responds to a duration between the start of one note and the end of that note (or the start of the next note) of 500ms. Compare to Figure 10.1 which had one additional feedback connection from the output to the input of the delay unit.

Note length is something that we are consciously aware of when we listen to music, so it is not surprising that there should exist a cortical map that responds directly to note length.

With note length, as with regular beat period, the information resulting from this processing layer is *not* time scaling invariant. By one means or another, the information from this layer must be processed in a way that measures *relative* note lengths.

Because the same invariance applies to regular beat period and note duration, it is possible that the processing which produces time scaling invariant characterisations may involve combinations of both, i.e. comparison of beat

periods to note durations in addition to comparison of beat periods to beat periods and comparison of note durations to note durations.

The similarity of neurons in the note duration map and the regular beat map suggests that the two maps evolved from a common ancestral cortical map.

10.9 The Melodic Contour Cortical Map

We have discussed four cortical maps relating to melody and pitch: a scale map, a harmonic map, a bass map and a home chord map. None of these maps actually contains any response to whether the tune is going up or down. In fact they all process pitch information modulo octaves, and an interval between two pitch values considered modulo octaves may be considered as either an interval going up or an interval going down.

We are, however, quite aware of whether a melody—be it speech or music—is going up or down. Speech melodies mostly go up and down in a smooth fashion. Musical melodies also tend to go up and down smoothly. They cannot go up and down completely smoothly, because the notes are taken from a discrete scale. But they do go up and down as smoothly as possible, in as much as the next note after a note in a melody is very often the same note, or just one step higher or lower.

Responding to the up and down motion is an easy way to produce a characterisation of melody that is pitch translation invariant. It has robust translation invariance, because it is invariant under any transformation that is **monotonic**. A transformation is monotonic if it preserves the distinction between intervals going up and intervals going down. A monotonic transformation does not necessarily preserve any particular notion of size of interval. It follows that characterisation of up and down motion of pitch does not require any special calibration, whereas (as we will see in Chapter 12) pitch translation invariant characterisations of melody depend on accurate comparisons of intervals between different pairs of notes.

The perception of melodic contour gives us an explanation of why melody is not **locally octave translation invariant**: we cannot translate individual notes of the melody by different numbers of octaves, in the way that we can do for notes in chords and bass, since adding octaves to some notes but not others would radically change the contour of the melody by changing the sizes of the steps between consecutive notes.

Chapter 11

Octave Translation Invariance

Octave translation invariance is a symmetry that applies both to musical scales and to individual notes within chords.

This invariance does not appear to satisfy any functional requirement. Rather, it appears to facilitate the efficient subtraction of one pitch value from another to calculate the size of the interval between them. In particular, the brain separates each pitch value into a precise pitch value modulo octaves and an imprecise absolute value, and performs subtraction separately on each of these components.

11.1 Octave Translation Invariant Aspects of Music

The following aspects of music are octave translation invariant:

- Chords and notes within chords can often be raised or lowered by an octave without significantly affecting the musical quality of a piece of music. The same applies for bass notes.
- All Western musical scales repeat themselves each octave. This rule also applies to most non-Western musical scales.
- Home chords and home notes are octave translation invariant.

- Two musical notes separated by an octave, or a whole number of octaves, have a similar perceived quality.

In all these cases we can suppose that the pitch value of musical notes is represented by the pitch value *modulo octaves*, in the sense that information about the position of the note within its octave is retained, but information about *which* octave the note is in is thrown away.

Information about octaves is not thrown away in *all* places where pitch information is processed: we know that the notes of a melody cannot be individually raised or lowered by an octave. This relates to the contour of the melody, which describes how the pitch goes up and down at different times. And, subjectively, we know that, although two notes separated by an octave sound partly the same, we can still tell that one of the notes is higher than the other.

11.2 Separation of Concerns

A common mode of operation in the brain is the separation of information into components. As previously mentioned, visual processing involves separation of information into components of position, motion, depth and colour, so that each component can be effectively processed by specialised processing areas.

We might suppose that something similar is going on with pitch: a separation into a component modulo octaves and a component that retains octave information. However, compared to other decompositions of information that occur in the brain, this particular decomposition has a rather unusual mathematical nature: an apparently simple continuum of possible pitch values is decomposed into a modulo value and a non-modulo value. What, if anything, is the point of such a decomposition?

11.3 Digital versus Analogue

How does an electronic computer represent values that can be represented as numbers from a continuum? Typically such values are represented as **floating point** values. A floating point value consists of a **mantissa**, which is a finite number of digits, and an **exponent**. In a computer the digits are normally base 2, i.e. either 0 or 1, but it will not matter too much if we pretend that they are actually decimal digits. The exponent can be thought of as telling us where the decimal point is in relation to the digits.

Examples:

- "1.023e6" means $1.023 \times 10^6 = 1{,}023{,}000$. "1.023" is the mantissa and "6" is the exponent.
- "2.54e-3" means $2.54 \times 10^{-3} = 0.00254$. "2.54" is the mantissa and "-3" is the exponent.

This floating point representation represents numbers with a certain precision determined by the number of digits. The range of values for the exponent allows for very small and very large numbers to be represented: the programmer of the computer can usually choose a standard floating point format which can represent all the numbers required to be represented and processed in their program, to a sufficient degree of accuracy for the purposes of the program.

The brain as computer must process perceptual values that, in software running on an electronic computer, would normally be represented by numbers, but, as we have already noted, the constraints of natural evolutionary design are not quite the same as those of human industrial design. In particular, the representation of numerical values in cortical maps is much more analogue than occurs in digital computers.

Firstly, there is never any recognisable division between mantissa and exponent. If the range of values required to be represented does not include very large or very small values, then there is no need for an exponent. In the cases where there is a large dynamic range (as with the perception of loudness), then the representation is effectively exponent only. This becomes a representation on a logarithmic scale.

Secondly, numerical values are not represented as finite sequences of digits. Most values are represented in terms of neurons that lie sequentially within a map, such that each neuron represents some particular value. "In-between" values are represented by means of population encoding.

Digital representations are very compact. High levels of precision can be represented in a small number of components. For example, the level of precision in human perception never exceeds 10000 values in a 1-dimensional range of values, and 4 decimal digits would be enough to store a value from a set of 10000 possible values.

In the case of pitch perception, there are about 10 octaves in the range of human hearing. Accuracy of pitch discrimination in those portions of the range with the most sensitivity (about 1000Hz to 4000Hz) is about 0.3%, or 1/240 of an octave. If this level of discrimination applied over the full range of hearing, we would be able to discriminate 2400 different pitch values. But the level of discrimination is reduced somewhat for higher and lower pitch levels, and the maximum number of distinguishable pitch values is closer to 1400.

If, at some point in the brain, the set of possible pitch values was represented by 1 neuron per pitch value, then we would need 1400 neurons to represent them. Now 1400 is not a large number of neurons. But the difficulty begins when we consider the need to calculate *relative pitch*. As we have already noted, many aspects of the perception of music are pitch translation invariant. To achieve pitch translation invariance, it is necessary, by one means or another, to compare different pitch values, and in particular to calculate the interval between two different pitch values.

A digital computer requires just 11 binary digits to represent a number from 0 to 1400. The computer can subtract an 11 bit number from an 11 bit number to get another 11 bit number (we'll ignore overflow here), using a subtraction circuit containing some small multiple of 11 bits, probably 22 or 33.

How much circuitry will it take our brain's analogue neural network to do subtraction between these values? The naïve answer is: $1400 \times 1400 = 1,960,000 \approx 2,000,000$. (I have rounded this to a simple 2,000,000, because all the numbers here are very rough.) Why so many? We need this many neurons because we need to wire up each pair of neurons representing a pair of input values to an intermediate neuron representing that particular subtraction problem, and then we need to connect each of these intermediate neurons to the corresponding neuron representing the answer. In effect the 2,000,000 neurons constitute a giant subtraction table. (Figure 11.1 shows a 4×4 subtraction table that implements subtraction of pitch values from a range of just 4 possible values, with $4 \times 4 = 16$ intermediate neurons and 7 output neurons.)

Now 2,000,000 is a non-trivial number of neurons. Perhaps not a large number in terms of the brain's total, but still large in terms of the calculation being performed. (There may also be a need for more than just one such subtraction table. We have already determined the existence of two musical cortical maps that process consonant relations between pitch values—the harmonic cortical map and the home chord cortical map—and each such map would require its own subtraction table.)

Even if providing 2,000,000 neurons is not a problem, correctly developing all the connections between the inputs and outputs and calibrating them might consume excessive resources. (More on the subject of **calibration** in the next chapter.)

In computer science terminology, we have $O(N^2)$ complexity[1] for a problem that really only requires $O(logN)$ amount of circuitry.

11.4 Digital Representations in the Brain

So, assuming that the required size of one or more subtraction tables for pitch values might impose a significant cost on the individual, can some of this complexity be reduced by using the digital solution?

To explore this possibility, I am going to analyse the problem of how to represent a series of values from 0 to 99 by separating each value into a first decimal digit and a second decimal digit.

We can assume that the original value would be represented by 100 neurons. The separate digit values would be represented by 10 neurons for the

[1] Reminder: **complexity** refers to usage of resources, *not* how complicated the problem is.

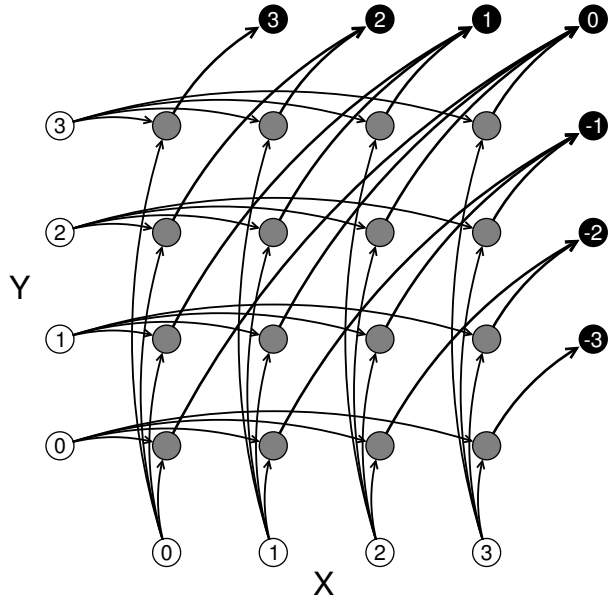

Figure 11.1. A neural subtraction table for the problem $Y - X$. Each circle represents one neuron. The white neurons are inputs representing values for X and Y. The black neurons represent the answer, and each gray neuron represents one subtraction problem. Population encoding allows the neural network to solve problems involving "in-between" values.

first digit and 10 neurons for the second digit. We have reduced the required circuitry from 100 neurons to just 20 neurons.

There is one basic problem with this simple separation, which is the general imprecision of representation of values by individual neurons. As discussed when I explained **population encoding**, each neuron represents a range of values, and each value is correspondingly represented by the activation of a range of neurons. This causes problems when we try to split the value from 0 to 99 into two values each from the range 0 to 9.

Consider, for example, a value 39.5. In the 100 neuron representation, the most active neurons will be those that maximally respond to 39 and 40, with lesser activation of those neurons active for 38 and 41, and even less for 37 and 42, and so on. No problem here: we can easily reconstruct the value 39.5 from this pattern of activity.

But now consider the separation into two digits. In the first digit, there will be neurons representing 3 and 4. Since 39.5 is in between the ranges of numbers with 3 as a first digit and 4 as a first digit, we would expect these 2 neurons to be equally active. Still no problem.

Now consider the second digit. The most active neurons will be those for 0 and 9. This represents a value between $X9$ and $Y0$, where Y is the next digit after X.

The problems begin when we try to reconstruct the original full value. The first digit is maybe 3 or 4, the second digit is maybe 9 or 0. This implies that the reconstructed number might be 39 or 40 or 30 or 49. Now 39 and 40 are good estimates, but the values of 30 and 49 are completely spurious, and nowhere near the real value.

One diagnosis of the cause of this problem is that the split of information between the first digit and the second digit is an exact split, with no sharing or overlap. This is fine in a digital computer, where the design relies on discrete components that represent discrete values with 100% reliability, but it doesn't work in neural networks where information is represented in a fuzzy manner shared between different components. If fuzzy information is to be split so that the original fuzzy information can be reliable reconstructed, then the splitting itself has to be fuzzy. This means that there has to be an overlap between what the first digit represents and what the second digit represents.

One way to do this for the 100 value example is to have the second digit be a number from 0 to 9 representing the original value modulo 10, as before, but have the first digit be a number from 0 to 19, representing the number of 5's. So 39 is represented by "79", and 40 is represented by "80".

What happens when we split and reconstruct? The reconstructed number becomes one of "70", "79", "80" or "89". In this case we still have two valid values, i.e. "79" and "80", and two spurious values "70" and "89". But this time the spurious values are intrinsically invalid, and the system can be wired to ignore them. For example, a first digit of 7 implies a number in the range from 35 to 39, and none of these numbers ends in 0, so "70" is an invalid number. Similarly a first digit of 8 implies a number in the range 40 to 45, so "89" does not represent a valid number.

This overlap between what the first digit represents and what the second digit represents introduces some redundancy, so there is less reduction in the number of neurons required. We have $20 + 10 = 30$ neurons, instead of $10 + 10 = 20$ neurons, but this is still less than 100 neurons.

We can now calculate the reduction of the size of the subtraction tables using the fuzzy split representation: ignoring the details of wrap-arounds and overflows, the original representation requires $100 \times 100 = 10000$ neurons to do subtraction, whereas the fuzzy split representation requires $20 \times 20 + 10 \times 10 = 400 + 100 = 500$ neurons, which is considerably fewer.

11.5 Split Representation of Pitch

The previous analysis suggests that the representation of pitch information is such that pitch values are split into two components:

- A pitch value modulo octaves, which has maximum precision.
- An absolute pitch value which is less precise.

Exactly how imprecise is the imprecise absolute pitch value representation? There is no obvious way to measure this, because the combined effect of the two representations is always equivalent to a representation of a single precise pitch value. From our analysis we would expect that the average error of the absolute pitch value representation is somewhat larger than the average error in the representation of the pitch value modulo octaves (because the absolute value representation is the imprecise first "digit") and somewhat smaller than an octave (because the split into two "digits" is fuzzy).

It is possible that neurological patients exist who have suffered some type of localised brain damage, and who can be identified as having lost the modulo-octaves representation of pitch. If these patients still have some degree of pitch perception, then the accuracy of their pitch discrimination could be an indicator of the accuracy of the absolute component of the split pitch value.

It might be supposed that our ability to detect up and down motions in pitch is tied to the absolute imprecise component. However, in 1964 Roger Shepard published a paper "Circularity in Judgments of Relative Pitch", which described a sequence of tones in which the pitch value modulo octaves rises forever. Such a sequence is indeed perceived as rising forever, even though it is completely repetitive. The basic trick in constructing these tones is that the only harmonics are those with frequencies which are multiples of the fundamental frequency by powers of 2, i.e. 1, 2, 4, 8, 16 etc. Also, the fundamental frequency is weak relative to the second harmonic. As a result, the absolute frequency of the sound is ambiguous, even though its value modulo octaves is unambiguous.

The implication of the perception of rising tones on these **Shepard scales** is that if the perceived fundamental frequency of a pitch value is ambiguous, small changes in the pitch value modulo octaves are preferentially interpreted (by the brain) as corresponding to small changes in absolute pitch value.

If small intervals modulo octaves are unambiguous in their direction, then we would expect larger intervals to be maximally ambiguous. The largest possible interval modulo octaves is half an octave, i.e. 6 semitones, also known as a **tritone**.

The **tritone paradox** refers to a phenomenon discovered by music psychologist Diana Deutsch, which is that the ambiguity in perception of direction of tritone intervals between Shepard tones is a function of absolute pitch modulo octaves, with the function being different for different individuals.[2] For each listener there is a particular position in the scale where the direction of a tritone interval is maximally unambiguous, and the ambiguity of other tritone intervals is a function of how close the notes defining those intervals

[2] *A Musical Paradox* Diana Deutsch (Music Perception 1986)

are to the maximally unambiguous tritone. For example, a given listener might have a maximally unambiguous tritone interval of $F\sharp$ to C such that change in pitch going from $F\sharp$ to C was unambiguously perceived as going upwards.

There are at least two possible interpretations of the observed pattern of ambiguity. One is that the neural representation of pitch value modulo octaves is circular, and that a particular direction in the brain is defined as being "upwards", for example as shown in Figure 11.2.

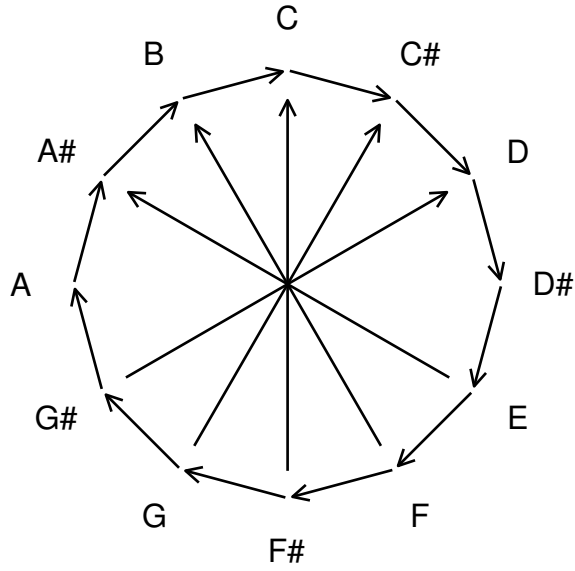

Figure 11.2. Circular tritone model. Direction for small intervals is clockwise (on the diagram) = upwards (perceived). Direction for tritones is upwards (on the diagram) = upwards (perceived). The tritone interval $F\sharp$ to C is the least ambiguous in its direction (definitely upwards); the interval A to $D\sharp$ is the most ambiguous (it could be either up or down).

A second possible interpretation is that the neural representation of pitch value modulo octaves is linear with overlap, and the maximally unambiguous tritone interval is located at the centre of this map, so that it is the least affected by ambiguous locations of neurons representing pitch values in the overlap. This interpretation is shown in Figure 11.3.

The advantage of the overlap model is that it simultaneously models perceived direction for both very small intervals and tritones. In the circular model, tritone direction is modelled by a fixed direction, whereas direction for small intervals is modelled by clockwise (or anticlockwise) motion around the circle.

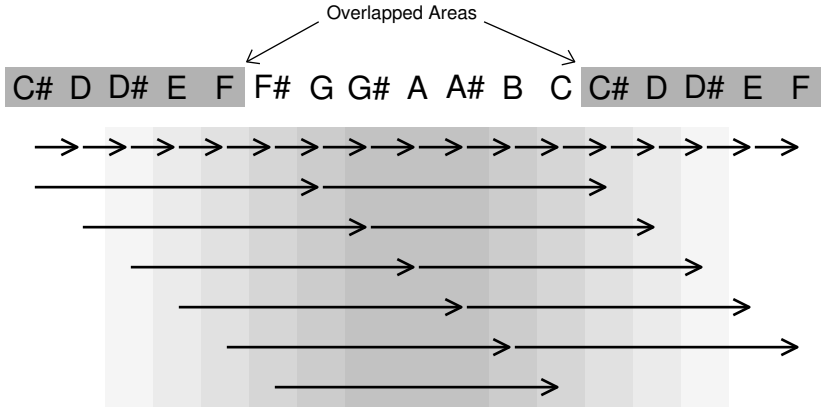

Figure 11.3. Overlap tritone model. Direction for all intervals is rightwards (on the diagram) = upwards (perceived). In this diagram only the interval $F\sharp$ to C has an unambiguous interpretation: all the others have two possible interpretations. This is a function of the size of the overlapped areas at the ends (in this case from $C\sharp$ to F). A variation on this theory is that greater priority is given to direction perceived from intervals represented in the central area of the map (as shown by the different shades of gray—darker means more weight is given to arrows lying in that region). In the example shown, $F\sharp$ to C would still be the most unambiguous upward tritone, and this would depend only on the midpoint of this interval (A) being at the centre of the map, and would not depend on how large the overlapped areas at the ends of the map were.

The other consideration making the linear overlap model more likely is that all other known cortical maps representing one-dimensional values map them in a linear fashion. In particular this applies to all known tonotopic[3] cortical maps.

The location of the maximally unambiguous tritone interval appears to be determined by the individual's exposure to spoken language, as correlations have been observed according to geographical location,[4] and also between mother and child.[5] This relationship between exposure to speech and the mechanics of octave translation invariance provides further evidence that octave translation invariance is relevant to speech perception (and not just to music perception).

[3]Reminder: a **tonotopic** map correlates position in one direction with frequency or pitch.

[4]*The Tritone Paradox: An Influence of Language on Music Perception* Diana Deutsch (Music Perception 1991)

[5]*Mothers and Their Children Hear a Musical Illusion in Strikingly Similar Ways* Diana Deutsch (Journal of the Acoustical Society of America 1996)

11.6 Octaves and Consonant Intervals

As already mentioned in Chapter 9, when discussing the relationship between invariances of pitch translation and octave translation, there is a correspondence between octave translation invariance and the importance of consonant intervals: all those aspects of music perception that depend strongly on consonant intervals are also octave translation invariant.

The one aspect of pitch perception which is not octave translation invariant, and which does not depend on perception of consonant intervals, is the perception of the up and down motion of melodic contours.

The next chapter on calibration suggests an explanation for all these observations, and also explains why octaves and other consonant intervals are so important in the first place.

Chapter 12

Calibration

The brain has the ability to perceive intervals between pairs of pitch values in such a way that intervals corresponding to the same frequency ratios are perceived as being the same.

How does the brain calibrate the perception of equality between frequency ratios? Careful consideration makes us realise that this calibration is non-trivial for a biological organism to achieve.

The answer seems to be that calibration is made against harmonic intervals observed to occur between the harmonic components of the human voice.

A similar type of calibration may underlie the time scaling invariance of rhythm perception.

12.1 A Four-Way Relationship

Pitch translation invariance involves an ability to perceive a four-way relationship between pitch values. For example, we can recognise that the interval from C to E is the same as the interval from F to A. This is a relationship between the notes C, E, F and A.

A naïve implementation of such a four-way relationship would involve connections between groups of 4 neurons. Taking into account all sets of pitch values related in this way, and even after reduction by means of splitting pitch into absolute and modulo octaves components, such an implementation would require a large number of connections. It would require $O(N^3)$, where N is the number of distinct pitch values (modulo octaves): for every 3 pitch

values X, Y and Z there is a 4th pitch value W determined by the equation $X - Y = Z - W$.

We know that we have a subjective perception of interval size. This can be interpreted as a three-way relationship between pairs of notes and interval sizes. Thus X, Y, Z and W are related as described above if there exists some interval Q such that the interval from X to Y equals Q and the interval from Z to W also equals Q.

This three-way relationship requires connections between sets of three neurons: two representing pitch values and one representing the interval between them, and this requires $O(N^2)$ connections (as already discussed in Chapter 11, when analysing the implementation of neural subtraction tables).

The ability of the combination of human ear, nervous system and brain to detect these relationships between quadruples of notes makes that combination into a reasonably precise measuring instrument. And seeing it as a measuring instrument, a simple question can be asked: how is the instrument *calibrated*?

12.2 Making Measurement Accurate

There are two main approaches to making sure that a measuring machine is as accurate as it is required to be:

- Construct the machine using precise construction methods that result in it having the required accuracy.
- Construct the machine less precisely, but include in the machine some mechanisms for adjustment which allow it to be calibrated against known standards for the type of measurement involved.

In the world of industry both of these methods are used. The first method is limited by the fact that a direct product from a manufacturing process is on average going to be less precise than the system used to manufacture it. The accuracy of most rulers and measuring sticks depends on the accuracy of the moulds and other factory machinery used to make them. But if we want a ruler that is really, really accurate, then a ruler stamped by a mould may not be good enough.

What does this have to do with our perception of intervals? The first type of calibration would involve the human ear, nervous system and auditory cortex all being pre-programmed to grow and develop in such a way that the intervals between different pairs of pitch values whose frequencies are in the same ratios are perceived as the same intervals.

I don't have a formal proof that this couldn't happen—but it seems very unlikely. Different people are all different shapes and sizes. Different parts of different people are different sizes. Everyone has differently shaped ears to everyone else. Much of the way that our body develops involves different

components growing in relation to other components, so that, for example, the lengths of our muscles and tendons match the lengths of our bones.

It seems implausible that, within this framework of variation and relative sizing, there could exist a system of measurement pre-programmed to develop to the accuracy exhibited by our ability to perceive and identify musical intervals.

This leaves us with the second possibility: approximate construction, followed by calibration against a naturally occurring standard.

Now we already know that intervals between pitch values which are simple fractional ratios play a significant role in our perception of music. And we know that these are the same ratios that occur between frequencies of harmonics of individual sounds, for certain types of sounds. And at least one type of sound having this property occurs naturally: the human voice.

This suggests an explanation as to why differences between the pitch values of *different* sounds are significant when they are equal to the differences between frequencies of harmonic components within the *same* sound: our auditory perception system uses the harmonic intervals between harmonic components of the same sound to calibrate its perception of intervals between the fundamental frequencies of different sounds.

In the world of industrial physical measuring instruments, we first calibrate our instrument to some degree of reliability, and having done that we then use our instrument to measure things, without concerning ourselves as to how the instrument was calibrated. The only lingering consequence of the method of calibration is that it adds to the expected error of our measurements.

In the world of biology, different components of functionality are often not as clearly separated from each other as we might expect from analogy with man-made artefacts and systems. With regard to the calibration of interval perception against harmonic intervals, there is one simple problem:

How do we calibrate our perception of non-harmonic intervals?

There are various ways that we might consider of doing this, and three main candidates are:

- Interpolate, i.e. relate non-harmonic intervals to harmonic intervals slightly larger than and slightly smaller than the non-harmonic intervals.
- Approximate non-harmonic intervals by harmonic intervals with fractions that contain numerators and denominators of greater size.
- Construct approximations to non-harmonic intervals by adding different harmonic intervals together.

There is a fourth option, and we will see that it may be the preferred one in many cases, which is *not to measure non-harmonic intervals.* But first

we will investigate what may be involved in the first three options, and how plausible they are as methods that could occur in practice.

12.2.1 Interpolation

The simplest form of interpolation would be to take two values that we know, and then identify a value that is half-way between those two values. The only technical difficulty we have to overcome is to devise a consistent way of determining what is "half-way". One way to do this involves observing pitch values that are rising in an approximately linear manner.

Figure 12.1 shows the calibration of an interval $(X+Y)/2$ by interpolation between two calibrated intervals X and Y. Pitch values f_0, f_1 and f_2 occur at times t_0, t_1 and t_2 respectively. $X = f_1 - f_0$, $Y = f_2 - f_0$, and the estimate for $(X+Y)/2$ is $f' - f_0$, where the pitch value f' occurs at time $t' = (t_1 + t_2)/2$, i.e. halfway between t_1 and t_2. The calibration depends on the assumption that log frequency is a linear function of time (during the period t_1 to t_2).

