20
Technique 4 Modulation

SECTION 20.1
Amplitude Modulation

Remember that modulation means changing something in accordance with something else. In this case we are changing the amplitude of one signal with another. To do amplitude modulation (AM), we multiply the two signals, call them A and B, to get a third signal C. That can be written simply in the time domain as

$$C = A \times B \tag{20.1}$$

We looked at modulation with slowly moving signals earlier while studying control envelopes. A familiar effect, often used with guitar, is *tremolo*, in which the audio signal is amplitude modulated with a slowly moving periodic wave of about 4Hz. In this section we consider what happens when modulating one audio signal with another. Let's begin by assuming that both are simple sinusoidal signals in the lower audible range of a few hundred Hertz.

Traditionally one of the input signals is called the *carrier* (at frequency f_c), the thing that is being modulated, and we call the other one the *modulator* (at frequency f_m), the thing that is doing the modulating. For the trivial case of amplitude modulation it doesn't matter which is which, because multiplication is commutative (symmetrical): $A \times B = B \times A$. Let's look at a patch to do this in figure 20.1, and the result in figure 20.2.

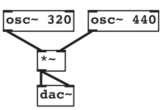

Figure 20.1
$A \times B$.

The patch is simple. We take two signals from cosinusoidal oscillators and combine them with a `*~` object. What will the resulting amplitude be if both signals are normalised? If signal A is in the range -1.0 to 1.0 and so is B, then the lowest the amplitude can be is $-1.0 \times 1.0 = -1.0$ and the highest it can be is 1.0×1.0 or -1.0×-1.0, both of which give 1.0, so we get a normalised signal back out. But what frequencies will we get? Figure 20.2 shows the answer, and maybe it isn't what you expect, since neither of the original frequencies is present. We see $f_c + f_m$ and $f_c - f_m$.

Keypoint
AM gives sum and difference components.

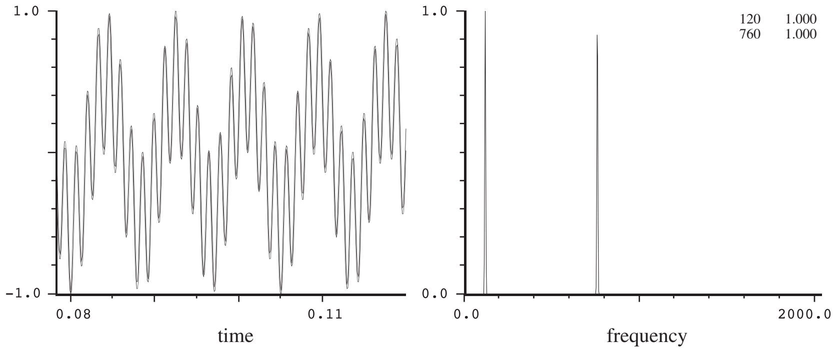

Figure 20.2
Multiplying two audio signals. The spectrum of the new signal is different from either of the inputs.

We multiplied two signals at 320Hz and 440Hz, and we got two frequencies, one at 760Hz and one at 120Hz. Multiplying two pure frequencies gives two new ones which are their sum and difference. We call these *sidebands* of the original frequencies. In this case the *upper sideband* or sum is $320\text{Hz} + 440\text{Hz} = 760\text{Hz}$, and the *lower sideband* or difference is $440\text{Hz} - 320\text{Hz} = 120\text{Hz}$. This can be seen mathematically from a trigonometric identity called the *cosine product to sum rule* which explains simple modulation.

$$\cos(a)\cos(b) = \frac{1}{2}\cos(a+b) + \frac{1}{2}\cos(a-b) \tag{20.2}$$

The amplitude of each input signal was 1.0, but since the output amplitude is 1.0 and there are two frequencies present, each must contribute an amplitude of 0.5. This can also be seen to follow from the cosine product equation. Note that the spectrograph in figure 20.2 shows the amplitudes as 1.0 because it performs normalisation during analysis to display the relative amplitudes; in actual fact these two frequencies are half the amplitude of the modulator input. So, what are the practical applications of simple modulation? As described above, neither of the original frequencies is present in the output, so it's a way of shifting a spectrum.

