Chapter 12

Calibration

The brain has the ability to perceive intervals between pairs of pitch values in such a way that intervals corresponding to the same frequency ratios are perceived as being the same.

How does the brain calibrate the perception of equality between frequency ratios? Careful consideration makes us realise that this calibration is non-trivial for a biological organism to achieve.

The answer seems to be that calibration is made against harmonic intervals observed to occur between the harmonic components of the human voice.

A similar type of calibration may underlie the time scaling invariance of rhythm perception.

12.1 A Four-Way Relationship

Pitch translation invariance involves an ability to perceive a four-way relationship between pitch values. For example, we can recognise that the interval from C to E is the same as the interval from F to A. This is a relationship between the notes C, E, F and A.

A naïve implementation of such a four-way relationship would involve connections between groups of 4 neurons. Taking into account all sets of pitch values related in this way, and even after reduction by means of splitting pitch into absolute and modulo octaves components, such an implementation would require a large number of connections. It would require $O(N^3)$, where N is the number of distinct pitch values (modulo octaves): for every 3 pitch values X, Y and Z there is a 4th pitch value W determined by the equation $X - Y = Z - W$.

We know that we have a subjective perception of interval size. This can be interpreted as a three-way relationship between pairs of notes and interval sizes. Thus X, Y, Z and W are related as described above if there exists some interval Q such that the interval from X to Y equals Q and the interval from Z to W also equals Q.

This three-way relationship requires connections between sets of three neurons: two representing pitch values and one representing the interval between them, and this requires $O(N^2)$ connections (as already discussed in Chapter 11, when analysing the implementation of neural subtraction tables).

The ability of the combination of human ear, nervous system and brain to detect these relationships between quadruples of notes makes that combination into a reasonably precise measuring instrument. And seeing it as a measuring instrument, a simple question can be asked: how is the instrument *calibrated*?

12.2 Making Measurement Accurate

There are two main approaches to making sure that a measuring machine is as accurate as it is required to be:

- Construct the machine using precise construction methods that result in it having the required accuracy.
- Construct the machine less precisely, but include in the machine some mechanisms for adjustment which allow it to be calibrated against known standards for the type of measurement involved.

In the world of industry both of these methods are used. The first method is limited by the fact that a direct product from a manufacturing process is on average going to be less precise than the system used to manufacture it. The accuracy of most rulers and measuring sticks depends on the accuracy of the moulds and other factory machinery used to make them. But if we want a ruler that is really, really accurate, then a ruler stamped by a mould may not be good enough.

What does this have to do with our perception of intervals? The first type of calibration would involve the human ear, nervous system and auditory cortex all being pre-programmed to grow and develop in such a way that the intervals between different pairs of pitch values whose frequencies are in the same ratios are perceived as the same intervals.

I don't have a formal proof that this couldn't happen—but it seems very unlikely. Different people are all different shapes and sizes. Different parts of different people are different sizes. Everyone has differently shaped ears to everyone else. Much of the way that our body develops involves different components growing in relation to other components, so that, for example, the lengths of our muscles and tendons match the lengths of our bones.

It seems implausible that, within this framework of variation and relative sizing, there could exist a system of measurement pre-programmed to develop to the accuracy exhibited by our ability to perceive and identify musical intervals.

This leaves us with the second possibility: approximate construction, followed by calibration against a naturally occurring standard.

Now we already know that intervals between pitch values which are simple fractional ratios play a significant role in our perception of music. And we know that these are the same ratios that occur between frequencies of harmonics of individual sounds, for certain types of sounds. And at least one type of sound having this property occurs naturally: the human voice.

This suggests an explanation as to why differences between the pitch values of *different* sounds are significant when they are equal to the differences between frequencies of harmonic components within the *same* sound: our auditory perception system uses the harmonic intervals between harmonic components of the same sound to calibrate its perception of intervals between the fundamental frequencies of different sounds.

In the world of industrial physical measuring instruments, we first calibrate our instrument to some degree of reliability, and having done that we then use our instrument to measure things, without concerning ourselves as to how the instrument was calibrated. The only lingering consequence of the method of calibration is that it adds to the expected error of our measurements.

In the world of biology, different components of functionality are often not as clearly separated from each other as we might expect from analogy with man-made artefacts and systems. With regard to the calibration of interval perception against harmonic intervals, there is one simple problem:

> How do we calibrate our perception of non-harmonic intervals?