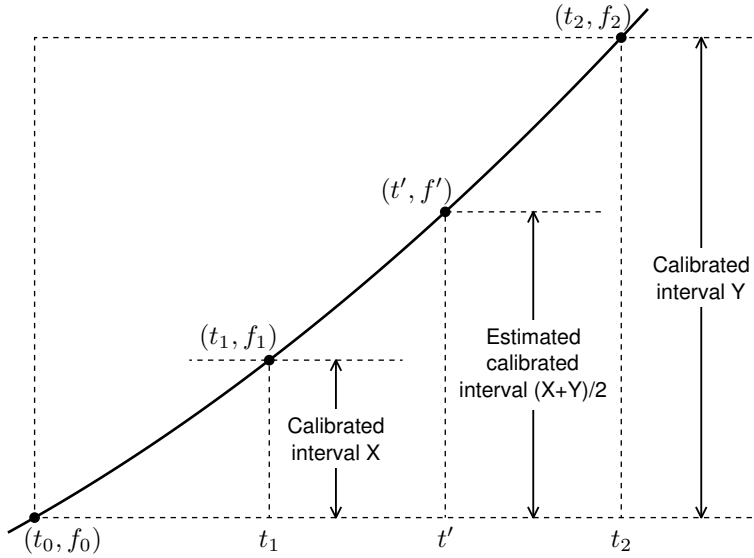

Figure 12.1. Interpolation of log frequency intervals on a smooth melodic contour. t' is exactly half-way between t_1 and t_2. The interval between f_0 and f_1 is the calibrated interval X, and the interval between f_0 and f_2 is the calibrated interval Y. The interval between f_0 and f' is an estimate for the size of $(X + Y)/2$, which would be exact if the contour was a straight line. But actually the contour curves upwards slightly, so the estimate is slightly too small.

In a pre-technological society, the major source of these rising (or falling) pitch values would be the melodic contours of speech. The contours of speech are not always straight-line contours, and linearity is an essential assumption in our interpolation procedure. In practice, however, the result may be adequate, for the following reasons:

- Any sufficiently smooth curve is linear for sufficiently small parts of that curve.
- If we average our interpolations over many different curves, they are likely to be linear on average.
- Even if average curves are not linear on average (e.g. they always curve one way or the other), the curvature that occurs in contours from different speakers with different pitch ranges may still be consistent enough to produce a calibration that is useful in practice. This will result in a calibration by interpolation which is not linear, but which is consistent among listeners exposed to a similar body of speech melodies (i.e. a group of individuals living in the same tribe).

12.2.2 Complex Fractions

There are two reasons to suppose that complex fractions[1] derived from comparing higher harmonics are not used as a means of calibrating our perception of intervals:

- It requires calibrations to be made against higher harmonics of very low frequency sounds, where the higher harmonics are in the range of the intervals that you are calibrating against. For example, to calibrate a ratio of 45:32 against the interval from 320Hz to 450Hz, you need a sound with a fundamental harmonic of 10Hz. There are not many natural sources of harmonic sound with this fundamental frequency. Certainly human speech does not go this low.
- Complex fractions are not observed to be significant in the perception of music. This suggests that the brain does not bother to use higher harmonics for the purpose of calibrating comparisons of interval sizes.

12.2.3 Arithmetic

Calibration by arithmetic is a common solution to the problem of calibrating the measurement of a value that can be defined as a sum of values that have already been calibrated. If I can calibrate a length of 1 metre, and I need

[1] By "complex" I just mean with a large numerator and denominator, in the sense that "complex" is the opposite of "simple", and "simple" fractions are fractions with small numerators and denominators.

to calibrate a length of 2 metres, then all I need to do is mark off 1 metre, and then mark off a second 1 metre that starts where the first one finished, and altogether I have marked off 2 metres. To make this work for interval perception, our perception of the interval between two notes X and Y would be mediated by the occurrence of an imaginary note Z such that $Z - X$ was a harmonic interval, and $Y - Z$ was also a harmonic interval.

Using arithmetic to calibrate non-harmonic intervals achieves a similar result to using higher harmonics, because it enables more complex fractions to be used.[2] The main reason to doubt that this type of arithmetic plays a significant role in the calibration of interval perception is the same as the second reason given above for supposing that higher harmonics are not involved in this calibration: complex fractions are not observed to be significant in music perception.

12.2.4 Not Measuring Non-Harmonic Intervals

In as much as we can perceive and compare non-harmonic intervals at all, interpolation seems the most likely of these three options to be used by the human auditory perception system for calibrating the perception of those intervals.

But the structure of music suggests major use of the fourth option: only measure harmonic intervals. Harmonic intervals are significant in music perception. Chords and harmony are the most obvious manifestation of this, but the determination of home chords also appears to be strongly tied to harmonic relationships between pitch values.

We can imagine the following means of identifying the harmonic characteristics of a smooth melodic contour:

- Identify the initial pitch value.
- Record the times at which pitch values occur that are harmonically related to the initial pitch value.

The recorded sequence of time values will then form a characterisation of the melodic contour such that this characterisation is pitch translation invariant. This means of characterisation comes very close to the operation of the harmonic cortical map, which we have previously identified as being the cortical map that responds to the occurrence of chords. (The main difference with the harmonic cortical map is that it only marks or records the times of pitch values which are harmonically related to the initial pitch value *and* to any other pitch values already marked or recorded.)

[2] Adding intervals is equivalent to multiplying fractions. If we add two intervals representing simple fractions to get an interval which does not itself represent a simple fraction, then it will necessarily represent a more complex fraction. Of course there are some complex fractions that cannot be derived from simpler fractions by means of multiplication—for example fractions containing a large prime in the numerator or denominator.

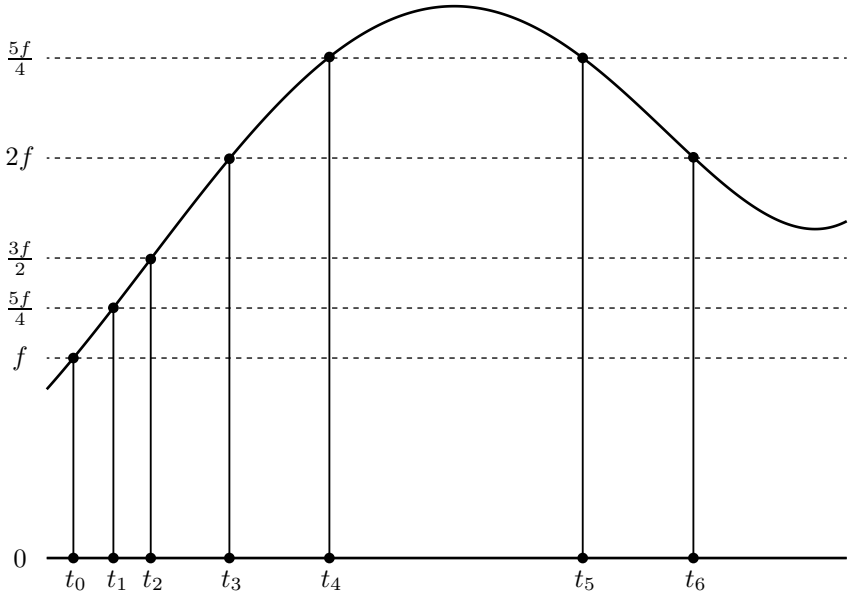

Figure 12.2. Recording the times of occurrences of some frequencies in a melodic contour harmonically related to the frequency f at time t_0. In this example, the frequencies recorded are those included in the major chord which has frequency f as its root note. (This is an unrealistic simplification, since the contour is not one that implies the major chord; however, it serves to demonstrate the general principle that the harmonic map creates a pitch invariant characterisation of the melody. In a more realistic example, the "chord" used to record pitch values in a continuous melody would be more fuzzy than a discrete set of pitch values, and the recorded time values would themselves be correspondingly fuzzy.)

From a mathematical point of view, this trick of only measuring harmonic intervals turns out to be an indirect form of interpolation. We plot a finite number of points in the melodic contour consisting of pitch values that are harmonically related, then we can fill in the rest of the contour if we need to by assuming that it is a smooth curve and joining up the points we plotted. There are occasions when the rest of the contour does need to be filled in: for example when we want to reproduce a particular contour in our own speech melodies.[3]

[3] A minor complication is being ignored here: the harmonic cortical map is only recording the times of occurrences of notes harmonically related to each other, without recording what the relations are. We must presume that the basic reconstruction of the melodic contour occurs via the melodic contour cortical map from the recorded ups and downs, and that the set of times derived from the harmonic cortical map is then used to more accurately position those points in the contour at the recorded times, by "snapping" them to the nearest frequency harmonically related to the frequency at time t_0.

12.3 Calibration Experiments

A strong test of the calibration theory would be to expose a subject to bad data over their lifetime, and see if predicted calibration errors could be observed. Bad data would consist of sounds with incorrect harmonic frequencies. On the assumption that the human voice is the main source of calibration data, all human speech that the subject heard would have to be appropriately altered. The subject would wear a microphone and headphones connected to a digital sound processor, such that all the sounds coming into the microphone were digitally altered, and the subject would hear only the altered sounds played through the headphones.

If, for example, all 2nd harmonics were increased in frequency by 5%, then we would predict that the subject's perception of octaves would be correspondingly altered. Alterations to other harmonics would alter the subject's perception of intervals. If harmonics were altered in a manner dependent on frequency, then this would be predicted to alter the subject's ability to accurately compare intervals at different pitch levels (i.e. to identify the interval from note W to note X as being the same as the interval from note Y to note Z).

It would not be ethical to carry out such an experiment on a person over their lifetime. But it is quite possible that calibration is not a once-in-a-lifetime event. As a person grows, the frequency response functions of locations in their ear (in the organ of Corti) are going to change slightly over time, and it is likely that adjustments have to be made continuously to keep the auditory cortex correctly calibrated.

If a willing subject can be exposed to altered speech for a period of days or weeks, it may be possible to observe adaptation to these alterations as a result of calibration against the contrived bad data.

This type of recalibration experiment has its precedents: experiments where subjects wear prismatic lenses which shift the image of the real world on their retinas. Subjects are observed to adapt over a period of time to this artificial shift. (And luckily the adaptation re-adapts back to normal once the subject stops wearing the special lenses.) Similar adaptation happens to anyone who starts wearing glasses, and can also happen to users of various types of virtual reality environment and augmented vision systems. Some of the research on adaptation to altered vision has been done to make sure that adaptation to virtual reality environments does not cause lasting perceptual impairment.[4]

If the recalibration of interval perception could be achieved, then a very interesting possibility arises: new types of music that are only perceived to be musical by someone whose perception of harmonic intervals has been artificially altered this way. To give a simple example, the locations of 3rd

[4]For example, *Virtual Eyes Can Rearrange Your Body: Adaptation to Visual Displacement in See-Through, Head-Mounted Displays* Frank Biocca and J.P. Rolland (Presence: Teleoperators & Virtual Environments 1998)

and 5th harmonics could be altered to exactly match the intervals that occur on the well-tempered scale. This would have the effect of making the well-tempered scale be (subjectively) perfect for these types of interval, and might increase the musicality of music played on that scale.

Experiments on calibration could also be carried out on animals. If, however, music is very human-specific, it will be difficult to find a useful animal model.

12.4 Temporal Coding

We might suppose that temporal coding plays a role in the calibration of the perception of harmonic intervals. If, for example, phase-locked neuron A was responding to a frequency of 200Hz, and phase-locked neuron B was responding to a frequency of 300Hz, then there would be exactly 2 firings of neuron A for every 3 firings of neuron B. If there was some way to count and compare how many times each neuron fired compared to the other, then this would give a natural way of knowing that the two frequencies were harmonically related.

It might also be possible to compare the times at which the two neurons fire, and record the intervals between those times, to determine whether or not the two frequencies are related to each other by a simple ratio. A basic difficulty with directly comparing the timings of individual firings is the level of accuracy required. For example, considering slightly higher frequencies, such as 1000Hz and 2000Hz (which are still within the range of musically significant frequencies), and assuming that we are required to achieve a 1% accuracy of interval perception, this implies that a 0.5% accuracy is required for each of two comparisons, which translates into 5 microseconds—a very short period of time. Although there are some known animal perceptions that operate on this time scale or even shorter, such as bat echo-location, there are severely non-trivial problems to overcome, including "jitter" and the sheer length of time it takes for an action potential to occur—typically 300 microseconds.[5] It's much easier to just calibrate against natural examples of sounds containing harmonics, which you already "know" have the correct relationships between their frequencies.

If calibration of harmonic intervals *was* based on direct comparisons of periods of vibration, then it would not be possible to mis-calibrate interval perception by exposing subjects to sounds with the "wrong" harmonics, in which case the experiments described in the previous section would give a negative result.

[5] "Bat Echolocation" by James Simmons, text box within *Neuroscience: Exploring the Brain* Bear, Connors and Paradiso (Williams & Wilkins 1996)

12.5 Other Calibrations

12.5.1 Calibration of Octave Perception

Octaves are a special sort of consonant interval. The split of pitch values into imprecise absolute pitch value and pitch value modulo octaves necessarily requires an accurate determination of the octave relationships between frequency values.

As in the case of consonant interval perception, this determination will need to be calibrated, and the most likely means of calibration is by comparison with the harmonic relationships that exist between the harmonic components of the sounds of the human voice. For calibrating octave perception it would be sufficient to consider just the fundamental frequency and the second harmonic.

12.5.2 Calibrating Ratios of Durations

Comparisons of intervals between different pairs of pitch values are required to achieve pitch translation invariant perception of melody. In a very analogous manner, comparisons of ratios between pairs of time durations are required to achieve time scaling invariant perception of rhythm.

Recall that the cortical maps relating to rhythm consist of groups of neurons that respond to percussive sounds separated by specific time intervals. Some neurons respond just to pairs of percussive sounds; these are the neurons that encode duration information. Other neurons respond to ongoing regular beats. In both cases, if there is to be an ability to perceive the same rhythm at different tempos, there needs to be a means of measuring the ratios between the time intervals that these rhythm-sensitive neurons respond to.

For example, an instance of a rhythm might consist of beats at times 0, 1 and 1.8. We will treat time as being in units of seconds. This results in two durations: 1 second and 0.8 seconds, and the ratio between them is 5:4. A slower version of the same rhythm might consist of beats at times 0, 1.5 and 2.7. The resulting durations are 1.5 seconds and 1.2 seconds, and their ratio is also 5:4. The identity of the two 5:4 ratios is what enables us to perceive that these are two versions of the same rhythm, with the second being a slowed down version of the first. Note that we are not particularly interested in the ratio between the two tempos; what matters is being able to identify the two rhythms. In fact the comparison may be between two occurrences of the rhythm at widely separated times, for example on different days. So there is no easy way to make comparisons between the durations of corresponding components of different occurrences of the rhythm; all comparisons must be made between durations occurring within each individual occurrence of the rhythm.

How can we calibrate the perception of ratios between durations and beat periods? The most obvious calibration is to compare durations where one

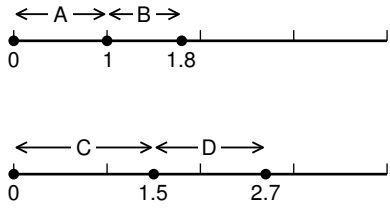

Figure 12.3. The "same" rhythm played on two different occasions at different tempos. The scale shows time in seconds. It is true that $C/A = D/B$, but to calculate C/A and D/B presumes an ability to accurately compare time durations perceived on different occasions. More realistic is to calculate the ratios in the equation $A/B = C/D$, as this only involves comparison of durations within each individual occurrence of the rhythm (and comparison of the ratios over the long-term).

duration is twice as long as another, i.e. a ratio of 1:2. If three beats occur, say X, Y and Z, with Y occurring halfway between X and Z, and one neuron A responds to the durations X to Y and Y to Z, and another neuron B responds to the duration from X to Z, then we can determine that the ratio between the duration periods of neurons A and B is 1:2.

A similar calibration could occur for other simple ratios, like 1:3, or 1:4 or even 2:3. But once again we can use our knowledge of observed aspects of music to guide us. Musical rhythm is strongly dominated by durations and beat periods related to each other by factors of 2. Factors of 3 come in a very distant second. A factor of 4 can be regarded as 2×2, and any larger factors are virtually non-existent.

So we can conclude that a calibration process occurs by which the brain identifies pairs of neurons in rhythm-oriented cortical maps that respond to durations related by a factor of 2, and to a lesser extent by a factor of 3.

12.5.3 Calibrating Against Regular Beats

There is one minor difficulty with this theory applied to time scaling invariance: calibration requires the occurrence of regular beats, such as the beats X, Y and Z given in the previous example, where the interval from X to Y is the same as the interval from Y to Z. Such regularity may not occur in natural speech or in other sounds that a developing child may hear.

Of course there is one situation where children will hear regular beats: when they are listening to music. This leads to a direct biological function for music, i.e. to assist calibration of cortical maps that process information about rhythm to produce time scaling invariant perceptions. This gives a counter-example to our working doctrine that it is only the perception of

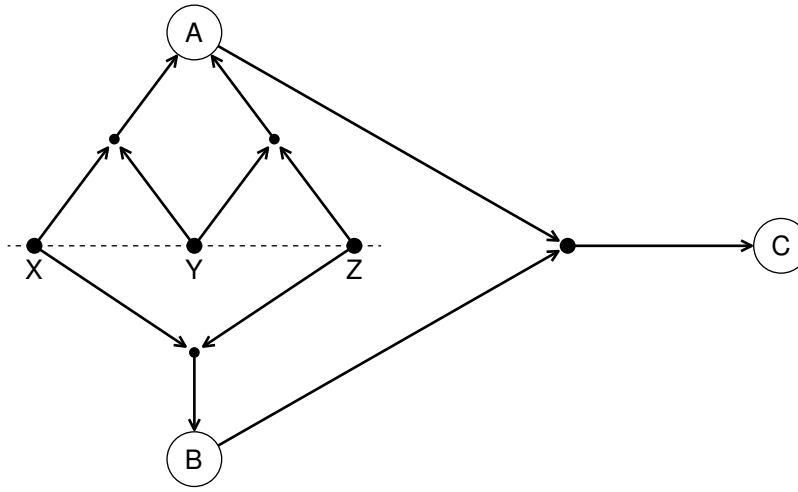

Figure 12.4. Relative tempo calibration. Events X, Y and Z represent percussive sounds perceived from a regular beat. The duration from X to Y and the duration from Y to Z both activate neuron A. The duration from X to Z activates neuron B. Neuron B therefore represents a duration twice as long as the duration represented by neuron A. Calibration results in neuron A and neuron B activating neuron C, with the consequence that neuron C encodes the perception of a ratio of 1:2 between different durations.

musicality that has a biological function, and that music in itself serves no biological function.

And if we can make this concession for rhythm perception, then we can also make it for melodic perception: the playing of melodies may assist in the calibration of the perception of harmonic intervals. For example, the calibration process may proceed more efficiently if the listener is exposed to different sounds such that the harmonic components of each sound are separated from each other by consonant intervals, and such that the different sounds have fundamental frequencies separated from each other by consonant intervals.

Chapter 13

Repetition

Repetition is a major aspect of music. The theory of musicality as an aspect of speech perception forces us to ask why exact repetition (free and non-free) occurs in music, even though it does not occur in normal speech.

Human perception of speech melody is intrinsically time translation invariant, and this creates problems when repetition or near-repetition occurs *within* a single speech melody. The solution is to maintain a **repetition count**, i.e. to distinguish the first occurrence from a second occurrence of a melodic fragment within a melody.

A secondary question is: When does the brain know *not* to keep count any more? A suggestion is that those features that normally come at the end of a melody may serve the function of resetting the repetition count to zero.

13.1 Repetition as a Super-Stimulus

Recall the relationship between aspects of music and the perception of speech:

- An aspect of speech is perceived by a cortical map.
- A corresponding aspect of music is a super-stimulus for that particular cortical map.

As a result, we sometimes see features of music that appear not to exist at all in speech, for example the occurrence of musical scales. Even when we

can recognise the similarity between a musical aspect and a speech aspect, such as the rhythms of speech and the rhythms of music, the musical version may have regularities not apparent in the speech version.

Repetition is an aspect of music where there is a high degree of regularity, with apparently no analogue in speech. Musical phrases are sometimes exactly repeated within a tune without any variation at all, and are often repeated an exact number of times—usually twice, sometimes more.

This kind of exact repetition does not normally occur in speech (although there is one major exception—see Section 13.7 which discusses **reduplication**), and speech would generally sound strange or contrived if it did occur. So what's going on?

I defined **free** and **non-free** repetition in Chapter 4. Free repetition is where the major components of music are repeated freely, such as choruses and verses of a song.

In some cases a tune exists in a cyclic time frame, in the sense that the end of one repetition blends directly into the beginning of the next repetition, and the musicality of each portion of the tune depends on what precedes it and what follows it, so the performer has no choice but to perform endless repetitions of the tune. In other cases the tune comes to a stop before starting again each time, but is still repeated an indefinite number of times as part of a single performance. Many recorded performances of popular songs come to an end by "fading out", suggesting that those producing the songs could not find a satisfactory way to end them. The only thing that prevents a song containing freely repeated components from repeating them forever is that the audience will get bored if the song goes on for too long.

Non-free repetition is perhaps of more interest. Components repeated non-freely are components *within* a major component of a song. They can range from single notes, to portions of a bar, to as much as a quarter of the song, e.g. a tune might take the form $[AABB]$.[1] The non-free aspect is that the repetitions occur a fixed number of times. Thus each occurrence of the repeated phrase has assigned to it some count of its location within the repetition. (With free repetition there is no sense of keeping count, unless perhaps we keep count consciously: freely repeated verses and choruses just go on and on.)

This suggests that somewhere there is a cortical map that keeps this count of the number of repetitions that have occurred for a phrase. In as much as repetitiveness is a perceived quality that we can be consciously aware of, there probably exists some cortical map representing it, since most perceptual qualities have corresponding cortical maps that process and represent them.

There is at least one difficulty with this theory of a cortical map that encodes a repetition count: it has to deal with **nested repetition**. This is where a non-freely repeated component itself contains non-freely repeated

[1] Here A and B etc. are used to refer to particular phrases (not notes), so that, for example, $[AA]$ refers to some phrase A being repeated twice.

components. For example, a tune might have structure $[ACBBACBB]$, where the component B is repeated twice inside the repeated component $[ACBB]$. Such nested repetition does occur within popular and traditional music—a good example is "Funiculi Funicula" (Denza & Turco, 1880). The representation of nested repetition in a cortical map would appear to require a separate dimension of count for each nested level of repetition.

Possibly even more common than exact non-free repetition is **partial** non-free repetition. A melody may have distinct components $[ABCD]$, but reduced to the rhythm only it may read $[A'A'C'D']$, where the first two phrases have different melody, but the same rhythm. There are many common variations on this partiality:

- The sequence of notes may be the same, but the rhythm may be different.
- The rhythm and the up and down contour may be the same, but the melody is translated up the scale, so that the exact intervals between corresponding pairs of notes in the partially repeated phrases are different.

Another variation is repetition of the beginning of a phrase, which may be either exact or partial, but then a variation occurs at the end of the phrase. This type of variation is often associated with a sense of progression.

Whereas non-free repetition requires the inclusion of a repetition count in the perceptual state, there is no such requirement for partial repetition, since the aspects of the music not being repeated provide the information that distinguishes the first repetition from the second.

13.2 Reasons for Perception of Repetition

If we were writing a computer program to perceive melodies as sequences of notes, there are two ways that we might deal with repeated sequences:

1. The program could ignore repetition. The melody is just treated as a sequence of notes, all of which are individually recorded and processed by the program. If repeated sequences happen to occur, the program doesn't care; it just processes each repetition in due course.

2. Or, the program could be written so as to recognise repeated sequences. The program would have to include some definition of what was a significant repetition, for example some minimum length of a repeated sub-sequence. When a repeated sequence was recognised, instead of processing it all over again, the program could just record that the repetition had occurred, and it would record which previous sub-sequence

of the melody it was that was repeated. Some file compression algorithms work this way.[2]

As I have stated previously, the human brain does not always solve a problem the same way that we might program an electronic computer to solve that problem. A computer programmer writing a program to process sequences of values in melodies would probably have the program write each sequence of note values that it was processing to a corresponding series of numerically indexed locations in memory. The index values would form an implicit global frame of reference against which the note values were indexed. The index values could be used to identify the location of a sequence of values that had already occurred and which was being repeated. It is, however, very unlikely that the brain uses any type of numerical indexing system to store the data it processes.

13.3 Perceptual State Machines

A **state machine** is a system that has a set S of possible states, a set E of possible events, and a **transition function** F which maps each pairing of input state s_{in} and event e to an output state s_{out}. We can use this concept to model how the brain processes sequences of values such as notes in a melody: the state corresponds to the state of activity in a cortical map, and the events correspond to the information coming into the cortical map about each musical note. For a given initial state s_0, we can model the perception of a melody as the updating of the state by the sequence of events representing the notes of the melody.

If the transition function is such that the state machine's current state has no dependence on more than the previous N values, then the state machine will automatically recognise repetitions of N or more notes, in the sense that it will always be in the same state at the end of two identical sub-sequences of N or more notes.

A state machine whose state depends on only a limited set of previous values is **forgetful**, because all past history eventually gets "forgotten" by the state machine.

A state machine with this forgetfulness property recognises a repeated sequence, but the disadvantage is that the machine is then completely incapable of knowing how many times the sequence has been repeated, since it is always in the same state at the end of a sequence, whether it be the first repetition or the second or the hundredth.

Forgetfulness is related to time translation invariance of perception. We wish the response to a sufficiently extended sequence of values to be the same

[2]See *A Universal Algorithm for Sequential Data Compression* Jacob Ziv and Abraham Lempel (IEEE Transactions on Information Theory 1977)

whenever that sequence occurs; thus the state of a system responding to the sequence must not maintain any state information too persistently.

13.3.1 A Neuronal State Machine

The following is a simple model of a neuronal state machine that responds to the occurrence of a sequence of information values representing the sequence of notes in a musical melody:

- Each note value is represented by a group of sub-values, where each sub-value relates to one aspect of the music. These sub-values are derived from the symmetry-invariant encodings of the notes in the melody.
- Simplifying slightly, each sub-value is represented by the activation of a neuron in a corresponding cortical map (simplified in that we are ignoring population encoding).
- There is a **state cortical map** within which individual neurons encode for individual states.
- The neuron for the current state and the neurons for the sub-values of the current value in the sequence activate the neuron in the state map representing the next state.

13.4 The Flow Model

Suppose that there are n sub-values per note value. Each value is therefore a point in an n-dimensional space. Suppose, hypothetically, that each value corresponding to a note in the melody is unique within that melody. And further suppose that each point in the n-dimensional space is represented by a neuron. Then in order to represent the sequential progress of the melody, all we need is a connection from each neuron to the next neuron in the n-dimensional space, such that each neuron activates the neuron representing the next step in the melody.

