Chapter 10

Musical Cortical Maps

Having found that the nature of music is determined by the response characteristics of cortical maps involved in speech perception, we can make intelligent guesses as to what some of these cortical maps are, based on observations about the structure of music.

Hypothetical maps "discovered" this way include the **regular beat cortical map**, the **harmonic (chord) cortical map**, the **bass cortical map**, the **scale cortical map**, the **home chord cortical map**, the **note duration cortical map** and the **melodic contour cortical map**.

We also discover a similarity between patterns of neural activity in pitch-related cortical maps and those in time-related cortical maps. This similarity is a hint of something deeper going on: a hint as to what musicality actually represents.

10.1 Cortical Plasticity

Cortical plasticity is a term that refers to the ability of areas in the brain to take on different functions. Cortical plasticity is related to a concept of **competitive recruitment**, where the processing functionality required "recruits" an area of neurons to perform that function. There is an ongoing battle between competing functionalities to recruit the most neurons, and the competition is decided by some measure of how important the different functionalities are.

Some well-known examples of cortical plasticity are:

- If a part of the body is lost, for example an arm, the cortical neurons that responded to sensation in that part of the body will be recruited by sensations in other parts of the body, for example a part of the face. The result is that when you touch a certain part of the person's face, they may report that you are touching their arm.[1]

- If a particular sensory input is lost or suppressed, then neurons that processed that sensory input may be recruited by other similar (but unsuppressed) inputs. This type of plasticity can depend strongly on age.

  For example, there are areas in the brain that process inputs from both eyes. A common problem is that some children have a "crooked" eye which fails to align with their direction of sight, and as a result that eye (the **weak eye**) fails to provide useful information to the visual processing areas in the brain, and neurons in the visual areas preferentially develop connections to the other eye (the **strong eye**).[2]

  A crooked eye can often be fixed by appropriate surgery, and the result is that the newly straightened eye is capable of re-forming connections to the visual processing areas, as long as the eye is fixed before the end of a **critical period**. After the critical period (which starts at about age 5 and ends about age 10), the cortical plasticity of these areas is lost, and it is no longer possible for the previously weak eye to recruit the neural connections required to make use of the useful information now coming in from it. (Recovery can be assisted by patching the strong eye for a period of time, which forces the child to make use of their weak eye and the information coming from it. This prevents the vicious circle where they only look at things with their strong eye because they can only "see" with that eye, and they only see with that eye because they only look at things with that eye.)

  It is also observed that temporary suppression of visual input from one eye *after* the critical period does not result in loss of neural connections from that eye.

There is a strong economic flavour to the concept of plasticity. We might imagine a town where all the bakers were killed by some disaster. The demand for bread would motivate some other people, perhaps the cake makers, to move into the business of bread making. And other food manufacturers might take up some of the resulting slack in the cake cooking business, and so on.

Similarly, if for some reason people stopped eating bread, then the bakers would have to consider a change of career, perhaps into the cake business, which in turn would put some of the cake makers out of business (due to increased competition), and so on.

---

[1] *The Emerging Mind* Vilayanur Ramachandran, page 14.

[2] The medical term for crooked eye is **strabismus**, and the loss of functionality in the weak eye is called **amblyopia**.

If the concept of plasticity was taken to a logical extreme, it would imply that any area of the brain could perform any function. But it is observed in practice that certain functions are always performed in certain areas. Processing of visual information always takes place in areas at the rear of the brain (in the **occipital cortex**). Processing of sound information always takes place in certain areas on the sides of the brain (within the **temporal lobes**). So cortical plasticity does not represent complete freedom to relocate functionality anywhere, but it does represent freedom to relocate functionality to some extent. This constrained relocation can be incorporated into the economic analogy: in principle bakers can bake bread anywhere, but in practice their preferred location is a function of the location of their supplies, like flour and cooking fuel, and the location of the customers who come in to buy their bread, and the locations of buildings that happen to have built-in ovens and gas connections.

We can consider cortical plasticity as a means of allocating resources to a fixed set of information processing functionalities, where the brain is in some way prepared to develop the ability to perform those information processing tasks. But we can also consider it a means of explaining how a brain can allocate resources to information processing tasks that did not exist in the previous evolutionary history of the owner of that brain.

A good example is the set of cortical maps that support the human ability to read. Most people are able to learn to read without too much difficulty, even if none of their ancestors have ever had the opportunity or the need to do any reading. And it seems reasonable to suppose that reading will end up having its own cortical maps devoted to the specific information processing tasks that make up reading, i.e. deciding where to direct the eyes, recognising shapes of letters, recognising sequences of letters as words, translating letters and words into sounds (maybe), and passing this information through to the those parts of the brain that process speech and language. (Although it is likely that some aspects of reading will be implemented in cortical maps that also implement similar aspects of other tasks.)

We invoke cortical plasticity to explain how the relevant areas of a person's brain are recruited in order to perform the information processing tasks related to reading. And we assume that if the person had not learnt to read, then those areas would have been devoted to other information processing tasks.

This relates to the issue of "hard-wiring" and "soft-wiring". Considering neural circuits as our "wires", the question is, for any given circuit, how predetermined is the nature of the information that is represented by activity in that circuit? Just how much plasticity is there? We have already considered the issue of representation of meaning, and now we are asking about how the representation of meaning *develops*, and how flexibly it can change.

The degree of plasticity is going to vary between different parts of the brain and nervous system, and between different functionalities. As we have already seen, the meaning of activity in a motor neuron connected to a muscle fibre is "contract that muscle fibre", and there isn't really any way that its meaning can change. Similarly for sensory neurons: the meaning of activity in a heat-detecting neuron in the skin at a certain position is "this position in the skin is hot", and there is no way that its meaning can change. In contrast, somewhere in the mysterious inner workings of the human brain there are regions that give us the ability to learn new ways of thinking and understanding, and the ability to develop skills that relate to circumstances that may be considerably different from anything in our past evolutionary history, such as reading. Some other activities that seem to involve information processing somewhat different to anything that our hunter-gatherer ancestors would have done include:

- Playing chess.
- Playing the piano.
- Driving a car.
- Doing mathematics.

The inner regions of the brain must be sufficiently plastic to be able to provide these new types of functionality. Given that our understanding of the organisation and operation of these regions is very limited, it is difficult to know with any certainty what sort of plasticity occurs, and how plastic the corresponding brain areas are. But we might expect that there is a continuum of plasticity, ranging from the hard-wired sensory and motor neurons, to the more flexibly soft-wired inner regions.

If one brain area A receives most of its inputs from another area B whose neurons are mostly hard-wired, the variation in meaning of neurons in area A will be limited by the nature of the connections from region B (where the neurons have relatively fixed meanings).

A similar limitation will occur if a brain area C sends most of its outputs to another area D whose neurons are mostly hard-wired. The general pattern suggested by this reasoning is that those functional maps closest to the external world, (i.e. sensory and motor maps) are the most hard-wired, and the maps connected to those maps are somewhat less hard-wired, and the next layer of maps can be even less hard-wired, and so on.

