CMPT 468: Lecture 7
Frequency Modulation (FM) Synthesis

Tamara Smyth, tamaras@cs.sfu.ca
School of Computing Science,
Simon Fraser University

January 26, 2009

Linear Frequency Modulation (FM)

- Till now we've seen signals that do not change in frequency over time. How do we modify the signal to obtain a time-varying frequency?
- A **chirp** signal is one that sweeps linearly from a low to a high frequency.
- Can we create such a signal by concatenating small sequences, each with a frequency that is higher than the last?
- This approach will likely lead to problems lining up the phase of each segment so that discontinuities aren't introduced in the resulting waveform (as seen below).

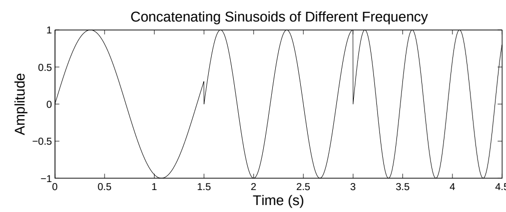

Figure 1: A signal made by concatenating sinusoids of different frequencies will result in discontinuities if care is not taken to match the initial phase.

Chirp Signal

- A better approach is to modify the equation for the sinusoid so that the frequency is time-varying.
- Recall that the original equation for a sinusoid is given by
$$x(t) = A\cos(\omega_0 t + \phi)$$
where the instantaneous phase, given by $(\omega_0 t + \phi)$, changes *linearly with time*.
- Notice that the time derivative of the phase is the radian frequency of the sinusoid ω_0, which in this case is a constant.
- More generally, if
$$x(t) = A\cos(\theta(t)),$$
the instantaneous frequency is given by
$$\omega(t) = \frac{d}{dt}\theta(t).$$

Chirp Signal cont.

- Now, let's make the phase *quadratic*, and thus non-linear, rather than linear with respect to time.
$$\theta(t) = 2\pi\mu t^2 + 2\pi f_0 t + \phi.$$
- The instantaneous radian frequency, which is the derivative of the phase θ, now becomes
$$\omega_i(t) = \frac{d}{dt}\theta(t) = 4\pi\mu t + 2\pi f_0$$
which in Hz becomes
$$f_i(t) = 2\mu t + f_0.$$
- Notice the frequency is no longer a constant but is changing linearly in time.
- To create a sweeping frequency from f_1 to f_2 therefore, we need only look at the equation for a line $y = mx + b$ to obtain a formula for the instantaneous frequency:
$$f(t) = \frac{f_2 - f_1}{T}t + f_1,$$
where T is the duration of the sweeping signal.

Sweeping Frequency

- We can then use this time varying value for the instantaneous frequency in our original equation for a sinusoid,

$$x(t) = A\cos(2\pi f(t)t + \phi).$$

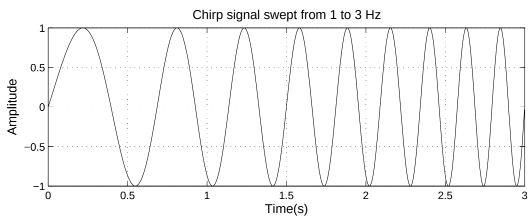

Figure 2: A chirp signal from 1 to 3 Hz.

- If the instantaneous phase $\theta(t)$ is constant, the frequency is zero.
- If $\theta(t)$ is linear, the frequency is fixed (constant).
- if $\theta(t)$ is quadratic, the frequency changes linearly with time.

Vibrato simulation

- Vibrato is a term used to describe a wavering of pitch.
- Vibrato occurs very naturally in the singing voice (though some may say a little exaggerated in some operatic performances), and in instruments where the musician has control after the note has been played (such as the violin, wind instruments, the theremin, etc.).
- In Vibrato, the frequency does not change linearly (as our last chirp signal example) but rather sinusoidally, creating a sense of a wavering pitch.
- Since the instantaneous frequency of the sinusoid is the derivative of the instantaneous phase, and the derivative of a sinusoid is a sinusoid, we merely apply a sinusoidal signal to the instantaneous phase of a carrier signal to create vibrato.