We can imagine these connections between neurons as being like a sequence of arrows in the n-dimensional space. In effect the n-dimensional space is the same as the space of possible states. We start at the state representing the beginning of the melody, we follow the arrows, and eventually we reach the state representing the end of the melody.

This model is somewhat idealised. There are at least two major objections to it as a realistic model of how the brain represents and processes information about melodic sequences:

- Cortical maps are not n-dimensional; in general they are no more than 2-dimensional with regard to representing numerical values.

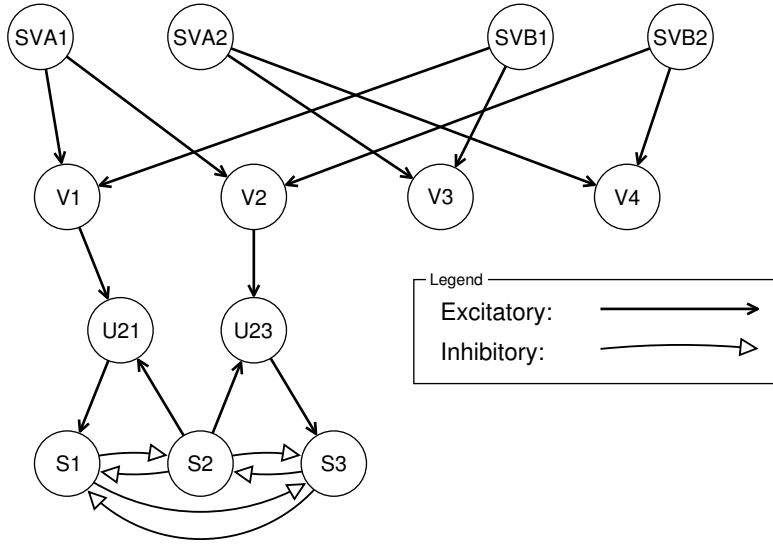

Figure 13.1. A neuronal state machine representing melody perception. Neurons SVA1 and SVA2 represent possible values of sub-value A; neurons SVB1 and SVB2 represent possible values of sub-value B. The different combinations of sub-values form the full values as represented by neurons V1, V2, V3 and V4. Neurons S1, S2 and S3 represent three possible states of the state machine. Update neuron U21 represents the rule that value V1 (i.e. sub-values A1 and B1 combined) should cause state S2 to transition to state S1. Similarly neuron U23 represents the rule that value V2 (sub-values A1 and B2 combined) causes state S2 to transition to state S3. Mutual inhibition between the state neurons ensures that only one state is active at a time, and that the transition to the new state deactivates the old state.

- If the state in the n-dimensional map ever repeats, then our model of the melodic sequence will get stuck—it will be forced to go around forever in the same loop. We can explain this better by considering what happens when we reach a particular state for the second time: the flow of state changes is entirely determined by the current state and the arrow from that state to the next, so the next state after visiting a state a second time has to be the same next state that happened when the state was visited the first time.

We will find a pragmatic way to solve the 2-dimensional limitation, but first I will look at the "stuck-in-a-loop" problem.

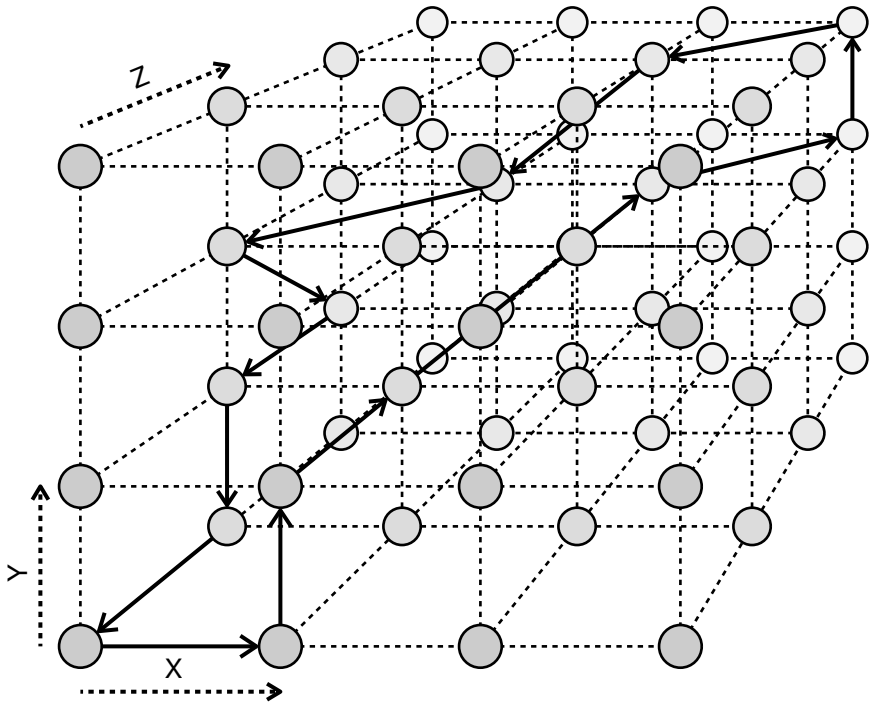

Figure 13.2. N-dimensional state. A hypothetical N-dimensional cortical map (in this case N=3, as that is the most that can practically be shown on a diagram). The arrows show a closed-loop path through the map. Comparing to Figure 13.1, the neurons shown simultaneously represent the "V", "U" and "S" neurons, and the arrows represent the connection from each "S" neuron via a "U" neuron to the next "S" neuron.

13.4.1 Breaking Out of the Loop

When we get to a state in the n-dimensional space for the second time, what we really want to know is that we have already been there before, and we want to provide this information about having been there before as *an extra dimension* of our state. The connections between neurons will then be able to take into account this extra dimension of information, to in effect say "we are now here for the second time, so don't follow the arrow we took the first time, instead follow this other arrow".

13.4.2 Almost Exact Repetitions

Musical repetitions are often exact repetitions. The repetition count (of how many times we have been here before) will be 0 the first time around, 1 the

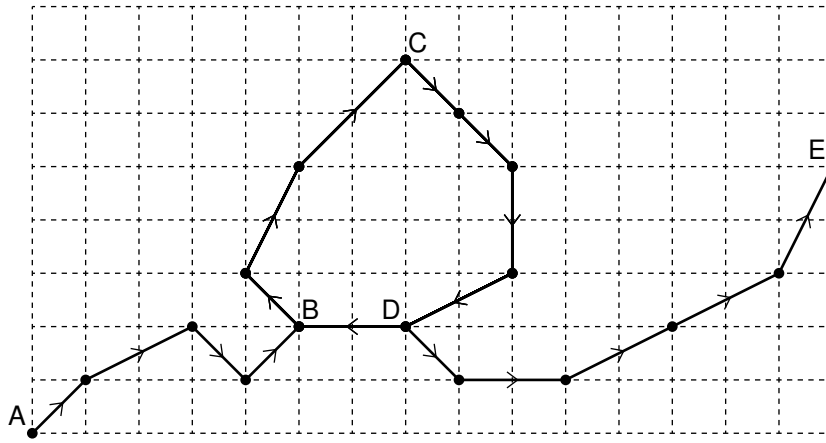

Figure 13.3. A path representing flow in a state space (here assumed to be just 2-dimensional). The flow starts at point A, goes to point B, goes around the loop B to C to D three times, and then exits the loop at D and finishes at E.

second time round, 2 the third time, and so on. These exact repetitions do not normally occur in speech,[3] yet if we suppose that there exist special mechanisms for the perception of repetition in melody, then those mechanisms must presumably exist for the purpose of perceiving speech melody.

Although speech melodies do not contain exact repetitions, it is entirely possible that they can contain repetitions that are close enough to being exact to cause a partial occurrence of the problems caused by exact repetition, and for which the addition of repetition count as an extra dimension of information is required. Thus the system not only represents "we have been here before", it also supports "we have been close enough to this spot before that it might cause confusion". This could be regarded as a fuzzy numerical attribute, i.e. 0 represents "we have not been here before", 1 represents "we have been here before", and values between 0 and 1 represent "we have been close to this point before".

13.4.3 Faking n Dimensions in 2-Dimensional Maps

The other problem with the flow theory is that cortical maps are only 2-dimensional (with a very thin 3rd dimension that would not be able to represent a continuous numerical attribute), whereas the flow is in an n-dimensional space of perceived values.

[3]Except for the very short exact repetitions caused by **reduplication**, as discussed in more detail in Section 13.7.

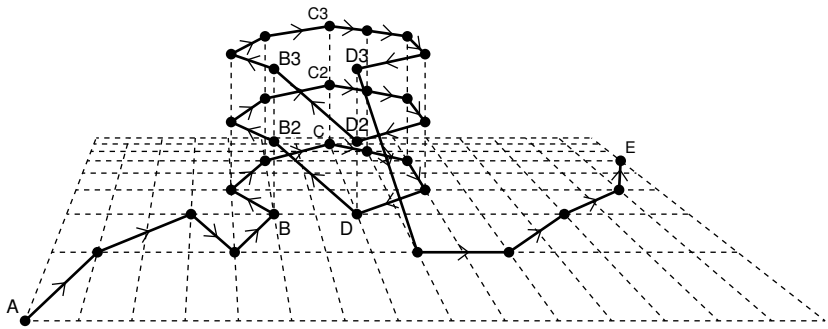

Figure 13.4. As in Figure 13.3, but now "Have we been here before?" is represented as an additional dimension. The flow goes from A to B to C to D. When it comes around to B a second time the repetition is represented by being lifted up to $B2$. From there it loops around $C2$ to $D2$ to $B3$ to $C3$ to $D3$. When it exits the loop at $D3$ the flow returns to "ground level" because the state is no longer a repetition of a previous state.

Given n dimensions, we could imagine all possible **coordinate projections**[4] to 2-dimensional subspaces representing all possible pairs of the n dimensions. There will be $n(n-1)/2$ such subspaces. For each of these subspaces, and for each melody, we can define a flow of motion in a hypothetical cortical map that represents the subspace. In each subspace there will be a path corresponding to the flow of the melody in that subspace. There will be many more collisions in the 2D subspaces than there are in the original n-dimensional space.

What happens if we leave the direction of flow undefined at the points where these collisions occur? Even if a collision occurs in one 2D subspace, there will not necessarily be a collision in all the other subspaces. It is likely that the flow will be defined in a sufficient number of other subspaces that the direction of flow can be fully reconstructed in the n-dimensional space.

The one occasion where a collision will occur in all the 2D subspaces is when there is an exact repetition, and this is precisely where some means of introducing repetition count as an extra dimension is needed in order to maintain enough state information to record the full path of the melody.

[4] A **coordinate projection** from an n-dimensional space to an m-dimensional space is a mapping defined by a subset of the numbers 1 to n, of size m, such that coordinate positions in the subset are retained, and coordinate positions not in the subset are not retained. For example, for $n = 5$ and $m = 2$, the subset $\{2, 4\}$ defines a projection that maps the point $(32, 67, -9, 21, 8)$ to $(67, 21)$.

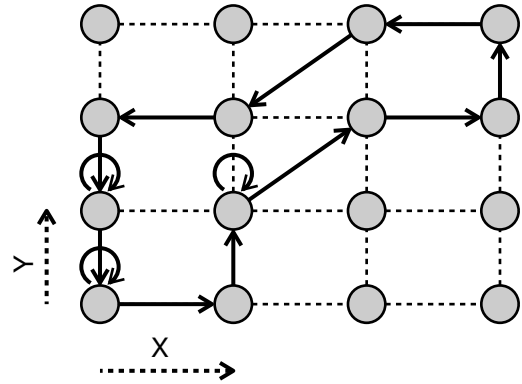

Figure 13.5. Projection of Figure 13.2 onto X and Y dimensions. In the original 3 dimensions the state loop shown formed a simple loop that defined a unique transition from each state in the path to the next one. This is no longer true in the projection because in some cases different points in the original path project down to the same points in the projection (because they only differ in their Z coordinate).

13.5 Non-Free Repetition: Summary

The theory of musical repetition given here is not as fully developed as the components of my theory relating to other aspects of music. A full understanding of repetition and the mechanisms of recording sequential information in cortical maps is one of the missing pieces of the puzzle that is required to properly complete the theory (see Chapter 15 for a fuller discussion of the incompleteness of the super-stimulus theory).

Within the framework of the super-stimulus theory, the following is a summary of my current understanding of repetition as it occurs in music:

- Exact non-free repetition is a common and well-defined aspect of music.
- Musical repetition must be a super-stimulus for an aspect of speech melody perception.
- Special perception of repetition is required in a system that is designed to work when perceiving information that is not repetitive.
- The characteristics of a perceptual system that enable it to automatically recognise repetitive sub-sequences also cause that system to fail when it is perceiving sequences that contain non-freely repeating sub-sequences.

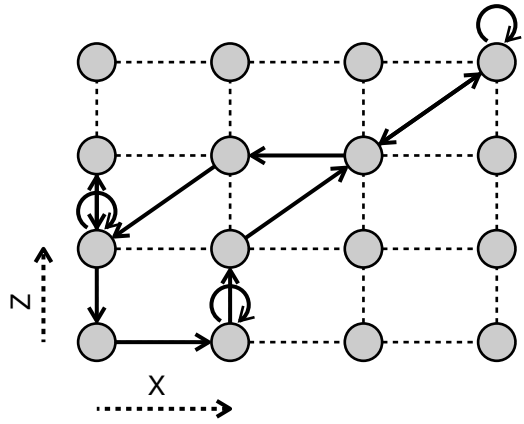

Figure 13.6. Projection of Figure 13.2 onto X and Z dimensions.

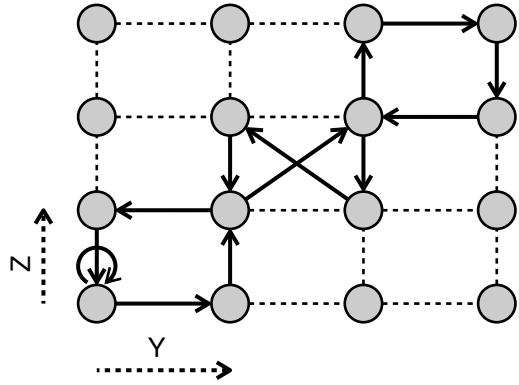

Figure 13.7. Projection of Figure 13.2 onto Y and Z dimensions.

13.6 Free Repetition and Home Chords

The theory of home chords given in this book explains how the home chord is determined, and what perceptual purpose is served by the cortical map that makes this determination. The theory also explains why a home chord occurs at the start of a melody: the diatonic scale causes more than one stable state to be possible in the home chord cortical map, and the map does not enter one of these states until the corresponding chord actually occurs.

This leaves unexplained the second major fact about home chords and notes, which is their occurrence at the *end* of a melody. Associated with the occurrence of the home note and chord at the end of the tune is the tendency

of the home chord to be preceded by a **dominant 7th** chord. Referring to examples on the white notes scale with home chord C major or A minor, the chord that precedes the home chord is the dominant 7th chord with root note a perfect fifth higher than the root note of the home chord, i.e. G7 = GBDF precedes C major, and E7 = EG♯BD precedes A minor.

Another common feature of the home note/home chord combination is the length of the last note. Very commonly the final home note is a single note that starts simultaneously with the final occurrence of the home chord at the start of the final bar, and continues for all or most of that bar.

Taken together, these features of a final note/chord combination that define the end of a tune are called a **cadence**.

I have not been able to discover any convincing explanation of why this combination of chords and a long final home note wants to occur at the end of a tune. But I can make one pertinent observation:

> A tune cannot freely repeat, until it has first *ended*.

So it may be that the purpose of a home chord, optionally preceded by a dominant 7th, is to *end* a tune by resetting the state of some or all of the cortical maps involved in perception of music/speech (in particular resetting any repetition counts), so that the tune can then be freely repeated.

We have seen that the perception of non-free repetition requires a keeping of the repetition count. If we consider a cortical map that is responding to a repeated sequence, especially one that repeats from the beginning of the tune so far, we might ask how the cortical map knows if it is meant to be a non-free repetition or a free repetition. In other words, is it meant to be keeping count or not? Keeping count for a free repetition would introduce a spurious dimension of perception into the perception of the music. Failing to keep count of a non-free repetition would remove a dimension of perception that was required.

It is possible that the default is to assume all repetition is non-free, and that the effect of a cadence is to reset the repetition count (of everything) back to zero, so that any following repetitions are perceived as free repetitions. A cadence might perhaps represent a prototypical sentence ending, prototypical in the sense that the human brain is partly predetermined to end a sentence with an intonation resembling a cadence, even though specific languages may adopt alternative intonations for the ends of sentences (in effect overriding the predetermined default).

In a musical cadence, the state of the harmonic cortical map matches the state of the home chord cortical map. It is possible that the same match occurs in a speech "cadence", even though, in the case of speech, neither map would be in a state corresponding to a musical chord.

13.7 Reduplication

There is one major exception to my earlier assertion that exact repetition does not occur in natural language. This is the phenomenon of **reduplication**. Reduplication is where all or part of a word is duplicated within the word to make a new word.

One family of languages where reduplicated words are common is the family of Polynesian languages. For example, in New Zealand Maori, "toru" means "three" and "torutoru" means "few".[5] In Hawaiian (another Polynesian language), "wiki" means "hurry" and "wikiwiki" means "quick".[6]

Reduplication is conspicuous by its absence in English and most other Indo-European languages. Reduplicated words sound strange to the English ear, and one could suppose that we positively avoid constructing them. Perhaps we are always in such a hurry to say what we want to say that saying something twice seems like a waste of time. The nearest we get to using reduplication is the use of phrases like "itsy-bitsy", "hodge-podge", "lovey-dovey", "shilly-shally" and "hoity-toity", all of which are highly colloquial and informal in their usage.

Exact non-free repetition occurs in music on a much larger time frame than lexical reduplication. But we cannot rule out the possibility that the same cortical map is responding to both forms of repetition. After all, we are supposing that music is a super-stimulus for speech perception, so the long non-free exact repetitions in music may be a super-stimulus for perception of the short non-free exact repetitions caused by reduplication.

[5] *The Reed Reference Grammar of Maori* Winifred Bauer (Reed Books 1997)
[6] *Hawaiian Dictionary* Pukui and Elbert (University Press of Hawaii 1977)

Chapter 14

Final Theory

Finally we have enough clues to provide a tentative answer to the big question: *What is music?*

And the answer is: musicality represents information about the internal mental state of the speaker. It is perceived via observation of **constant activity patterns** in those cortical maps of the listener that respond to aspects of speech rhythm and speech melody. Constant activity patterns in the *listener's* brain echo constant activity patterns in the *speaker's* brain, which are a function of the level of conscious arousal of the speaker.

Perceived musicality *confirms* the listener's emotional response to the content of the speaker's speech. This perception must be subtly but constantly affecting our response to all speech that we hear, even though we are not consciously aware of it.

14.1 The Story So Far

Here is a summary of the important points in the theory developed so far:

- It is necessary to develop a theory of music that gives a satisfactory evolutionary explanation of music in terms of biological purpose.
- But music itself doesn't have to have a purpose: perhaps it is only the *response* to music that has a purpose.
- The human response to music can be described as the perception of **musicality**.

- Musicality is a perceived aspect of speech, and in particular it is a perceived aspect of the speech of a single speaker speaking to the listener.

- Music is a **super-stimulus** for musicality.

- We can identify plausible cortical maps that respond to the observed aspects of music. These cortical maps include the regular beat cortical map, the note duration cortical map, the scale cortical map, the harmonic cortical map, the melodic contour up/down cortical map and the home chord cortical map. The representation of repetition is uncertain: it may be an aspect of other cortical maps, or it may have a cortical map of its own.

- The various aspects of music are super-stimuli for the perception of components of musicality in these and other as yet unidentified cortical maps involved in the perception of corresponding aspects of speech melody and rhythm.

- Music and speech perception have various symmetries of perception. Some of these are functional, including pitch translation invariance, time scaling invariance, time translation invariance and amplitude scaling invariance. Others are non-functional but play a role in the efficient implementation of perceptual functions; these include octave translation invariance and pitch reflection invariance.

- Some of the cortical maps identified from consideration of aspects of music perception can be interpreted as satisfying the requirements of perception invariant under these symmetries. For example, the scale map, the home chord map and the harmonic map all provide pitch translation invariant characterisations of speech melody. The regular beat map and the note duration map play roles in providing time scaling invariant characterisations of rhythm.

14.2 So What is Musicality?

The development of the theory so far rests very strongly on the concept of "musicality" as a perceived aspect of speech and music. But I have not said very much about what musicality actually is, how it is perceived, what it means, and what purpose is served by the perception of it.

We have identified cortical maps that respond to aspects of music and speech. All the maps identified so far have their own identifiable purpose, i.e. the perception of speech melody and rhythm invariant under the required symmetries. To put it another way: these maps are involved in the perception of music, and therefore they must contribute information to be processed so as to calculate musicality, but their major purpose is something else, i.e.

the perception of speech melody and rhythm. Musicality is an extra output extracted from the information processed by these cortical maps.

14.2.1 A List of Clues

This seems to leave musicality as a mystery nowhere near being solved. The only result of our endeavours is to be mildly confident that there is such a thing as musicality, and that it is perceived by the musical parts of the brain.

But we do have a number of significant clues as to what musicality might be and what it might mean:

- Musicality is an attribute of the speech of a single speaker. And it is distinct from the attributes that we already know about: speaker identity, syntax, speech melody, speech rhythm and semantic content.
- Music has an emotional effect on the listener, and the intensity of the emotional effect is a function of the level of perceived musicality.
- Music has many different aspects: melody, rhythm, harmony, bass and repetition. Some musical genres put more emphasis on some aspects, and other musical genres put more emphasis on other aspects.
- The patterns of neural activity in cortical maps responding to music contain regularities which are not found in the responses of the same cortical maps to speech.

We will consider each of these clues in more detail.

14.2.2 Musicality is an Attribute of Speech

The perceived musicality of speech tells us something about the speech of an individual speaker. What is this something? We can start with the negatives, what musicality is *not*:

- Music does not have any semantics in the usual sense. A tune does not tell us information about the world in the way that speech or writing do. Some like to say that music is a universal language. It is true that people of different cultures and nationalities have a reasonably consistent response to the same music.[1] But no one is claiming that we can use music to communicate specific information about the world.
- Speech contains rhythm and melody. Speech melody must be processed by the listener in as much as lexical speech melody and intonational speech melody provide part of the semantic content of speech. Speech rhythm must be recognised in order to efficiently and reliably identify

[1] Although an ability to respond to a given type of music may depend on being exposed to that type of music at a sufficient early stage of one's life.

syllable boundaries. But there is no particular evidence that the perception of musicality contributes to these processing steps. It is more the case that melody and rhythm provide information required to calculate musicality.

- Music does not have syntax. Some researchers talk about the "grammar of music". The grammars of natural languages are approximately equal to what are technically known as **context-free grammars**. (Most computer programming languages can be defined by context-free grammars.) The main objection to the idea of musical grammar is that no one has actually written one down. Context-free grammars are reasonably straightforward to describe. Anyone who claims that there is such a thing as musical grammar or syntax should be prepared to support their claim by writing the supposed grammar down in a notation such as **Backus-Naur normal form** (a standard notation for describing context-free grammars).

- Speaker identity. Musicality would appear to be largely independent of speaker identity. The musical correlate of speaker identity would be the identity of a singer, or the identity and timbre of musical instruments. Different types of music and song do suit certain types of voice and singer and certain types of musical instrument. But, subjectively, we would still say that speaker identity and musicality are quite distinct percepts: knowing *who* is singing is different (and *feels* different) from knowing *what* they are singing.

14.2.3 The Emotional Effect of Music

Music has two major effects on the listener: emotional and pleasurable. The two are somewhat interconnected: the more intense the emotional effect, the greater the pleasure. It is paradoxical that music always makes us feel good, even though sometimes the emotions it evokes are those associated with feeling bad, e.g. sadness or loneliness.

If the perceived musicality of speech is telling us something about that speech or its speaker, then this emotional effect should be the major clue as to what we are being told.

The most obvious explanation is that musicality tells us something about the emotions of the speaker. Certainly the internal emotional state of the speaker is an important thing to know about. There are many clues in the manner of someone's speech as to what their emotions are, and listeners do pick up on these clues. In some cases we can identify a speaker's emotional state even if they are making some effort to conceal their emotions. Musicality may represent some portion of this perception of the speaker's emotions.

This notion is an attractive one, but there are a couple of objections to it:

- Emotion is a multi-dimensional attribute. We will see that there are reasons to suppose that musicality is purely one-dimensional. If musicality represents information about some component of the speaker's emotional state, then this component would have to be restricted to a single dimension. But musicality interacts with different types of emotion, so it cannot be restricted to just one type of emotion. It could be that musicality determines the *intensity* of emotion, independently of the *quality* of emotion, and that if other aspects of music determine the quality of emotion, those aspects are distinct from the aspects that determine musicality.

- Musicality seems to interact with the *listener's* emotions rather than with the listener's perception of the *speaker's* emotions. The primary effect of listening to a sad song is that it gives the listener a feeling of *being* sad (a feeling that paradoxically they may enjoy), as opposed to a feeling that the *singer* is sad (although usually there is also some perception that the singer feels the emotion of the song). If musicality is supposed to represent information about the state of the speaker's brain, then why is it telling us about the emotional state of the listener?

14.2.4 Different Aspects and Genres

A major assumption that I have made so far, and which is generally made by most music researchers, is that music is a single phenomenon. We assume that we have not accidentally grouped several distinct phenomena together and called them all "music".

Yet there are many aspects of music, and there are many different genres of music. Different genres emphasise different aspects of music. For example, some genres have complex syncopated rhythms, whereas others have very simple rhythms. If music is a single unified phenomenon, then we need to understand how all these different aspects and genres of music can be explained within a unified explanation of what music is.

We can almost dissociate the musical aspects of rhythm and melody. Although pure rhythmical percussion seems a bit boring to most Western ears, there are types of musical performance that only involve percussion. And these percussive performances do have a definite musicality, even if it is not as intense as what we get from the usual combinations of rhythm, melody and harmony.

Most melodies have some unevenness of rhythm. But I know of at least one popular classical tune that has a very even rhythm: the main theme of "Jesu, Joy of Man's Desiring" by Johann Sebastian Bach, where the rhythm is a continuous 1-2-3 which only stops when the tune stops. (And, subjectively, I would say that the main strength of the tune is contained in the portion before the end.)

Modern popular music makes heavy use of syncopation. In contrast, most popular classical music items and traditional Western European folk songs (in particular those traditional folk songs that remain popular today) are relatively un-syncopated. (The syncopation that exists in modern pop and rock music is a descendant of West African rhythms which originally came to North America in the musical cultures of African slaves.)

The first example—percussive music devoid of melody or harmony—is perhaps the most challenging to any theory of musicality. We must presume that the information coming out of all the pitch-related cortical maps plays a major role in the brain's calculation of musicality. Yet the brain is fully capable of perceiving non-zero musicality in music that has *no melody at all*.

So if we formulate a definition of musicality where melody is an essential component of that definition, then our definition must be wrong. At the same time our definition must explain the fact that the strongest values of musicality can only be achieved if there is both melody and rhythm.