However, even when there are multiple processing layers, the inner layers may still be quite hard-wired, especially if their function is largely predetermined by specific requirements that have evolved under natural selection over a long period. These areas will have neurons specialised to process the particular types of information they are intended to process. For example, the areas that process specific types of visual information like colour, motion and depth will have evolved to perform the processing of those types of information.

Even an area of the brain that performs a function like recognising faces is likely to have its function predetermined, because of the importance of this task. But the mapping of neurons to individual faces must necessarily be soft-wired, because the set of faces that any individual has to recognise is going to be different for each individual.

10.1.1 Plasticity and Theories of Music

So what does all this talk of cortical plasticity lead to?

Having distinguished different degrees of plasticity, we can ask a basic question about the cortical maps that respond to music:

> How plastic are those maps?

Because music is so strange and unlike anything else, it is easy to fall into the assumption that the cortical maps that respond to it are developed from scratch in response to the patterns that occur in music. These cortical maps are assumed to develop in parts of the brain that have a high degree of plasticity.

This describes my thinking when I developed the 2D/3D theory of music. I assumed that the patterns of music determined the development of cortical maps that responded to those patterns. It seemed unlikely that there were pre-existing cortical maps to process things like scales and chords and hierarchical systems of rhythm and tempo.

And it was not implausible to me that if musical intervals had natural 3D representations and 2D representations related by a linear 3D to 2D projection, then the requirements of music perception could recruit neurons in the brain already designed to process that type of projection, regardless of the fact that those neurons were "designed" (by evolution) for visual processing and not auditory processing.

Since (under this assumption) the cortical maps that processed music were not pre-existing maps designed for that purpose, I supposed that musicality was some very generic property of music that translated somehow into a generic property of the response of those musical cortical maps to the music. The 2D/3D theory seemed to supply a plausible candidate in the form of the "80 = 81" paradox (corresponding to the syntonic comma). Somehow, I supposed, the paradox of the syntonic comma gave rise to pleasure and emotionality. I will have to admit, however, that I was never able to plausibly provide any details of that "somehow".

As it happens, the super-stimulus theory of music, based on musicality perception as part of speech perception, also explains musicality as resulting from a generic property of neural responses. But the new theory is quite specific as to the why and how of this property. And, what is more interesting, the new theory does not depend on cortical plasticity to explain how we perceive those features of music that appear to only exist in music.

10.2 Musicality in Cortical Maps

If musicality is a perceived attribute of speech, then it follows that all of the cortical maps that respond to music *are also cortical maps that respond to speech.* Since speech is a significant component of human behaviour, and listening to speech is a significant component of human perception and cognition, it is very likely that many of the cortical maps that perceive and process universal aspects of speech are substantially predetermined in both their location and their functionality.

Thus each identifiable aspect of music is a super-stimulus for one of these predetermined cortical maps that plays a role in speech perception. But, as already discussed in Chapter 8, the corresponding speech aspect may lack some of the qualitative features of the musical aspect. To give an example: a cortical map might respond to multiple notes in music, but that does not mean that its purpose is to respond to simultaneous speech from multiple speakers.

With these considerations in mind, we can proceed to the next major steps in the analysis:

- For each musical aspect, make an intelligent guess as to what type of cortical map would respond to that aspect.

- And, having made such a guess, try to discover a plausible purpose that such a cortical map would have in the perception of speech.

When I started making these guesses about cortical maps, it took me a while to realise the importance of symmetry, and initially I did not take it into account. Perhaps this was a good thing: some of my hypothesised cortical maps seemed to represent information in ways that were unnecessarily obscure and indirect, but then later I realised that these representations made sense given the requirement for perception of speech melody that was invariant under both pitch translation and time scaling. It was a good thing because one hypothesis—that music is a super-stimulus for musicality which is an aspect of speech perception—had caused me to suppose the existence of cortical maps for calculating representations of speech which happened to be invariant under certain symmetries, and this prediction coincided with the implications of a second hypothesis, i.e. that speech and music perception include non-trivial mechanisms to achieve invariance under these symmetries, where the second hypothesis had been made for reasons independent of those for making the first hypothesis. When you are formulating speculative scientific theories based on a limited supply of hard facts and evidence, it's always comforting to discover that different and independent lines of thinking succeed in arriving at the same destination.

10.3 The Regular Beat Cortical Map

Underlying the human perception of rhythm is a basic response to a regular beat, where a regular beat is defined as some constant sound repeated at an exact and constant interval.

It is not too hard to imagine a neuron within a cortical map that specifically responds to such a beat. The following is one possible arrangement of inputs and outputs for a neuron that is intended to respond to a regular beat with a period of 500 milliseconds:

- A direct input of the current sound.
- An input of the current sound delayed by 500ms.
- An input of the neuron's own output delayed by 500ms.

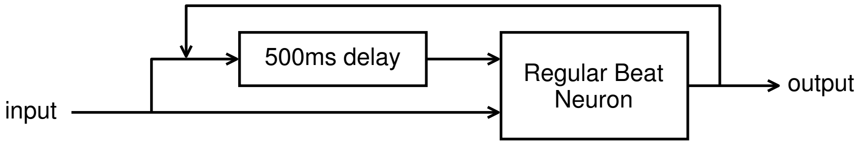

Figure 10.1. A regular beat neuron. The neuron is only activated if both inputs are active: this happens when an input is followed by an input delayed by 500ms, or when an input occurs 500ms after the neuron's own output.

The first two inputs activate the neuron when two beats occur separated by an interval of 500ms. The third input reinforces the neuron's firing when the beat occurs regularly. (Since the second and third inputs both require a 500ms delay, it is possible that they are combined *before* entering the delay, so that only one delay unit is required instead of two. This is how the delayed inputs are shown in Figure 10.1.)

Our perception of a regular beat is able to "jump over" missing beats. For example, if we program a drum machine to emit beats at a particular regular interval, we can detect the regularity of the beat, and are sensitive to any errors or changes in the timing. If we further program the machine to randomly omit a certain portion of the beats, we can still detect changes or errors in the underlying regular beat. This implies that we maintain an internal beat that helps us to fill in the missing beats.

If we are listening to a regular beat with a period of 500 milliseconds, and then one beat is omitted, there will be a gap of 1000 milliseconds. Can the regular beat neuron fill in the missing beat somehow? It is the third input in the list above that enables the neuron to maintain its own internal beat: if an external beat is omitted, the neuron will still have responded to the input from the previous beat, and at the time of the omitted beat, the delayed input from the neuron's own output will stand in for that omitted beat.

One problem with this simple model is that if a regular beat neuron can fire in response to its own delayed input, then once it starts firing it will continue firing forever, once every beat period, whether or not there is any further input signal.

A solution to this problem is to be able to generate an output signal of varying intensity. That way there can be an output signal in response to a delayed previous output signal, whether or not there is a current input signal, but at the same time the output signal will be stronger if there is both a current input signal and a delayed output signal (delayed by the correct period). If no further input signals occur, then the output signal will repeat by itself, but will gradually fade away.

This solution almost works. But individual neurons usually represent different intensities of signal by how often they fire, and unfortunately the regular beat neurons are constrained to fire on the beat.

One way to solve that problem is to represent an output signal of varying intensity by a group of neurons, such that intensity is represented by the percentage of neurons in the group that fire at one time.