There are various ways that we might consider of doing this, and three main candidates are:

- Interpolate, i.e. relate non-harmonic intervals to harmonic intervals slightly larger than and slightly smaller than the non-harmonic intervals.

- Approximate non-harmonic intervals by harmonic intervals with fractions that contain numerators and denominators of greater size.

- Construct approximations to non-harmonic intervals by adding different harmonic intervals together.

There is a fourth option, and we will see that it may be the preferred one in many cases, which is *not to measure non-harmonic intervals*. But first we will investigate what may be involved in the first three options, and how plausible they are as methods that could occur in practice.

12.2.1 Interpolation

The simplest form of interpolation would be to take two values that we know, and then identify a value that is half-way between those two values. The only technical difficulty we have to overcome is to devise a consistent way of determining what is "half-way". One way to do this involves observing pitch values that are rising in an approximately linear manner.

Figure 12.1 shows the calibration of an interval $(X+Y)/2$ by interpolation between two calibrated intervals X and Y. Pitch values f_0, f_1 and f_2 occur at times t_0, t_1 and t_2 respectively. $X = f_1 - f_0$, $Y = f_2 - f_0$, and the estimate for $(X+Y)/2$ is $f' - f_0$, where the pitch value f' occurs at time $t' = (t_1 + t_2)/2$, i.e. halfway between t_1 and t_2. The calibration depends on the assumption that log frequency is a linear function of time (during the period t_1 to t_2).

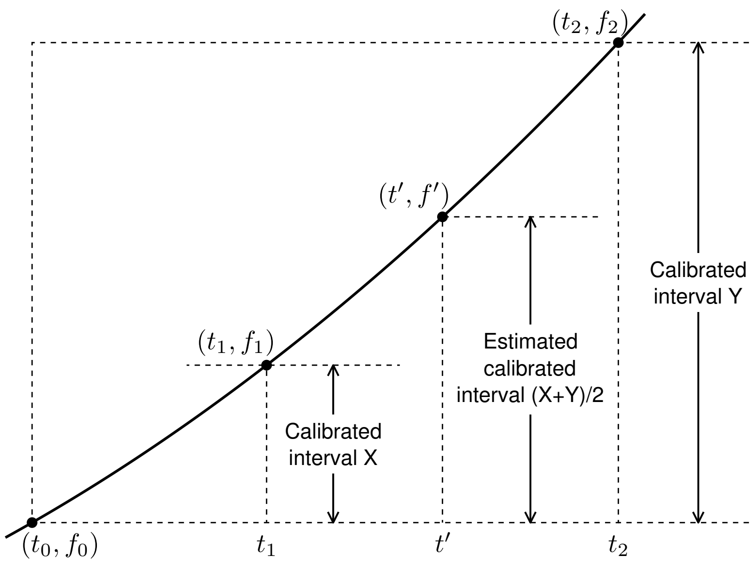

Figure 12.1. Interpolation of log frequency intervals on a smooth melodic contour. t' is exactly half-way between t_1 and t_2. The interval between f_0 and f_1 is the calibrated interval X, and the interval between f_0 and f_2 is the calibrated interval Y. The interval between f_0 and f' is an estimate for the size of $(X + Y)/2$, which would be exact if the contour was a straight line. But actually the contour curves upwards slightly, so the estimate is slightly too small.

In a pre-technological society, the major source of these rising (or falling) pitch values would be the melodic contours of speech. The contours of speech are not always straight-line contours, and linearity is an essential assumption in our interpolation procedure. In practice, however, the result may be adequate, for the following reasons:

- Any sufficiently smooth curve is linear for sufficiently small parts of that curve.
- If we average our interpolations over many different curves, they are likely to be linear on average.
- Even if average curves are not linear on average (e.g. they always curve one way or the other), the curvature that occurs in contours from different speakers with different pitch ranges may still be consistent enough to produce a calibration that is useful in practice. This will result in a calibration by interpolation which is not linear, but which is consistent among listeners exposed to a similar body of speech melodies (i.e. a group of individuals living in the same tribe).