This dissociation across different aspects suggests that musicality reflects some fairly general property of music (or speech) which can be measured across different aspects both individually and in combination. The greatest musicality will be found if the general property holds for all or most aspects, but some musicality will be detectable even from a lesser set of aspects.

The idea of a general property gets more support from the next clue.

14.2.5 Constant Activity Patterns

Recall some of the activity patterns that we have observed in cortical maps that respond to music:

- The scale map responds to pitch values modulo octaves that have occurred in the immediate past. If musical melody is on a scale, then the neurons corresponding to pitch values on the scale are active, and the neurons representing in-between values are not active. Therefore there are zones of activity and zones of inactivity. The number of active zones corresponds to the number of notes in the scale (in each octave), typically 5 to 7.

- The harmonic map responds to pitch values modulo octaves in the current chord. Neurons representing notes in the chord are active, notes not in the chord are inactive. The pattern of active neurons remains constant for the duration of each chord. There are typically 3 or 4 notes in a chord, so the harmonic map will have 3 to 4 zones of activity.

- The home chord map responds to pitch values in the home chord, where the choice of home chord is determined by the relationships between notes in the scale and the occurrence of the home chord in the music before the occurrence of any other likely choice of home chord. A home chord has 3 notes, representing 3 zones of activity.

- The regular beat map responds to regular beats in a tune. These typically include regular beats whose period is one bar, one count, the shortest beat period (usually corresponding to the smallest fractional note length) and other durations in between. (This assumes that periods that are multiples of a bar length are outside the range that the regular beat map responds to.) A typical tune in 4/4 time (4 counts per bar) with sixteenth notes will have 5 identifiable regular beat periods: 1 bar (= 4 counts), 1/2 bar = (2 counts), 1 count (= a quarter note), 1/2 count (= an eighth note) and 1/4 count (= a sixteenth note). These 5 regular beats would generate 5 zones of activity in the map.
- The note duration map responds to lengths of notes. The set of note lengths will include all the periods active in the regular beat map, and may also include additional small multiples, i.e. three times a beat period, or (for a beat period that occurs in multiples of three) two times a beat period.

There is a common pattern occurring in all these cases:

- A cortical map that responds to music in the following manner:
  - Activation of neurons within a number of active zones.
  - The number of active zones ranges from 3 to 7.
  - Little or no activity outside the active zones.
  - The location of the active zones remains constant for the whole tune, or in the case of the harmonic map, for substantial portions of the tune.
- These **constant activity patterns** only occur in response to music—they do not occur in response to normal speech melody and speech rhythm.

We might suppose that similarities between activity patterns in the scale map, the harmonic map and home chord map are caused by the similarity that they have in their rules of activation—they all respond to pitch values, and the latter two both include mutual activation between harmonically related pitch values.

But the regular beat and note duration maps respond to a completely different type of information: duration rather than pitch. This similarity between activity patterns in the pitch-valued maps and those in the duration-valued maps is too great to be ignored—it seems to be telling us something fundamental about the nature of music and musicality.

So here is a hypothesis about musicality:

- Overall musicality is calculated from the musicality of activity in individual cortical maps that respond to the speech of an individual speaker speaking to the listener.

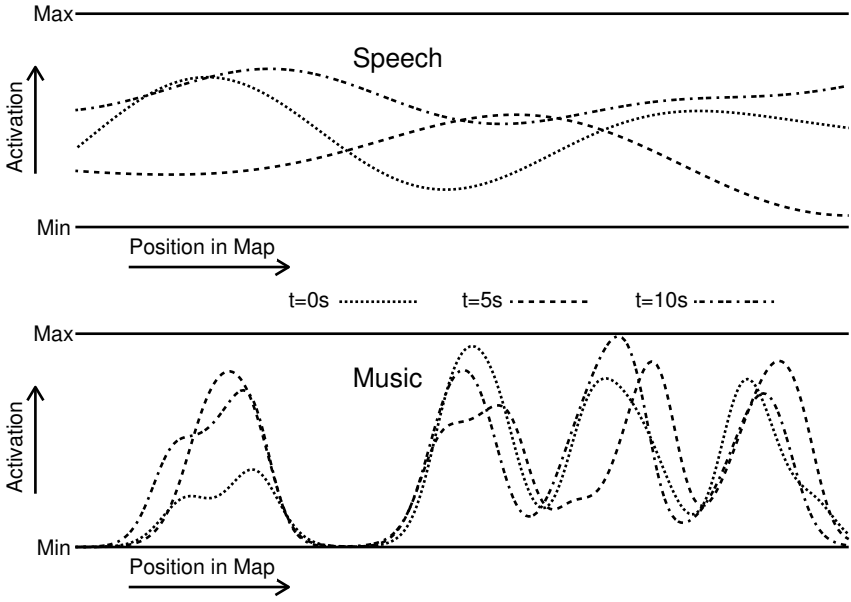

Figure 14.1. Activity levels across one dimension of a hypothetical cortical map at three moments in time (t = 0, 5 and 10 seconds), while perceiving a spoken sentence, and while perceiving a musical item. Neural activity when perceiving the speech is more spread out, and the activity is not restricted to any portion of the cortical map. Neural activity when perceiving the music is restricted to four (in this instance) fixed zones of activity, and more often reaches maximum levels within those zones.

- Musicality for an individual cortical map is a function of activity in that cortical map such that the function takes on a maximum value for a pattern of activity in which all the activity takes place within certain zones, and there are several such zones in the cortical map, "several" being from 3 to 7, depending on the cortical map.

14.3 The Musicality Neuron

Thus each cortical map relevant to the perception of musicality has its own **musicality function**. We have defined this function as being maximised when the cortical map has constant activity patterns. This leaves unstated what the function actually is.

It seems reasonable to assume that the function might have some degree of locality, i.e. that it is calculated by individual neurons that detect the constant

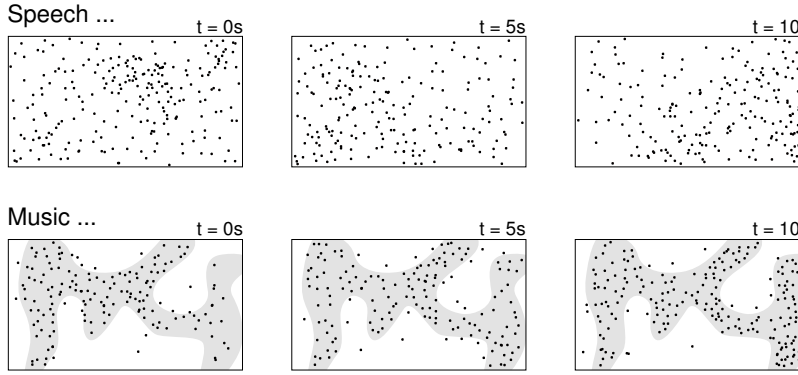

Figure 14.2. 2-dimensional views of the response of a hypothetical cortical map to a spoken sentence (top row) and a musical item (bottom row), at three moments in time, each 5 seconds apart. Each view of the map shows active neurons as black dots. When speech is being processed, the neural activity is spread across the map, whereas when music is being processed, most of the activity is restricted to a particular active zone (shown in gray).

activity patterns over local portions of a cortical map. If we consider what a constant pattern of activity looks like over a small region, there are three main possibilities:

1. All of the neurons in the region are active.
2. None of the neurons in the region are active.
3. Some of the neurons in the region are active, and some are inactive.

Given that the constant activity patterns caused by music often contain more than just one active zone and one inactive zone within each cortical map where they occur, we might suppose that there is a tendency to maximise the number of small regions in which the third pattern of activity occurs, i.e. where some neurons are active and some remain inactive.

So we can suppose the existence of a **musicality neuron**, which detects the occurrence of constant activity and inactivity of those neurons within a region that it is connected to. Since, in practice, neurons that input into a neuron must have either inhibitory or excitatory connections, each musicality neuron must have a fixed division of its inputs into those expected to be active and those expected to be inactive, and the musicality neuron will only be activated when the actual activity of the neurons that it receives input from takes on this pattern. In effect the musicality neuron is an "edge detector", which detects a particular edge between an active zone and an inactive zone within a larger pattern of constant activity in the cortical map.

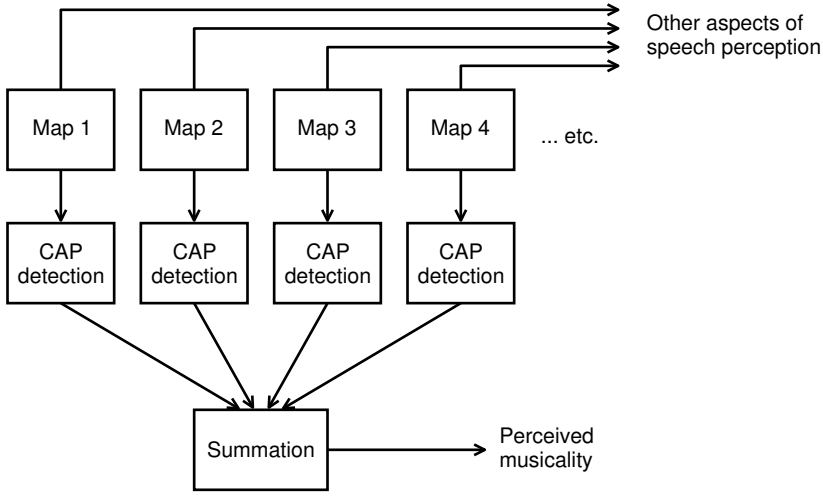

Figure 14.3. Detection of constant activity patterns (CAP) in multiple cortical maps. Maps 1, 2, 3, 4 etc. produce information relevant to speech perception. In addition, CAP-detecting neurons in all these maps detect CAP for each map, and then combine the individual CAP levels to generate an overall value that determines the perceived musicality of speech (or music).

Of course, at any point in time, some of the inputs to *any* neuron will be active, and others will be inactive. In order to respond only to constant activity patterns, the musicality neuron will have to be strongly and quickly inhibited by its inhibitory inputs, and weakly and slowly excited by its excitatory inputs.

It is possible that the musicality neurons receive all their inhibitory connections from intermediate neurons that receive excitatory connections from the perceptual neurons in the cortical map. Neurons on both sides of the "edge" being detected are generally going to be the same type of neuron, usually excitatory neurons, and if excitatory neurons can only form excitatory (out-going) connections, then an intermediate neuron is required to translate excitation into inhibition. (One alternative is that the musicality neuron has input synapses through which it is directly inhibited by what are normally excitatory neurotransmitters, but this seems anatomically less plausible.)

Since each musicality neuron of this type only detects activity and inactivity from one particular division of its inputs, there will need to be more than one musicality neuron to detect activity and inactivity in a given set of perceptual neurons according to different divisions of those that might be active and those that might be inactive.

The required characteristics of these musicality neurons are such that they

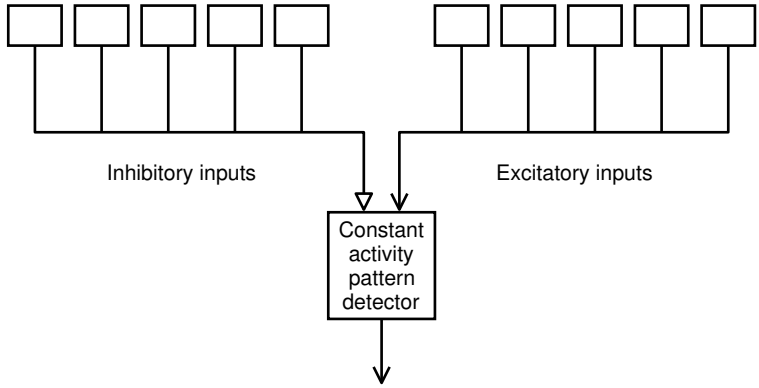

Figure 14.4. A CAP-detecting neuron that detects an "edge" between inactive neurons in the group on the left and active neurons in the group on the right. Inhibition from the left group is long-lasting in the sense that any input from the left causes inhibition of the CAP-detecting neuron for some period of time. Excitation from the right group is slow-acting in the sense that activity must occur for a while in the right group neurons, without any inhibition from the left group neurons, before the CAP-detecting neuron becomes active.

are quite likely to have a unique anatomical form, with a unique population of synapses and synapse types. Furthermore, the output of the musicality neurons will be routed to one particular location, where information about musicality from different cortical maps is combined to calculate a final musicality value, and from that location the information will be sent to those parts of the brain that can influence a listener's emotions. These expected characteristics may make it possible to identify musicality neurons according to their form, their intrinsic response characteristics, and the patterns of connections they form to other neurons (both for incoming information and for outgoing information).

If musicality neurons detect only the edges within a constant activity pattern, then the more edges in the pattern, the greater will be the perceived musicality. For example, if a piece of music is in a 7 note scale, then the scale cortical map will respond to that music with 7 active zones and 7 inactive zones, which will give rise to 14 edges (along the dimension representing pitch).

Why not have even more than 7 active zones? Why not go for 20 or 100? Part of the answer to this question has to do with constraints on the operation of the cortical maps themselves. For example, the regular beat cortical map will only respond over a limited range of beat periods, and neurons for two different beat periods can only be active simultaneously if one of the beat periods is a multiple of the other. This constrains the total

set of beat periods to be a sequence of values such that each one is a small integral multiple of the previous one.[2]

In the case of chord-related maps, the maps themselves inhibit the activation of neurons corresponding to pitch values that are not harmonically related to notes already active in the maps. Thus most chords in popular music do not have more than 4 notes in them, and the home chord never has more than 3 notes.

But there appears to be no such restriction that would apply to the scale cortical map. Why not have scales with much more than 7 distinct notes, and thereby achieve greater musicality? The answer in this case may come from considering population encoding. If we consider the neurons in a pitch-valued cortical map to be very precisely tuned, then the active and inactive zones can be thought of as a series of sharply defined black and white stripes. But if the tuning is not so sharp, then the stripes will be fuzzier. If the stripes are sufficiently fuzzy, and there are too many stripes crowded into the cortical map, then there will not be a clear separation of active zones and inactive zones. This may explain why each cortical map has an optimal number of stripes to achieve maximum possible musicality: too few and there is not as much border between active and inactive zones as there could be, too many and there cease to be distinct regions of activity and inactivity. Figure 14.5 shows this in a simplified 1-dimensional model of constant activity patterns in a cortical map.

There are some musical cultures that use scales with many more notes than 7. For example, a 22 note scale is used for traditional Indian music. However, not all notes from the scale are used at one time. **Ragas** are sets of 5 to 7 notes chosen for an individual composition, so in effect the ragas are the actual scales. (This is an over-simplification, as a raga may contain additional rules about how and where notes are played, and there may be different sets of notes going up and going down, so it might be more accurate to regard a raga as a mini-genre, rather than just a scale.)

14.4 Discount Factors

Most of the cortical maps involved in perceiving music respond to one item of music with an activity pattern that is (usually) constant for the duration of the item. A major exception to this is the harmonic cortical map. The activity patterns in this map must change suddenly each time a new chord is introduced. Presumably this would temporarily reduce the perception of musicality at that time, because musicality neurons respond less when the cortical map is changing its activity pattern. Yet our subjective feeling when

[2] Although, as already mentioned in Chapter 3, some non-Western music makes extensive use of polyrhythm, i.e. music with multiple simultaneous tempos that are not multiples of each other. Polyrhythmic music would create a finer pattern of active zones in the regular beat cortical map.

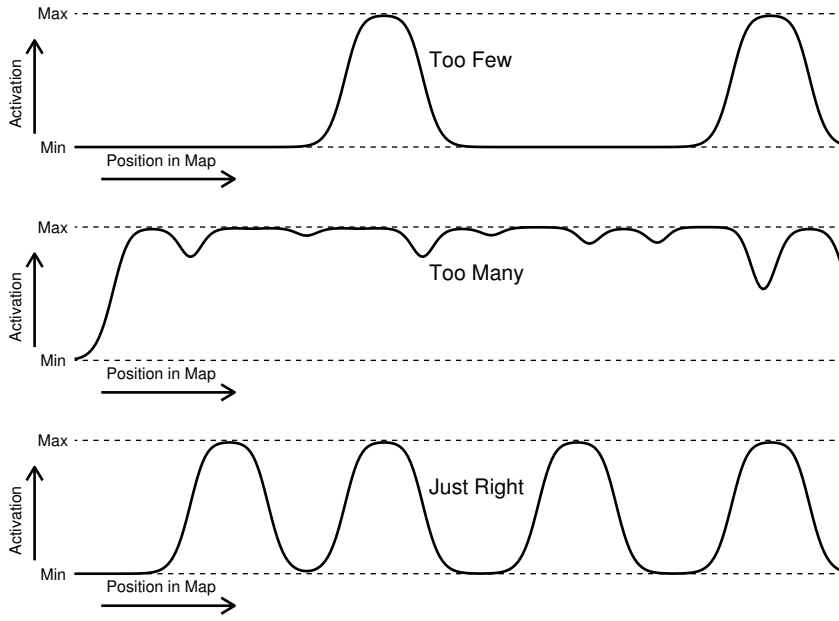

Figure 14.5. Too few, too many, just right. Activity patterns across a 1-dimensional section of a speech-perception-related cortical map, in response to three different tunes. The first creates 2 clear peaks which results in 4 detectable edges. This does not maximise the measured musicality of the activity patterns, because there is room for more edges in the map. The second has 12 peaks, which, due to population encoding, merge together, and do not create any well-defined edges. The last one has 4 peaks, which gives 8 well-defined edges—probably close to the maximum possible for this map, given the peak width caused by population encoding.

listening to music is often that a chord change is a point of *maximum* musicality and emotional effect.

The harmonic cortical map has the property that the activity pattern tends to stabilise, because neurons corresponding to harmonically related pitch values mutually reinforce each other, and they also inhibit other pitch values not active in the map. The map therefore reaches a stable activity pattern, regardless of any particular properties of the perceived melody. This pattern remains until the map is reset according to melodic features that cause it to be reset, i.e. a strong beat coinciding with occurrence of pitch values distinct from those currently active in the map, perhaps reinforced by occurrence of a low pitch value as processed by the bass cortical map.

The strong musicality at the start of a new chord can be explained if we assume that the calculation of musicality is linked to the operation of the

reset function of the cortical map. If the map has not been reset for a while, then the observation of a constant activity pattern has less significance and is **discounted**. The period immediately after a reset is the time when this cortical map is not normally expected to show a constant activity pattern, because it is still settling in to a new stabilised activity pattern, so occurrence of a constant activity pattern at this time is *not* discounted. And when responding to music, the harmonic map achieves a constant activity pattern almost immediately after a reset, so the activity pattern becomes constant at a time when it is not expected to be constant in the speech case, and since the constancy is not discounted at that time, it contributes to the level of perceived musicality.

But what if the notes of a chord are not all played immediately at the point where the chord change occurs? In this circumstance, the activity pattern in the harmonic cortical map will not become constant until all of the notes in the chord have occurred. This would seem to imply that the musical effect will be weaker if the discount factor applies to any point in time that does not occur immediately after a reset of the map.

One possible solution is that the discount factor is reduced not just after a full reset, but after any sudden change to the state of the harmonic map. So if the notes of a chord are played sequentially, then each new note of the chord will count as a sudden change to the state of the harmonic map, and the discount factor will be correspondingly reduced for some period of time after that new note is played.

14.5 The Meaning of Musicality

We have identified a plausible neural correlate of musicality, and suggested that this relates to something that influences the listener's emotions. But we still haven't said what the *meaning* of this perceived musicality is, and why it matters so much to perceive it.

The property of musicality as based on activity patterns in cortical maps is a property of the state of the *listener's* brain. But in as much as musicality is a perceived attribute of the *speaker*, it seems to be an attribute of the state of the wrong brain.

One plausible way out of this difficulty is the **echoing hypothesis**:

> The state of activity of those cortical maps in the *listener's* brain concerned with *perceiving* speech **echoes** the state of activity of those cortical maps in the *speaker's* brain concerned with *generating* speech.

By "echo", I mean that the activity patterns in the listener's brain are a partial copy of the activity patterns in the speaker's brain. In particular, if the speaker's cortical maps for generating speech have an increased level of constancy of activity patterns, the listener's cortical maps for perceiving

speech will have a corresponding increased level of constancy of activity patterns.

The correspondence between the state of the listener's brain and the state of the speaker's brain will not be perfect, but it may be sufficient that the listener can perceive with some degree of confidence the occurrence of constant activity patterns in the speaker's brain. We have already seen that musicality appears to be measured separately over a range of cortical maps, with these separate measurements then being combined into an overall perceived musicality. This combination of multiple measurements may be sufficient to see past the noise caused by the imperfection of the correspondence between the speaker's brain state and the listener's brain state.

The echoing hypothesis shifts the perception of musicality from being the perception of the listener's brain state to being the perception of the speaker's brain state. But we also have the problem that the emotional effect of musicality seems to relate to the *listener's* emotions rather than the *speaker's* emotions. And why should constant activity patterns tell us anything useful about the state of the speaker's brain?

14.5.1 The Conscious Arousal Hypothesis

The theory becomes more speculative at this point. Having supposed that musicality is the indirect perception of constant activity patterns in the speaker's brain, we need to develop a plausible hypothesis as to what the constant activity patterns would be caused by, and therefore what it is that the perception of them actually tells the listener about the speaker.

So far we have the following ideas:

- Musicality is the perception of *something*.
- Musicality is the perception of constant activity patterns across cortical maps.
- Musicality has something to do with emotion.

A plausible conclusion is that the echoing of constant activity patterns amounts to an echoing of information about the emotional state of the speaker. But the appropriate emotional response of a *listener* to particular content is not necessarily the same as the emotions that the *speaker* may be feeling when delivering that content. Even if the content of the speaker's speech is emotionally significant to both speaker and listener, there are many reasons why the specific emotions are unlikely to be the same for both parties.

A good example of a sentence which causes different emotions in speaker and listener is: "I don't love you any more". This has emotional significance for speaker and listener, but different in each case. However, although the *emotional responses* of the speaker and listener are different, we might expect that the speaker would be **consciously aroused** when delivering any speech

that is emotionally charged, independently of which particular emotion is appropriate for either speaker or listener. So perhaps perception of constant activity patterns in the speaker's brain is a means of perceiving the speaker's arousal level.

I therefore propose that musicality is a measure of the *conscious arousal* of the speaker, and the result of the perception of a high level of musicality—implying a high level of arousal in the speaker—is for the listener to accept the emotions that they have in response to the content of the speech. Figure 14.6 summarises the flow of information according to this theory. The implication is that the perception of musicality is a "truth detector":[3] if the speaker says something of emotional significance, and the perceived musicality of their speech indicates that they are consciously aroused, then this perceived arousal is at least consistent with the truth of what they say. If they make an emotionally significant statement, but are not themselves consciously aroused, then it is likely that what they say is not true, or is not as significant as it seems.

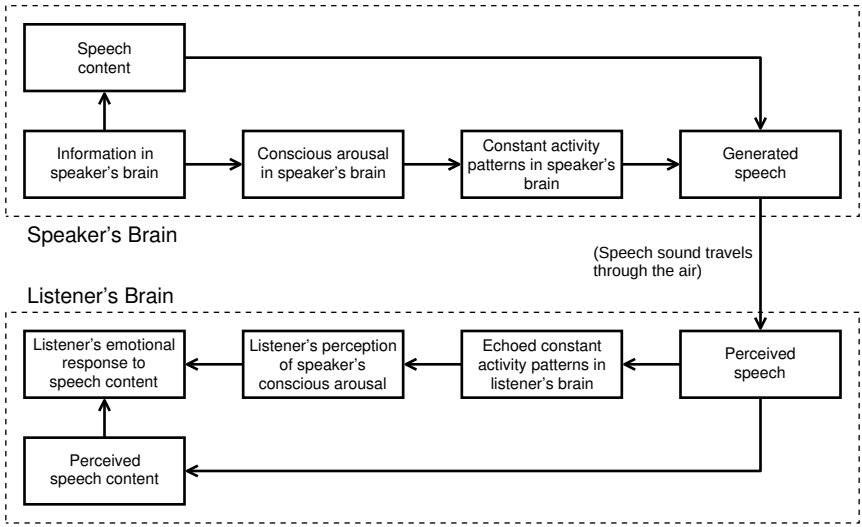

Figure 14.6. Information flow in and between the brains of the speaker and the listener. What the speaker knows determines both their level of conscious arousal (which in turn determines their CAP level) and the content of their speech. When the listener listens to the speech, they extract both speech content and information about the level of CAP in the speaker's brain. The estimated CAP level influences the listener's emotional response to the information in the speech content.

[3]More or less the opposite of a "lie detector", with the qualification that the musicality-based truth detector only detects the *likely* truth of what is said under *some* circumstances (i.e. where what is said is of emotional significance).

What exactly do we mean by "conscious arousal"? It is difficult to give a precise definition, partly because it depends on our (very limited) understanding of what consciousness is. If there is some alteration of brain state that causes neurons in many locations to consistently behave in a certain way (giving rise to constant activity patterns), then this alteration is likely to be mediated by a neurotransmitter that is non-specific in its effect on target neurons. There are a number of neurotransmitters that undergo such non-specific transmission—major ones include **norepinephrine**, **acetylcholine** and **serotonin**.[4]

These neurotransmitters **modulate** the activity and responses of neurons in the cortex. A full understanding of the meaning and purpose of such modulation may only come with a full understanding of what consciousness is: suffice to say that the mystery of consciousness is perhaps even more mysterious than the mystery of music.

One hypothesis that can explain the meaning of modulatory transmitters is the **non-routineness hypothesis**.[5] This hypothesis asserts that the short-term purpose of consciousness is to deal with non-routine circumstances. "Non-routine" circumstances can be defined (somewhat circularly) as those circumstances sufficiently out of the ordinary that they cannot be dealt with by non-conscious information processing. The processing of "non-routineness" occurs in several stages:

1. Detect non-routineness of current circumstance.

2. Use one or more modulatory neurotransmitters to broadcast a message to all relevant parts of the brain that a non-routine circumstance is occurring.

3. Neurons respond to the "this is not routine" message by altering their mode of operation accordingly. In the non-routine mode the neurons give preference to the use of learned information that is general rather than specific. The mode also causes neurons to perform calculations in such a way that the results might be less certain (given the greater difficulty of calculating the correct response to a situation not so similar to what you have previously experienced), on the assumption that the result of such calculations will be subject to further checking (i.e. see next step). So the response of neurons in this mode is more of a suggestion than a definite decision.

4. A secondary checking procedure is applied to the suggested response to the non-routine circumstance. This secondary check corresponds to our subjective experience of conscious judgement.

[4] **Dopamine** is another modulatory neurotransmitter, but its targets are non-cortical and not as widely distributed as those for norepinephrine, acetylcholine and serotonin.

[5] More information about this theory can be found on the author's website at **http://www.1729.com/consciousness/.**

5. A final "yes" or "no" decision is made about the suggested response, with a "yes" acting to confirm that the response should be put into action.

