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.

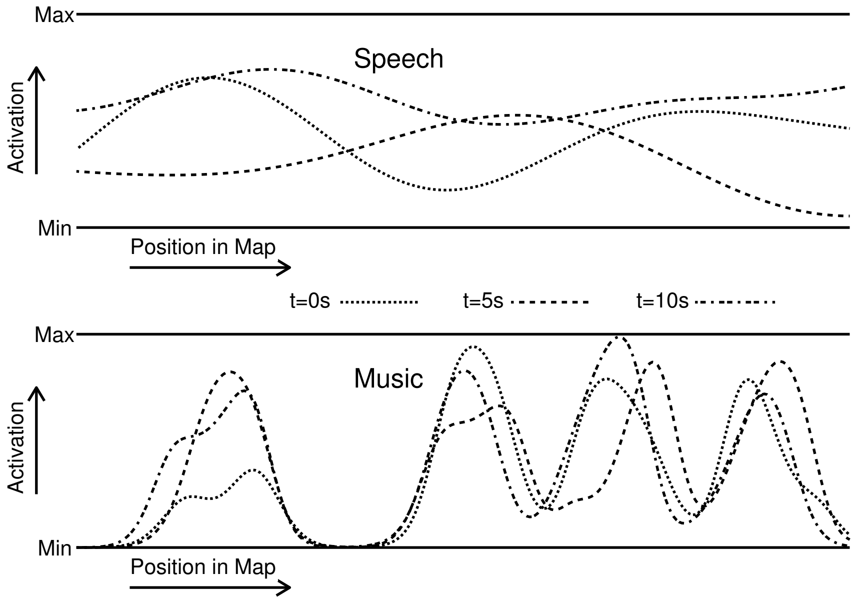

Figure 14.1. Activity levels across one dimension of a hypothetical cortical map at three moments in time (t = 0, 5 and 10 seconds), while perceiving a spoken sentence, and while perceiving a musical item. Neural activity when perceiving the speech is more spread out, and the activity is not restricted 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

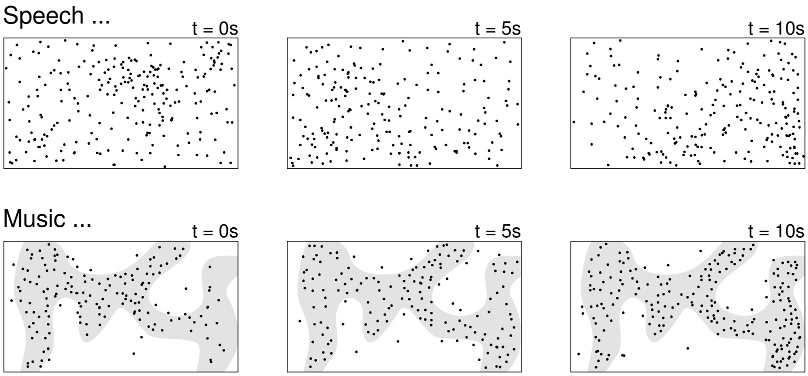

Figure 14.2. 2-dimensional views of the response of a hypothetical cortical map to a spoken sentence (top row) and a musical item (bottom row), at three moments in time, each 5 seconds apart. Each view of the map shows active neurons as black dots. When speech is being processed, the neural activity 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.

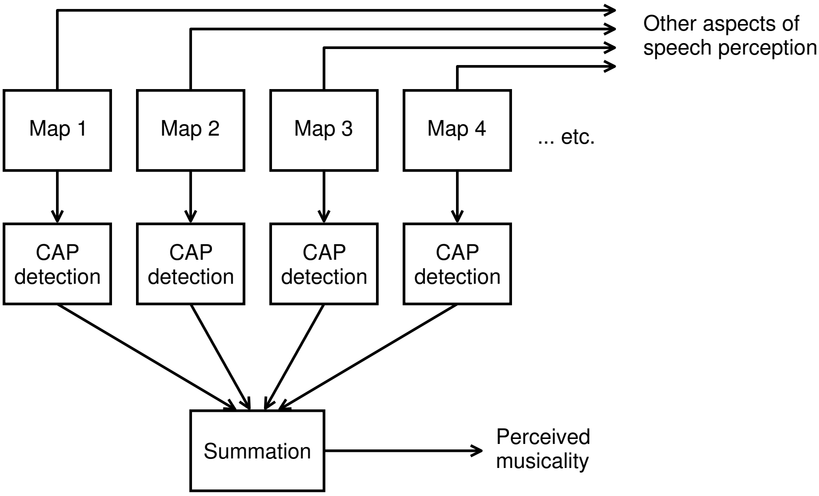

Figure 14.3. Detection of constant activity patterns (CAP) in multiple cortical maps. Maps 1, 2, 3, 4 etc. produce information relevant to speech perception. In addition, CAP-detecting neurons in all these maps detect CAP for each map, and then combine the individual CAP levels to generate an overall 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

