**Linear Algebra in FM Synthesis**

**What Is Sound?**

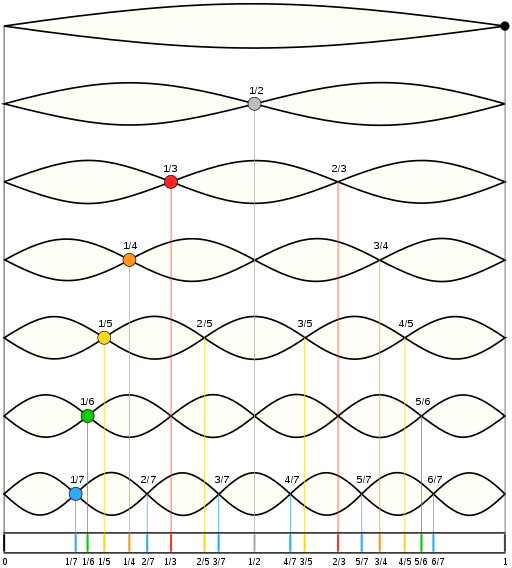
All sounds are vibrations traveling through the air as sound waves. Sound waves are caused by the vibrations of objects and radiate outward from their source in all directions. A vibrating object compresses the surrounding air molecules (squeezing them closer together) and then rarefies them (pulling them farther apart). Although the fluctuations in air pressure travel outward from the object, the air molecules themselves stay in the same average position. As sound travels, it reflects off objects in its path, creating further disturbances in the surrounding air. When these changes in air pressure vibrate your eardrum, nerve signals are sent to your brain and are interpreted as sound.

**What is Pitch?**

Pitch is an auditory sensation in which a listener assigns musical tones to relative positions on a musical scale based primarily on their perception of the frequency of vibration. Pitch is closely related to frequency, but the two are not equivalent. Frequency is an objective, scientific attribute that can be measured. Pitch is each person's subjective perception of a sound wave, which cannot be directly measured. However, this does not necessarily mean that most people won't agree on which notes are higher and lower.

Sound waves themselves do not have pitch, but their oscillations can often be characterized in terms of frequency. Pitches are usually associated with, and thus quantified as frequencies in cycles per second, or hertz, by comparing sounds with pure tones, which have periodic, sinusoidal waveforms. Complex and aperiodic sound waves can often be assigned a pitch by this method.

**Application**

There are many devices that we commonly use that are able to interpret a sound wave, and decompose it into its essential elements. One of the most relatable examples is the smart phone. The principal that is used is called Fourier series. Your smart phone not only has the ability to record your voice, encode it, transmit, and decode back to sound for the person on the other end, but it can also recognize your voice (think of Siri and Ok Google), while negating verbal commands given by other individuals. It does this by analyzing the fundamental frequency and formants of your voice. One may ask, what does these terms mean? To break it down in laymen terms, a fundamental frequency is the lowest frequency produced by any particular instrument, voice, or sound. This fundamental frequency is the core of a certain pitch we perceive. For example, if I strike an A key on a piano, it produces a fundamental frequency of 440hz. Now a piano is a rather complex sound, since it consists not only of the strings vibrations, but also the sound of the hammers, mechanical key action, and sympathetic resonance of the other strings in the piano. A pure fundamental sound would be a sine wave. Ever take a hearing test? That is the sound of a sine wave; very smooth with no harmonics. This leads to another important definition, which we must first establish before one can understand what a formant is. A harmonic is any member of the harmonic series, a divergent infinite series. Its name derives from the concept of overtones, or harmonics in musical instruments: the wavelengths of the overtones of a vibrating string or a column of air (as with a tuba) are derived from the string's (or air column's) fundamental wavelength. Essentially, they are other frequencies above the fundamental that resonate and help build a complex sound like a guitar string. Certain sounds like organs have odd harmonics while others have even.

Finally, we can now define what a formant is. Formants are areas of emphasis or attenuation in the frequency spectrum of a sound that are independent of the pitch of the fundamental note but are found always in the same frequency ranges. They are characteristic of the tone color or ‘timbre’ of each sound source. This is where the magic is.