The altered non-routine mode of neural operation gives more preference to learning that is applicable to a wider set of circumstances, and less preference to the use of learned information that is only applicable to specific circumstances. It may be that both types of learned information are encoded in the connections and strengths of connections found on each individual neuron, but in a way which allows one or the other type of information to dominate the response of the neuron to its inputs, according to the effects of modulatory neurotransmitters on those connections. So neurotransmitters representing a message of "non-routineness" would increase the effect of those connections representing learned information of greater generality.

A change in mode of neural operation into a more non-routine mode is presumed to correspond to increased conscious arousal. A further presumption, to tie everything into the super-stimulus theory, is that the change in mode somehow results in increased occurrence of constant activity patterns, but I do not currently have any concrete ideas about why this should be so.

There are too many unknowns here to have any confidence in a particular account of what is going on. So the reader will have to be satisfied with the following assumptions:

- There exists some global change of state in the brain of a speaker that represents information about the mental state of that speaker.
- This global change of state can result in the occurrence of constant activity patterns in the speaker's brain.
- It is useful for the listener to know something about the occurrence of this change of mental state in the speaker's brain.
- The listener perceives constant activity patterns in the speaker's brain by detecting echoes of those constant activity patterns in corresponding cortical maps in their own brain that process the speech generated by the speaker.

14.5.2 Arousal, Emotion and Emphasis

Because music is a super-stimulus for the perception of musicality, the observed effects of the musicality of music are always the effects of very high (in effect unnaturally high) levels of musicality. Thus we can conclude that the emotional effect results from the perception of a high level of musicality.

But the perception of musicality may satisfy other purposes, even when the perceived level of musicality is not sufficient to generate a significant emotional effect. In particular, it is likely that a person's level of conscious

arousal is constantly varying, according to the routineness or non-routineness of whatever information they are processing at any given moment. Thus the perception of musicality in speech may provide ongoing clues to the listener as to how non-routine (to the speaker) the content of speech is from moment to moment, providing the listener with important information about the relative significance of different things that the speaker says to them.

14.6 Other Cortical Maps

The general nature of the concept of constant activity patterns allows the theory to be extended immediately to other cortical maps involved in the perception of speech, even if we don't know how those cortical maps work, or how they represent meaning:

> Musicality is measured within each relevant cortical map according to the occurrence of constant activity patterns within that cortical map.

It is difficult to interpret the implications for any particular cortical map if we don't know how it represents information about music or speech. But we can look at other aspects of music for which we have not yet identified cortical maps that respond to those aspects, and we can at least see if it is plausible that the CAP theory applies to those aspects.

For example, we can look at repetition. We have observed that non-free exact repetition occurs in music, but does not occur in speech. We have hypothesised that there might exist some cortical map that encodes a repetition count, and that in ordinary speech this repetition count can take on non-integral values, corresponding to repetitions that are close but not exact. It may be that the relationships of similarity and difference between different phrases in a tune cause constant activity patterns in the cortical map that represents values of repetition count. In other words, only certain values of repetition count occur, and in-between values of repetition count do not occur. The values that do occur may or may not be integral values only, depending on the music in question.

There is also no particular reason why the cortical maps involved in the perception of musicality only involve the perception and processing of sound. The concept of musicality may apply to any cortical map that perceives something about the speaker that is associated with the delivery of speech.

The major non-sound-related aspect of speech perception is the visual perception of the speaker—their facial expressions, gestures, and other body language. It is therefore possible that a component of musicality is calculated from activity patterns in cortical maps that process information about the movement of the speaker. And it is quite plausible that *dance* may be a super-stimulus for these cortical maps, and that dance is therefore *an aspect of*

music, and not just something that happens to accompany or be accompanied by music.

Another aspect of music is rhyme. For the most part, particular choices of words in lyrics appear to play no significance in the musicality of music, other than the need for consistency between the rhythm of a tune and the rhythm of its lyrics, and the need for lyrics to have emotional significance that can interact with the musicality of the melody. But rhyme is one aspect of song lyrics which makes them consistently different from normal speech, indeed rhyme is ubiquitous in modern popular song.

To explain rhyme within the framework of the CAP theory, we must hypothesise that somewhere there is a cortical map involved in the perception of music, such that rhyme causes this cortical map to have constant activity patterns, or at least more constant than would be the case without the rhyme.

Almost certainly there are other cortical maps which play a role in the perception of speech, and which are involved in the perception of musicality. Until all the cortical maps involved in the perception of musicality are known and understood, we won't have a complete description and understanding of musicality. This is therefore one of the major challenges arising from the development of the theory so far:

- Identify all of the cortical maps relevant to the perception of musicality.
- Understand the representation of meaning in those cortical maps, and what this implies about super-stimuli that would cause constant activity patterns in them.
- Identify any discount factors that may apply (as in the case of the harmonic cortical map).

14.7 Implication of Identified CAP

A significant implication of the CAP theory is the following:

> If, for a given cortical map, constant activity patterns in that cortical map are identified with the perception of musicality, *the major purpose of that cortical map must be the perception of something other than musicality.*

Musicality is a perceived attribute of the operation of cortical maps that play a role in the perception and generation of speech. If the purpose of one of those cortical maps was solely to perceive musicality, then the logic of the explanation would be too circular: it would imply that the cortical map processed the musicality of the activity patterns of neurons whose purpose it was to detect musicality (so the musicality neurons would be detecting their own musicality).

14.8 Can CAP be Consciously Controlled?

If the perception of constant activity patterns serves to alter the listener's response to speech, it might be useful if the speaker could control the constancy of activity patterns in their own speech-related cortical maps, so that they could influence the emotional response of their listeners.

In one sense you can do this easily, by choosing to deliver your speech in the form of a song. However, assuming that you and the listener live in a culture that explicitly acknowledges the concept of music, the listener will not necessarily be fooled—they will be aware that the musical nature of your delivery will alter their emotional response to the content of what you say.

If there was some non-obvious way to fake the musicality of speech, then less purpose would be served by the existence of a system to detect it in the first place. The persistence (in evolutionary terms) of systems in the brain for perceiving musicality suggests that in fact it is not easy to fake.

Whatever makes it difficult to fake musicality may also explain why it is so difficult to compose music. If your perception of musicality is based on perception of constant activity patterns in your own brain, then one way to compose music might be to define arbitrary constant patterns, activate your neurons accordingly, and this would cause corresponding music to be realised (i.e. composed) within your brain. Unfortunately, we do not have any ability to specify the geometry of neural activity patterns by the direct power of thought alone (although we cannot rule out the possibility that such control could be learnt in the future with the help of suitable feedback devices).

Assuming that it is true that constant activity patterns in your brain are caused by conscious arousal, then one way to control them would be to control your own level of conscious arousal. But if arousal is something that controls consciousness, then it follows that consciousness must be prevented from being able to directly control the level of conscious arousal, otherwise the logic of control becomes too circular, and positive feedback is likely to occur. (Although indirect conscious control of conscious arousal may be achieved by subjecting yourself to circumstances likely to result in more or less arousal. I cannot easily choose to become consciously aroused just by thinking myself into it, but I could, for example, choose to ride on a roller-coaster, which is likely to make me consciously aroused.)

14.9 Constraints

How large is the set of melodies, and how large is the set of *musical* melodies?

If we consider *arbitrary* melodies—musical, speech or otherwise—then the set of possible melodies is a very large set. The exact size of the set depends on the precision of perception and how long we might allow a melody to be.

If we imagine listening to random melodies selected from this set of possible melodies, then most of those melodies would not have any musical merit

whatsoever.

Starting from the full set of possible melodies, we can apply various constraints, one at a time, to reduce it down to the set of musical melodies. For each constraint applied, there will be a corresponding reduction in the number of melodies in the remaining set.

Before we apply any constraints, we can reduce the size of the set by deeming melodies related to each other by symmetry transformations to be the *same* melody. In particular, a group of melodies related to each other by pitch translation and/or time scaling are to be considered a single melody. Pitch translation can be regarded as not significantly altering musical quality unless the translation is very large. Musicality is somewhat more sensitive to time scaling, i.e. there is usually a preferred tempo for performing a given item of music. Applying both symmetries, we can consider the canonical representative of a melody to be the pitch-translated time-scaled version of the melody that has the greatest observed musicality.

Next we can apply a series of constraints that follow from the basics of music theory:

- That pitch contours consist of notes, where each note consists of a constant pitch value that starts at a certain time and ends at a certain time.

- That the pitch values of the notes are taken from a finite set of values from a scale defined modulo octaves.

- That at least some of the intervals between pairs of notes in the scale are approximate or exact consonant intervals.

- That the range of steps in the scale is such that the largest step is not more than three times the size of the smallest step (and usually not more than twice the size).

- That the scale is sufficiently uneven that it is not invariant under any translation by an interval smaller than an octave.

- That the times of the note boundaries belong to a set of discrete times defined by a regular tempo (i.e. corresponding to the shortest beat period).

- That there is a series of tempos which define beat periods, each one a multiple of the previous one in the series, starting with the shortest beat period, and finishing with the bar length, such that notes starting at the beginnings of the longer beat periods are given more emphasis.

With these constraints applied, we are left (approximately) with the set of melodies that can be written using standard musical notation. A random

member of this set is like a melody defined by a random page of sheet music written in a particular key.

This set of melodies that we have defined is much smaller and more constrained than our original set of all possible melodies. But it still mostly consists of melodies with limited musical merit. There will be some degree of musicality as a consequence of the constraints that notes belong to scales and that rhythm exists within regular hierarchical tempo, but it is not going to be enough to guarantee a record deal.

At this point we can attempt to impose further constraints, based on our understanding of aspects of music theory not already implied by the constraints we already have. These further constraints are unfortunately somewhat more vaguely defined, but they will noticeably increase the average musicality of melodies in the constrained set:

- That each note duration is a length equal to one of the beat periods, or a very small multiple thereof (almost always 2 or 3).
- That there should be an identifiable chord for each bar, and the notes in the bar should relate to this identifiable chord: the notes on the main beats should almost always be members of the chord.
- That there should be a chord at the start (or maybe very near the start) which is one of the preferred home chords for the scale (A minor or C major on the white notes scale).
- That the tune should finish on the home chord with a long note which is a member of the home chord (and most likely the root note of the home chord), and the final home chord should probably be preceded by the associated dominant 7th chord.
- That in most cases the note following a note should be either the same note or one step above or below in the scale. Where the step between notes is a larger interval, it should not be more than an octave, and it should be a consonant interval. If the consonant interval is within the bar it should be part of the chord associated with that bar.
- That the tune should be divisible into phrases of consecutive notes, with some type of identifiable similarity between different phrases. This similarity may consist of exact repetition, or repetition of certain aspects only, such as the rhythmic structure or the melodic contour (or both of those two aspects).

Some of these constraints are fuzzy and probabilistic: they describe rules that should apply "most of the time", or "as often as possible". Pure binary constraints (either it is satisfied or it is not) define a set such that a melody is either in it (all the constraints are satisfied) or it is not (at least one constraint is not satisfied). Probabilistic constraints define a probability of

the melody being acceptable, and the probabilities calculated for all of the constraints should then be multiplied together to get an overall probability, where 1.0 means the constraints are definitely satisfied, and 0.0 means they are definitely not satisfied.

Applying this second set of constraints corresponds to applying some of the rules of composition specified by music theory. But anyone who has sat down at a piano and tried to compose music according to these so-called "rules" (and maybe some additional rules) will know that the enormous majority of melodies in this set do not have significant musical merit, and they will still not be good enough for that record deal that we hope to get. There is a gap between the rules and constraints that we know of and which can be described formally and objectively, and the full set of constraints that define what our brain is prepared to consider as musical. This gap corresponds to the incompleteness of existing music "theory".

14.9.1 The Implications of Constraint

We have established that the set of musical melodies is much smaller than the set of all possible melodies. Some of this smallness can be explained in terms of known mathematical constraints as contained in well-known aspects of music theory. There is still a remaining degree of constraint which is not explained by existing music theory, and which can be considered a measure of our ignorance about what music is.

At the same time, the degree of constraint is not so great as to restrict the set of musical melodies to only a small number of melodies. There are indeed thousands upon thousands of musical compositions and songs which are considered to be high quality by a significant number of listeners. Given that currently music is composed by a variety of ad hoc processes, we have no real way of knowing how many possible musical compositions could exist for any quality criterion that we might wish to set.

Not only are new musical compositions being constantly produced, but occasionally whole new genres evolve. The twentieth century saw the invention of jazz, blues, rock and roll, heavy metal, and rap, just to name a few.

The conclusion is that the set of musical melodies is a large set, which is defined by applying a set of constraints to a much larger set of possible melodies. Some of these constraints are represented by the known facts of music theory; others represent the things about music that we don't know (yet).

The observations made here about the constraints of musicality are relevant to any theory claiming to explain the existence music. Any such theory must explain the observed properties of the constraints that distinguish music from non-music.

Does the CAP theory successfully explain these properties? We will see that it does, because the maximisation of constant activity patterns in dif-

ferent cortical maps translates into corresponding constraints on different aspects of music.

The activity patterns in cortical maps responding to arbitrary melodies (such as non-musical speech melodies) are not normally going to be constant. Assuming that the maps are arranged so as to make full use of their components under normal conditions of use, we would expect most neurons to be active some of the time and not active some of the time in response to incoming information. Any tendency for activity patterns to become fixed would imply that a cortical map contained redundant components that were not performing any useful information processing.

We have shown that some of the musical constraints relating to known aspects of musical theory arise from the constraint that there should be constant activity patterns in particular cortical maps for all or parts of the duration of a musical item.

It is reasonable to suppose that the same explanation applies to the unknown constraints: that there are corresponding cortical maps involved in the perception of melody, and each constraint is determined by the requirement that constant activity patterns have to occur in the corresponding cortical map if the musicality of the melody is to be maximised.

For each cortical map, applying the constraint that activity patterns be constant in that cortical map reduces the number of melodies in the set of possible melodies by a certain factor. Applying this constraint to all relevant cortical maps gives a combined reduction factor that is the product of all the individual reduction factors. This overall reduction factor will be a very large number, corresponding to the rarity of good music, i.e. the very low probability that a random melodic contour will be highly musical.

At the same time, the CAP theory still allows the set of musical melodies to be very large. For each cortical map, the number of ways the cortical map can be activated over a period of time is very large compared to the number of ways it can be activated that produce a constant activity pattern. But even when the activity is constrained to occur within a constant pattern, there are many different possible constant patterns to choose from. Remember also that the activity pattern refers to the pattern of maximum activity over a medium time frame, and that the level of activity in the active zones within a pattern can vary over the short term. Thus for each cortical map there is a considerable number of possible constant activity patterns, and sometimes there is a considerable number of histories of neural activity consistent with any given activity pattern. We can multiply together all the numbers representing choices of activity patterns for each of the relevant cortical maps, to arrive at an estimate of the total number of choices.[6] The final result of multiplying all these choice factors will be quite a large number. This number

[6] A further complication is that the activity patterns of distinct music-related cortical maps are not all determined independently of each other, so straight multiplication of the number of choices for each cortical map will overestimate the total number of possible musical melodies.

represents the set of all possible musical compositions of high musicality, both those already composed, and those yet to be composed, some of them yet to be composed in genres that are yet to be invented.

In conclusion, the CAP theory successfully explains the observed "constrainedness" of music. If you have your own theory about what music is, make sure that you include the issue of constraints in the list of things that your theory explains.

14.10 Compromises and Rule-Breaking

One of the annoying features of musical "rules" is that no sooner has one formulated some rule that is observed to apply to a wide range of music, one finds that there is always some case where the rule gets broken. Not only does the rule get broken, but it gets broken in a way that subjectively appears to contribute to the musicality of the music that breaks the rule. This inability to find any rules that apply to all music is part of the difficulty of discovering musical "universals".

For example, one rule tells us that musical notes have discrete values taken from a finite set of values in a scale. This is in contrast to speech melody where pitch values vary continuously. But then we have music that contains **note bending**. A note is "bent" when its pitch is altered from its normal value on the scale before, during or after it is played. Certain musical instruments favour the bending of notes: guitar notes can be bent by pushing strings sideways, or by using a slide to define the note. (The electric guitar is the most common source of bent notes in modern popular music.) Other instruments, such as the human voice, the violin and the trombone, allow the musician to play notes at arbitrary pitches, and continuously alter the pitch if desired. An example of an instrument which does not allow any note bending at all is the piano.

There are other rules that get broken. Sometimes time signatures change. Sometimes individual bars have different numbers of notes in them. Examples of well-known popular songs with irregular bar lengths are "Memory" (Andrew Lloyd Weber) and "Money" (Pink Floyd).

Another type of rule-breaking with respect to musical time is where the bar length remains unchanged, but the bar or some part of the bar is divided into a different number of components. By far the most common example is the occurrence of **triplets**, which is where a period of time normally divided into 2 halves is occasionally divided into portions of 3. The opposite of this is **doublets**, where time normally divided into 3 instead gets divided into 2. The similarity of these two variations is somewhat concealed by standard musical notation: a triplet requires a special notation, because 2 is always the default factor for dividing time into smaller portions, whereas a doublet can be notated using a combination of dotted notes and tied notes.

Syncopation can also be regarded as a form of rule-breaking—where the

rule being broken is one that says "minor beats only appear where the corresponding major beat also appears".

And then there are accidentals. Notionally there is a rule saying "only use notes from the diatonic scale". An accidental thus breaks this rule. If you try composing music on a diatonic scale, and then start inserting random accidentals (or changing notes into accidentals), the odds are that you will make your tune sound worse.

With the CAP theory, we can understand the rules of music as arising from optimisation of the constancy of activity patterns within individual cortical maps. For example, playing notes from a fixed scale and not playing any pitch values in between the values from the scale is the *only* way to maximise the constancy of the activity pattern in the scale cortical map. Bending notes would cause activity of neurons corresponding to pitch values that are meant to be in an inactive zone. Playing accidentals would cause sudden activity in what was previously an inactive zone. A change in time signature has a similar effect on the regular beat cortical map: it will change the constant activity pattern that existed before the change of time signature occurred. All of these "rule-breaking" aspects would be expected to reduce musicality.

But the CAP theory also tells us that musicality is summed over the musicality from a number of cortical maps. It is therefore entirely possible that a change to a melody that decreases perceived musicality from one cortical map may more than make up for it by an increase in perceived musicality from another cortical map. So note bending may slightly decrease perceived musicality from the scale cortical map, but may increase musicality perceived in some other cortical map. For example, there might be a cortical map that responds to the rate of change of pitch, and appropriate note bending will cause this map to have active and inactive zones, corresponding to which rates of pitch change occur and which rates don't occur.

This is the concept of **compromise theory**. A **compromise** occurs where the optimal result against a criterion that is a sum of a set of individual sub-criteria may not be optimal for each of the sub-criteria. It provides a reasonable explanation of why there are rules, and yet why at the same time the rules are sometimes broken.

Compromise is not the only possible explanation for musical rule-breaking. There is a general observation that music listeners can develop a taste for more difficult types of music, "difficult" in the sense that other music listeners might struggle to enjoy or appreciate those types of music. Different listeners develop tastes for different types of difficulty; for example, some learn to appreciate more extreme forms of syncopation, others develop a taste for the (somewhat) out-of-tune melodies of "bluesy" music. However, the phenomenon of musical "difficulty" may turn out to be a manifestation of compromise—the "development" of the listener's taste may simply correspond to the wiring up of musicality neurons in the cortical maps whose musicality is increased by the compromise in question.

14.11 Aspectual Cross-Talk

According to the super-stimulus theory of music, the fundamental component of music is the melody. The melody by itself causes activity in all the cortical maps that respond to the different aspects of music: the scale map, the melodic contour map, the home chord map, the bass map, the harmony map, the regular beat map, the note duration map and almost certainly others that we don't know about yet.

The super-stimulus theory deals with various forms of accompaniment by supposing that they increase the musicality of the music by causing (or enhancing) a response within particular cortical maps that respond to those aspects of music manifested by that accompaniment.

Thus the bass accompaniment acts on the bass map. The bass map preferentially responds to the lowest notes in the melody, and the bass accompaniment takes this to extremes by consisting of *very* low notes.

The chordal accompaniment acts on the harmonic map. The harmonic map responds to groups of notes that are related to each other by consonant intervals, and tends to reset itself on a strong beat. This is reflected in the structure of chords and the way they are used in music: chords are groups of notes related to each other by consonant intervals, and they normally change at the start of a bar.

The rhythmic accompaniment consists of purely percussive sounds (with no identifiable pitch value) which act on the regular beat map and also on the note duration map.

Each of these musical components exists primarily in order to affect the activity in particular cortical maps. So the bass accompaniment is designed to act on the bass cortical map, which in turn influences activity in the harmonic map. It is not the primary purpose of the bass accompaniment to activate those maps activated by other features of melody, or to activate those maps that respond to rhythmical features. But in real music we observe that bass can acquire a melodic nature of its own, and also that it often has a rhythmic aspect. Similarly, chordal accompaniments are often embellished to contain their own intrinsic melody, and may be played in a way that provides part of the rhythm of the music.

Now the bass line may exist primarily to act on the bass map, but the brain doesn't know that only the bass map is meant to respond to the bass line. Other maps will respond to some extent to the melody and rhythm of the bass, and this may explain why bass lines tend to acquire their own melody and rhythm.

I call this phenomenon **aspectual cross-talk**. The simplified model of music explains each component of music in terms of one primary aspect of music perception. For each component the model identifies which cortical maps respond to that component. But the model fails to explain why the component has features relevant to other aspects. For example, the theory predicts a bass line consisting only of notes corresponding to the root note of

the current chord. The concept of cross-talk fills this gap in the explanation by admitting that each component is going to cause some degree of response in cortical maps related to other aspects of music. Thus the root note of the current chord is still the most important part of the bass component, in relation to the primary role of the bass component, but at the same time the bass line can contain other notes that give it a melodic character, or a rhythmic character, or both.

14.12 Music/Speech Specialisation

According to the CAP theory, cortical maps involved in the perception of speech are performing two separate tasks:

- Perceiving an aspect of speech.
- Attempting to detect constant activity patterns within the equivalent cortical map of the speaker.

The presumption that a cortical map performs the second task is based on the assumption that activity in a listener's cortical maps in some way copies activity within the speaker's cortical maps.

Even if this assumption is true, it is likely that there is some conflict between the requirements for the first task (doing the actual speech perception), and those of the second task (detecting constant activity patterns in the speaker's brain).

A related issue has to do with what we might call the **overkill factor**. The ability of the brain to perceive and discriminate musical melodies appears to far exceed what is required for the perception of speech melodies. The perception of pitch is far more accurate than what is required for the perception of either lexical or intonation melodies. Linguists argue over how many distinct pitch levels are required to properly describe intonation melodies, but it is generally assumed not to be more than 4 or 5. Intonation melodies are sometimes described in terms of rising and falling pitch contours, and other times they are described as combinations of specific levels and rising and falling contours. But either way, the precision required to perceive intonation melody is much less than the precision of an average person's ability to perceive musical pitch values and melodic contours. The average person can distinguish at least 200 different pitch values within one octave, which is a lot more than 4 or 5.

The number of known distinct musical melodies is in the tens (or maybe hundreds) of thousands. Go to a record store, or go to a karaoke bar and read through the menu book. And this only counts those melodies deemed to be of commercial quality within modern Western systems of music. The total number of melodies (musical or otherwise) that could be distinguished by an average music listener could easily be in the millions.

The number of lexical melodies or intonation melodies that must be distinguished as part of the processing of speech is much lower than this. We would like to be able to prove this by counting the number of such melodies. Unfortunately, intonation melodies do not have distinct identities like those of musical melodies. It is more accurate to say that intonation has different aspects, and each of these aspects bears a relation to the semantics of what is being said, and each aspect has some specific effect on the intonation. Aspects can include things like the contrast between old and new information, and between what is expected and what is not expected, and the distinction between statement, command and question (and also between various sub-types of each of these types of sentence). As far as I can tell, experts in speech intonation are still arguing among themselves about what is the best way to describe intonation in the different languages they study (and across languages as well), so any attempt to actually count intonation patterns is fraught with difficulty. I will just make a weak assertion that it seems to me that the number of recognisably distinct intonation patterns relevant to the perception of speech is somewhat lower than the number of recognisably distinct musical melodies.

These considerations suggest the following **split map** theory:

- Cortical maps used to perceive speech play a corresponding role in the generation of speech; each cortical map concerned with an aspect of speech perception is also concerned with the correctness of that aspect in the generation of speech.
- Musicality is an aspect of speech perception whereby constant activity patterns in the speaker's cortical maps are detected by observation of corresponding constant activity patterns in the listener's corresponding cortical maps.
- Originally each cortical map for speech perception in the listener performed two roles: direct speech perception, and indirect perception of speaker's constant activity patterns in the same cortical map.
- At some point in the evolutionary history of the human species, some or all of these cortical maps evolved into two separate cortical maps: the first a **perceptual map** specialised for perception (and generation) of speech content, the second a **musicality map** specialised for perception of constant activity patterns in the speaker's corresponding perceptual map.

As soon as this split occurred, the musicality maps were free to evolve so as to optimise the perception of constant activity patterns, although they would still have been constrained to correctly echo the activity of their corresponding perceptual maps.

The task of content perception is the most important perceptual task—it matters more often to know what the speaker is saying than it does to

determine subtleties of the speaker's internal mental state. But the task of attempting to perceive internal mental state of another person is perhaps a more difficult task, and we might suppose for this reason that the processing capabilities of the human ear and auditory cortex have evolved to provide the required level of information processing capability.

14.12.1 Double Dissociation Revisited

I have previously mentioned the interpretation of experimental and clinical observations of dissociation between speech perception and music perception. It was observed that one cannot correctly dissociate speech perception from music perception if in fact perception of musicality is an unknown aspect of speech perception.

The split map theory provides an alternative possible explanation of dissociation, since the perception of musicality of activity patterns in the speaker's cortical map X by echoing in the listener's cortical map X has been replaced by perception of musicality in the speaker's cortical map X_p by echoing in the listener's cortical map X_m, where evolution has split cortical map X into perceptual map X_p and musicality map X_m. Dissociation will occur whenever one but not both of X_p and X_m suffer damage in a patient.

Because we now have a theory of music perception being speech perception that explains any possible observed dissociation between the two, the theory is less falsifiable in this regard. However, we don't get everything for free, because the split map theory raises the stakes: we are now hypothesising the devotion of a larger portion of the brain's resources to the task of the perception of musicality. The more resources devoted to solving a problem, the more important the solution of that problem must be, if the cost of those resources is to be justified in evolutionary terms.

14.12.2 The Implied Importance of Musicality

I have already considered the possibility that musicality perception measures the speaker's mental state for two different reasons:

- Validation of the listener's emotional reaction if the speaker is judged to have a high level of conscious arousal when saying something.
- Continuous monitoring of the speaker's level of conscious arousal (not just when it is at a high level), to provide relevant clues about the relative significance of what the speaker is saying.

If major resources are devoted to the perception of musicality, then we are forced to conclude that this perception is useful and important all or most of the time, and not just on rare occasions when someone says something emotionally significant.