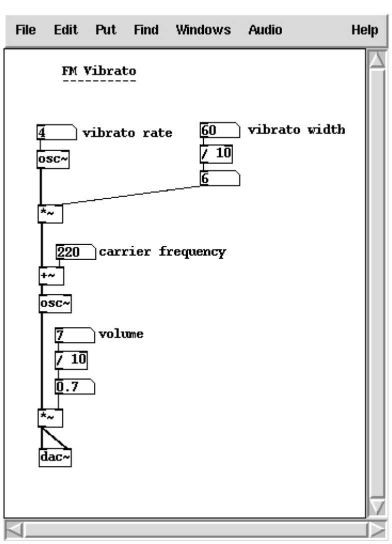

Figure 3: A simple FM vibrato instrument in Pd.

Vibrato cont.

- We can therefore use FM synthesis to create a vibrato effect, where the instantaneous frequency of the carrier oscillator varies over time according to the parameters controlling
  - the *width* of the vibrato (the deviation from the carrier frequency)
  - the *rate* of the vibrato.
- The *width* of the vibrato is determined by the amplitude of the modulating signal.
- The *rate* of vibrato is determined by the frequency of the modulating signal.
- In order for the effect to be perceived as vibrato, the vibrato rate must be below the audible frequency range and the width made quite small.

FM Synthesis of Musical Instruments

- When the vibrato rate is in the audio frequency range, and when the width is made larger, the technique can be used to create a broad range of distinctive timbres.
- Frequency modulation (FM) synthesis was invented by John Chowning at Stanford University's *Center for Computer Research in Music and Acoustics* (CCRMA).
- FM synthesis uses fewer oscillators than either additive or AM synthesis to introduce more frequency components in the spectrum.
- Where AM synthesis uses a signal to modulate the *amplitude* of a carrier oscillator, FM synthesis uses a signal to modulate the *frequency* of a carrier oscillator.

Frequency Modulation

- The general equation for an FM sound synthesizer is given by

  $$x(t) = A(t)\left[\cos(2\pi f_c t + I(t)\cos(2\pi f_m t + \phi_m) + \phi_c\right],$$

  where

  $$\begin{aligned} A(t) &\triangleq \text{ the time varying amplitude} \\ f_c &\triangleq \text{ the carrier frequency} \\ I(t) &\triangleq \text{ the modulation index} \\ f_m &\triangleq \text{ the modulating frequency} \\ \phi_m, \phi_c &\triangleq \text{ arbitrary phase constants.} \end{aligned}$$

Modulation Index

- The function $I(t)$, called the *modulation index envelope*, determines significantly the harmonic content of the sound.
- Given the general FM equation

  $$x(t) = A(t)\left[\cos(2\pi f_c t + I(t)\cos(2\pi f_m t + \phi_m) + \phi_c\right],$$

  the instantaneous frequency $f_i(t)$ (in Hz) is given by

  $$\begin{aligned} f_i(t) &= \frac{1}{2\pi}\frac{d}{dt}\theta(t) \\ &= \frac{1}{2\pi}\frac{d}{dt}\left[2\pi f_c t + I(t)\cos(2\pi f_m t + \phi_m) + \phi_c\right] \\ &= \frac{1}{2\pi}[2\pi f_c - I(t)\sin(2\pi f_m t + \phi_m)2\pi f_m + \\ &\qquad \frac{d}{dt}I(t)\cos(2\pi f_m t + \phi_m)] \\ &= f_c - I(t) f_m \sin(2\pi f_m t + \phi_m) + \\ &\qquad \frac{1}{2\pi}\frac{d}{dt}I(t)\cos(2\pi f_m t + \phi_m). \end{aligned}$$

Modulation Index cont.