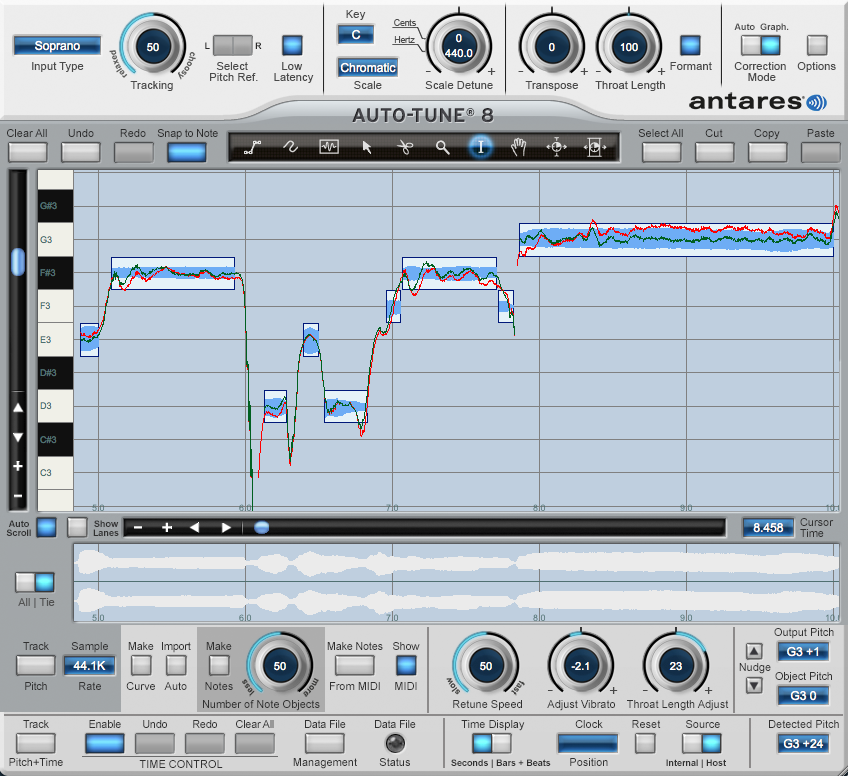
For example, say I have a piano, a flute, and a bell. For sake of simplicity, each one is tuned to the musical note of A4. This means that each instrument has a fundamental frequency of 440hz. If one has a musical ear, and I play each instrument, you will be able to immediately tell that they are all indeed playing the note A440. Our ears perceive the same pitch, but how do we distinguish the different sounds of each instrument? Obviously a piano, guitar, and bell do not sound the same, but very different. Even though they are all playing the same musical note, our ears can easily distinguish these different instruments. How? Formants. Formants consist of all the harmonics above the fundamental. These create the tone or timbre which helps us easily differentiate each instrument. The same principle applies to the human voice. If your mother or father is speaking to you, you can easily tell who is who. A smart phone works in the same way by sampling your voice, encoding it into digital form, and then performing filtering and other signal processing methods along with the Fourier series to match up its record that it indeed you speaking to it.

**The Fourier Transform**

The principle of the Fourier transform is that any signal, such as the sound produced by a musical instrument, such as the piano, violin, trumpet, or drum and any sound recording can be represented as the sum of a collection of sine and cosine waves with different frequencies and amplitudes. This collection of waves can then be manipulated with relative ease—for example, allowing a recording to be compressed or noise to be suppressed. This Fourier decomposition lies at the heart of modern electronic music; a synthesizer combines pure sine and cosine tones to reproduce the diverse sounds of instruments, both natural and artificial, according to Fourier’s general preposition. If anyone has ever witnessed the sheer marvel of the tiny size of an MP3 file compared with the same recording in an uncompressed form has seen the power of the Fourier transform at work.