Thus a given beat period is represented by a group of neurons, all of whose outputs feed into each other as delayed inputs, and all of which also take the current input signal as a direct input. The probability of each neuron in the group firing is then a function of how much input signal there is, and how many delayed output signals are received from other neurons in the group. (This idea is similar to the **volley principle** that applies to phase-locked neurons representing a frequency of sound. However, given that the frequency of regular beats is much lower than the frequency that neurons can fire at, it may not be absolutely necessary for neurons representing beats to fire at the exact millisecond the beat occurs: it may, for example, be sufficient to represent a beat by a short burst of firings, in which case the number of firings in the burst represents the perceived intensity of the beat.)

To properly recognise a regular beat of a given period, the regular beat neurons have to do more than be excited by a delayed output signal and a direct input signal—they have to be inhibited by an output signal delayed by the wrong amount (i.e. by a period which is not a multiple of the beat period). This inhibition means that a regular beat neuron can only respond to one phase of a regular beat at any particular time, because if it responds to two different phases of beat, one phase will inhibit the response to the other phase.

One way to achieve the required inhibition is to have an additional inhibitory connection from the output of the regular beat neuron which inhibits the neuron's response to new inputs that occur too soon after the output activation, i.e. less than 500ms afterwards (see Figure 10.2). In effect the neuron chooses its phase when it fires, and the additional inhibitory connection has the effect of making it insensitive to inputs with the wrong phase.

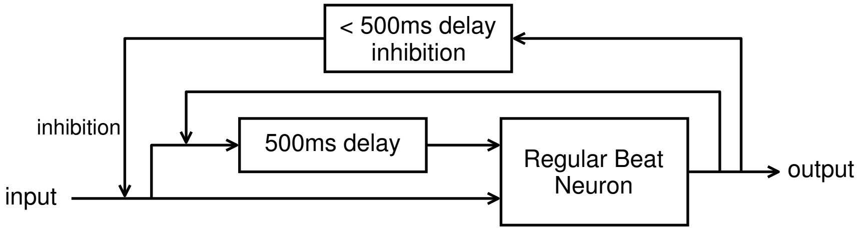

Figure 10.2. Adding inhibition derived from output signal of regular beat neuron to suppress response to out-of-phase inputs.

The neuron in the example given above is one that responds most strongly to a beat with period 500ms. To explain our perception of all the components of musical rhythm, from bar length down to fractions of a note, and our perception of music that plays at different tempos, we will have to suppose the existence of an array of neurons, all maximally responsive to different beat periods.

And, as for all cortical maps, we must take population encoding into account: there cannot be one neuron for each exact beat period, since the set of possible periods forms a continuum. As in other cases of population encoding, the response of a neuron peaks for a particular perceived value, but the response is still strong for values close to but not equal to the neuron's value of maximal response. Thus a particular beat will be represented by the firings of neurons with peak responses near to that beat period, even though none of those neurons has a peak period exactly equal to the beat period in question.

It is interesting to speculate about the mechanism that implements the required delay function, and this might involve a circuit of several neurons specialised for creating fixed delay periods between input and output signals. However, for our current purposes it is sufficient to suppose that a delay *can* be achieved somehow, without worrying too much about exactly *how* it is achieved.

The period with maximal response defines one dimension of the regular beat cortical map. It is possible that the second dimension relates to the timbre of the beat sound. Perception of a regular beat is affected by the similarity or not of the sounds in the beat; for this reason drummers have drum kits with many different drum sounds, and a given accompaniment will contain several percussive timbres, each defining a regular or semi-regular beat at a particular tempo.

Music has a hierarchical structure of regular beats. What does this imply about the pattern of activity in the regular beat cortical map when a person listens to music? There will be a series of active zones in the cortical map, with each zone corresponding to one period in the beat hierarchy. For example, with music in 4/4 time and containing sixteenth notes, there will be an active zone for each of the following periods:

- 1 bar
- 1/2 bar = 2 counts
- 1 count = a quarter note = a crotchet
- 1/2 count = an eighth note = a quaver
- 1/4 count = a sixteenth note = a semi-quaver

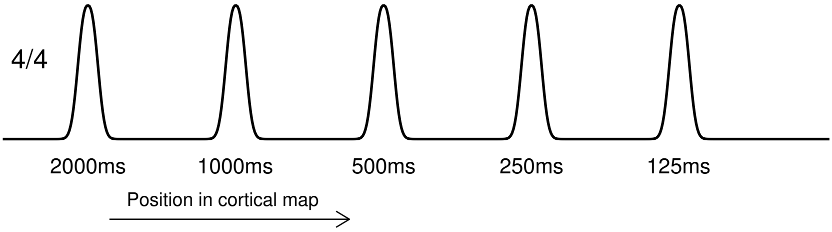

Figure 10.3. Response of regular beat neurons to 4/4 time. The graph shows response of neurons in the map to the beat periods of 1 bar length (2000ms), 1/2 bar, 1/4 bar = 1 count, 1/2 count and 1/4 count.

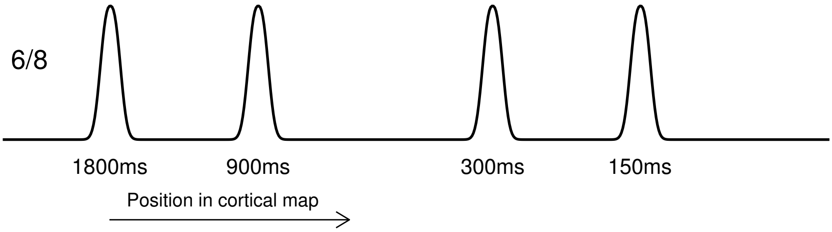

Figure 10.4. Response of regular beat neurons to 6/8 time. The graph shows response of neurons in the map to the beat periods of 1 bar length (1800ms), 1/2 bar, 1/6 bar = 1 count, and 1/2 count.

If we imagine the beat periods to be arranged on a logarithmic scale, then these zones will form a regular pattern, as in Figure 10.3. Not all time signatures, however, are based on powers of 2. There are time signatures where the beat hierarchy contains one or two factors of 3, and in these cases the spacing between zones will not be completely even, as in Figure 10.4.

What will the response of this cortical map be to the rhythms of speech? Since the rhythms of speech are not regular like those of music, there will not be a set of fixed active and inactive zones. Figure 10.5 shows a typical response that might occur to speech rhythm. The peaks are much less pronounced than in the case of the response to musical rhythm. Also the response pattern will change over time, whereas the response to musical rhythm remains constant (except in as much as the tempo gradually changes).

There is one slight simplification that I have made in the diagrams showing response to musical rhythm: all the peaks are the same height. In practice we would expect the peaks to be different heights, depending on how much the rhythm of a piece of music emphasises the different periods in the beat hierarchy. There is, however, a further complication that counteracts this variation, which is that of **saturation**, where any very high peaks get trimmed down to a size that reflects the dynamic range of the neurons in the cortical map. I explain this in more detail when I describe the scale cortical map in Section 10.6.

Figure 10.5. Response of regular beat cortical map to an irregular speech rhythm.

10.3.1 Symmetries of Regular Beat Perception

What are the symmetries of this cortical map? It is certainly time translation invariant, because the frame of reference used to define the response of neurons in the map is created by the immediate past activity of each neuron and other neurons close to it in the map—there is no global frame of reference.