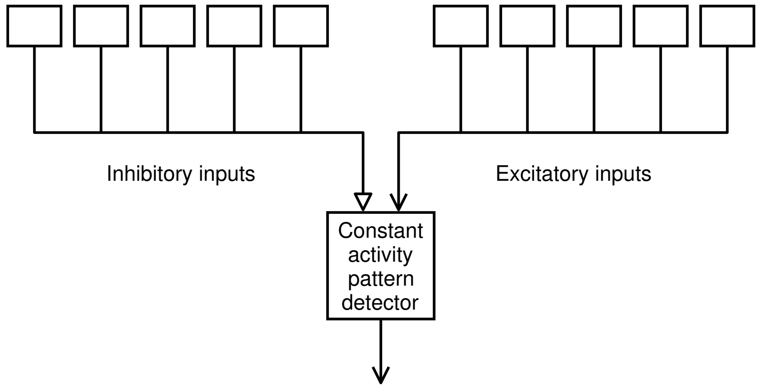

Figure 14.4. A CAP-detecting neuron that detects an "edge" between inactive neurons in the group on the left and active neurons in the group on the right. Inhibition from the left group is long-lasting in the sense that any input from the left causes inhibition of the CAP-detecting neuron for some period 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.

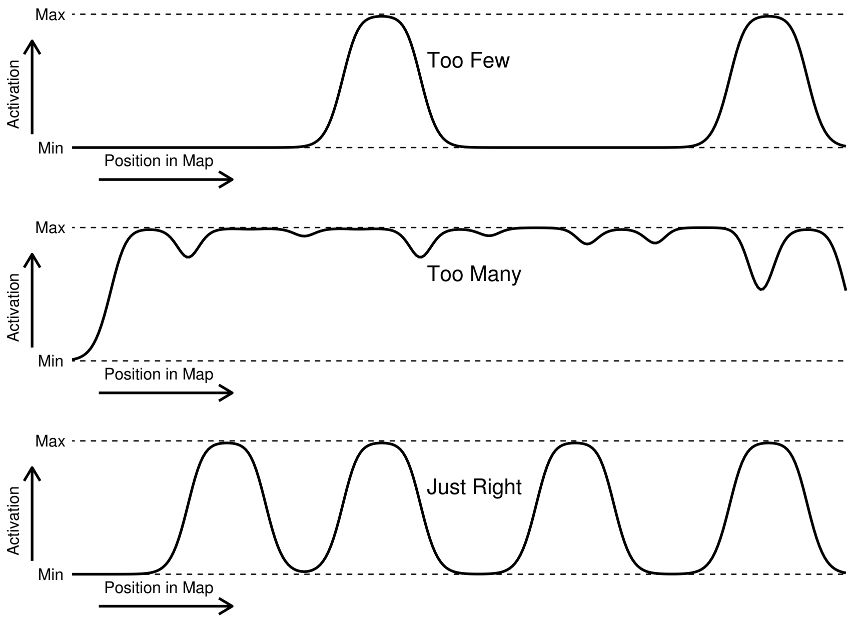

Figure 14.5. Too few, too many, just right. Activity patterns across a 1-dimensional section of a speech-perception-related cortical map, in response to three different tunes. The first creates 2 clear peaks which results in 4 detectable edges. This does not maximise the measured musicality of the activity 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.

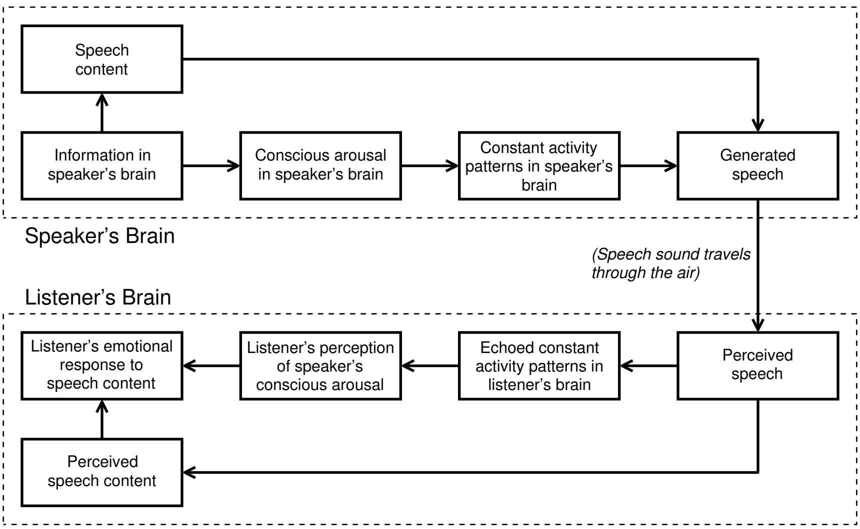

Figure 14.6. Information flow in and between the brains of the speaker and the listener. What the speaker knows determines both their level of conscious arousal (which in turn determines their CAP level) and the content of their speech. When the listener listens to the speech, they extract both speech 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 different 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.