12.2.2 Complex Fractions

There are two reasons to suppose that complex fractions[1] derived from comparing higher harmonics are not used as a means of calibrating our perception of intervals:

- It requires calibrations to be made against higher harmonics of very low frequency sounds, where the higher harmonics are in the range of the intervals that you are calibrating against. For example, to calibrate a ratio of 45:32 against the interval from 320Hz to 450Hz, you need a sound with a fundamental harmonic of 10Hz. There are not many natural sources of harmonic sound with this fundamental frequency. Certainly human speech does not go this low.
- Complex fractions are not observed to be significant in the perception of music. This suggests that the brain does not bother to use higher harmonics for the purpose of calibrating comparisons of interval sizes.

12.2.3 Arithmetic

Calibration by arithmetic is a common solution to the problem of calibrating the measurement of a value that can be defined as a sum of values that have already been calibrated. If I can calibrate a length of 1 metre, and I need

[1] By "complex" I just mean with a large numerator and denominator, in the sense that "complex" is the opposite of "simple", and "simple" fractions are fractions with small numerators and denominators.

to calibrate a length of 2 metres, then all I need to do is mark off 1 metre, and then mark off a second 1 metre that starts where the first one finished, and altogether I have marked off 2 metres. To make this work for interval perception, our perception of the interval between two notes X and Y would be mediated by the occurrence of an imaginary note Z such that $Z - X$ was a harmonic interval, and $Y - Z$ was also a harmonic interval.

Using arithmetic to calibrate non-harmonic intervals achieves a similar result to using higher harmonics, because it enables more complex fractions to be used.[2] The main reason to doubt that this type of arithmetic plays a significant role in the calibration of interval perception is the same as the second reason given above for supposing that higher harmonics are not involved in this calibration: complex fractions are not observed to be significant in music perception.

12.2.4 Not Measuring Non-Harmonic Intervals

In as much as we can perceive and compare non-harmonic intervals at all, interpolation seems the most likely of these three options to be used by the human auditory perception system for calibrating the perception of those intervals.

But the structure of music suggests major use of the fourth option: only measure harmonic intervals. Harmonic intervals are significant in music perception. Chords and harmony are the most obvious manifestation of this, but the determination of home chords also appears to be strongly tied to harmonic relationships between pitch values.

We can imagine the following means of identifying the harmonic characteristics of a smooth melodic contour:

- Identify the initial pitch value.
- Record the times at which pitch values occur that are harmonically related to the initial pitch value.

The recorded sequence of time values will then form a characterisation of the melodic contour such that this characterisation is pitch translation invariant. This means of characterisation comes very close to the operation of the harmonic cortical map, which we have previously identified as being the cortical map that responds to the occurrence of chords. (The main difference with the harmonic cortical map is that it only marks or records the times of pitch values which are harmonically related to the initial pitch value *and* to any other pitch values already marked or recorded.)

[2] Adding intervals is equivalent to multiplying fractions. If we add two intervals representing simple fractions to get an interval which does not itself represent a simple fraction, then it will necessarily represent a more complex fraction. Of course there are some complex fractions that cannot be derived from simpler fractions by means of multiplication—for example fractions containing a large prime in the numerator or denominator.