This cortical map is *not* time scaling invariant. Slower and faster versions of the same rhythm will activate different neurons in the regular beat cortical map.

But, if we imagine the map laid out on a logarithmic scale (i.e. neurons separated by the same ratio of peak response beat period are separated by the same physical distance), then the effect of speeding up a rhythm by a certain factor will be to shift the activity pattern of regular beat neurons by a corresponding distance along the cortical map. This will happen for both regular musical rhythms and irregular speech rhythms.

So to achieve time scaling invariance, what is required is a second layer of processing that relates activity of each neuron in the regular beat map to activity in other neurons in the map that are fixed distances from that

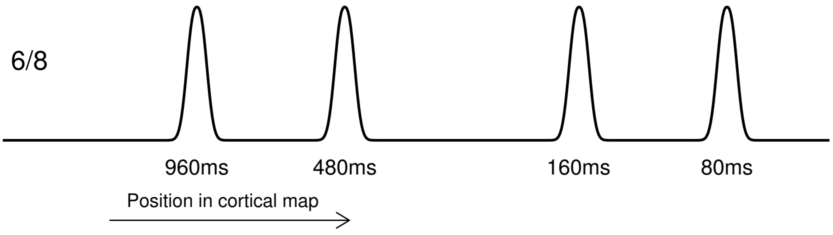

Figure 10.6. Response of regular beat neurons to 6/8 time but with faster tempo. The pattern of responses is the same as in Figure 10.4 except that it has been translated a fixed distance to the right.

neuron. There are various ways this might be done, but our strongest clue will come when we look at the brain's response to pitch information, because the strategies applied to achieve pitch translation invariance can also be applied to achieve time scaling invariance.

10.3.2 Unification

Scientists always like a good **unified theory**. Newton managed to unify falling apples and orbiting planets. Einstein unified time and space and he tried to unify gravity and electromagnetism. Modern physicists have unified two out of the four basic forces,[3] and consider it almost axiomatic that some day they should be able to unify everything.

The most satisfying aspect of my theory of music is that it achieves a convincing unification of time scaling invariance of rhythm perception and pitch translation invariance of melody perception, and does this even though the implied analogies between melody and rhythm are not immediately obvious.

The details of this unification will become apparent as we look at more cortical maps that respond to music.

10.4 The Harmonic Cortical Map

What type of cortical map would respond to chords? From the properties that chords have, we might suppose the following:

- Neurons in the map respond to notes as a function of their pitch.
- The response of the neurons to pitch is octave translation invariant.

---

[3] The four forces are the strong force, the weak force, the electromagnetic force (itself previously identified as a unification of the electric and magnetic forces) and the gravitational force. The weak and electromagnetic forces have been unified into an "electro-weak" force.

- The response of the neurons to a given pitch value is a function of whether that pitch value is harmonically related to other pitch values that the cortical map is already responding to, and whether those pitch values came from sounds with the same timbre as the sound with the new pitch value.
- Once neurons in the map respond to a given pitch value, they continue to respond to it, until activity in the map is reset (i.e. setting all neurons inactive) in some way.
- Neural activity in the cortical map is reset by a strong beat and by occurrence of a relevant low-pitch value.

These properties of the harmonic cortical map are all reasonable guesses that follow from the observed properties of chords:

- A chord consists of a particular set of pitch values.
- The occurrence of notes in chords is octave translation invariant.
- The notes occurring in a chord tend to be related to each other by consonant intervals, and are usually all played on the same instrument.
- A chord exists within a piece of music for a certain duration, even if not all notes of the chord are being played simultaneously for all of that duration.
- New chords generally start at the beginning of a bar.
- Chords are accompanied by a bass line where the dominant bass note for each bar corresponds to the root note of the chord.

What could be the purpose of this cortical map? We have already noted that, even though chords can be played as simultaneous notes, it is very unlikely that the purpose of the harmonic cortical map is to respond to simultaneous speech from different speakers. It is more likely that its purpose is to identify harmonic relationships between frequencies occurring at different times in the speech melody of *one* speaker.

Why does the harmonic cortical map have a reset function? We can say that specific notes **enter** the map when the neurons that represent those notes become (and remain) active. For example, suppose the map is initially empty (i.e. all neurons are inactive), and we play the notes of the chord C major in sequence. First we play the note C, so the neurons for the note C become active, and C has entered the map. Then we play the note G. Because G is harmonically related to C, it also enters the map, and the neurons for the note G become active. Finally we play the note E. This is harmonically related to both of the notes already in the map, so it also enters. Now we have the map in a state where neurons for the notes C, E and G are all active. If any other notes try to enter the map, they will not do so easily because they will not be harmonically related to all the existing notes in the map. Of course any re-occurrences of C, E and G will continue to reactivate the corresponding neurons.

(There are additional factors that determine which notes count as part of the chord and which do not—in particular the notes of a chord are typically played on one instrument as notes with a common timbre, which has the effect of grouping those notes together, and presumably this affects the extent to which the harmonic cortical map recognises the mutual relationships between those notes and ignores relationships among notes in the melody which are not in the current chord. Without this grouping effect, chords containing groups of notes not so strongly related to each other by consonant intervals might not be recognised as chords at all. Continuity of timbre across different chords also helps to define which notes are perceived as being part of the chord, so that, for example, notes from the melody but not in the chord do not incorrectly enter the harmonic map.)

As a result of the mutual reinforcement between neurons representing C, E and G, and the inhibition of neurons representing pitch values for other notes, the map will become stabilised into a pattern of activity involving those neurons representing C, E and G. If we assume that information is to be derived from the changing patterns of activity in the harmonic cortical map, then no more information is going to be derived once the pattern of activity becomes stabilised. An easy solution to this problem is to start again: clear the map of all activity, allowing a new set of notes to enter the map.

What event triggers the reset? Empirically we observe that a reset tends to happen when there is a strong beat at the beginning of a long beat period (i.e. the beginning of a bar). Probably it doesn't matter too much exactly what causes a reset,[4] as long as it is an event that happens occasionally, and as long as it is an event that is defined in a manner which makes it invariant under all the relevant symmetries. A beat-based trigger is unaffected by all pitch-related symmetry transformations, and it is time translation invariant and time scaling invariant.

I have described the operation of the harmonic cortical map in terms of how it responds to notes and chords in music. In speech there are no notes held at constant frequency, so the pattern of discrete notes entering the map will not occur with speech melody. What will happen is that, at any point in the speech melody, the current pitch will activate corresponding neurons in the harmonic map, but in a way that depends on which neurons are already active in the map, and on the harmonic relationships between the pitch values of the neurons already active and the current pitch value.

---

[4] When I say it doesn't matter too much, I mean that evolution could perhaps have chosen some other criterion for resetting, and the map would serve its purpose just as well.