When using slowly moving envelope signals to modulate a signal we take its spectrum to be fixed and assume the amplitudes of all the frequencies rise and fall together. Most of the time that's true, but as is apparent from the previous equations, changing the amplitude of a signal rapidly changes its spectrum.

This seems a bit weird to begin with. But where have we seen this before? It is implied by Gabor and Fourier ...

> Keypoint
> As we make shorter and sharper changes to a signal it gains higher frequencies.

SECTION 20.2

Adding Sidebands

Above, we started with two oscillators producing two frequencies, and we ended up with two new frequencies. It seems a long way round to get rather little advantage. If we had wanted 760Hz and 120Hz, why not just set the oscillators to those frequencies? But of course we still have the two original sine signals to play with. We could add those in and end up with four frequencies in total. So, one of the main uses of AM in synthesis is to construct new and more complex spectra by adding sidebands.

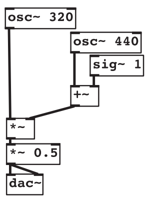

Figure 20.3
Ring modulator.

Figure 51.4 shows a patch called a *ring modulator* which is a common idiom in synthesisers and effects. This time it matters which we call the carrier and modulator. The carrier is the 320Hz signal connecting to the left of `*~`, and the modulator is the 440Hz one connecting to the right side. Notice that we add a constant DC offset to the modulator. This means that some amount of the carrier signal will appear in the output unaltered, but the modulator frequency will not appear directly. Instead we will get two sidebands of $carrier + 440$Hz and $carrier - 440$Hz added to the original carrier.

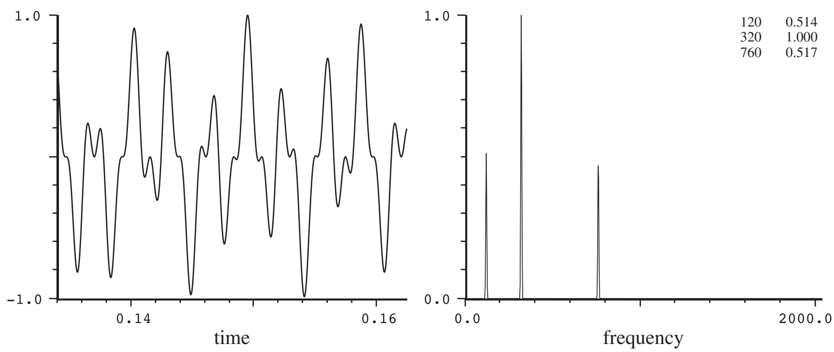

Figure 20.4
Ring modulator, showing the carrier plus two sidebands produced by modulation.

In the spectrograph of figure 20.4 you can see the relative amplitudes of the carrier and sidebands, with the sidebands having half the amplitude. No signal is present at 440Hz.

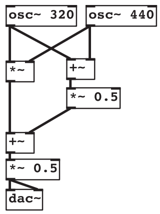

Figure 20.5
All band modulator.

If we want to get as many components in the spectrum as possible, the patch of figure 20.5 can be used. There are four possible frequencies: the carrier, the modulator, and two sidebands. The spectrum is shown on the right of figure 20.6 in which all bands have equal amplitude. Because the amplitude sum of the carrier and modulator will be twice that of the modulated signal we use half of it so that all the harmonics are of equal amplitude. So far we haven't said anything about the phases of sidebands, but you might notice that the time domain waveform is raised by 0.5 because of the way the signals combine.