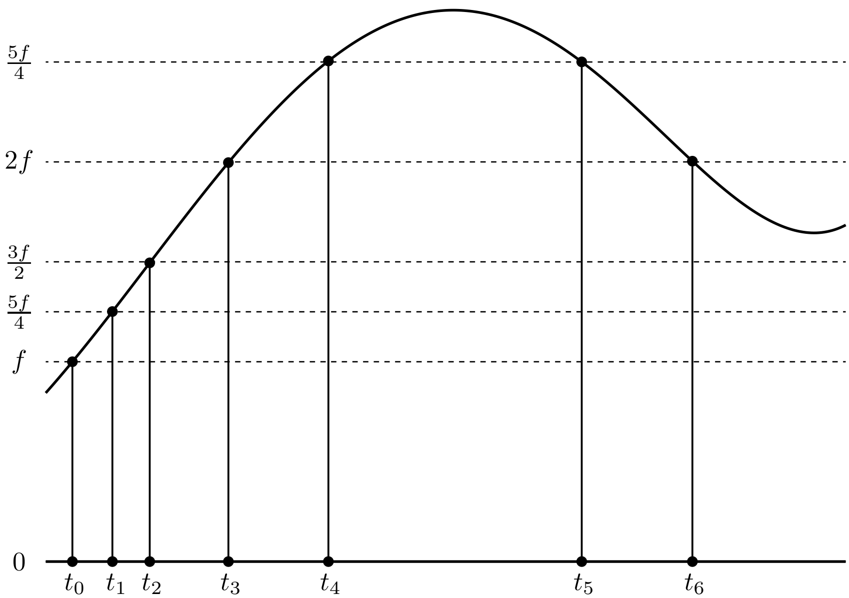

Figure 12.2. Recording the times of occurrences of some frequencies in a melodic contour harmonically related to the frequency f at time t_0. In this example, the frequencies recorded are those included in the major chord which has frequency f as its root note. (This is an unrealistic simplification, since the contour is not one that implies the major chord; however, it serves to demonstrate the general principle that the harmonic map creates a pitch invariant characterisation of the melody. In a more realistic example, the "chord" used to record pitch values in a continuous melody would be more fuzzy than a discrete set of pitch values, and the recorded time values would themselves be correspondingly fuzzy.)

From a mathematical point of view, this trick of only measuring harmonic intervals turns out to be an indirect form of interpolation. We plot a finite number of points in the melodic contour consisting of pitch values that are harmonically related, then we can fill in the rest of the contour if we need to by assuming that it is a smooth curve and joining up the points we plotted. There are occasions when the rest of the contour does need to be filled in: for example when we want to reproduce a particular contour in our own speech melodies.[3]

[3]A minor complication is being ignored here: the harmonic cortical map is only recording the times of occurrences of notes harmonically related to each other, without recording what the relations are. We must presume that the basic reconstruction of the melodic contour occurs via the melodic contour cortical map from the recorded ups and downs, and that the set of times derived from the harmonic cortical map is then used to more accurately position those points in the contour at the recorded times, by "snapping" them to the nearest frequency harmonically related to the frequency at time t_0.

12.3 Calibration Experiments

A strong test of the calibration theory would be to expose a subject to bad data over their lifetime, and see if predicted calibration errors could be observed. Bad data would consist of sounds with incorrect harmonic frequencies. On the assumption that the human voice is the main source of calibration data, all human speech that the subject heard would have to be appropriately altered. The subject would wear a microphone and headphones connected to a digital sound processor, such that all the sounds coming into the microphone were digitally altered, and the subject would hear only the altered sounds played through the headphones.

If, for example, all 2nd harmonics were increased in frequency by 5%, then we would predict that the subject's perception of octaves would be correspondingly altered. Alterations to other harmonics would alter the subject's perception of intervals. If harmonics were altered in a manner dependent on frequency, then this would be predicted to alter the subject's ability to accurately compare intervals at different pitch levels (i.e. to identify the interval from note W to note X as being the same as the interval from note Y to note Z).

It would not be ethical to carry out such an experiment on a person over their lifetime. But it is quite possible that calibration is not a once-in-a-lifetime event. As a person grows, the frequency response functions of locations in their ear (in the organ of Corti) are going to change slightly over time, and it is likely that adjustments have to be made continuously to keep the auditory cortex correctly calibrated.

If a willing subject can be exposed to altered speech for a period of days or weeks, it may be possible to observe adaptation to these alterations as a result of calibration against the contrived bad data.

This type of recalibration experiment has its precedents: experiments where subjects wear prismatic lenses which shift the image of the real world on their retinas. Subjects are observed to adapt over a period of time to this artificial shift. (And luckily the adaptation re-adapts back to normal once the subject stops wearing the special lenses.) Similar adaptation happens to anyone who starts wearing glasses, and can also happen to users of various types of virtual reality environment and augmented vision systems. Some of the research on adaptation to altered vision has been done to make sure that adaptation to virtual reality environments does not cause lasting perceptual impairment.[4]

If the recalibration of interval perception could be achieved, then a very interesting possibility arises: new types of music that are only perceived to be musical by someone whose perception of harmonic intervals has been artificially altered this way. To give a simple example, the locations of 3rd

[4]For example, *Virtual Eyes Can Rearrange Your Body: Adaptation to Visual Displacement in See-Through, Head-Mounted Displays* Frank Biocca and J.P. Rolland (Presence: Teleoperators & Virtual Environments 1998)

and 5th harmonics could be altered to exactly match the intervals that occur on the well-tempered scale. This would have the effect of making the well-tempered scale be (subjectively) perfect for these types of interval, and might increase the musicality of music played on that scale.