In other words, our system of musicality perception is constantly processing information about the perceived mental state of any person talking to us, and the result of this processing is constantly influencing our reaction to the content of their speech, even though we are not consciously aware of this influence.

It seems radical to claim that the solution to the mystery of music is an aspect of perception that –

- is happening all the time,
- but we are not consciously aware of it.

But the amount of time, effort and money that people put into composing, playing and listening to music already suggests that the systems in the brain that process music matter for some reason. Even if all that composition, performance and enjoyment of music is just a wasteful side-effect, evolution must have some good reason to tolerate the waste, implying that the real purpose (whatever it is) of the music-processing systems is something important. Given that no one currently has any idea what music is, when we find out what the important something is, it will necessarily be something that we didn't realise what it was.

Chapter 15

Questions and Further Research

The theories of musicality perception, symmetry and constant activity patterns, together with hypotheses about specific cortical maps involved in music perception, successfully explain many of the observed features of music. But there is still a lot left to explain. This chapter considers some of the unanswered questions. These include questions about musical taste and variations thereof, intensity/position conversion, choruses and verses, and the relationship between emotion and pleasure in music.

15.1 Questions Answered by the Theory

The theory developed in this book provides plausible answers to basic questions about music:

- What is music? *(Music is a super-stimulus for the perception of musicality.)*
- What determines the musicality of music? *(Musicality is determined by the occurrence of constant activity patterns in cortical maps involved in the perception of speech.)*
- Why has music evolved? *(Musicality provides information about the mental state of a speaker, which usefully influences the listener's response to speech.)*

- What is going on inside our brains when we listen to music? *(Cortical maps designed to perceive aspects of speech melody and rhythm respond to music with constant activity patterns. The brain measures the overall constancy of these activity patterns and the result influences the listener's emotions.)*

- What is the relationship between language and music? *(The perception of musicality is an aspect of speech perception; music is a super-stimulus for musicality.)*

The theory also answers questions about the more technical details of music:

- Why are melodies constructed from notes in scales? *(There is a cortical map which responds to the past occurrence of pitch values modulo octaves; this map is active in a constant pattern if, and only if, the melody is constructed from notes of constant pitch taken from a fixed set of values modulo octaves.)*

- Why do notes differing by an octave sound the same? *(The brain splits pitch value information into an imprecise absolute value and a precise value modulo octaves so that it can more efficiently process pitch values.)*

- Why is the musical quality of music independent of the key it is played in? *(Musical quality is invariant under pitch translation because different people speak at different pitch levels.)*

- Why do consonant intervals play such a major role in music? *(Consonant intervals occur naturally as intervals between harmonic components of the human voice; they are used to calibrate the brain's ability to compare intervals between distinct pairs of pitch values.)*

- Why do tunes have home notes and home chords? *(Home notes and home chords are a result of a need for pitch value characterisation that defines characteristics of pitch values in relation to other pitch values that have occurred in a melody, such that the characterisation is pitch translation invariant.)*

- Why is musical time both regular and structured? *(Musical time is regular because there is a cortical map that responds to regular beat. Activity in this cortical map has a constant pattern if and only if musical time consists of beats from a hierarchical sequence of beat periods such that each beat period is a multiple of the next period in the sequence.)*

- Why does harmony occur in music? *(Harmony occurs because there is a cortical map that responds to consonant relationships between different pitch values that occur within a speech melody. This same map happens to be capable of responding to consonant relationships between simultaneous pitch values.)*

- Why is there a bass line? *(The activation and deactivation of neurons in the harmonic cortical map is influenced by a bass cortical map which is activated by the occurrence of pitch values with lower pitch values than other pitch values in a melody.)*

There are still many questions we can ask about music that the theory does not answer, or does not answer fully. The development of the theory also raises some new questions.

The Constant Activity Patterns theory asserts the occurrence of constant activity patterns in cortical maps relevant to the perception of speech (or in cortical maps analogous to those maps, if the **split map** theory is correct). Neither I, nor anyone else, claim to fully understand all of the cortical maps involved in the perception of speech, music and other sounds, so it is easy to explain any feature that has not yet been explained by just saying that there is some cortical map we don't know about, such that the said feature of music causes a constant activity pattern to occur in that cortical map.

For each feature of music that might be explained in this way, the following questions remain to be answered:

- What is the cortical map which responds to that feature?
- How does that cortical map represent meaning?
- How do constant activity patterns in that cortical map relate to the musical feature that we are explaining?

15.2 Outstanding Questions

15.2.1 The Effect of Loudness

The musical quality of music is mostly invariant under amplitude scaling. In fact the brain separates the perceived quality of all sounds into loudness and other qualities independent of loudness (and therefore independent of amplitude).

One basic reason for this separation is that perceived loudness is a function of distance from the source of the sound. It would be wrong for the brain to regard two sounds as being different just because the source of one of the sounds happens to be farther away. (And at the same time it is important to be able to estimate the distance from the source of a sound.)

Loudness is sometimes a semantic aspect of speech, in that a person will speak more loudly because they feel more strongly about what they say, or they want their audience to take more notice of them, or because they want more people to hear them. And some people just have louder voices.

The emotional and pleasurable effect of music is increased if the music is played louder, although there is usually some point of diminishing returns.

Is the effect of loudness on musicality a side-effect of the more obvious aspects of the perception of loudness in relation to the perception of speech, or is there some other reason for it?

One possible answer is that musicality is a rather subtle effect (at least in normal speech), and the perception of musicality is likely to be more accurate if the speech being perceived is being heard loud and clear. Thus the influence of perceived musicality is greater if the sound with musical attributes is louder, because the perception of the musicality is judged to be more reliable.

Another possible explanation is that the musicality of auditory cortical activity patterns is increased as a side-effect of the processing that the brain performs to separate perceived loudness from all other perceived characteristics of sound. For example, if louder sound is more broadly encoded coming into the brain, then the cortical maps within the brain may sharpen their encoding functions to compensate, which might result in sharper boundaries between active and inactive zones in some cases.

15.2.2 Stereo versus Mono

Music sounds better in stereo than in mono. But this effect is not restricted to music—for example, we prefer to listen to television and movies in stereo, or even in surround sound. Stereo sound gives us a more natural representation, than mono sound, of the experience of being there and hearing the original sound, mostly because we have two ears. Given this general preference for stereo over mono, there may be no need for a theory of music to specifically explain the preference for stereo in the case of music.

15.2.3 Rhyme

Rhyme is a ubiquitous component of popular song, which implies that it almost certainly has a direct effect on perceived musicality. Is it because the response of some cortical map to rhymed speech has greater constancy of activity patterns? Or is there some more indirect manner in which rhyme alters our perception of music?

15.2.4 Timbre

Different types of music tend to be played with different types of instruments. There is a converse to this: give musicians and composers new instruments with new timbres, and there is a good chance they will invent new types of music.

The most obvious example of this in modern times is that the development of electric guitars and over-driven valve amplifiers has resulted in the creation of new genres: **hard rock** and **heavy metal**.

We can conclude that there is a definite relationship between timbre and musicality, and that the optimal timbre or set of timbres for playing an item

of music is a function of the specific melody, harmony and rhythm of that music.

The models of pitch-related cortical maps that I have developed in this book have mostly ignored timbre: response to music has been defined entirely in terms of response to pitch values, the only exception being that musical notes and percussive sounds with a common timbre are presumed to be grouped together by the brain when it processes harmonic and temporal relationships. If timbre does affect musicality in other ways (in addition to said grouping effect), then these models must be over-simplifications.

There are two major effects that timbre has on musicality:

- We expect a consistency of timbre within a melody, and within any component of the accompaniment.
- Music works better if appropriate timbres are chosen for the different components of melody, harmony, bass and rhythm accompaniments.

The first thing we can note about the consistency requirement is that the speech of a single human speaker will have a consistent timbre, so it is not surprising that the perception of musicality is a function of consistency of timbre.

There are two possible reasons for the consistency requirement:

- The speech perception system expects consistent timbre from a single speaker speaking, and if timbre is inconsistent then this implies that what you are listening to is *not* the speech of a single speaker.
- The speech perception system *relies* on the consistency of timbre to optimise its processing of speech. That is, it performs calculations in a way that depends on consistency of timbre (and the results would be wrong if in fact timbre was *not* consistent).

The second issue, of why different types of music work better with different timbres, has implications for the operation of the various pitch-valued cortical maps. An over-simplified model of processing is as follows:

- Sound comes into the ears and is initially encoded in terms of separate harmonic components.
- Groups of harmonic components are analysed into separate information about timbre and pitch (i.e. fundamental frequency).
- Only the pitch information is passed on to those cortical maps that process melody (speech or musical).

This model could be called the **full abstraction** model, as it assumes that pitch and timbre values are fully abstracted from the raw sound data

containing all harmonic frequencies, and that only the abstracted pitch values are input into further melody processing areas in the brain (and the abstracted timbre values go somewhere as well, for example to calculate speaker identity and vowel identity). The obvious problem with this model is that it fails to explain how timbre can affect musicality.

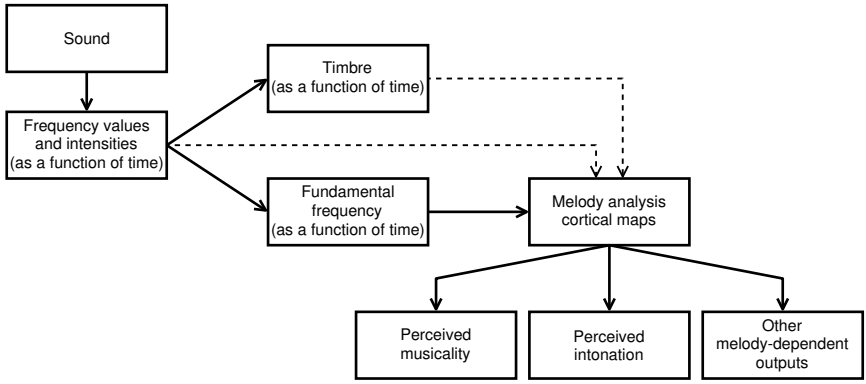

Figure 15.1. Raw frequency values, timbre, pitch and melody analysis. In the flow shown by the solid arrows, incoming description of sound as multiple frequency values is analysed into timbre and pitch (fundamental frequency), and only pitch is passed onto those cortical maps that process melody (speech or musical). The flows shown by dashed arrows represent raw information about individual harmonics and processed information about timbre being included in the inputs to cortical maps that process melody.

An alternative is a **partial abstraction** model, where pitch and timbre values are still calculated, but at the same time raw information about harmonic frequencies is retained and passed on to facilitate calculations whose results are nominally functions of pitch value only.

We have already determined two main types of calculation (relevant to perception of music) that are performed on perceived pitch values:

1. The melodic contour map identifies pitch values that are rising and falling.

2. Different pitch values are compared by subtracting one from the other to determine the interval size between them, and these subtractions are performed on pitch values that have been reduced modulo octaves.

In the full abstraction model, these calculations are performed using just the abstracted pitch values, i.e. the fundamental frequencies of the sounds.[1]

[1]The fundamental frequency may not occur at all in the actual sound, but its value is determined from perception of the corresponding multiples of itself that occur in the sound.

In the partial abstraction model, we must presume that information about the non-fundamental harmonics is still included in the information presented to the pitch-valued cortical maps, and that the maps use this information to help them perform the calculations that they perform, even though we could specify the results of the calculations as if they were done using input values representing pitch values only.

Consider first the melodic contour map. In the full abstraction model, comparison is made between a pitch value at a certain time and its value at a slightly later time, where the pitch value is observed to be higher or lower.

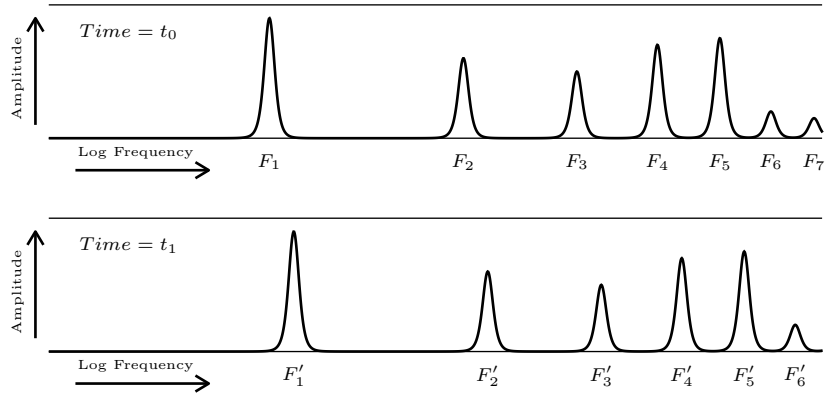

Figure 15.2. Observations of the frequency spectrum of a sound at a time t_0 and at a slightly later time t_1. We can see that the pitch of the sound rises. But there are two ways to measure the rise in pitch: either we estimate each pitch value first, from the observed values of all the harmonics, and then calculate the ratio between the pitch value estimates, or, we calculate the corresponding ratio between each pair of corresponding harmonic frequencies, and then average these ratios to get our estimate of the change in pitch. (The latter option is more consistent with the perception of ever-increasing tones in Shepard scales.)

In the partial abstraction model, comparison can be made simultaneously between all of the harmonics at a certain time, and the corresponding values at a slightly later time, where the pitch value is observed to be higher or lower in each case.

Something similar will happen with subtraction of pitch-values to calculate interval sizes: determination of the interval between the fundamental frequencies of a sound X and sound Y can be performed by comparing all the harmonics of sound X and all the harmonics of sound Y. This will result in a large set of calculated intervals, but the pattern of calculated intervals will be a function of the interval between the fundamental frequencies of the two sounds, so it will be possible for the cortical map to extract this basic interval value from the calculated set of intervals.

15.2.5 Home Chords

The existence and choice of home chords is explained as a result of the need to characterise pitch values based on their relationship to each other, independently of absolute pitch. This also explains why a home chord occurs at the start (or near the start) of a tune, but not why it has to occur at the end. We have speculated that the occurrence of a home chord at the end relates to resetting repetition count, and that the final home chord may satisfy a perceptual criterion for identifying the end of a sentence.

15.3 Further Research

The CAP theory raises questions that could be answered by further research in the "lab".

15.3.1 Brain Studies

A major theme underlying research on the brain (human or animal) is the representation of meaning, even if this isn't always explicitly stated. A number of brain mapping technologies and techniques exist that can be used to determine relationships between the location of brain activity and the supposed occurrence of various mental activities:

- Brain imaging technologies measure brain activity when subjects perform particular tasks. Techniques for measuring brain activity include:
  - Electroencephalography: making recordings from scalp electrodes. These recordings are called electroencephalograms (EEGs).
  - Recording electrical activity from micro-electrodes inserted into individual neurons.
  - Positron Emission Tomography (PET) and Functional Magnetic Resonance Imaging (fMRI), which both measure blood flow. Increased blood flow is an indirect indication of areas in the brain that are more active at any one time.
- Portions of the brain can be artificially activated or deactivated by various means, and the effects of this observed. The size of a portion activated or deactivated can range from an individual neuron to half of the entire cortex on one side of the brain (deactivated by sodium amytal injections—this technique can be used to determine which side of a person's brain processes language).

For ethical reasons some of the more intrusive techniques can only be carried out on animals. Studying animal brains can certainly help scientists understand the general principles of how human brains work. Some animals

also make good models for aspects of human perception; for example, human visual perception is not much different from that of monkeys and apes. But if music is purely a human phenomenon, then animal studies will be of limited use in helping us to understand how the human brain processes music.

Sometimes deliberate destruction of portions of a human brain is required for medical reasons—in an attempt to prevent epileptic seizures, or to remove a tumour. In many other cases a part of the brain is damaged by accident or disease. The relationship between position of damage and lost functionality can help researchers understand what information is processed where.

Another type of research is carried out (mostly) on brains that are no longer alive: this consists of investigating the anatomy of the brain, and finding out what types of neurons exist in different parts of the brain, and how they are all connected to each other.

15.3.2 Musical Brain Studies

Many brain studies have been done by researchers interested in understanding more about music. The design and interpretation of such studies are, however, often strongly influenced by the explicit hypotheses and implicit assumptions made by the researchers doing the studies.

The theory given in this book contains its own set of hypotheses, and it challenges many of the assumptions that are made by those working in the field. The following sections suggest experiments that could be done in relation to different aspects of the theory.

15.3.3 Constant Activity Patterns

The most direct confirmation of the theory would be to observe constant activity patterns in the cortical maps of someone listening to music. Are current imaging technologies precise enough to make this observation? I suspect not. Both PET and fMRI observe blood flow rather than the activity of individual neurons. Electroencephalography measures electrical activity directly, but not in a way that enables the location of the activity within the brain to be determined to any great degree of accuracy.

Could the occurrence of constant activity patterns be observed by some indirect means? I leave this question as an open challenge.

If constant activity patterns were directly observable, then we could observe the overall constancy of activity patterns when a person perceives music. In addition, by playing music that was only musical in some aspects, we should be able to observe constancy of activity patterns in specific musical cortical maps. For example, if a subject was made to listen to random notes played from a musical scale, there should be a constant activity pattern observable in the subject's scale cortical map, and perhaps in the home chord cortical map (since it receives most of its input from the scale cortical map), but probably not in any other cortical map.

15.3.4 Calibration

The theory of calibration of interval perception against harmonic intervals could be tested by causing a person to be exposed only to sounds with "incorrect" harmonic frequencies. (I described the details of this experiment in Section 12.3.)

15.3.5 Symmetries

If a certain perception is invariant under a set of transformations associated with a symmetry, then it should be possible to observe which cortical maps in the listener's brain have activity that is invariant under those transformations. As with attempting to perceive constant activity patterns, the feasibility of this type of experiment depends on the accuracy of the relevant imaging technique used. For each symmetry there are ways to test the effect of the associated transformations on the listener's brain activity:

- **Octave Translation**: If the listener hears a rising sequence of notes C, $F\sharp$ and C (an octave higher than the first C), there will be some cortical maps where the sequence of activity follows the linear pattern of the absolute frequencies. There will be other cortical maps which represent the pitch values modulo octaves, and for these cortical maps the responses to the first C and the second C should be identical to each other and different from the response to $F\sharp$.

- **Pitch Translation**: The simplest test input in this case would consist of pairs of notes played sequentially. The experiment would vary both the absolute pitch values of pairs of notes and the intervals between them. The aim is to look for cortical maps whose activity was only a function of the intervals between the notes.

- **Time Scaling**: Following a similar methodology as for pitch translation invariance, play different rhythms at different tempos, and look for cortical maps whose response is time scaling invariant.

- **Amplitude Scaling**: The test inputs would consist of a set of tunes played at different volume levels. It would also be a good idea to research the general effect of different volume levels on brain activity in response to other types of sound (especially speech), so as to distinguish general effects of amplitude on sound perception from the specific effects that it has on music perception.

- **Time Translation**: If an experiment can be repeatedly carried out on the same subject and gets a repeatable result, then the result is by definition time translation invariant. The second aspect of time translation invariance relates to repetition, and to what extent the state

of a listener's cortical maps is independent of repetition count of a non-freely repeated component of a tune. Sufficiently precise brain imaging could shed light on this issue.

- **Pitch Reflection**: It would be more difficult to directly observe the effects of pitch reflection invariance on brain activity, because it is a symmetry of interactions between neurons representing different pitch values (whereas the other symmetries cause constancy of perception invariant under transformations of input data).

15.3.6 Repetition: Free and Non-Free

One prediction of the theory of repetition is that there will be some part of a listener's cortex that keeps count of non-free repetition, but not of free repetition. The theory also predicts an interaction between the occurrence of home chords and the resetting of any cortical state that represents repetition count.

15.3.7 Cortical Maps

The various hypothetical cortical maps—scale map, harmonic map, home chord map, bass map, regular beat map and note duration map—should all have activity that is an observable function of the relevant musical components and aspects. The main difficulty in observing them directly is likely to be that current imaging technology does not have the required degree of precision. But one could at least try. Experimental design should be driven by assumptions as to what these cortical maps represent and how their activity is driven by different musical inputs.

For example, the scale cortical map should have different responses to each of the following types of melody:

- Speech-like melody, with continuous melodic contours.
- Melodies consisting of notes of constant pitch, but not selected from a fixed set of pitch values.
- Melodies consisting of notes of constant pitch, with pitch values selected from a finite set of values (i.e. from a scale).

15.3.8 Musicality

Perhaps the most important musical cortical map to look for is the one that represents musicality. Musicality may not be mapped as such; it may consist of a set of neurons spread across different cortical maps such that their overall activity represents the current level of musicality. We would expect activity in this set of neurons to be correlated with two features of music being listened to:

- How much the listener likes the music.
- How loudly the music is being played.

A well-known physiological correlate of musicality is the so-called "chills", which corresponds to an emotional effect sufficiently intense as to cause physiological changes such as the occurrence of goosebumps.[2] There may be some way of measuring this, or it may be enough to rely on verbal reports by the subject. Either way it should be possible to determine which parts of the brain are involved in translating the effect of musicality into goosebumps (and other physiological effects). One would expect the neurons representing musicality to be connected with those neurons that represent emotion, and in particular with those neurons that are active when there is an emotional response to the content of speech.

15.3.9 Non-Typical Musical Aspects

There are certain performance features which often accompany music, but which are not usually regarded as representing musicality in themselves. These non-typical musical aspects include rhyme and dance. Particular features of dance that seem to contribute to its emotional effect are the visibility of the rhythm of the dance and the synchronisation of motion of multiple dancers.

Another possible non-typical music feature is the device of stop/start slow motion that is sometimes used in music videos, and which seems to have an emotional effect that matches or enhances the emotional effect of the music. As is the case for dancing, the emotional response is only significant if the effect is applied to the motion of people.

There are many other gimmicks and techniques that are used in music videos, but most of them come into the "Wow look at this!" or "interesting" category, and they do not seem (at least not to me) to create an emotional response like that caused by slow motion.

If a neural correlate of musicality can be determined, then it becomes possible to determine by direct observation whether any particular performance feature is or is not an aspect of musicality. For example, a dance could be performed by one dancer and then by multiple dancers. And comparison could be made between multiple dancers dancing the same dance and multiple dancers dancing different dances (all to the same music). These different dance options could be performed while imaging brain activity in a subject viewing the dances. The results of brain imaging would indicate which particular locations in the brain responded to the dancers' multiplicity and synchronisation, and whether at least one of those locations was one that is active when perceiving the musicality of music.

[2] A technical term for goosebumps is **horripilation**.

15.3.10 Mathematical Models

Studying the activity of the human brain is one way to test theories about music. Another approach is to construct mathematical models. One goal of music science is to discover a formal computable algorithm that calculates musicality (parameterised for variations in musical taste etc.). Creating models for individual aspects of music perception is an initial step in this direction.

The most interesting models to create would be neural network models of the cortical maps hypothesised to perceive the different aspects of music. To test each model in isolation it is necessary to provide it with inputs equivalent to those which are (according to the hypothesis) output from the cortical maps that provide the inputs to the cortical map that we are modelling. For example:

- For the models of the regular beat and note duration cortical maps, the inputs should consist of series of impulses, representing both speech and musical rhythms. Neurons in the network would have activity that is a function of current input, delayed input and delayed output (i.e. recycled back into an input). Neurons in the model of the note duration map would lack input from delayed output. One could also experiment with variations in the strength of the input connection from the delayed output, i.e. full strength simulates response to regular beat, zero strength simulates response to individual note length, and intermediate strength represents something in between.

- The model of the scale map would accept an input of a continuous pitch contour. Neurons in the network would become activated and reactivated by corresponding pitch values, and their activity would decay slowly. The model of the home chord map would accept as input the output from the scale map.

- A model of the harmonic map would accept an input of the melodic pitch contour, and also an input from the output of the bass map. And it would have a reset feature, activated by strong beats.

- Models for musicality neurons could be added to each perceptual cortical map model.

The ultimate goal of this type of modelling is to create a neural network that "knows" good music when it hears it. Such a network would be an implementation of a predictive algorithm as described in Chapter 2.

15.4 Musical Taste

15.4.1 Why Does Musical Taste Vary?

If the perception of musicality represents an attempt to perceive the internal mental state of a speaker, then this perception should be independent of whoever is doing the perceiving. In other words, there should be a tendency for everyone's perception of musicality to be the same as everyone else's perception of musicality.

I have already remarked (in Chapter 2, when discussing universality) that the musical tastes of different people have much in common, but at the same time there are also many differences. Does the amount of observed variation in musical taste contradict the hypothesis that musical taste represents a perception of something external to the perceiver?

15.4.2 Variation in Super-Stimuli

It is important to remember that, according to the hypothesis that the perception of musicality is an aspect of speech perception, music is the *super*-stimulus, not the normal stimulus. The normal stimulus is speech. Two music listeners may have perceptions of musicality that are approximately the same when applied to the perception of speech, but for which the optimal super-stimuli are considerably different.

Remember the herring gull super-stimulus example in Chapter 8: a herring gull chick perceives the quality of its parent's beak as being long and yellow with a red spot, and a super-stimulus for this perception is a long yellow stick with three red stripes on it. Now we could imagine that there might be two herring gull chicks with slightly different perceptions of a parent's beak, such that the first herring gull's optimal super-stimulus was a yellow broomstick with three red stripes, but the second herring gull's super-stimulus was a yellow wooden spoon with two red stripes.

If we paid too much attention to the differences between their respective super-stimuli, we might suppose that these chicks' perceptions of long yellow things with red stripes were in some way intrinsically different, and perhaps served an intrinsically different purpose in each case. But we would know that really they are both trying to recognise the same thing: an adult herring gull beak. The super-stimuli can be somewhat different, even though there is no significant difference when the perception is applied to the normal stimulus.

15.4.3 Variation in Musicality Perception

If musicality perception depends on the occurrence of constant activity patterns in the listener's brain, then there are several possible explanations as to why one listener—let us call her Alice—likes an item of music, and another listener—let us call him Bob—does not like the same item:

- At least one aspect of music that Alice responds to does not occur at all in Bob's perception of the music. (There is a limit as to how much this explanation can be applied, as we have already seen that many of the cortical maps that respond to music are close to being hard-wired and predetermined.)

- Alice and Bob both perceive a particular aspect of the music, and a constant activity pattern occurs for Alice in the cortical map that perceives that aspect, but a constant activity pattern does not occur for Bob in his corresponding cortical map.

- Alice and Bob both perceive a particular aspect of the music, and a constant activity pattern occurs identically for both listeners in the corresponding cortical map, but Alice's constant activity pattern is detected by her relevant CAP-detecting neurons, whereas Bob's constant activity pattern is not detected by his corresponding CAP-detecting neurons.

- Alice and Bob both perceive a particular aspect of the music, and a constant activity pattern occurs identically for both listeners in the corresponding cortical map, and both of them have the same number of CAP-detecting neurons which detect this constant activity pattern, but, there is a discount factor for CAP detection in this aspect, which applies to Bob but not to Alice.

The above list refers to the low-level mechanics of musicality perception. We can also look at differences from a higher-level view:

- Musicality perception is an attempt to perceive internal state of the speaker's brain as echoed in the state of the listener's brain. The listener's brain can only be an approximate model of the speaker's brain—since everyone's brain is different.