- Without delving into rigorous mathematics, it is possible to determine the relationship of the modulation index $I(t)$ to the harmonic content.
- Given the result for instantaneous frequency

  $$\begin{aligned} f_i(t) &= f_c - I(t) f_m \sin(2\pi f_m t + \phi_m) + \\ &\qquad \frac{dI(t)}{dt}\cos(2\pi f_m t + \phi_m)/2\pi, \end{aligned}$$

  we may see that if $I(t)$ is a constant (and it's derivative is zero), the third term goes away and the instantaneous frequency becomes

  $$f_i(t) = f_c - I(t) f_m \sin(2\pi f_m t + \phi_m).$$

- Notice now that in the second term, the quantity $I(t)f_m$ multiplies a sinusoidal variation of frequency f_m, indicating that $I(t)$ determines the maximum amount by which the instantaneous frequency deviates from the carrier frequency f_c.
- Since $I(t)$ is a function of time, the harmonic content, and thus the timbre, of the synthesized sound may vary with time.

FM Sidebands

- The upper and lower sidebands produced by FM are grouped in pairs according to the harmonic number of f_m, that is, frequencies present are given by $f_c \pm k f_m$.

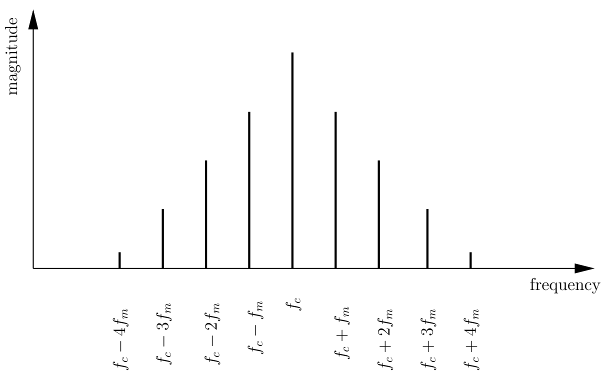

Figure 4: Sidebands produced by FM synthesis.

Determining the Modulation index

- In *Computer Music*, the modulation index I is given by

$$I = \frac{d}{f_m}$$

  where d is the amount of frequency deviation produced by the modulating oscillator.
- When $d = 0$, the index I is also zero, and no modulation occurs. Increasing d causes the sidebands to acquire more power at the expense of the power in the carrier frequency.
- The deviation d can therefore act as a control on FM bandwidth.

Bessel Functions of the First Kind

- The amplitude of the k^{th} sideband is given by $J_k(I)$, where J_k is a Bessel function[1] of the first kind, of order k.

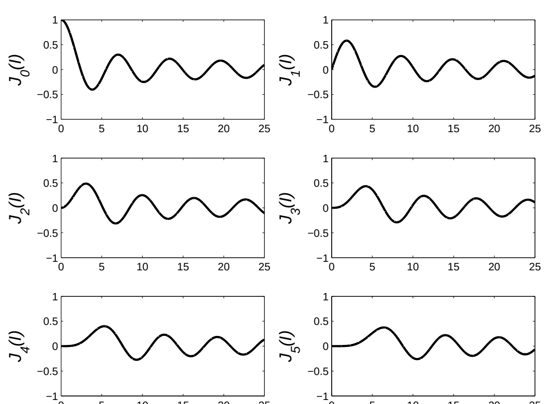

Figure 5: Bessel functions of the first kind, plotted for orders 0 through 5.

- From this plot we see that higher order Bessel functions, and thus higher order sidebands, do not have significant amplitude when I is small.

[1] Bessel functions are solutions to Bessel's differential equation.

- Higher values of I therefore produce higher order sidebands.
- In general, the highest-ordered sideband that has significant amplitude is given by the approximate expression $k = I + 1$.