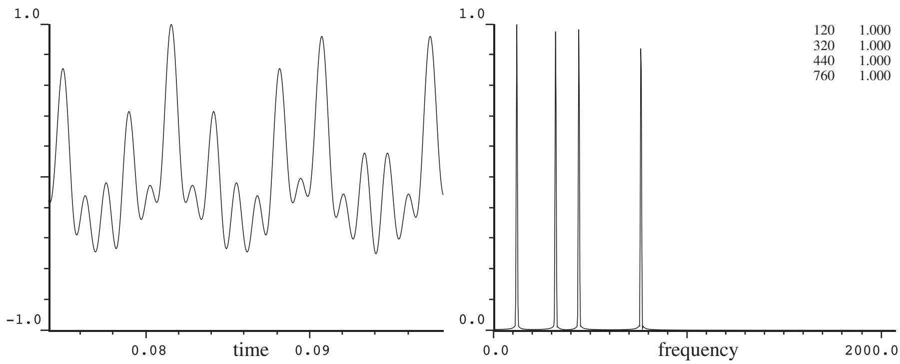

Figure 20.6
All band amplitude modulation giving sum, difference, and both originals.

SECTION 20.3

Cascade AM, with Other Spectra

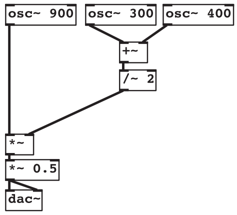

Figure 20.7
AM with two harmonics.

This process can be repeated two or more times to add more harmonics. If a signal containing more than one frequency, let's call them f_a and f_b, is modulated with a new signal of frequency f_m, as shown by the patch in figure 20.7, then we get sidebands at $f_a + f_m, f_a - f_m, f_b + f_m, f_b - f_m$, which can be seen in figure 20.8. Starting with one signal containing 300Hz and 400Hz, and modulating with 900Hz we obtain 900Hz + 400Hz = 1300Hz, 900Hz − 400Hz = 500Hz, 900Hz + 300Hz = 1200Hz, and 900Hz − 300Hz = 600Hz. We can chain ring modulators or all sideband modulators to multiply harmonics and get ever denser spectra. Starting with two oscillators we can get 4 harmonics, then add another oscillator to get 8, and so on.

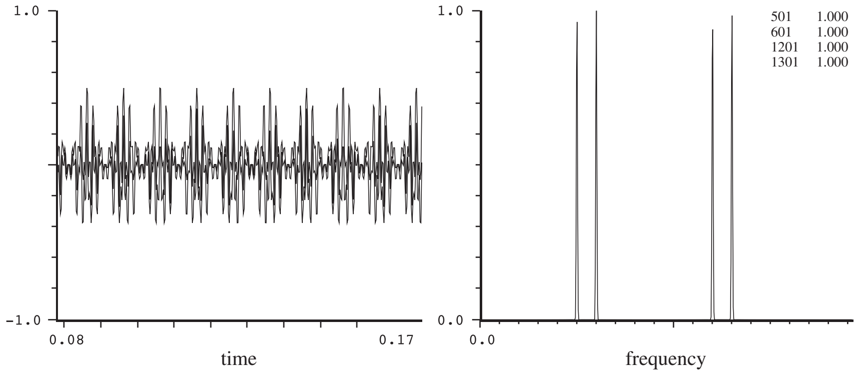

Figure 20.8
Modulating a signal containing more than one harmonic.

SECTION 20.4

Single Sideband Modulation

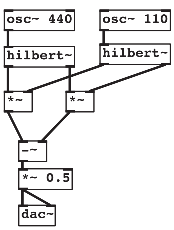

Figure 20.9
Single sideband modulation.

One of the problems with simple AM-like ring modulation is that we often get more harmonics than are required, and often they appear in places we don't want. Sometimes it would be nice if we could obtain only one extra sideband. It would be useful to make a frequency shifter, a patch that would move all the harmonics in a signal up or down by a fixed interval like that shown in figure 20.9.

The *Hilbert transform*, sometimes called the *singular integral*, is an operation that shifts the phase of a signal by 90° or $\frac{\pi}{2}$, and we can write it as $H(f)(t)$ for a function of time, f. So, $H(\sin(t)) = -\cos(t)$. In Pure Data we have an abstraction `hilbert~` that provides two outputs separated in phase by $\frac{\pi}{2}$, called a *quadrature shift*. What it enables us to do is cancel out one of the sidebands when doing modulation. In figure 20.9 we are performing a normal multiplication to get two shifted version of the carrier, an upper and lower sideband, but we also perform this on a quadrature version of the signal. Because of phase shifting the lower sideband in the left branch of the patch will be 180° or π out of phase with the one from the right branch. When we combine the two by subtraction the lower sideband vanishes, leaving only the upper one. The result is seen in figure 20.10, showing that we end up with a pure 550Hz sinusoidal wave after modulating 440Hz and 110Hz signals. Frequency shifting of this kind can be used to create harmony and chorus effects.