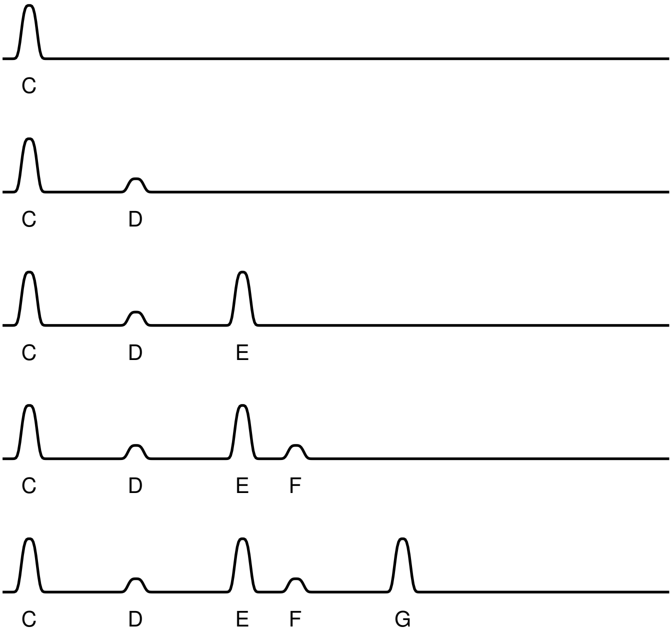

Figure 10.7. Notes entering the harmonic map. First the note C enters the map. No other note is active yet, so it enters unopposed. Then D tries to enter. Because C is already active in the map, and D is not harmonically related to C, the activation of D is suppressed. Next E enters. E is harmonically related to C but not to D. Since C is active in the map, and D is not very active, the consonant relationship with C causes E to be activated in the map. When F tries to enter the map, the dissonance between E and F suppresses the activation of F, even though F is harmonically related to C. Finally G enters the map and becomes activated because it is harmonically related to both the notes already active.

The rhythm of speech will affect the activity in the map in such a way that stronger beats of longer period will cause a reset of activity in the map back to zero (or back to a lower level). To explain the effect of bass, we must suppose that a lower frequency pitch value X activates corresponding neurons in a related bass cortical map, and these neurons in turn predispose activation in the harmonic cortical map of neurons representing pitch values that correspond to low harmonics (1st, 2nd, 3rd, 4th, 5th and 6th) of the pitch value X.

10.4.1 Active Zones

We have already observed the occurrence of active and inactive zones in the activation of the regular beat cortical map by music. We see something similar happening with the harmonic cortical map, but with the difference that the pattern of activity changes suddenly each time the map resets and responds to a new chord. And, as is the case for the regular beat map, the pattern of active and inactive zones in the harmonic cortical map does not occur in response to speech melody—the response to speech consists of changing patterns of activity with a continuous range of activity levels found across the map.

10.4.2 Octave Translation Invariant Representations

We have seen that the harmonic cortical map represents pitch in an octave translation invariant manner. This means that if a C occurs, then no matter which C it is, the same neurons will be activated. To avoid saying "octave translation invariant representation", I will use the simpler terminology that the map represents pitch values **modulo octaves**. "Modulo" is a mathematical term meaning "ignoring multiples of". For example, two numbers are equal **modulo 10** if they have the same last digit. Two musical notes are equal modulo octaves if the interval between them is a whole number of octaves.

10.4.3 Intensity Invariance

The patterns of activity in the harmonic cortical map are not pitch translation invariant. But, if we measure the *intensity* of activation of the currently entering pitch, then this intensity (as a function of time) *is* pitch translation invariant. In effect the degree of activation of a given note in the map constitutes a pitch translation invariant encoding of that note, derived from its relationship to the occurrence of other notes in the melody.

Invariance of intensity can also be found in the regular beat cortical map. In this case we are looking for time scaling invariance. We have noted that the effect of time scaling is to cause the pattern of activity to be translated within the cortical map. The intensity of the response to a given beat reflects the current dominance of the beat period whose corresponding neurons are activated by that beat. In as much as the cortical map consistently relates this dominance to intensity of activation (for different beat periods), the sequence of intensity values will be time scaling invariant.

The analogy to the harmonic cortical map is quite strong—the regular beat cortical map has time scaling invariant intensity and the harmonic cortical map has pitch translation invariant intensity. In the harmonic cortical map, translating the pitch will correspondingly translate the pattern of activity in the map. Because the map represents notes modulo octaves, there will

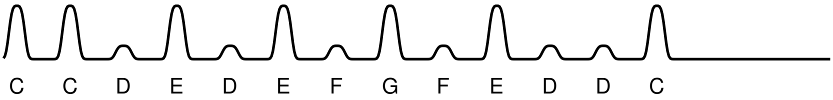

Figure 10.8. An encoding of the melody CCDEDEFGFEDDC, where each note is encoded according to its degree of activation in the harmonic cortical map as in Figure 10.7. Because the activations are based on the relationships between notes, the sequence of intensities is a pitch translation invariant characterisation of the melody. It is not a *complete* characterisation: in the current example there is no distinction between the encodings of C, E and G which all have the same level of activation in the harmonic cortical map when it is responding to the chord of C major, and there is no distinction between the non-chord notes D and F, which have the same level of activation as each other.

be a wrap-around with this translation, i.e. values translated off one end of the map will reappear in corresponding positions at the other end of the map. But as long as the map consistently represents the strengths of activation and the effects of consonant relationships for different pitch values, the intensity as a function of time will be pitch translation invariant.

10.5 The Bass Cortical Map

Bass notes are tightly coupled to chords. The major function of a bass note appears to be to emphasise the identity of the current chord which has that bass note as its root note. So we can guess the existence of a corresponding cortical map with the following properties:

- It responds most strongly to notes of lower pitch.
- It affects the entry of notes into the chord map, in particular favouring the entry of notes that are equal (modulo octaves) to the bass note, and of notes that are equal (modulo octaves) to other low harmonics of the bass note.

Because a new bass note representing the root note of the chord starts at the same time as the chord, the bass can be understood as helping to trigger the reset of the harmonic cortical map, to deactivate the old chord and start activating a new chord.

The bass cortical map sits in between being octave translation invariant and not being octave translation invariant. The effect of the input is not necessarily octave translation invariant, as the bass cortical map responds most strongly to the lowest note. If I add an octave or two to the lowest note, then it is not going to be the lowest note anymore. But the effect that the output of the bass cortical map has on the harmonic cortical map *is* octave translation invariant. For example, if C is identified as a bass note, then this reinforces any chord with a root note C, and it does not matter which octave the bass C note was in.

As is the case for other aspects of music, we must remember that music is a super-stimulus. Bass notes in music are often played much lower than the notes in the melody. The only constraint that seems to exist on how low bass notes can go is that we be able to hear them. Speech melody does not include extra notes so much lower than the main speech melody. The response of the bass cortical map to speech melody is such that the response to lower pitch values is greater than the response to higher pitch values. The consequence of this response function is that the super-stimulus for the map consists of notes with very low pitch values.

10.6 The Scale Cortical Map

Scales are a major component of almost all music. This includes not just modern Western music and traditional Western music, but the music of most cultures. As has already been explained in the chapter on music theory, a scale is a set of notes from which a melody is constructed, and scales are normally octave translation invariant.

There are two basic difficulties we encounter when trying to understand how scales relate to perception of speech melody:

- There are no notes in speech melody.
- There are no scales in speech melody (if there aren't any notes, then there cannot exist a scale from which notes are chosen).