Odd-Numbered Lower Sidebands

- The amplitude of the odd-numbered lower sidebands is the appropriate Bessel function multipled by -1, since odd-ordered Bessel functions are **odd functions**. That is
$$J_{-k}(I) = -J_k(I).$$

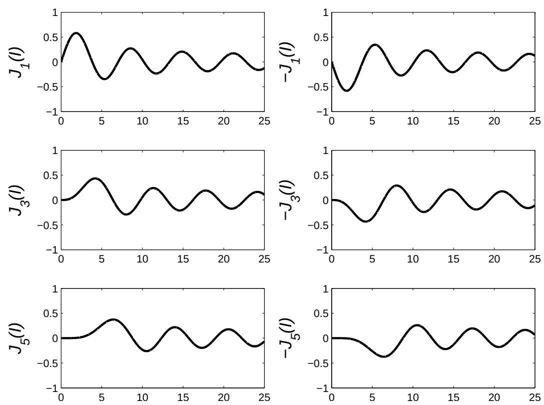

Figure 6: Bessel functions of the first kind, plotted for odd orders.

Effect of Phase in FM

- The phase of a spectral component does not have an audible effect unless other spectral components of the same frequency are present. In the case of frequency overlap, the amplitudes will either add or subtract and the tone of the sound will change as a result.
- If the FM spectrum contains frequency components below 0 Hz, they are folded over the 0 Hz axis to their corresponding positive frequencies.
- The act of folding reverses the phase. A sideband with a negative frequency is equivalent to a component with the corresponding positive frequency but with the opposite phase.

FM Spectrum

The frequencies present in a simple FM spectrum are $f_c \pm k f_m$, where k is an integer greater than zero. The carrier frequency component is at $k = 0$.

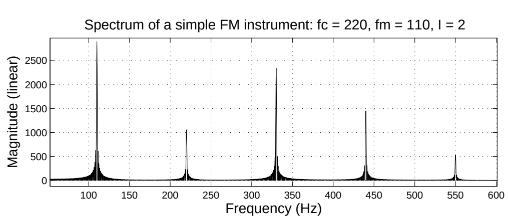

Figure 7: Spectrum of a simple FM instrument, where $f_c = 220$, $f_m = 110$, and $I = 2$.

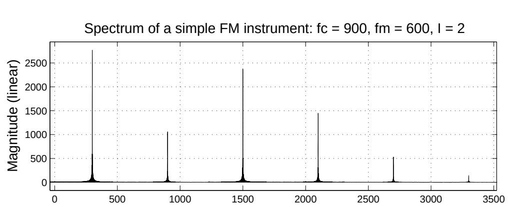

Figure 8: Spectrum of a simple FM instrument, where $f_c = 900$, $f_m = 600$, and $I = 2$.

Fundamental Frequency in FM

- In determining the fundamental frequency of your FM sound, it is useful to define the following ratio:
$$\frac{f_c}{f_m} = \frac{N_1}{N_2}$$
where N_1 and N_2 are integers with no common factors.
- The fundamental frequency is then given by
$$f_0 = \frac{f_c}{N_1} = \frac{f_m}{N_2}.$$
- As in the previous plot for example, a carrier frequency $f_c = 220$ and modulator frequency $f_m = 110$ yields the ratio of
$$\frac{f_c}{f_m} = \frac{220}{110} = \frac{2}{1} = \frac{N_1}{N_2}.$$
and a fundamental frequency of
$$f_0 = \frac{220}{2} = \frac{110}{1} = 110.$$
- Likewise the ratio of $f_c = 900$ to $f_m = 600$ is 3:2 and the fundamental frequency is given by
$$f_0 = \frac{900}{3} = \frac{600}{2} = 300.$$

- If $N_2 = M$ where M is an integer greater than 1, then every M^{th} harmonic of f_0 is missing in the spectrum.
- This can be seen in the plot below where the ratio of the carrier to the modulator is 4:3 and $N_2 = 3$. Notice the fundamental frequency f_0 is 100, but every third multiple of f_0 is missing from the spectrum.