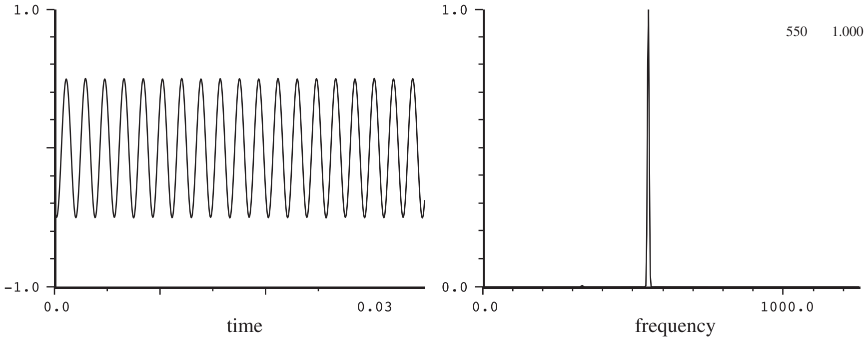

Figure 20.10
Using a Hilbert transform to obtain a single sideband.

SECTION 20.5

Frequency Modulation

Frequency modulation is another way of synthesising complex spectra. When we modulate the frequency of a signal very slowly it's called *vibrato*. As the modulating frequency increases into the audio range it causes new sidebands to appear, a bit like AM. In some ways it is more flexible than AM, in some ways less so. Let's look at a few configurations and spectra to see how it differs and learn where it might be useful for sound design.

Figure 20.11
FM.

Figure 20.11 shows the simplest form of the FM idea. This time we are not multiplying a signal by the modulator, but changing the frequency of another oscillator. The output of the top oscillator connects to the frequency inlet of the bottom one, so the top oscillator is modulating the frequency of the bottom one. As it stands this is a useless patch, but it shows the essential principle.

A more realistic demonstration of FM is shown in figure 20.12. The modulator and carrier are output to left and right channels so we can see their relationship in figure 20.13. This time we provide an offset which sets the *carrier frequency* to 100Hz, and add another signal on top of this. The signal we add is the modulator scaled by a new number which we call the *frequency deviation*. In this case the deviation is 30, so the carrier will wobble around between 70Hz and 130Hz. I've added a number to control the modulator frequency too,

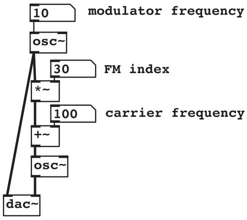

Figure 20.12
Real FM patch.

so we have three parameters to play with in a basic FM patch: the *carrier frequency* (f_c), the *modulation frequency* (f_m), and the *deviation*, which is sometimes called the *FM amount*. The *FM index* is often given as a small number, which is the ratio of the *frequency deviation* (Δf) to the modulation frequency, so $i = \Delta f / f_m$, but it is sometimes given in percent. Strictly it should not be measured in Hertz, but in some of our discussion we will talk about the index as a *frequency deviation*, which isn't really correct, since the unit amplitude of the modulator is 1Hz. Notice in figure 20.13 that the modulator is always positive. The carrier gets squashed and stretched in frequency. Where the modulator is at a maximum or minimum the carrier frequency is a maximum or minimum.