- The normal stimulus for musicality is speech. It seems plausible that the ways in which constant activity patterns can be perceived will depend very much on which language is being spoken. It may also vary depending on the individual speakers that the listener listens to on a regular basis. And it is plausible that the CAP-detecting neurons adapt themselves to optimally detect the musicality of the speech that the listener is normally exposed to.

The last point leads us to ask:

- To what extent is musical taste determined by exposure to language?

- And to what extent is musical taste determined by exposure to music?

15.4.4 Dependence on Exposure to Language

It is difficult to determine how much (if at all) musical taste is affected by one's native language. There are a number of reasons why people who speak different languages might listen to different kinds of music, even if language does not directly influence musical taste:

- Different languages correspond (at least historically) to different cultures, and different cultures have different music. Thus the speaker of a Chinese language likes Chinese music because, as a speaker of a Chinese language, they live in China and, in China, one is exposed to Chinese music.

- People strongly prefer to listen to songs sung in their own language. This creates a barrier to exposure to foreign music, because people speaking one language will generally only listen to songs written in other languages if they are first translated into their own language.

There is no hard evidence of any correlation between language and musical taste that cannot be explained by these effects. And in the modern world there is a tendency for populations to become exposed to Western music and adopt it as the popular music form, with perhaps some elements of non-Western music also being assimilated into Western music.[3] The result is that we are all listening to very much the same music, whatever language we happen to speak.

One might also suppose that, even for people speaking the same language, there could be a correlation between accent (or dialect) and musical taste, but there is no evidence of this occurring either.

15.4.5 Dependence on Exposure to Music

Exposure to music mostly appears to determine musical taste according to the criterion that people only like the type of music that they have previously listened to. There is some evidence for what is called a **critical period**: a period of development of musical taste, after which it is not possible to develop new musical tastes. This period seems to occur sometime during adolescence or early adulthood, and is analogous to other critical periods that occur in human development. For example, it is very difficult to become fully fluent in a second language that is learned too late in life, in particular from teenage years onwards (so the critical period for development of musical taste comes after that for learning a language fluently).

[3]This is not to deny that many cultures continue to have their own unique genres of music, but many of these unique genres fall entirely within the confines of Western music. In particular they use the diatonic scale (mostly well-tempered except perhaps for instruments that give the performer a choice), they have regular hierarchical tempo, and they are constructed from the usual combination of melody, chords, bass and percussion.

The best evidence for the existence of a critical period for music is that the musical tastes of many older people reflect the music that they were exposed to when they were young. We might suppose that musical taste is a function of age, but as newer generations grow up they carry on listening to the newer music that they were exposed to (in *their* critical period).

If a critical period exists for some component of information processing in the brain, this implies that the details of how that information is processed become fixed when the critical period has ended. Critical periods are in some sense the opposite of cortical plasticity, because cortical plasticity means an ability to change, whereas a critical period represents an inability to change once the period is over.

Related to critical periods is the phenomenon of **imprinting**, which is a particular form of fixation. The classical example of imprinting is that of the newly-hatched gosling. This bird starts life with a very rough idea of what its mother looks like, and as soon as it sees some actual object that matches this mother "template", it determines that object to in fact be its mother. The criterion for the template appears to be any large moving object. Most of the time this works well, because the first object matching the template that most goslings see is indeed their mother. But with sufficient contrivance from determined scientists, it is possible to get goslings to treat a range of objects as their mother, the most famous example of a fake goose mother being ethologist Konrad Lorenz (who caused goslings to become imprinted on him as their mother in the course of his studies of the phenomenon). The critical period for a gosling's perception of its mother (the "motheriness" of something) is only a few days.

A more human example of imprinting is "falling in love", which can be interpreted as a person being imprinted with a decision about who should be their partner (for the next few years, if not forever). There is an element of "falling in love" in the way that musical tastes are formed, in that a person may at a certain age strongly react to some new music that they have heard, and this same music appears to determine their musical taste for the rest of their life. (The author's experience of this was around the ages of 17 and 18, when I first left home and was exposed to what everyone else was listening to, and I remember in particular "Hotel California" (The Eagles), "Stairway to Heaven" (Led Zeppelin), various David Bowie songs and Neil Young's "Live Rust" album, which probably accounts for my mainstream pop/rock preferences with a slight tendency toward hard rock/heavy metal.)

If there is a critical period, then it must be a critical period *for* something, i.e. for something that matters. The timing of this critical period might even help us understand the importance of this something. There might be some significance in the observation that the critical period for musical taste comes after the critical period for language fluency—first we have to learn to listen to speech, and then having done that we can learn to perceive other secondary information contained within that speech. It may also be that although the

primary components of language fluency must be developed before that age period, there are secondary aspects of language ability, including aspects of intonation, which are developed during that period, and which depend on the development of the same aspects of language in the person's peer group. (So it would be premature to imprint on criteria for musicality of intonation melodies *before* that age period.)

15.4.6 Adaptation and CAP-Detectors

In Section 14.3 we hypothesised the existence of **musicality neurons**, i.e. special neurons for detecting the occurrence of constant activity patterns in cortical maps. These neurons have inhibitory synapses and excitatory synapses representing the two sides of the edge between active and inactive zones in the activity pattern of a cortical map. If there is some type of adaptation that occurs in the perception of musicality (as a result of exposure to speech or music), then it must involve the wiring or re-wiring of these neurons. We might suppose that an undeveloped CAP-detecting neuron contains many inhibitory and excitatory inputs, and that the neuron gradually determines which of those inputs to drop and which to keep, based on the pattern of input signals that it receives over time. A CAP-detecting neuron may even start out with excitatory and inhibitory inputs coming from the same neuron, giving it the option of deciding whether to be excited or inhibited by activity in said neuron—the decision being made by dropping the type of input not desired.

And we might suppose that the following criteria determine which connections in a CAP-detecting neuron are retained:

- Firstly, look for an edge which occurs at least sometimes. An edge consists of a division of input synapses into one half that is active and one half that is not active. If a particular edge never occurs, then there is no point in maintaining a musicality neuron to detect that edge.
- Secondly, look for an edge that is not *always* constant. If an edge between active and inactive zones is always constant—regardless of any possible variations in the speaker's mental state—then detection of that edge does not represent any useful information about anything.

15.4.7 Why Language Makes Little Difference

Given that musicality is supposed to be a perceived property of language, and further given that the development of musical taste appears to depend to some degree on exposure to music before a certain age, we might expect that musical taste should also depend on the type of language that a person is exposed to. Yet this appears not to be the case.

The simplest explanation that I can give in response to this difficulty is to suppose that, because music is a super-stimulus, it completely dominates any

learning process (during the critical period) that determines the development of ability to perceive different aspects of musicality.

Another possibility is that language exposure can have some influence on the development of musical taste, but that the effect is very limited, and that because the development process is a positive one, i.e. you learn to like what you hear, the nature of the language you are exposed to will never *prevent* you from acquiring a taste for the type of music that you are exposed to.

15.5 Intensity/Position Conversion

The scale cortical map and the home chord cortical map both produce outputs encoded in the form of the *intensity* of firing of neurons, where the intensity represents pitch values in a manner that is pitch translation invariant (here "intensity" simply means frequency of firing). For example, considering the home chord cortical map, notes in the home chord cause greater activity in the neurons representing the notes in the home chord than that caused by notes not in the home chord. The degree of activity represents the "home" quality (or "resolvedness") of the note.

There is one basic problem with this representation: given a variable representing a perceived quantity, the brain needs to be able to represent the variable in such a way that different variable values are represented by activity in *different* neurons. So if the value of the variable is encoded by intensity of activity, the intensity-based encoding has to be re-encoded into a positional encoding. (By "positional encoding" I mean encoding by means of different neurons representing different information values, which is positional in the sense that different neurons in a cortical map occupy different positions.)

There are situations where we know that the brain can convert information represented by intensity into information represented by position. A simple example would be our ability to make verbal descriptions of how bright a light is. In the sensory neurons, the only difference between dim and bright is the degree of activation. When we speak the words "dim" and "bright", at some point we must be activating completely different sets of motor neurons. Somewhere in between the point where the sensory cells in our eyes are activated, and the point where neurons in the speech areas of the brain select which words to say, there must have been a conversion from intensity-based encoding to position-based encoding.

So let us consider two cortical maps: an intensity map I and a position map P. A particular neuron in map P fires if and only if the rate of activity in the currently active neurons in map I is at a certain level. The firing of the neuron in map P must be independent of *which* neurons in map I are currently active. This would seem to imply that every neuron in map I must be connected to every neuron in map P.

The brain may or may not support this level of interconnection between neurons in two cortical maps. It may stage the conversion in some manner,

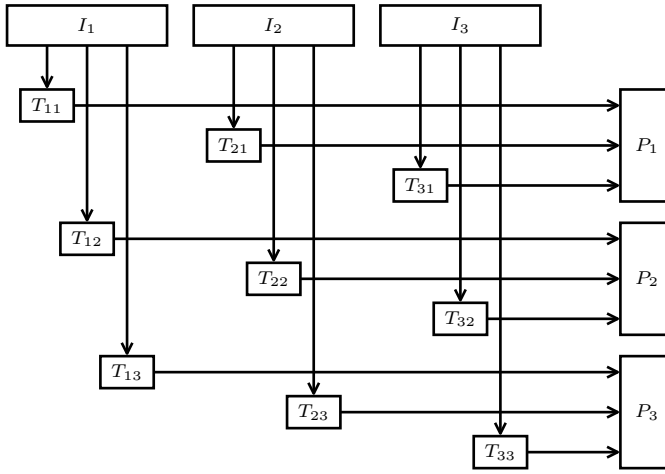

Figure 15.3. A very simple model for an intensity-to-position conversion cortical map. There are 3 intensity-encoded neurons I_1, I_2 and I_3. For each neuron I_i and for each of three intensity levels $j = 1, 2, 3$, there is a threshold neuron T_{ij} which accepts input from I_i and only fires an output if the intensity for neuron I_i exceeds the jth threshold. If any of the neurons I_i exceeds the jth threshold, then the neuron P_j is activated. The pattern of activity in the positional neurons is therefore a function of the set of intensities of activation in the intensity neurons, invariant under any permutation of the intensity neurons.

so that some intermediate cortical map PI has a mixed representation, and every neuron in map I is connected to every neuron in map P indirectly through one or more neurons in map PI.

If intensity to positional encoding conversion plays a role in those aspects of music perception invariant under pitch translation and time scaling, it may also play a role in other types of perception invariant under various transformations, such as visual perception, which as we have already noted, is invariant under scaling, rotation and translation of 2D images (although visual perception is not 100% invariant under rotation: some things look different at different orientations).

15.6 Choruses and Verses

The simplest type of music consists of a single tune or melody. This tune will be repeated freely within a performance of the music, but any repetitions within the tune are strictly non-free. Thus the tune constitutes a sort of indivisible atom of musical data. We can break the tune into smaller pieces—into

phrases, bars or individual notes—but these components lack the musicality of the complete tune. The whole is greater than the sum of its parts.

A slight increase in complexity of music occurs when there is a **verse** and **chorus**. Each of these consists notionally of a separate tune, but the overall effect is greater if the two are combined together within a performance. In the simplest case the combination can be a sequence of verse, chorus, verse, chorus, verse, chorus and so on until it stops (or fades away). In such a case we might say that really the combination of verse plus chorus is the tune, and this combination is repeated freely in the musical performance.

But the somewhat independent existence of the chorus and verse often betrays itself, in that the verse may repeat itself one or more times (i.e. freely) in between occurrences of the chorus. Similarly the chorus may repeat freely. Thus there is no component of the music larger than the chorus and verse which is indivisible in the sense already mentioned.

At the same time, the verse and chorus are not completely self-sufficient. In some sense they go together. We cannot freely mix verses and choruses from different songs and get a musically satisfactory result.

What can we say about the relationship between verse and chorus from a subjective point of view? What do they feel like? The verse is generally the quieter part of the tune, which usually comes first (often twice), and is followed by the chorus, which is louder and more exciting in some way. Sometimes it seems that the verse creates a tension which is resolved by the chorus. There is some similarity between this tension and the tension caused by a dominant 7th chord which is resolved by the occurrence of the home note and chord.

Applying the CAP theory to the concepts of verse and chorus, we can ask some questions:

- What constant activity patterns occur when listening to the verse?
- What constant activity patterns occur when listening to the chorus?
- What is the relationship between constant activity patterns when listening to the verse and constant activity patterns when listening to the chorus? Are there activity patterns in some cortical maps that remain constant as the listener responds to both the verse and the chorus?

We might suppose, for example, that the constant activity patterns in the verse and chorus are similar, but perhaps not exactly the same. I have already hypothesised that the final home chord in a tune may reset the repetition count, so that the tune can then be repeated freely. It may be that the chorus and verse reset each other's repetition count in a similar way.

As the reader may realise, this all rather vague speculation, so I leave the chorus/verse problem as one that requires further investigation.

15.7 The Pleasure of Music

The theory of musicality as a measurement of conscious arousal in the speaker can plausibly explain the *emotional* effect of music. The other effect of music, which also needs to be explained, is the *pleasurable* effect.

The first observation to make is that these two effects cannot really be distinguished. Although the particular emotions evoked by different musical items can vary (in as much as specific emotions can be reliably and consistently identified), in general the *intensity* of the emotional effect of music is tightly correlated with the amount of pleasure that the music causes in the listener.

Another clue to understanding musical pleasure is the phenomenon of "goosebumps" (already briefly mentioned earlier in this chapter), where music that strongly affects us causes the hairs on our skin to stick up. Goosebumps are also a human reaction to cold, fear and emotion. Now music doesn't particular make us cold. Nor does it make us fearful (although music is used in horror movies, where it enhances the fear, but this would appear to be just a special case of using music to influence the movie-watcher's emotions). The goosebumps of music seem to correspond to the goosebumps of emotion.

Is there a general association of emotionally-caused goosebumps with pleasure? Living the easy life in a modern technological society, most of us do not have daily encounters sufficiently emotional or scary to cause goosebumps. If we do get the emotional goosebumps, it is from some contrived entertainment, such as music or film.

Can we develop a theory of musical pleasure based on an analogy with our enjoyment of movies?

The last intensely emotional movie watched by the author was the horror movie "The Ring" (Universal Studios, and actually a remake of a Japanese movie called "Ringu"), which did indeed cause me a few goosebumps. We watch horror movies because we want to be horrified. This seems paradoxical, because if something is horrifying, then almost by definition it's something we want to avoid. The resolution to the paradox lies in the fact that the movie is not something really happening to us. We enjoy the excitement and the fear, while at the same time being relaxed because we know it is not real.

This is not a terribly precise explanation, but, in as much as it works, we can presume that something similar might apply to music. For example, even if the music we are listening to evokes a sad feeling, we enjoy feeling the effects of the sadness, knowing that actually there isn't anything to feel sad about.

Chapter 16

Review of Assumptions

It is likely that reading this book has challenged some of the assumptions that you might have made about music and what music is. My theory not only challenges common sense assumptions, it also challenges many assumptions made by music scientists, music philosophers and music theorists in their efforts to understand what music is.

The theory contains its own assumptions. They do not necessarily all stand or fall together. So in this chapter I list them individually, discuss the consequences of each one, and where relevant list some of the alternatives.

16.1 General Assumptions

16.1.1 Information Processing

Assumption: The human brain is an information processing system. Functional components of an information processing system are also information processing systems. An information processing system can be characterised in terms of inputs, outputs, calculation and storage of information. Processing of information can and should be described in terms of the meaning of information.

Unless one is disposed towards mystical understandings of the human mind (and even perhaps if one is so disposed), there is not much reason to doubt that the human brain is an information processing system. But discussions of the deeper mysteries of human existence sometimes lose sight of this fact.

Music is not an information processing system. Music is the information. Those parts of the human brain that respond to music are the information processing system. So we can ask:

- What is the input? (The obvious answer is music, but we might ask if that is the only input, or if it is the intended input.)
- What is the output? (Emotions?)
- What calculation is being performed?
- What information (if any) is stored by the system?

Some theories of music suppose that the importance of music has to do with its creation. In which case the input/output questions are the other way around:

- What is the input? (Emotions?)
- What is the output? (Music?) What does the output mean?

16.1.2 The Importance of Musicality

Assumption: The perception of musicality is the major result of the perception of music.

For anyone who studies music scientifically with commercial motivations, the importance of musicality is obvious. If you can discover an algorithm that composes music that a large number of people are going to like a lot, then you are going to get rich.

Existing music science does not completely ignore the issue of musicality, but there is a tendency for it to slip below the radar. One reason for this has to do with universality and political correctness. If you ask why some music is better than other music, a common response is that different people have different ideas about what is good, and it wouldn't be fair to pick a criterion of goodness for one person, and the notion of goodness is so variable and culturally determined that it is impossible to measure, so you might as well forget about it.

The difference between what is moderately good and what is really good is also the difference between the rules of music and the mysterious elements of genius and inspiration—which are assumed to be beyond the reach of scientific investigation.

But to deny or ignore the existence of musicality for these reasons is to give up too easily. If music isn't musical enough, then in practice we don't bother to listen to it, and performers don't bother to perform it. Unmusical music isn't really music, so if we don't understand what "musical" is, then we don't understand what music is either.

The notion of musicality implies one-dimensionality: that it is a number which can be defined and measured on a simple linear scale. We know that the emotional effect of music is not entirely one dimensional—some music has a happy feeling, and some has a sad feeling. It is hard to determine exactly how many different emotions have to be enumerated to account for all the different effects that music can have. The concept of musicality does not deny these extra dimensions, but it does relegate them to a secondary role in the effort to understand what music is and what function it serves.

16.1.3 We Need to Explain Perception of Musicality

Assumption: The biological problem of explaining the purpose of music can be reduced to the problem of explaining the purpose of our ability to perceive musicality.

This has a strong negative corollary: that music in itself *does not necessarily serve any biological purpose*. A large portion of the literature of music science, philosophy and theory assumes one or more of the following:

- That listening to music serves some useful purpose
- That performing music serves some useful purpose
- That composing music serves some useful purpose

Some authors discuss purpose without properly analysing it within the framework of theoretical biology, i.e. having more grandchildren, as discussed in Chapter 2. Among those who stick to a Darwinian framework, there is an endless variety of creative ideas about how listening to, performing and composing music can serve one purpose or another.

My theory says that all of these ideas are both wrong and irrelevant. The only reason that composers compose music is so that it can be performed and listened to. The only reason that performers perform music is so that they and their audience can listen to it. The only reason that we listen to music is because we perceive the musicality of music, and this perception of musicality makes us feel good.

We listen to music because we perceive its musicality, but the purpose of our ability to perceive musicality does not have to be the perception of the musicality of music.

16.1.4 Musicality of Speech

Assumption: Musicality is an attribute of the speech of a person speaking. This attribute is perceived by the listener.

If the purpose of perceiving musicality is not to perceive the musicality of music, then it has to be the perception of the musicality of something else.

The only thing that is similar to music other than music itself (and poetry, as discussed earlier) is *speech*.

Assuming that musicality is an attribute of speech does not immediately tell us what the meaning of musicality is, but it does imply that it is something important for the listener to know. And we can go a long way without necessarily knowing what that important something is.

How similar are music and speech? There is a sense in which most music *is* speech. Turn on your radio, and you will hear songs. Songs are music whose major component is an individual speaker speaking to you the listener. The speech (normally called **lyrics**) doesn't always make a lot of sense, and its melodies and rhythms contain unnatural regularities, but it is identifiable as speech.

Even looking at the differences between speech and music, we can see analogies between the relevant aspects:

- Musical rhythm has regularities that are not found in normal speech rhythm, but rhythm is still an important perceived aspect of speech.
- Musical melody has regularities that are not found in normal speech melody, but melody is still an important perceived aspect of speech.
- Music contains instruments other than the human voice, but where those instruments produce pitch values, the timbres of the instruments have characteristics analogous to the human voice, and in particular to human vowel sounds, because they have harmonic components whose frequencies are integral multiples of the fundamental frequency of the sound. The sounds of percussive instruments (which either don't have harmonic components, or they have harmonics that are not integral multiples of the fundamental frequency) may in some cases be considered analogous to consonants.
- The musicality of music causes an emotional response in the listener. The content of speech can also cause an emotional response in the listener. Where music is song, the emotional effect of the musicality interacts with the emotional response to the content of the lyrics.

16.1.5 Music is a Super-Stimulus

Music is a super-stimulus for the perception of musicality.

Musicality is a perceived aspect of speech, and music is highly contrived "speech", which is contrived in such a manner as to maximise its musicality, and which may also lack those features of speech (such as coherent plausible semantic content) which are not relevant to the determination of musicality.

One consequence of this contrivance is the appearance of the observed regularities of rhythm and melody that occur in music but not in normal speech.

16.1.6 Emotions

Assumption: The perception of musicality is an attempt to determine the emotions of the speaker.

The major effects of listening to music are emotion and pleasure. If we already assume that musicality is a property of speech, then it seems plausible that this perceived property tells us something about the emotions of the speaker.

There is, however, an alternative to this assumption ...

16.1.7 Our Emotions, Not the Speaker's

Assumption: The perception of musicality affects our emotional response to the content of the speech. It is a measurement of the speaker's internal mental state, but the emotions that we feel in response to music are *not* an estimate of the speaker's emotions.

This is a rather subtle point. The idea is that something about the speaker's internal mental state *justifies* or *confirms* the emotions that we (as listeners) feel in response to the content of the speech. The most likely candidate for the relevant aspect of the speaker's internal mental state is *conscious arousal*—if the content of speech has some emotional significance for us (and assuming that the speaker is aware of this), then most likely the speaker will be consciously aroused to some degree, even if their emotional response to the content of the speech may be different to ours.

16.1.8 Musicality is Not Emotion-Specific

Assumption: The musicality of music does not determine the specific emotion contained in the listener's emotional response to music.

This follows from the assumption that musicality measures conscious arousal, as conscious arousal is an essentially one-dimensional quantity. If an item of music does seem to specify a particular emotion, the assumption is that something else other than musicality is specifying the particular emotion. For example, it is well known that major chords tend to sound happy and minor chords tend to sound sad. A theory of non-emotion-specific musicality does not have anything to say on the causes of this association—it treats it as a separate problem from the problem of explaining what musicality is. The most that can be said, is that if the characteristics of a tune cause it to express a particular emotion, such as sadness, and the tune has a high level of musicality, then the result will be a strong level of that particular emotion (sadness) felt by the listener.

16.1.9 Musical Cortical Maps

Assumption: Cortical maps that respond to music do not exist for the purpose of perceiving music. They exist for the purpose of perceiving speech.

This assumption is implied by the assumption that musicality is a perceived attribute of speech. Those cortical maps that provide input to the perception of musicality must actually be cortical maps that process and perceive the sounds of speech.

There are various corollaries to this assumption:

- The cortical maps that respond to music are more hard-wired than soft-wired, since they are designed to respond to speech. The importance of speech suggests that many cortical maps would have evolved to serve specific and well-defined purposes in the perception of speech.

- For each cortical map that can be identified as perceiving a particular aspect of speech, the purpose of that cortical map is not to perceive that aspect of music.

- At least some of the cortical maps that respond to music may have a primary function in the perception of speech *other than* the perception of musicality. (However, the **split map** theory weakens this assumption: some of the maps may be evolved copies of other cortical maps. In each case, the original cortical map serves a purpose unrelated to the perception of musicality; the evolved copy remains analogous to the original cortical map, but it has evolved to optimise the perception of musicality.)

The implications of this assumption are most radical when applied to those musical aspects which do not exist in speech, or which are quite different to the analogous aspects of speech.

The most significant of these implications are:

- The cortical map that responds to musical scales has a purpose which is not the perception of musical scales because *there are no musical scales in speech.*

- The cortical map that responds to harmony and chords has a purpose which is not the perception of different notes played simultaneously because *there is only ever one current pitch value in the speech of a single speaker.* (And it seems highly implausible that the human brain has evolved a specific capability for perceiving relationships between the speech sounds of simultaneous speakers.)

- The cortical map that defines and perceives the "home" quality of home notes and home chords does not have the purpose of perceiving home

chords, because there are no chords (and therefore no home chords) in speech. Also, the home chord for a melody normally belongs to the set of preferred home chords for the scale that the melody exists in, and (as already noted) there are no scales in normal speech.

- The cortical map that responds to regular beat patterns has a purpose, and this purpose is not the perception of regular beat patterns because *there are no regular beat patterns in speech.* There are *irregular* beat patterns in speech rhythm, and it is likely that the purpose of the cortical map that responds to regular beat patterns is actually to respond to irregular beat patterns.

- If there is a cortical map that responds to the occurrence of exact **non-free** repetition in music, then the purpose of that map cannot be to respond to exact non-free repetition, since such repetition does not occur in normal speech (with the possible exception of **reduplication**, which consists of non-free repetition of components within a word). The purpose of the cortical map that responds to non-free repetition in music may be related to the occurrence of *approximate* repetitions within speech intonation patterns.

- The cortical map which responds to (or is affected by) rhymes at the ends of lines must have some purpose other than the perception of rhymes at the ends of lines because *normal speech does not rhyme.*

- The cortical map that responds to multiple dancers dancing in synchrony must have some purpose other than the perception of multiple dancers dancing, because *normal speech is only spoken by a single speaker.* (Here we are assuming that the perception of dance is an aspect of musicality perception which has to be included in the theory as an aspect of speech perception.)

16.1.10 Symmetries

There are various symmetries in the perception of music and speech, where a perceived aspect is invariant under some set of transformations applied to the input information.

For each symmetry we can ask the following questions:

- What does the symmetry apply to?

- Does the symmetry serve a **functional** perceptual requirement of invariance, i.e. is the perception required to be invariant under the symmetry, or, is the symmetry an **implementation** symmetry—a consequence of internal mechanisms underlying the perception?

- Is the symmetry a symmetry of both music and speech perception?

Each of the six observed symmetries is defined by a set of transformations:

- Time Translation: Music played later or earlier.
- Time Scaling: Music played faster or slower.
- Pitch Translation: Music transposed into a different key.
- Amplitude Scaling: Music played louder or softer.
- Octave Translation: Adding or subtracting octaves to individual chords and notes within chords (including bass notes). Most musical scales are invariant under translation by an octave.
- Pitch Reflection: Reflecting notes in a scale about a pivot note.

Each symmetry consists of invariance of some aspect of the musical quality of music when the music is transformed by any member of the corresponding set of transformations.

Time translation invariance is seemingly trivial in the sense that the response to an item of music is fairly independent of the time it is played at. A slightly less trivial aspect of time translation invariance applies to our perception of components within a single item, where repeated sub-sequences within a melody are always recognised as such. Thus the invariance applies to both whole melodies and portions of a melody. It applies to both speech and music. It serves the functional requirement that the same speech melody has the same meaning on different occasions, and that meanings can be assigned to portions of speech melody within an overall speech melody.