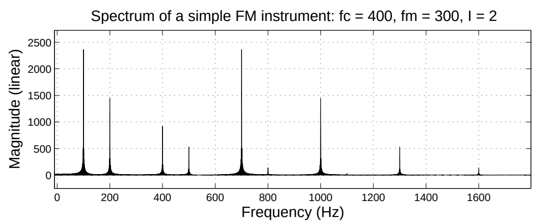

Figure 9: Spectrum of a simple FM instrument, where $f_c = 400$, $f_m = 300$, and $I = 2$.

Some FM instrument examples

- When implementing simple FM instruments, we have several basic parameters that will effect the overall sound:
  1. The duration,
  2. The carrier and modulating frequencies
  3. The maximum (and in some cases minimum) modulating index scalar
  4. The envelopes that define how the amplitude and modulating index evolve over time.
- Using the information taken from John Chowning's article on FM (details of which appear in the text *Computer Music* (pp. 125-127)), we may develope envelopes for the following simple FM instruments:
  – bell-like tones,
  – wood-drum
  – brass-like tones
  – clarinet-like tones

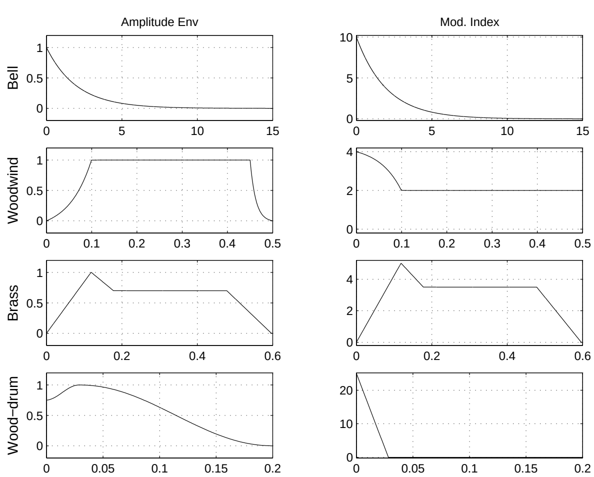

Figure 10: Envelopes for FM bell-like tones, wood-drum tones, brass-like tones and clarinet tones.

Formants

- Another characteristic of sound, in addition to its spectrum, is the presence of formants.
- The formants describe certain regions in the spectrum where there are strong resonances (where the amplitude of the spectral components is considerably higher).
- We may view formants as the peaks in the spectral envelope.
- As an example, pronounce aloud the vowels "a", "e", "i", "o", "u" while keeping the same pitch for each. Since the pitch is the same, we know the integer relationship of the spectral components is the same.
- The formants are what allows us to hear a difference between the vowel sounds.

Two Carrier Oscillators

- In FM synthesis, the peaks in the spectral envelop can be controlled using an additional carrier oscillator.
- In the case of a single oscillator, the spectrum is centered around a carrier frequency. With an additional oscillator, an additional spectrum may be generated that is centered around a formant frequency.
- When the two signals are added, their spectra are combined.
- If the same oscillator is used to modulate both carriers (though likely using seperate modulation indeces), and the formant frequency is an integer multiple of the fundamental, the spectra of both carriers will combine in such a way the the components will overlap, and a peak will be created at the formant frequency.