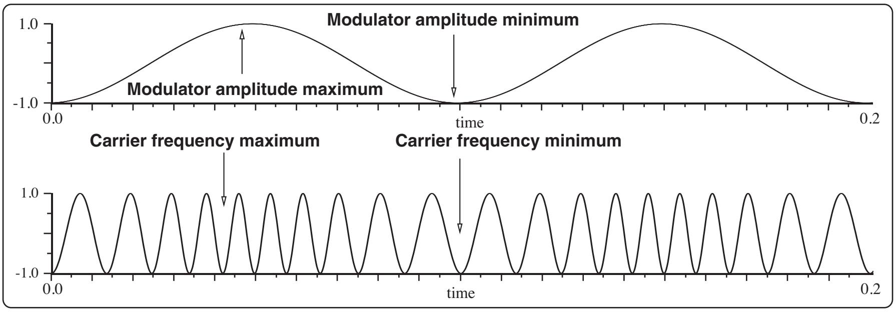

Figure 20.13
FM with a carrier of 100Hz, modulator of 10Hz, and an index of 30Hz.

If you listen to the patch above you will hear an effect more like a fast vibrato. As the modulator frequency increases the wobbling starts to fuse into the carrier frequency, creating a richer timbre. Increasing the index will make the sound brighter. So what is happening to the spectrum?

In figure 20.14 we see the first patch that demonstrates the sidebands introduced by FM. The modulator is 200Hz and the carrier is 600Hz, but the index is zero. On the right in figure 20.14 the only harmonic is the sinusoidal carrier, and the spectrum has a single component at 600Hz.

Keypoint
If the FM index is zero we only get the carrier.

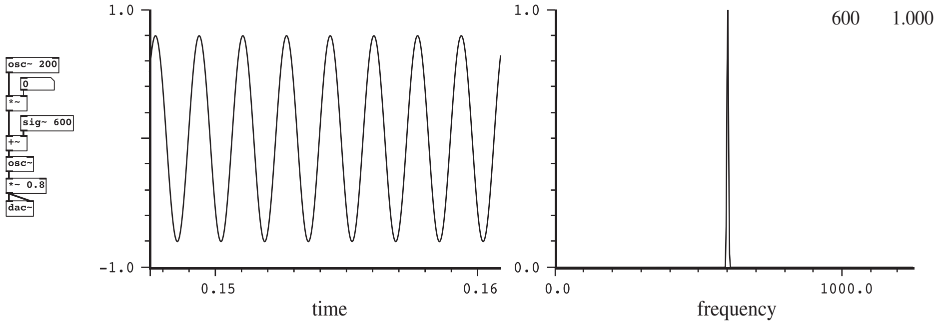

Figure 20.14
FM with a carrier of 600Hz, modulator of 200Hz, and an index of 0Hz.

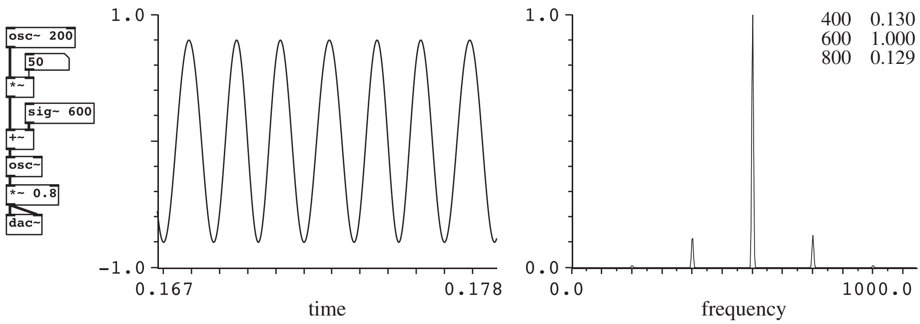

Figure 20.15
FM with a carrier of 600Hz, modulator of 200Hz, and an index of 50Hz.

Now we start to increase the index, adding a 50Hz excursion to either side of the carrier. You can see in figure 20.15 that two sidebands have emerged at 400Hz and 800Hz. At the moment this looks rather like AM with sidebands at $f_c + f_m$ and $f_c - f_m$.

> Keypoint
> In FM, the sidebands spread out on either side of the carrier at integer multiples of the modulator frequency.