Experiments on calibration could also be carried out on animals. If, however, music is very human-specific, it will be difficult to find a useful animal model.

12.4 Temporal Coding

We might suppose that temporal coding plays a role in the calibration of the perception of harmonic intervals. If, for example, phase-locked neuron A was responding to a frequency of 200Hz, and phase-locked neuron B was responding to a frequency of 300Hz, then there would be exactly 2 firings of neuron A for every 3 firings of neuron B. If there was some way to count and compare how many times each neuron fired compared to the other, then this would give a natural way of knowing that the two frequencies were harmonically related.

It might also be possible to compare the times at which the two neurons fire, and record the intervals between those times, to determine whether or not the two frequencies are related to each other by a simple ratio. A basic difficulty with directly comparing the timings of individual firings is the level of accuracy required. For example, considering slightly higher frequencies, such as 1000Hz and 2000Hz (which are still within the range of musically significant frequencies), and assuming that we are required to achieve a 1% accuracy of interval perception, this implies that a 0.5% accuracy is required for each of two comparisons, which translates into 5 microseconds—a very short period of time. Although there are some known animal perceptions that operate on this time scale or even shorter, such as bat echo-location, there are severely non-trivial problems to overcome, including "jitter" and the sheer length of time it takes for an action potential to occur—typically 300 microseconds.[5] It's much easier to just calibrate against natural examples of sounds containing harmonics, which you already "know" have the correct relationships between their frequencies.

If calibration of harmonic intervals *was* based on direct comparisons of periods of vibration, then it would not be possible to mis-calibrate interval perception by exposing subjects to sounds with the "wrong" harmonics, in which case the experiments described in the previous section would give a negative result.

[5] "Bat Echolocation" by James Simmons, text box within *Neuroscience: Exploring the Brain* Bear, Connors and Paradiso (Williams & Wilkins 1996)

12.5 Other Calibrations

12.5.1 Calibration of Octave Perception

Octaves are a special sort of consonant interval. The split of pitch values into imprecise absolute pitch value and pitch value modulo octaves necessarily requires an accurate determination of the octave relationships between frequency values.

As in the case of consonant interval perception, this determination will need to be calibrated, and the most likely means of calibration is by comparison with the harmonic relationships that exist between the harmonic components of the sounds of the human voice. For calibrating octave perception it would be sufficient to consider just the fundamental frequency and the second harmonic.

12.5.2 Calibrating Ratios of Durations

Comparisons of intervals between different pairs of pitch values are required to achieve pitch translation invariant perception of melody. In a very analogous manner, comparisons of ratios between pairs of time durations are required to achieve time scaling invariant perception of rhythm.

Recall that the cortical maps relating to rhythm consist of groups of neurons that respond to percussive sounds separated by specific time intervals. Some neurons respond just to pairs of percussive sounds; these are the neurons that encode duration information. Other neurons respond to ongoing regular beats. In both cases, if there is to be an ability to perceive the same rhythm at different tempos, there needs to be a means of measuring the ratios between the time intervals that these rhythm-sensitive neurons respond to.

For example, an instance of a rhythm might consist of beats at times 0, 1 and 1.8. We will treat time as being in units of seconds. This results in two durations: 1 second and 0.8 seconds, and the ratio between them is 5:4. A slower version of the same rhythm might consist of beats at times 0, 1.5 and 2.7. The resulting durations are 1.5 seconds and 1.2 seconds, and their ratio is also 5:4. The identity of the two 5:4 ratios is what enables us to perceive that these are two versions of the same rhythm, with the second being a slowed down version of the first. Note that we are not particularly interested in the ratio between the two tempos; what matters is being able to identify the two rhythms. In fact the comparison may be between two occurrences of the rhythm at widely separated times, for example on different days. So there is no easy way to make comparisons between the durations of corresponding components of different occurrences of the rhythm; all comparisons must be made between durations occurring within each individual occurrence of the rhythm.