By contrast:

- Speech melody consists of pitch (or frequency) as a smoothly varying function of time.
- The only jumps in pitch occur when there is a gap in the voiced sound.

But, as is the case for all other aspects of music, the dissimilarity between the musical aspect and the speech aspect is not a fatal obstacle in our plan to relate the musical aspect to the perception of speech. We must try to discover a cortical map that responds to the musical aspect, and then see if the same cortical map serves a useful purpose in the perception of speech.

If we regard the scale as being a property of a melody, then we can ask what is required to perceive this property. The scale is determined according to the occurrence of notes in the melody, but independently of the times at which they occur. This suggests a cortical map that responds to the occurrence of pitch values, and continues responding indefinitely, so that it builds up a picture of the full set of pitch values that have occurred in the tune. In a continuous speech melody, each pitch value only occurs for an infinitesimal period of time, so the "set" of pitch values is better thought of as a continuous density function. Another consideration is that we would not expect the neural response to last too long, since the details of any melody (speech or musical) eventually become irrelevant (and the cortical map needs to clear itself so that it can process new melodies). The response will have to "fade away" at some finite rate.

So we have a cortical map with the following characteristics:

- Neurons in the scale map have a response to pitch modulo octaves. (This immediately explains why scales are octave translation invariant.)
- Neurons are activated by pitch, in inverse proportion to the speed at which the pitch in the melody is changing.
- The activation of neurons by incoming pitch values decays slowly. We might expect the rate of decay to correspond to the time scale of a spoken sentence.

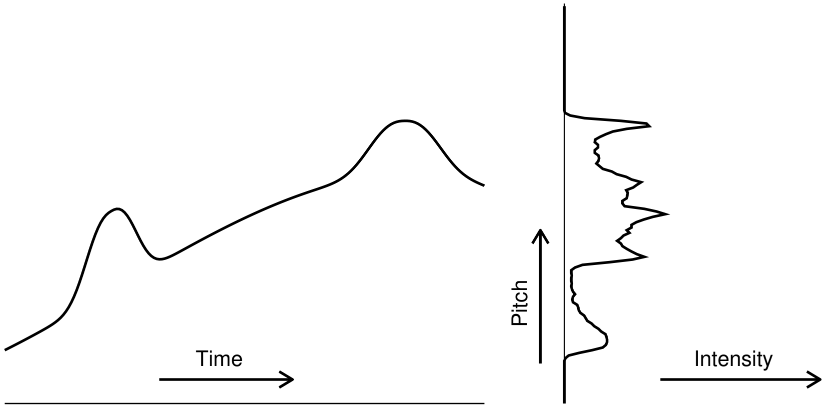

Figure 10.9. Response of the scale cortical map to a smooth melodic contour. The graph on the left is an arbitrary melodic contour as a function of pitch against time. On the right is the graph of intensity of activation in the scale cortical map as a function of pitch. The graph of intensity against pitch is rotated and reflected so that it shares a common pitch axis with the first graph. The intensity is greater for those pitch values where the contour is increasing or decreasing more slowly, and for those pitch values that occur more than once in the melody.

As in other descriptions of cortical maps, the statement that neurons respond to particular pitch values is subject to the caveat of population encoding: i.e. each individual neuron has a response that is a function of how close the incoming pitch value is to the pitch value that the neuron has a peak response to.

So what will be the response of this scale map to music which is composed from notes on a scale?

- Because musical notes do not move up or down, but remain constant for their duration until they change to a new note, the activation of the neurons for those pitch values will be at a maximum rate.
- Because pitch values between the notes on the scale do not occur, the neurons for those in-between pitch values will not be activated.

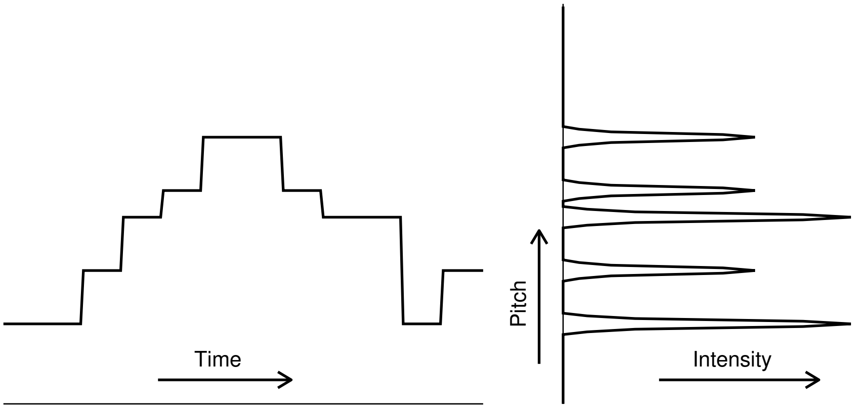

Figure 10.10. Response of scale map to musical melody. The same algorithm is used to calculate the response of the cortical map to the melody as in Figure 10.9. However, because the musical melody consists of notes held at fixed pitches selected from a finite set of values, the response of the cortical map consists of high activity of some neurons and very low or zero activity of other neurons.

The result will be a series of active zones separated by inactive zones. We have already seen this pattern with the regular beat cortical map and the harmonic cortical map. Seeing the same pattern occur in maps that relate to qualitatively different aspects of music strongly suggests to us that there is something deeper going on here. We may not be so far from a full answer to the question "What is music?". But for the moment I will carry on with the analysis of individual maps.

The active zones in the scale cortical map are very likely to be **saturated**. The response to pitch values is inversely proportional to the rate of change (of pitch): in effect the cortical map is measuring how much each pitch value has occurred as integrated over time. The constant notes of music are unnatural, and it is therefore likely that the degree of activation will go outside the dynamic range of the cortical map, since the cortical map was not designed (by natural selection) to deal with the contrived extreme patterns of musical melody. What happens when a neuron attempts to encode a numerical value that goes outside its normal dynamic range? The most that could be expected of a neuron in this situation is that it fire at its maximum possible rate.

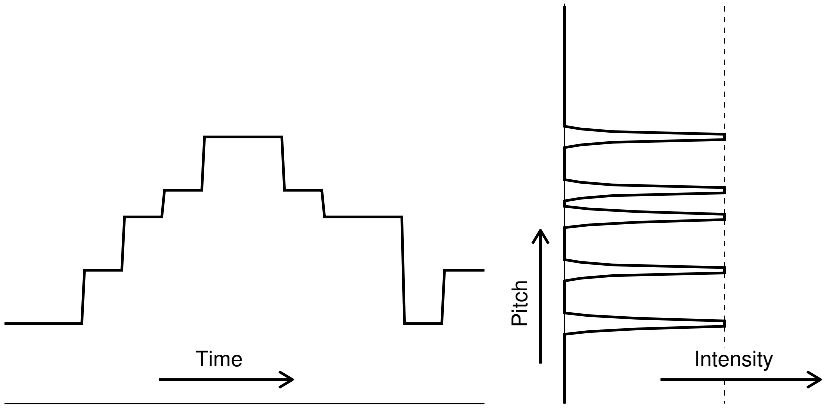

Figure 10.11. The effect of saturation on the response of the scale cortical map to music. The dashed line on the right shows the maximum intensity represented in the map (corresponding to the maximum rate of neural firing). Saturation occurs when the intensity function is capped by this maximum value. Saturation does not occur with smooth speech melodic contours because the measured intensities are lower.