What happens as we increase the index further? In figure 20.16 we have a modulation index of 200Hz, and you can see four sidebands. As well as the previous two at 400Hz and 800Hz, we now have two more at 200Hz and 1000Hz

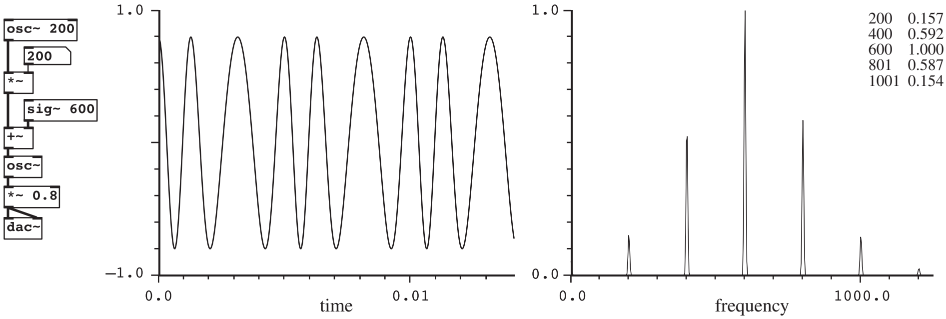

Figure 20.16
FM with a carrier of 600Hz, modulator of 200Hz, and an index of 200Hz.

(ignoring the small FFT error in the plot). Notice the distance between these sidebands.

We can express this result by noting the sidebands are at $f_c + f_m$, $f_c - f_m$, $f_c + 2f_m$, and $f_c - 2f_m$. Is this a general rule that can be extrapolated? Yes, in fact, the formula for FM gives the sidebands as being at integer ratios of the modulator above and below the carrier. As for amplitude modulation, we can see how this arises if we look at some slightly scary-looking equations. Starting with something we already know, a sinusoidal or cosinusoidal wave is a periodic function of time given by

$$f(t) = \cos(\omega t) \tag{20.3}$$

or by

$$f(t) = \sin(\omega t) \tag{20.4}$$

in which ω is the angular frequency and t is time. The value of t is the phasor or increment in our oscillator, and in Pure Data we can basically ignore ω or its expansion to $2\pi f$ because of rotation normalised ranges. We can express the FM process as another similar equation for a new function of time where an extra value is added to the phasor.

$$f(t) = \cos(\omega_c t + f(\omega_m t)) \tag{20.5}$$

The new thing is another function of time. In other words, a new oscillator with angular frequency ω_m. So, let's make that explicit by filling out the new time variant function to get

$$f(t) = \cos(\omega_c t + i \sin(\omega_m t)) \tag{20.6}$$

The value i is the FM index since it scales how much the $\sin(\omega t)$ part affects the outer cosine term. If it is used as a rate of change of increment, then we call the process FM; if it is a change that is merely added to the phase (which is done by rearranging the formula) then we call it PM, meaning *phase modulation*. The two are essentially equivalent, but I will show an example of PM later for completeness. Now, to see what spectrum this gives, a few tricks using trigonometric identities are applied. We use the sum to product (opposite of the previously seen product to sum rule)

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b) \tag{20.7}$$

with

$$\cos(a)\cos(b) = \frac{1}{2}(\cos(a-b) + \cos(a+b)) \tag{20.8}$$

and

$$\sin(a)\sin(b) = \frac{1}{2}(\cos(a-b) - \cos(a+b)) \tag{20.9}$$

and by substitution and expansion obtain the full FM formula

$$\begin{aligned}
& \cos(\omega_c t + i \sin \omega_m t) \\
= \;& J_0(i)\cos(\omega_c t) && (20.10)\\
- \;& J_1(i)(\cos((\omega_c - \omega_m)t) - \cos((\omega_c + \omega_m)t)) && (20.11)\\
+ \;& J_2(i)(\cos((\omega_c - 2\omega_m)t) + \cos((\omega_c + 2\omega_m)t)) && (20.12)\\
- \;& J_3(i)(\cos((\omega_c - 3\omega_m)t) - \cos((\omega_c + 3\omega_m)t)) && (20.13)\\
+ \;& \ldots && (20.14)
\end{aligned}$$

So, you can see where the series of components $f_c \pm nf_m$ comes from, and also note that components are alternately in different phases. But what are the functions $J_0 \ldots J_n$ all about? They are called *Bessel functions* of the first kind. Their appearance is a bit too complicated to explain in this context, but each is a continuous function defined for an integer that looks a bit like a damped oscillation (see fig. 20.17) and each has a different phase relationship from its neighbours. In practice they scale the sideband amplitude according to the modulation index, so as we increase the index the sidebands wobble up and down in a fairly complex way.