How can we calibrate the perception of ratios between durations and beat periods? The most obvious calibration is to compare durations where one

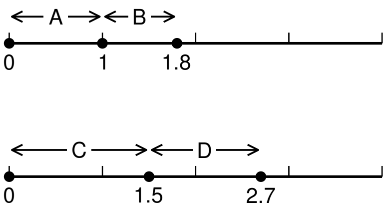

Figure 12.3. The "same" rhythm played on two different occasions at different tempos. The scale shows time in seconds. It is true that $C/A = D/B$, but to calculate C/A and D/B presumes an ability to accurately compare time durations perceived on different occasions. More realistic is to calculate the ratios in the equation $A/B = C/D$, as this only involves comparison of durations within each individual occurrence of the rhythm (and comparison of the ratios over the long-term).

duration is twice as long as another, i.e. a ratio of 1:2. If three beats occur, say X, Y and Z, with Y occurring halfway between X and Z, and one neuron A responds to the durations X to Y and Y to Z, and another neuron B responds to the duration from X to Z, then we can determine that the ratio between the duration periods of neurons A and B is 1:2.

A similar calibration could occur for other simple ratios, like 1:3, or 1:4 or even 2:3. But once again we can use our knowledge of observed aspects of music to guide us. Musical rhythm is strongly dominated by durations and beat periods related to each other by factors of 2. Factors of 3 come in a very distant second. A factor of 4 can be regarded as 2×2, and any larger factors are virtually non-existent.

So we can conclude that a calibration process occurs by which the brain identifies pairs of neurons in rhythm-oriented cortical maps that respond to durations related by a factor of 2, and to a lesser extent by a factor of 3.

12.5.3 Calibrating Against Regular Beats

There is one minor difficulty with this theory applied to time scaling invariance: calibration requires the occurrence of regular beats, such as the beats X, Y and Z given in the previous example, where the interval from X to Y is the same as the interval from Y to Z. Such regularity may not occur in natural speech or in other sounds that a developing child may hear.

Of course there is one situation where children will hear regular beats: when they are listening to music. This leads to a direct biological function for music, i.e. to assist calibration of cortical maps that process information about rhythm to produce time scaling invariant perceptions. This gives a counter-example to our working doctrine that it is only the perception of

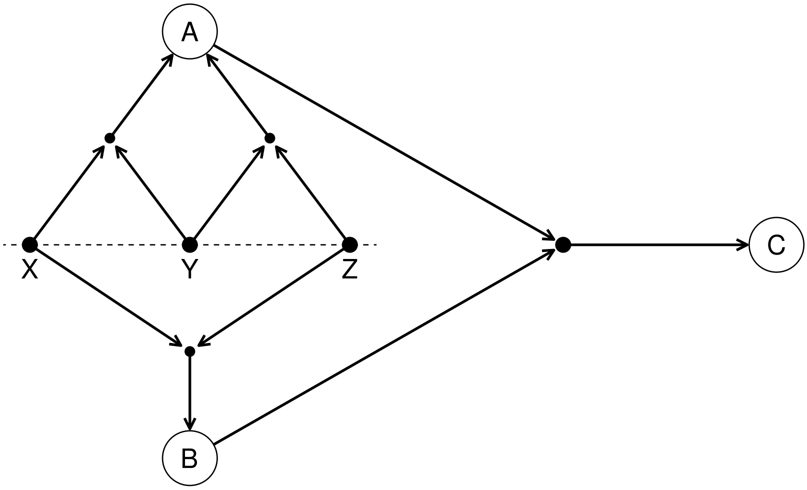

Figure 12.4. Relative tempo calibration. Events X, Y and Z represent percussive sounds perceived from a regular beat. The duration from X to Y and the duration from Y to Z both activate neuron A. The duration from X to Z activates neuron B. Neuron B therefore represents a duration twice as long as the duration represented by neuron A. Calibration results in neuron A and neuron B activating neuron C, with the consequence that neuron C encodes the perception of a ratio of 1:2 between different durations.

musicality that has a biological function, and that music in itself serves no biological function.

And if we can make this concession for rhythm perception, then we can also make it for melodic perception: the playing of melodies may assist in the calibration of the perception of harmonic intervals. For example, the calibration process may proceed more efficiently if the listener is exposed to different sounds such that the harmonic components of each sound are separated from each other by consonant intervals, and such that the different sounds have fundamental frequencies separated from each other by consonant intervals.