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.