It is likely that similar phenomena of **saturation** will be found in the regular beat cortical map and in the harmonic cortical map.

We can also consider the invariances of intensity for the output from the scale map. Intensity will be pitch translation invariant, for similar reasons to those that cause invariance of intensity in other maps. But because activation for a given pitch value depends on the rate of change in pitch, the intensity of the output will not be time scaling invariant for speech melody (for musical melody it will be invariant due to the saturation, but that is not relevant to how the cortical map provides invariant perception for the case of speech). Only the *relative* intensities will be time scaling invariant; i.e. there will need to be further processing that provides a final output value invariant under the operation of scaling the intensity of the scale map's output by a constant factor.

10.7 The Home Chord Cortical Map

Most simple tunes in Western music have a home note and a home chord. The tune starts with the home note and chord, moves on to other notes and chords, and eventually returns back to the home note and chord.[5]

The home chord for a tune always contains notes from the scale that the tune exists in, but different tunes on the same scale can have different home chords. However, for a given scale, most tunes on that scale have a home chord taken from a very limited set of choices. In particular, on the diatonic scale, the home chord is almost always one of two choices. On the white notes scale, these choices are A minor or C major.[6] We can postulate that the tendency to have one of these two chords as the home chord is an intrinsic property of the scale itself. This makes it much easier to speculate about the forces that determine which chord becomes the home chord, because a scale is a much simpler thing than a piece of music.

Given that the home chord of a tune played on the white notes scale is either A minor or C major, what determines which of these two it is? A very simple rule appears to work in all cases that I know of: whichever of the two chords occurs first.

The set of possible home chords for a scale is obviously a pitch translation invariant function of that scale. To put it another way, the choice of notes from the scale to make a home chord must depend entirely on the relationships that the notes in the scale have with each other. We can also mention that since the home chord and the scale are both octave translation invariant, the processes that determine the home chord must also be octave translation invariant.

An important determinant of the home chord of a scale must be the very unevenness of the scale. The diatonic scale is invariant under an octave translation (or multiple thereof), and *it is not invariant under any other translation.* If, for example, it consisted of two identical halves, so that it was invariant under a translation of half an octave, then the set of possible home chords would have to be invariant under the same translation (of half an octave).

We could assume, as a first approximation, that the choice of home chord is determined by the relationships between notes considered pairwise: it is always simpler to connect items of information two at a time. Relationships between pairs of notes can be identified by two main criteria:

1. How close the notes are to each other, i.e. proximity relationships.
2. If the notes are related by harmonic intervals.

---

[5] However, both the home note and home chord may occur in the middle of the tune, separately or together, without the tune being finished at that point.

[6] And on some of the variants of the white notes scale, where G is replaced by G♯ and optionally *F* is replaced by F♯, there is really only one choice, which is A minor.

Which of these two criteria has the most influence on the choice of home chord?

If we look at the white notes scale, the strongest (i.e. most common) choice of home chord is C major. If we look at the immediate environment of the note C, the three steps below it (going upwards from G to C) are tone, tone, semitone, and the two steps above it are tone, tone. The only other note that has this environment is the note F. But F does not occur as a home note for melodies in the white notes scale. It seems that we can therefore rule out proximity relationships as a major contributor to determining the choice of home note or home chord.

The second possibility is to consider harmonic relationships. We have already found a concise way to present and view all harmonic relationships between notes on the white notes scale: the **Harmonic Heptagon**. When we look at the location of the two possible home chords on this diagram (as in Figure 10.12), the following aspects are very suggestive:

- The notes of the two possible home chords exist on one side of the diagram, opposite the location of the note D.
- The Harmonic Heptagon has a reflective symmetry, and A minor and C major are mirror images of each other under this symmetry.
- The note D, as well as being the centre of symmetry, and opposite the home chords, is surrounded by the notes B and F, which are the two notes that have fewer harmonic relationships between themselves and other notes. The interval between B and F is not a consonant interval, whereas all other intervals between notes two steps apart on the heptagon are a perfect fifth.

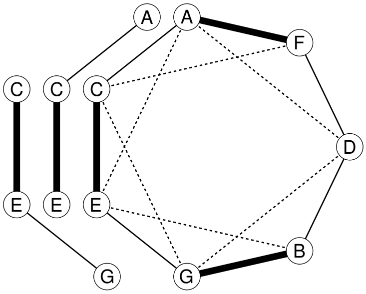

Figure 10.12. The preferred home chords (for the white notes scale) C major (CEG) and A minor (ACE) and their positions in the Harmonic Heptagon.

These observations suggest the existence of a home chord detecting cortical map with the following properties:

- Neurons in the map represent pitch modulo octaves.
- Neurons representing one pitch value will reinforce the activity of neurons representing another pitch value if the interval between the two pitch values is consonant.
- The level of reinforcement (as a function of interval size) is the same between two notes in each direction—this underlies the reflective symmetry.
- Notes that are not harmonically related to each other mutually inhibit one another in the map.

This results in the observed home chords of C major or A minor as follows:

- The notes B, F and D are weak in the map (i.e. the neurons representing those notes are only weakly activated) because the missing consonant interval between B and F weakens the mutual reinforcement between those notes and other notes.
- The mutual inhibition between neurons representing notes not harmonically related to each other means that only one of A or G can be in the home chord. Thus, given that the home chord tends to contain the notes A, C, E and G, it will be either A, C and E or C, E and G. This is the broken symmetry referred to earlier—"broken" in the sense that the set of possible home chord notes A, C, E and G has full reflective symmetry, but one of two non-symmetric subsets of this set must be chosen to be the actual home chord.
- Given that the tune is being played on the white notes scale, the home chord map can stabilise in one of two states: activation of neurons representing A minor, or activation of neurons representing C major. The competition for the two states is won by whichever chord has its neurons activated first, i.e. which of the two chords is played first.

As already mentioned, the choice of home chord is very much a function of the scale, and not of any details of the tune, except for determining the choice between the two possible chords. In particular, the strength of activation of neurons for a given note in the home chord map is independent of how many times that particular note occurs, which suggests that the inputs to the home chord map are filtered through some cortical map that responds to the occurrence of notes, while ignoring how many times or for how long those notes occur. But this describes the scale cortical map whose existence we have already hypothesised. We conclude that the home chord map probably receives its input from the output of the scale map.

10.7.1 Why Reflective Symmetry?

If there really does exist a musical symmetry of pitch reflection, it appears to be derived from a symmetry of mutual reinforcement between neurons representing notes as a function of the interval between them being consonant, or not, as the case may be. That is, the reinforcement (or inhibition) of activity in neurons representing note X by activity in neurons representing note Y is equal to the reinforcement (or inhibition) of activity in neurons representing note Y by activity in neurons representing note X.

Is there any particular reason why this symmetry needs to exist? We can consider mutual reinforcement as giving rise to an iterative voting system. Notes vote for and against each other, and the more votes a note receives in its favour, the more its own votes count for. Asymmetrical reinforcement could give rise to instabilities in this iterative process, and these instabilities would result in artefacts of changing activation independent of the incoming data.