Keypoint

The amplitude of the *n*th FM sideband is determined by the *n*+1th Bessel function of the the modulation index.

For small index values, FM provides a regular double sided, symmetrical spectrum much like AM, but instead of only producing the sum and difference it yields a series of new partials that decay away on either side of the carrier. When we say they *decay away*, what does this mean? Well, in fact there are really more partials than we can see. Those at $f_c \pm 3f_m$ are also present, but are too small to be detected. As the index increases they will start to appear much stronger, along with others at $f_c \pm 4f_m$, $f_c \pm 5f_m$, $f_c \pm 6f_m$, and so on.

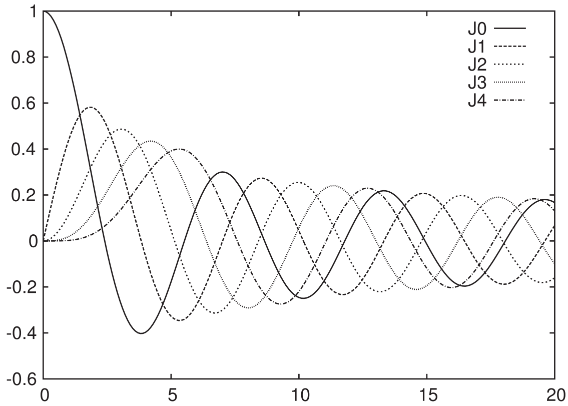

Figure 20.17
The first five Bessel functions of the first kind.

The ones that are loud enough to be considered part of the spectrum, say above -40dB, can be described as the *bandwidth* of the spectrum. As an estimate of the bandwidth you can use Carson's rule, which says the sidebands will extend outwards to twice the sum of the frequency deviation and the modulation frequency, $B = 2(\Delta f + f_m)$.

Another thing to take note of is the amplitude of the time domain waveform. It remains at a steady level. If we had composed this same spectrum additively there would be bumps in the amplitude due to the relative phases of the components, but with FM we get a uniformly "loud" signal that always retains the amplitude of the carrier signal. This is useful to remember for when FM is used in a hybrid method, such as in combination with waveshaping or granular synthesis.

Looking at figure 20.18, we are ready to take a deeper look at FM in order to explain what is happening to the spectrum. It no longer appears to be symmetrical around the carrier, and the regular double-sided decay of the sidebands seems to have changed. For an index greater than 1.0 (when $\Delta f \geq f_m$) we see a new behaviour.

Negative Frequencies

Let's break it down again and look at a simplified FM patch in which the modulation can produce negative frequencies.

What do we mean by a negative frequency? To answer that let's plug some numbers into the patch, setting the first modulating oscillator to 10Hz and making the sweep carrier be 100Hz. In figure 20.19 I have sent the modulator to one output channel and the modulated carrier to the other. Take a look at figure 20.20 where these are shown together. When the amplitude of the 10Hz modulator is 1.0, the frequency of the carrier is 100Hz. This is true at the

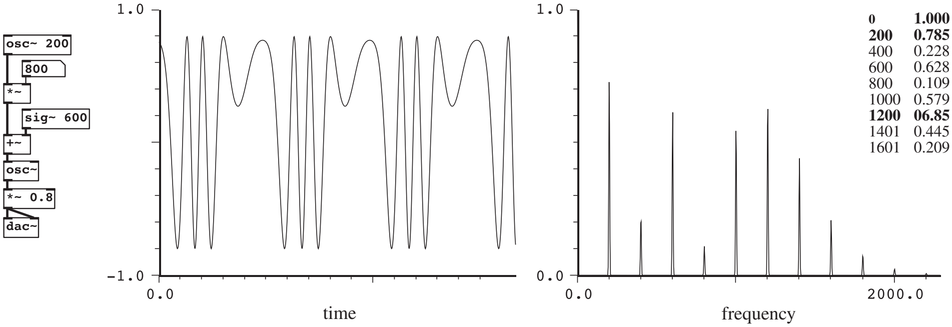

Figure 20.18
FM with a carrier of 600Hz, modulator of 200Hz, and an index of 800Hz.