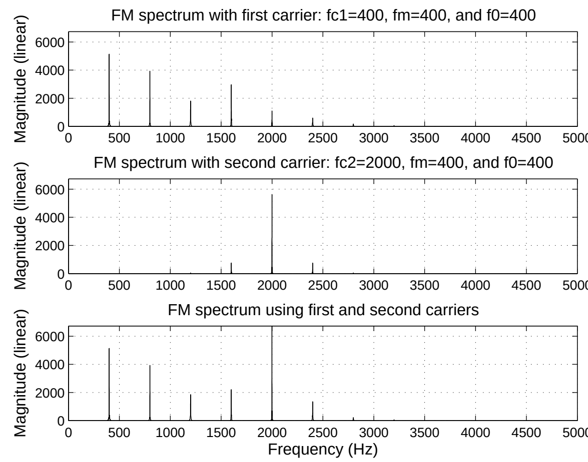

Figure 11: The spectrum of individiual and combined FM signals.

Two Carriers cont.

- In Figure 11, both carriers are modulated by the same oscillator with a frequency f_m.
- The index of modulation for the first and second carrier is given by I_1 and I_2/I_1 respectively.
- The value I_2 is usually less than I_1, so that the ratio I_2/I_1 is small and the spectrum does not spread too far beyond the region of the formant.
- The frequency of the second carrier f_{c2} is chosen to be a harmonic of the fundamental frequency f_0 that is close to the desired formant frequency f_f (from *Computer Music*). That is

$$f_{c2} = nf_0 = \text{int}(f_f/f_0 + 0.5)f_0.$$

- This ensures that the second carrier frequency remains harmonically related to f_0. If f_0 changes, the scond carrier frequency will remain as close as possible to the desired formant frequency f_f while remaining an integer multiple of the fundamental frequency f_0.

Two Modulating Oscillators

- Just as the number of carriers can be increased, so can the number of modulating oscillators.
- To create even more spectral variety, the modulating waveform may consist of the sum of several sinusoids.
- If the carrier frequency is f_c and the modulating frequencies are f_{m1} and f_{m2}, then the resulting spectrum will contain components at the frequency given by $f_c \pm if_{m1} \pm kf_{m2}$, where i and k are integers greater than or equal to 0.
- For example, when $f_c = 100$ Hz, $f_{m1} = 100$ Hz, and $f_{m2} = 300$ Hz, the spectral component present in the sound at 400 Hz is the combination of sidebands given by the pairs: $i = 3, k = 0; i = 0, k = 1; i = 3^-, k = 2$; and so on (see Figure 12).
- In fact, there are an infinite number of i, k dyads that produce a sideband at ± 400 Hz. The actual number that contribute to the overall amplitude is determined by the modulation indeces.

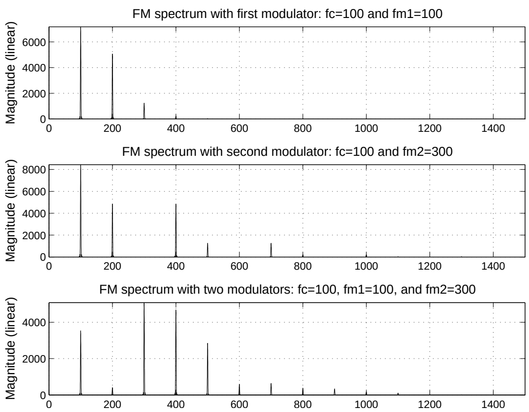

Figure 12: The FM spectrum produced by a modultor with two frequency components.

Two Modulators cont.

- Modulation indeces are defined for each component: I_1 is the index that charaterizes the spectrum produced by the first modulating oscillator, and I_2 is that of the second.
- The amplitude of the i^{th}, k^{th} sideband ($A_{i,k}$) is given by the product of the Bessel functions

$$A_{i,k} = J_i(I_1)J_k(I_2).$$

- Like in the previous case of a single modulator, when i, k is odd, the Bessel functions assume the opposite sign. For example, if $i = 2$ and $k = 3^-$ (where the negative subscript means that k is subtracted), the amplitude is $A_{2,3} = -J_2(I_1)J_3(I_2)$.
- In a harmonic spectrum, the net amplitude of a component at any frequency is the combination of many sidebands, where negative frequencies "foldover" the 0 Hz bin (*Computer Music*).