For example, instead of the Harmonic Heptagon, we might have a scale that was approximately even, with all notes separated by the same interval. No note would occupy a special position relative to the others, and on average each note would be equally reinforced by all the other notes. But if reinforcement was stronger in one direction around the circle than the other, e.g. reinforcement of notes separated by a certain interval was stronger going clockwise than anticlockwise, then we would get “waves” of activation travelling clockwise around the circle. As a result the patterns of activity in the cortical map would fail to stabilise, or would take an undesirably long time to stabilise, and the travelling “waves” would be sensitive to minor variations in the input data. These properties conflict with the requirement that a frame of reference should be stable and insensitive to small changes in the data.

10.7.2 Alternative Theory: The Dominant 7th

An alternative theory of the preferred home chord is based on the observation that the white notes scale contains the dominant 7th (GBDF) and the chord it normally resolves to, C major (CEG), which is the home chord. This theory possibly better accounts for the choice of A minor as home chord for the harmonic minor scale (A, B, C, D, E, F, G♯) and the melodic minor scale (A, B, C, D, E, F♯, G♯), since these scales contain both the dominant 7th (EG♯BD) and the chord A minor (ACE).

In effect this alternative theory states that the preferred home chord is determined by more than just pairwise relationships. The full story may be a mixture of this theory and the theory based on pairwise consonant relationships. Note that A minor can be the home chord for tunes on the white notes scale that do not have the chord (EG♯BD) appear at all, so the dominant 7th is not essential to the “hominess” of the home chord.

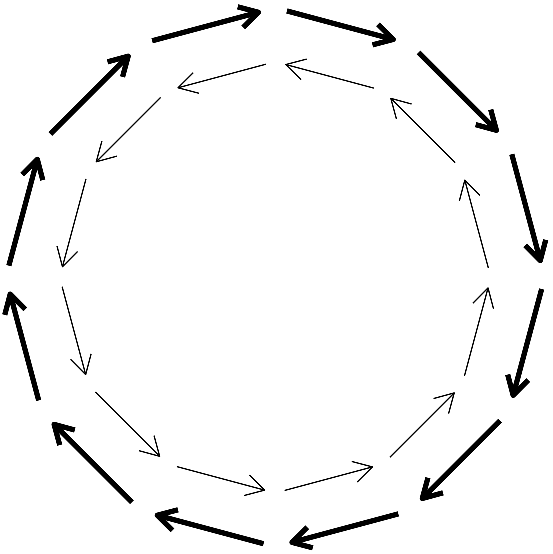

Figure 10.13. A pattern of mutual reinforcement which is pitch translation invariant but not pitch reflection invariant (the cycle represents one octave modulo octaves). Reinforcement is stronger in one direction than in the opposite direction. This will cause waves of activation to travel in the dominant direction.

10.7.3 The Evolution of Cortical Maps

The home chord cortical map is somewhat similar to the harmonic cortical map. Both maps represent pitch modulo octaves, both have an activation of neurons that persists after the occurrence of the relevant pitch, and both have mutual reinforcement between consonantly related notes and mutual inhibition between notes not consonantly related.

Here is a list of differences:

- The home chord map does not have any reset function. Thus the home chord of a simple tune played on a constant scale remains constant for the whole tune.
- The mutual inhibition between notes not consonantly related to each other is much stronger in the home chord map. It is not possible to have four notes in a home chord, whereas in general four note chords occur quite commonly in music, and five note chords are not unknown.

When we see groups of cortical maps that are not exactly the same, but somewhat similar in their properties, then this has an obvious evolutionary interpretation: at one point in the history of our species there only existed one map, but then this map evolved into multiple copies of itself, and each copy evolved useful variations in the activation characteristics of its neurons.

This is not the only example of musical cortical maps that may be relatives of each other in evolutionary terms. We will see this next when we investigate the perception of note length.

10.8 The Note Duration Cortical Map

Remember that our first model of the regular beat cortical map contained neurons with three inputs:

- Current input signal
- Current input signal delayed by beat period
- Current output delayed by beat period (feedback input)

It was the last input which gave these neurons the ability to detect ongoing regular beats, in a manner robust to omission of occasional beats.

If we omit the feedback input, then what we are left with is a neuron that responds to a particular duration of note, if we regard the duration of a note as being delimited by a beat at the beginning of the note and a beat at the end of the note (which is usually also the start of the next note).

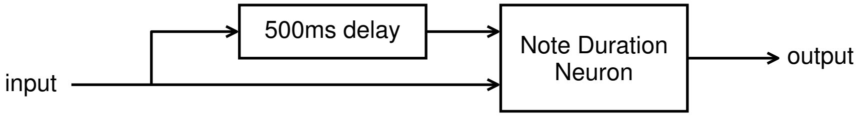

Figure 10.14. A neuron that responds to a duration between the start of one note and the end of that note (or the start of the next note) of 500ms. Compare to Figure 10.1 which had one additional feedback connection from the output to the input of the delay unit.

Note length is something that we are consciously aware of when we listen to music, so it is not surprising that there should exist a cortical map that responds directly to note length.

With note length, as with regular beat period, the information resulting from this processing layer is *not* time scaling invariant. By one means or another, the information from this layer must be processed in a way that measures *relative* note lengths.

Because the same invariance applies to regular beat period and note duration, it is possible that the processing which produces time scaling invariant characterisations may involve combinations of both, i.e. comparison of beat periods to note durations in addition to comparison of beat periods to beat periods and comparison of note durations to note durations.

The similarity of neurons in the note duration map and the regular beat map suggests that the two maps evolved from a common ancestral cortical map.

10.9 The Melodic Contour Cortical Map

We have discussed four cortical maps relating to melody and pitch: a scale map, a harmonic map, a bass map and a home chord map. None of these maps actually contains any response to whether the tune is going up or down. In fact they all process pitch information modulo octaves, and an interval between two pitch values considered modulo octaves may be considered as either an interval going up or an interval going down.

We are, however, quite aware of whether a melody—be it speech or music—is going up or down. Speech melodies mostly go up and down in a smooth fashion. Musical melodies also tend to go up and down smoothly. They cannot go up and down completely smoothly, because the notes are taken from a discrete scale. But they do go up and down as smoothly as possible, in as much as the next note after a note in a melody is very often the same note, or just one step higher or lower.

Responding to the up and down motion is an easy way to produce a characterisation of melody that is pitch translation invariant. It has robust translation invariance, because it is invariant under any transformation that is **monotonic**. A transformation is monotonic if it preserves the distinction between intervals going up and intervals going down. A monotonic transformation does not necessarily preserve any particular notion of size of interval. It follows that characterisation of up and down motion of pitch does not require any special calibration, whereas (as we will see in Chapter 12) pitch translation invariant characterisations of melody depend on accurate comparisons of intervals between different pairs of notes.

The perception of melodic contour gives us an explanation of why melody is not **locally octave translation invariant**: we cannot translate individual notes of the melody by different numbers of octaves, in the way that we can do for notes in chords and bass, since adding octaves to some notes but not others would radically change the contour of the melody by changing the sizes of the steps between consecutive notes.