point where the top waveform hits a maximum, which corresponds to the middle cycle of the first group of three in the bottom trace. When the modulator amplitude is somewhere about halfway the carrier is oscillating at about 50Hz.

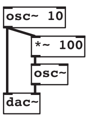

Figure 20.19
Basic FM patch.

It's not easy to pick any point on the lower waveform and say that the oscillator has a precise frequency there, because the modulator is continuously changing its frequency. The result is that carrier becomes distorted, squashed, and then stretched in frequency. You can see what happens as the modulator reaches zero, the carrier reaches a frequency of 0Hz and comes to a halt. But look what happens as the modulator swings negative towards −1.0. The carrier changes direction. It still reaches a frequency of 100Hz when the modulator amplitude hits −1.0, but its phase has flipped.

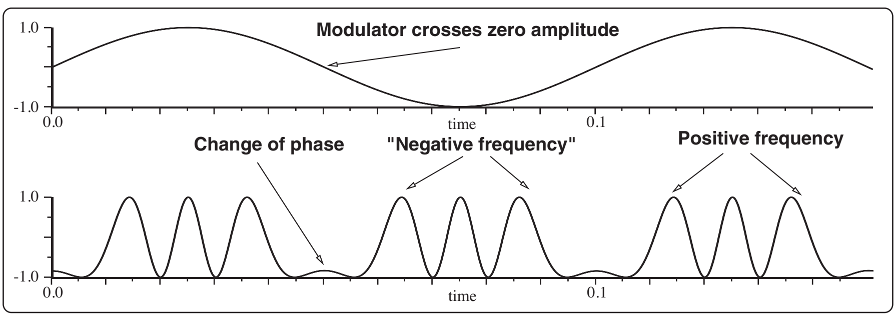

Figure 20.20
Negative frequencies cause a change of phase.

Negative frequencies are folded back into the spectrum with their phase inverted. Much like aliasing caused by frequencies that *fold-over* above the Nyquist point we say ones that get reflected at the bottom *fold-under*. If they combine with real, positive phase components, they cancel out, so we start to get holes in the spectrum.

Phase Modulation

If you compare figure 20.21 to figure 20.12 the similarities should be obvious. But ponder the subtle difference for a moment and think about the FM formula.

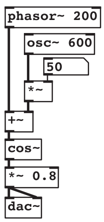

Figure 20.21
Phase modulation.

Instead of supplying a steady signal via sig~ to a osc~ oscillator that already contains a phase increment we have a separate phasor~ which indexes a cos~ function. This does exactly the same thing as the combined oscillator. But instead of changing the carrier frequency we are adding a new time variant signal to the phase. Since a change in the rate of change of phase is the same as a change in frequency, we are doing the same thing as FM. However, we have the advantage that the phase accumulator is available separately. This means we can derive other time variant functions from it which will maintain the same overall phase coherence. The upshot is to greatly simplify the design of complex FM patches in which we have more than one modulator signal combined.

Keypoint
"Negative frequencies" produce harmonics inverted in phase.

References

Carson, J. R. (1922) "Notes on the theory of modulation." *Proc. IRE* 10, no. 1: 57–64.

Chowning, J. M. (1973) "The synthesis of complex audio spectra by means of frequency modulation." *J. Audio Eng. Soc.* 21: 526–534.

Chowning, J., and Bristow, D. (1986). *FM Theory and Applications by Musicians for Musicians.* Yamaha Music Foundation, Tokyo.

Truax, B. (1977). "Organizational techniques for C:M ratios in frequency modulation." *CMJ* 1–4: 39–45.