Time scaling invariance is the ability to recognise the identity of a melody or rhythm played at different tempos. It applies to whole melodies. It applies to both speech and music. It serves the functional requirement that a listener can understand speakers who speak at different speeds.

Pitch translation invariance is the preservation of musical quality when music is shifted up or down by a fixed interval. It applies to whole melodies. It applies to both speech and music. It serves the functional requirement that a listener is able to understand speech spoken by people whose voices have different pitch ranges. It is a very exact musical symmetry in the sense that the musical quality of music is almost unchanged by moderate pitch translations.

Amplitude scaling invariance is the preservation of musical quality when music is heard louder or softer. There is one aspect of music perception that is *not* amplitude scaling invariant: the intensity of the emotional and pleasurable effect of music is often increased if music is played louder, up to some level which is generally a function of how much the listener likes that particular item of music.

Octave translation invariance does not appear to serve any major functional requirement, and it is not particular evident in the perception of speech

(although this may simply be because individual speech melodies do not normally contain components separated from each other by an octave). The most likely explanation for this invariance is that it provides the brain a means to economise the representation and processing of pitch value information, in that *different* pitch values separated by one or more octaves are represented by activity in the *same* neurons within certain cortical maps. This economy is likely to be particularly important for the implementation of "subtraction tables" that calculate the sizes of intervals between pairs of pitch values.

Where pitch values are reduced to pitch values modulo octaves, information about absolute pitch values is still retained, but at a lower level of precision. There is enough overlap between information in the modulo-octaves representation of pitch values and the information in the absolute representation of pitch values to allow unambiguous representation of the full absolute pitch value.

Pitch reflection invariance is a plausible symmetry, but we cannot be as certain of its existence as we can for the other five symmetries. If it does exist, it results from a symmetry of mutual interaction of neurons in certain cortical maps representing pairs of pitch values separated by certain intervals, where the interaction is a function of the interval size. Pitch reflection invariance might explain the fact that the preferred home chords A minor and C major on the diatonic white notes scale are reflections of each other about the point of reflective symmetry of the scale itself (which is the note D).

Symmetries are an important concept in the study of music for several reasons:

- If a symmetry is non-trivial to implement (as seems to be the case for pitch translation invariance and time scaling invariance), then there must be significant neural machinery involved in its implementation, and there must be a very good reason why it needs to be implemented.

- Occurrence of a given symmetry in both speech perception and music perception is strongly suggestive of a significant relationship between the two.

- The indirect manner in which some of the hypothesised musical cortical maps represent information about melody makes more sense if we realise that these cortical maps are designed to produce characterisations of input information that are invariant under the required symmetries.

- The brain may employ similar mechanisms to achieve perceptual invariance for different types of symmetry. This extends beyond music and speech: many aspects of visual perception are invariant under certain transformations, such as image translation, scaling and rotation. Understanding the perceptual invariances involved in the perception of speech and music may help us understand other types of perceptual

invariance—it is likely that the brain reuses the same solution to the same problem.

16.2 Individual Cortical Maps

16.2.1 Scale Map

Assumption: The cortical map that responds to musical scales is a map that responds to recent occurrences of pitch values modulo octaves in a speech melody.

This map has the property that it develops a pattern of activity in which some areas are constantly active and other areas are inactive if, and only if, the melody is formed from notes taken from a scale which is a fixed set of pitch values modulo octaves.

Musical scales are not like any other sound stimulus that a person is normally going to be exposed to. Alternative explanations of how and why the brain perceives musical scales fall into two main groups:

- **Hard-wired:** Musical portions of the brain are presumed to exist for the purpose of listening to music. Cortical maps that respond to notes in scales exist in order to respond to music because music is constructed from scales. The problem with this theory is its obvious circularity: it presumes the existence of music played on musical scales in order to explain why we have cortical maps that respond to scales.

- **Soft-wired:** The alternative is to assume that certain cortical maps organise themselves in response to the scales that music has. This theory implicitly assumes a sufficient degree of cortical plasticity: that the brain does not expect to encounter anything like musical scales, but given that they occur it is sufficiently flexible to organise parts of itself to respond to those scales. This level of cortical plasticity can explain potentially any feature of music. The difficulty remains to explain why music takes on the attributes that it does. For example, cortical plasticity should allow the brain to perceive and process music that is *not* constructed from notes in a scale. However, such music does not occur in practice.

16.2.2 Harmonic Map

Assumption: The cortical map that responds to chords is one that is activated by pitch values that are consonantly related to pitch values already active in the map. Activity in this map tends to be reset (to zero) by a combination of a strong beat in the rhythm and a low pitch value (representing the bass) corresponding to the root note of a new chord. The purpose of this map is to detect the

occurrence of sequential (i.e. not simultaneous) pitch values that are harmonically related.

Much of the literature on harmony assumes that to understand harmony one must understand how the brain responds to simultaneous pitch values. This seems reasonable enough: the very definition of **harmony** is that it consists of simultaneous occurrence of different notes.

But if music perception is really speech perception, then there are no simultaneous pitch values to be perceived. Any cortical map that happens to be able to respond to simultaneous pitch values must have as its primary function a response to pitch values that are *not* simultaneous.

16.2.3 Home Chord Map

Assumption: There is a cortical map that characterises pitch values in a melody according to the relationships between each pitch value and the other pitch values in the same melody. The perception of home notes and home chords is a consequence of the operation of this map. The result of this perception is a characterisation of notes that is pitch translation invariant.

Because there are two strongly preferred home chords for the diatonic scale, it seems reasonable to assume that the preference for these home chords is an intrinsic function of the structure of the scale itself. Since home chords are pitch translation invariant, the processes that determine the preferred home chord for a scale must be pitch translation invariant, and therefore must be a function of the intervals between pitch values in the scale. The most likely candidates are (1) proximity of pitch values (i.e. small intervals) and (2) consonance of intervals.

Assumption: The preference for home chord in a scale is largely a function of consonance relationships between notes on the scale.

If we define the **neighbourhood** of a note to be the pattern of notes close to that note, it is observed that notes with similar neighbourhoods do not have similar levels of "home" quality. For example, comparing C and F in the white notes scale, the neighbourhood of C consisting of G, A, B, C, D and E is an exact transposition of the neighbourhood of F consisting of C, D, E, F, G and A. But the "home" quality of the two notes is quite different, as C is the root note of the possible home chord C major, whereas F does not belong to any possible home chord on the scale.

However, if the notes in the white notes scale are unfolded into the **Harmonic Heptagon**, then there is an obvious relationship between the sets of notes that are most strongly connected to other notes by consonant relationships, and the sets of notes that form the preferred home chords. In particular, the notes B and F have one less connection to other notes, and D is connected to both of these. Assuming a pattern of mutual reinforcement via connections to other notes in the heptagon, this leaves A, C, E and

G as being the most strongly reinforced, and these happen to be the notes contained in the two preferred home chords.

To explain why the home chord has to be *either* A minor or C major, we have to assume that there is some reason why the home chord cannot be A minor 7th, i.e. ACEG. The simplest explanation is that the home chord cortical map disallows activation of notes not consonantly related, i.e. A and G, so that one or the other of those two has to be left out.

16.2.4 Regular Beat Map

Assumption: There is a cortical map that contains a map of neurons activated by the occurrence of regular beats of fixed tempos. These neurons tolerate omissions of individual beats (implying they maintain their own "internal" beat), and they do not respond to beats occurring out of phase. The purpose of this map is to respond to the *irregular* rhythms of speech.

The regular beat map responds to the basic musical feature of regular hierarchical time structure.

16.2.5 Note Duration Map

Assumption: There is a cortical map that contains a map of neurons activated by the lengths of the durations between beats occurring within a rhythm.

16.2.6 Melodic Contour Map

Assumption: There is a cortical map that contains a map of neurons activated by the rising and falling of pitch values in a melody.

16.3 Repetition

Assumption: Repetition is an explicit aspect of music, and this implies that there are specific mechanisms in the brain that play a role in the perception of repetition. There is a distinction between the perception of free and non-free repetition, where a repetition count is maintained only in the case of non-free repetition.

16.4 Assumptions of the Final Theory

16.4.1 General Principle of Music

Assumption: There exists a general principle of musicality.

There are many different aspects of music, i.e. melody, harmony, bass, home notes, home chords, rhythm, tempo and repetition. Different types of music emphasise different aspects to different extents. It is possible to perceive musicality even when some aspects are completely missing. The best example of this is the musicality of purely percussive music, which has temporal aspects but no pitch-related aspects (all rhythm, no melody).

This aspectual nature of music suggests that there may be some general principle of musicality which is perceived separately in each cortical map that responds to a different aspect of music, such that the musicality perceived in different cortical maps is then combined to calculate an overall musicality.

16.4.2 Echoing

Assumption: The listener can detect certain patterns of neural activity in cortical maps in the speaker's brain involved in the generation of speech, echoed in patterns of neural activity in their own corresponding cortical maps when they perceive speech.

The assumption is that there is some correspondence between neural activity when you generate speech and neural activity when you listen to the same speech spoken by someone else.

16.4.3 General Principle and Conscious Arousal

Assumption: The general principle of musicality reflects some aspect of the internal mental state of the speaker.

The echoing assumption allows us to connect the perception of patterns of activity in the listener's own brain to the perception of patterns of activity in the speaker's brain.

Conscious arousal is assumed to be a general aspect of mental state which causes an alteration in the mode of neural activity in large areas of the brain, and which is modulated by means of **non-specific neurotransmission**, where certain neurotransmitters are released non-specifically into the brain medium such that their action is distributed over neurons in a large area. It is assumed that this modulation causes an alteration in certain average characteristics of neural activity, and this characteristic is echoed in the brain of the listener, and the echoed characteristic can be measured by those neurons detecting musicality.

16.4.4 Constant Activity Patterns

Assumption: The specific patterns of neural activity corresponding to perceived musicality are constant zones of activity and constant zones of inactivity, with a maximal border perimeter between the two.

This assumption is consistent with the distinct zones of activity and inactivity that occur in both the scale cortical map and the regular beat cortical map, in response to music, but which do not occur in response to the smooth melodic contours and irregular rhythms of normal speech. These **constant activity patterns** can be detected by edge-detecting neurons which respond to the occurrence of persistent edges between zones of active neurons and zones of inactive neurons in a cortical map.

Chapter 17

The Future of Music

Technology has already had a significant effect on the business of creating and distributing music. But the development of a complete theory of music will have an even more dramatic impact.

In the future, to get new music, you won't have to wait for someone to write it—you'll just push a button on your computer. This will completely change the economics of the music industry—musical composition and song-writing will cease to be an activity anyone can get paid for, and there will be a premium on artists who can learn and play newly composed music on the fly. Or perhaps we won't bother to wait for human performers to keep up with the machine, and we'll just tell our computers to compose the music and then perform it for us as well.

A musical composition algorithm based on a scientific understanding of music could make music into the newest drug—one that cannot be banned.

17.1 Music as a Commercial Enterprise

Music, like many other aspects of human culture, has been transformed by Western capitalism into a giant business enterprise. The informational nature of music means that small numbers of composers and performers can potentially supply the musical needs of hundreds of millions of music-listening consumers. Enormous resources can be put into the production of music. A record company might spend hundreds of thousands of dollars (US) to produce an album for a band or performer that is expected to sell well.

The corollary of this few-to-many relationship is that consumers can expect to get the best possible music, and only the best, as long as their personal taste is not too distinct from everyone else's.

Cynical commentators will say that the capitalistic nature of the modern Western music industry has some negative effects, that record companies become concentrations of marketing power with an inertia of their own that can resist innovation, that the music industry is distorted by the demographics of who spends on albums and who doesn't, and that the industry creates talentless "manufactured" bands and artists according to a marketing formula.

Despite these criticisms, I am prepared to believe that a basic consequence of the economics of the modern Western music industry and its technologies of production and distribution is that the average modern Western music consumer is exposed to music that contains a much more intense musicality than what would have been heard by a person living in a small tribe thousands of years ago. In the prehistoric tribal environment, "production" would have consisted of people singing and playing hand-made instruments in a small village environment, and "distribution" would have consisted of live performances, and perhaps some transference of musical skills and repertoire from teacher to pupil by means of direct teaching.

The specific features of the modern music industry that enable the development and distribution of high quality music include the following:

- Electronic and computer systems for recording, playing, composing and altering musical sounds.
- Consumers with a high level of disposable income, some of which is available to fund a pool of composers and musicians, from which the crème-de-la-crème of composition and performance can be selected.
- Efficient mechanisms of distribution, including radio, printing of sheet music and production of recorded music, which allow almost everyone in society, including those working in the music business, to be exposed to a large body of existing work.
- Enormous potential rewards (for the composers and performers), which encourage the development of ever-improving music for the listening audience. This can sometimes be counterproductive, as too much money and luxury allows successful musicians to either kill themselves from overindulgence, or permanently retire from full-time work. But these effects of excessive wealth are more than offset by the benefits of having the freedom to work full-time (if an artist can make enough money from their music to support themselves), and by the relentless human tendency to create and to achieve (which continues even when one has made enough money to retire on).

The most recent development in this mix is the Internet. The Internet has the potential to put a damper on some aspects of the music industry.

Hundreds of millions of people have ready access to computers connected to a global public network. The invention of efficient lossy compression algorithms—such as **MPEG Audio Layer 3** (more commonly known as **MP3**)—has enabled music fans to quickly and easily (and illegally) transfer near-perfect copies of commercially produced music recordings between themselves, denying performers, composers and producers income for their work.

But it is not all bad news:

- The Internet lowers the bar for distribution and advertising costs. Internet technologies such as email, blogs, RSS, search engines and "social software" (and probably a few more technologies yet to be invented), can streamline word-of-mouth communications so that a new performer becomes known in the shortest possible time to the largest possible audience, with near-zero advertising cost. The problem then remains of how to collect any money after you have become famous. Artists may have to rely to a larger extent on income derived from live performances. Or society as a whole may have to determine a means of paying content producers that does not depend on the ability to enforce copy protection.

- The Internet can help to lower the initial production costs of making music, by making it easy for people creating music to share ideas, content and software with each other. Consumers may be prepared to accept slightly lower production values in return for access to a much larger variety of music. (They may have no choice if their persistence in using file-sharing systems breaks the existing business models.) Performers and producers may need to find ways to better exploit the possibilities for building on each other's work. Some have suggested an open-source movement for music, similar to existing open-source software.

17.1.1 Composition Technology

The profits of the modern Western music industry have funded (and encouraged) the development and use of music-related technology. Technology contributes to recording, editing, and distribution; it allows the development of new musical instruments, such as the electric guitar and the synthesizer; and it enables computerised performance, where the musician is replaced by a programmer typing performance instructions into computer software.

But there is one major component of the music production process where technology still plays a very subsidiary role to the efforts of the human musician: composition. There does exist software that can help the musician to compose music, but such software can no more compose good music by itself than a word-processor can write a best-selling novel.

As already discussed in some detail in this book, the non-existence of any effective technology to automate the composition of music is a direct consequence of our lack of understanding of what music is. The process of musical composition remains ad hoc and intuitive. This connection is summed up in Chapter 2 in the "luxury yacht test": if you knew what music was you would be able to develop an algorithm for musical composition, and from sales of the music composed by the algorithm you would have enough money to buy a luxury yacht.

17.1.2 Profiting from a Complete Theory

The time may come when someone realises how much money could be made from a complete scientific theory of music, and commits a significant investment to the development of such a theory.

A simple business plan is as follows:

- Decide to solve music problem.
- Hire suitable employees and swear them all to secrecy.
- Acquire database of music and perform research thereon.
- Research response to music on human and animal subjects.
- Solve problem sufficiently to algorithmically generate music.
- Use algorithm to compose new strong music.
- Produce and sell the music.
- Rake in profits.

An essential component of this plan is secrecy. Once everyone knows what the composition algorithm is, or enough is known about music to easily derive a composition algorithm, anyone will be able to compose music algorithmically, and the commercial value of individual musical compositions will be reduced to zero.

Given the requirement for secrecy, it is entirely possible that *someone is already carrying out this plan.* It is even possible that multiple parties are carrying out the same plan, each in ignorance of the others.

But unless the release of algorithmically composed music is very carefully controlled, there will be some obvious signs and symptoms of music composed algorithmically:

- Some of the music will be radically distinct from existing genres in ways identifiable even to musically naïve listeners.

- The quality of the music will be noticeably better than music composed the traditional way. This will lead to almost total domination of the commercial music industry by a small number of composers and songwriters.

It is slightly possible that neither of these signs will become apparent:

- It may be that we have already discovered all possible musical genres, so it is not possible to create any new genres.

- It may be that there is some sort of "ceiling" of musical strength, and ad hoc musical composition has already reached this ceiling. Algorithmic composition might increase the number of very strong songs, but the strength of the songs may be no greater than anything in the existing body of music.

17.2 A Post-Music-Theory World

If one person or one group of people can discover the secret of music, then so can others. So even if the first discoverers keep it a secret for commercial reasons, the secret will eventually get out. There are the temptations of fame and credit—once one has made one's first few millions, one might want to claim the credit for a major scientific discovery before someone else makes the same claim.

The standard legal answer to the problem of commercially exploiting a new discovery which cannot be kept secret is to apply for a patent.[1] The disadvantage of a patent is that it requires disclosure. If the applicant is lucky, there is a gap between initial application (after which commercial exploitation is permitted), and granting of the patent (when disclosure is compulsory). Depending on which country you are in (or more precisely, depending on which country or countries you wish to apply for a patent in), you may or may not be permitted to commercially exploit an invention *before* your application for a patent on the invention. But if a dishonest inventor of a musical composition algorithm secretly used their algorithm to compose music, published the music, received royalties, and then they made a patent application (claiming that they had not yet used the invention commercially), it would be difficult to prove that they had indeed used their invention prior to the application (and therefore were not entitled to receive a patent).

Different countries also vary in whether or not they allow patents on algorithms.

[1] Legal Disclaimer: I am not a lawyer. The contents of this book do not constitute legal advice; they are simply my own understanding of the legal situation. If you do solve the music problem, and you are unclear about legal issues, my advice is to get proper legal advice from a lawyer.

Once a patented composition algorithm has been disclosed, preventing infringement may not be easy. For example, it may be possible for a competent software developer to implement the invention with a few hundred lines of code in a high-level programming language. The file-sharing saga has shown that many people will do whatever it takes to get direct and free access to music that they want. The instant pleasure of music can overcome anxieties about illegality, especially if the risk of being caught is low enough.

Whether a musical composition algorithm gets patented or not, the discovery of such an algorithm will have a major impact on the economics of the music industry.

Here is a rough sequence of events that currently happen when someone composes new good quality music:

- The composer composes the music.
- The composer, or a separate lyricist, writes some lyrics, because singing is the preferred form of music for most listeners.
- The song is taken to a publisher.
- The publisher accepts the song, and looks for a performer to perform the song.
- The performer (perhaps already signed up to a record company) signs up to perform the song and make a recording.
- The performer performs and records the song in a recording studio.
- The recorded performance gets mixed by a mixer.
- The mixed recorded performance gets mastered.
- The record company decides to sell the recording.
- Someone makes a video of the performer pretending to sing the same song live against a soundtrack of the mastered recording.
- The record company's promoters promote the song to radio stations and TV music channels.
- The performer undertakes a world tour, playing the new song and any others they happen to have on their new album.
- Consumers hear the song on the radio, see the video and perhaps go to a concert.
- Eventually the song appears in other forms: bands play it in pubs, other well-known performers do cover versions, sheet music becomes available, and lots of people sing it in karaoke bars.

- Consumers have listened to the song so many times that most are moderately bored by it. The song is included in various low cost compilation albums. The original CD appears on sale tables in CD shops.

- Consumers crave their next "fix" of new music.

There are variations on some of these steps, like "the composer also performs the song themselves", or "few consumers buy the song until a second performer performs a cover version".

In general the duration from initial composition to the stage of being played excessively on the radio is at least a few months.

Now consider the sequence of events in a post-music-theory world:

- The consumer installs some musical composition software on their computer. It is assumed that they have already installed a good sound-card, good quality speakers and a high quality set of **sound fonts** (i.e. digital versions of instrument sounds).

- The consumer fiddles with input parameters on the composition software, and uses the mouse to click the "Compose" button.

- The software creates a new composition and then plays it through the computer's sound system.

- If the consumer likes the composition, they save the details, and play it some more.

- If the consumer particularly likes the results of the current composition, they may post it to their web-log, or email it to their friends.

- The consumer gets bored, fiddles a bit more with input parameters for composition, and clicks the "Compose" button again ...

In as much as "consumers" are people who buy things that others have produced, our "consumer" was only really "consuming" when they set up their computer hardware and installed the composition software (and the software might have been free anyway). After that initial step it doesn't make so much sense to refer to them as a "consumer", since they are now doing all of their own production.

How long will it take to go from pressing the "Compose" button to hearing the song? If the composition algorithm is computationally intensive, then there might be some delay. In the worst case music lovers might be forced to leave their computers running overnight and see what comes up in the morning. But it is quite possible that gratification will be totally immediate. And the software will be configurable to automate all the above steps: the computer composes a new tune, plays it several times, composes another tune, plays it several times, then repeats a few tunes composed on some

earlier occasion, and so on, with configuration options chosen to match the listener's preferred rate of exposure to new and old music.

There are some things missing from this do-it-yourself (on your own computer) approach to musical composition:

- The quality of performance, mixing and mastering will only be as good as what is programmed into the software, or what the user is able to do for themselves by adjusting options in the software.
- The composition algorithm will probably not generate good lyrics: lyrics are natural language, and generating interesting natural language is a whole separate difficult problem in itself.
- Even if lyrics can be written, the software might not be able to sing very well.
- And even if the software can sing as well as a person, we actually like to know that a person is singing a song to us.

Writing lyrics is a non-trivial skill: if you don't believe this, pick a well-known tune and try writing your own lyrics to it, and see if they sound as good as the original lyrics.

Difficulties with lyrics and singing will create new markets in the music industry, specifically for:

- Lyricists who can write good lyrics quickly.
- Singers who can quickly learn to sing new songs. It may be hard for some singers to do this. (But singers in the new post-music-theory economy will at least be spared the unbearable boredom of having to sing the same hit song over and over and over again at all their concerts.)
- Ditto with instrumentalists, given that there will still be a demand for live performances of music.

17.2.1 Music Junkies?

A cornucopia of music from a composition algorithm based on a complete scientific understanding of music perception may not be an entirely good thing. It may, as the saying goes, be too much of a good thing.

Technology constantly threatens us with new and dangerous addictions: fast cars, television, designer drugs, video games and Internet pornography. Algorithmically generated music may be the next addition to this list. Come home from work, turn on the computer, bring up the software, compose and play some new music. Or just download the latest hot compositions that have been posted on the Internet.

As is the case for many other forms of entertainment, music "addiction" may be self-limiting for most people. If there is some 5 percent of the population (the "music junkies") whose life is ruined and corrupted by this new pleasure, then so be it, and the rest of us enjoy it in moderation, and life goes on.

We can only hope that the power of rationally composed music over us will not be too great. There might be a campaign to get musical composition algorithms classified as a "Class A" drug, but it seems unlikely that those in power could successfully enforce a law against citizens composing music in the privacy of their own computers.

17.2.2 The Future

Futurology is a difficult enterprise. The things that make the future most interesting are the ideas that become known in the future that were not known in the past. By definition these ideas are not known at the time the prediction is made.

Sometimes even a small technological change has profound consequences for everything, and the full extent of these consequences is not immediately obvious. The intrinsic conservativeness of our thinking makes us reluctant to throw away assumptions about how the world is and how it should be, even if we have observed a change and we know logically that it breaks many of those same assumptions.

Faced with these difficulties, and not wanting to appear too much of a fool to future generations, I will risk just one more prediction about the future of music: the next step in improving our understanding of music may be taken by a reader of this book.

Bibliography

Books

[1] Bear, Connors and Paradiso *Neuroscience: Exploring the Brain*
Williams & Wilkins 1996

[2] Francis Crick *The Astonishing Hypothesis*
Simon & Schuster 1994

[3] Charles Darwin *The Origin of Species* 1859

[4] Charles Darwin *The Descent of Man* 1871

[5] Diana Deutsch *The Psychology of Music, 2nd Edition*
Academic Press, 1999

[6] Donald A. Hodges *Handbook of Music Psychology, 2nd Edition*
Institute for Music Research, 1996

[7] Robert Jourdain *Music, the Brain and Ecstasy: How Music Captures Our Imagination*
William Morrow, 1997

[8] Patrik N. Juslin and John A. Sloboda *Music and Emotion: Theory and Research*
Oxford University Press, 2001

[9] Fred Lerdahl and Ray Jackendoff *A Generative Theory of Tonal Music*
MIT Press, 1983

[10] Paul Lorrain and Dale Corson *Electromagnetic Fields and Waves*
WH Freeman and Co, 1970

[11] Leonard B. Meyer *Emotion and Meaning in Music*
University of Chicago Press, 1956

[12] Isabelle Peretz and Robert Zatorre *The Cognitive Neuroscience of Music*
Oxford University Press, 2003

[13] Vilayanur Ramachandran *The Emerging Mind*
Profile Books, 2003

[14] John S. Rigden *Hydrogen: The Essential Element*
Harvard University Press, 2002

[15] Oliver Sacks *An Anthropologist On Mars: Seven Paradoxical Tales*
Random House Inc, 1995

[16] Anthony Storr *Music and the Mind*
Free Press, 1992

[17] Nils L.Wallin, Björn Merker and Steven Brown *The Origins of Music*
MIT Press, 2000

[18] Semir Zeki *A Vision of the Brain*
Blackwell Science, 1993

Papers

[19] Frank Biocca and J.P. Rolland "Virtual Eyes Can Rearrange Your Body: Adaptation to Visual Displacement in See-Through, Head-Mounted Displays"
Presence: Teleoperators & Virtual Environments 7(3) 1998

[20] Diana Deutsch "A Musical Paradox"
Music Perception 3(3) 1986

[21] Diana Deutsch "The Tritone Paradox: An Influence of Language on Music Perception"
Music Perception 8 1991

[22] Diana Deutsch "Mothers and Their Children Hear a Musical Illusion in Strikingly Similar Ways"
Journal of the Acoustical Society of America 99(4) 1996

[23] Roger Shepard "Circularity in Judgments of Relative Pitch"
Journal of Acoustical Society of America 36(12) 1964

[24] Jacob Ziv and Abraham Lempel "A Universal Algorithm for Sequential Data Compression"
IEEE Transactions on Information Theory 23(3) 1977

Web Sites

[25] *Consciousness* http://www.1729.com/consciousness/ (section of the author's website on the subject of consciousness)

[26] *Hit Song Science* http://www.hitsongscience.com/

[27] *M.C. Escher: The Official Website* http://www.mcescher.com/

[28] *PRAAT* http://www.praat.org/

[29] *2003 Reith Lecture* http://www.bbc.co.uk/radio4/reith2003/

[30] *University of Tübingen press release*
http://www.uni-tuebingen.de/uni/qvo/pm/pm2004/pm824.html
(Geissenklösterle ivory flute)

[31] *What is Music?* http://whatismusic.info/ (the official website for this book)

Index