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.

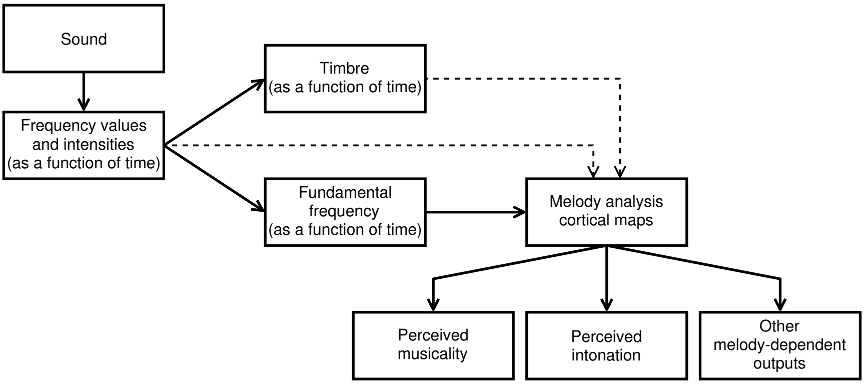

Figure 15.1. Raw frequency values, timbre, pitch and melody analysis. In the flow shown by the solid arrows, incoming description of sound as multiple frequency values is analysed into timbre and pitch (fundamental frequency), and only pitch is passed onto those cortical maps that process melody (speech 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.

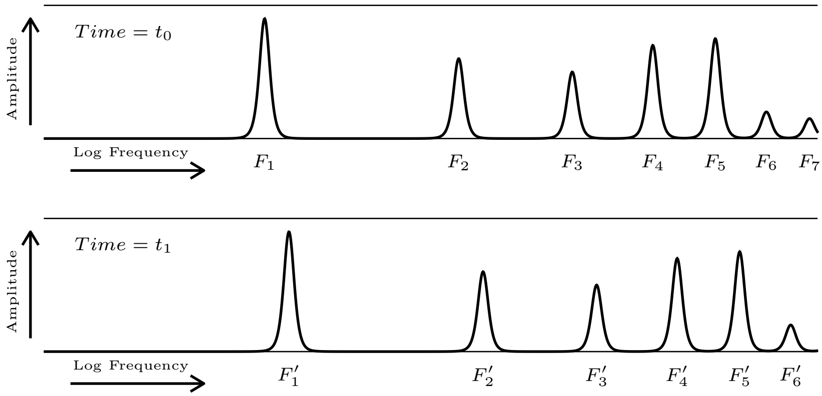

Figure 15.2. Observations of the frequency spectrum of a sound at a time t_0 and at a slightly later time t_1. We can see that the pitch of the sound rises. But there are two ways to measure the rise in pitch: either we estimate each pitch value first, from the observed values of all the harmonics, and 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,

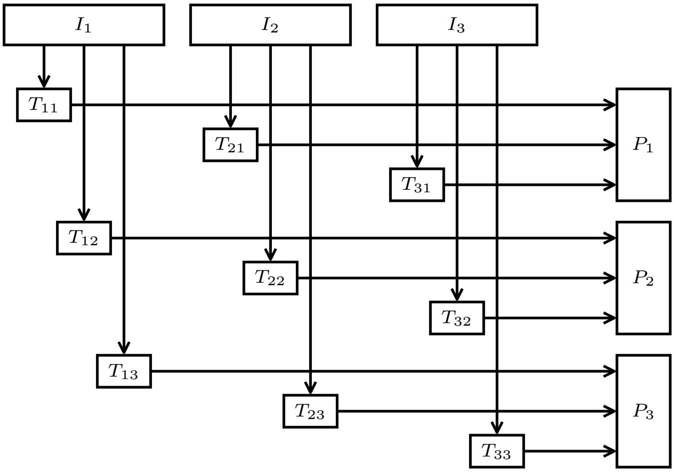

Figure 15.3. A very simple model for an intensity-to-position conversion cortical map. There are 3 intensity-encoded neurons I_1, I_2 and I_3. For each neuron I_i and for each of three intensity levels $j = 1, 2, 3$, there is a threshold neuron T_{ij} which accepts input from I_i and only fires an output 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.