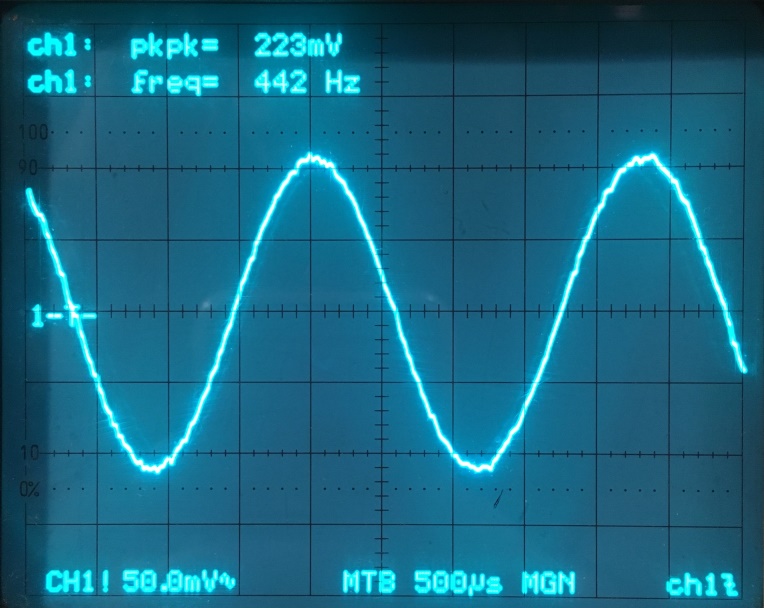
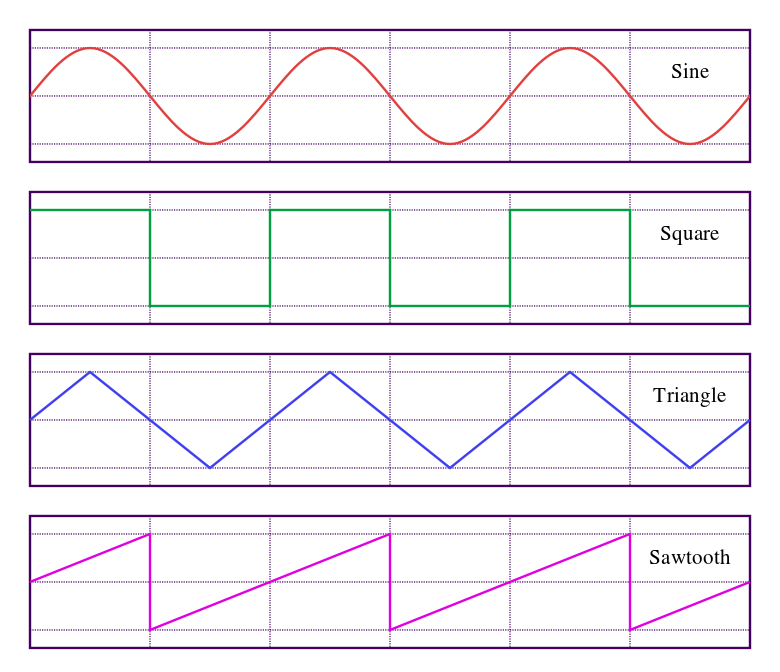
The algorithms that are used to sample your voice take advantage of something called a DAC, which converts the analog signal recorded from the microphone of your phone to a digital one. Your voice is “sampled” meaning that for each peak in volume and change in frequency, a matrix is created with an integer 0-9 to represent each value. Of course since time is linear, and processors only run at certain clock frequencies, only certain intervals of the audio transmission are sampled. You can easily hear this when talking on the phone. Have you ever noticed how it sounds very muffled, and there are no low frequencies present as well? This is because phones are only sampled at 8khz. Your ears can roughly receive and interpret frequencies from 20hz up to 20,000hz, but phones only encode 300hz to 3000hz. Isn’t it a miracle that you can still fully understand someone’s conversation on the phone? And for the most part always be able to tell who they are? That is due to the fact that enough formants fall in that range for our brains to be able to still interpret it. Since cellular bandwidth is limited, the Fourier series makes it possible to only sample and transmit the audio data necessary for us to hear.

**Pitch Correction**

Auto-tune works in the same principle, with algorithms analyzing the pitch of your voice and mapping it over time. Auto tune was originally created by Exxon engineer Andy Hildebrand who developed methods for interpreting seismic data in 1979. In 1990, Hildebrand (now retired) created his final company Antares Audio technologies, where he developed programs to sample orchestra sounds for synthesizers. In 1995, Hildebrand realized he could use his seismic analysis algorithms to analyze, detect, and modify the fundamental pitch of audio files. The algorithms used implemented the Fourier series and matrices. Hildebrand built Antares Auto Tune in 1996 and released it in the spring of 1997. It was originally designed and intended to be used to analyze the pitch and note from a vocal recording of a singer, and it allow it to be “tuned” and corrected to either a scale, or note input through midi. This would allow an audio engineer in the studio to edit the vocal track and correct all the mistakes (sharp or flat notes, also time errors) and make the vocal track sound perfect. This would cut down on time and the amount of money spent on studio usage.

**What is a Synthesizer?**

A synthesizer is an electronic musical instrument that generates electric signals that are converted to sound through instrument amplifiers and loudspeakers or headphones. Synthesizers may either imitate traditional musical instruments like piano, Hammond organ, flute, vocals; natural sounds like ocean waves, etc.; or generate novel electronic timbres. They are often played with a musical keyboard, but they can be controlled via a variety of other input devices. They are used in music everywhere, and would be easily recognized to even the most inexperienced listener. Synthesizers became popular in the 70s and especially the 80s, were used in anything from disco, to pop, progressive rock, reggae, funk, jazz, movie soundtracks, 90s, 2000s, and continuously in modern times. Any song you can think of that had “electronic” sounds (even Justin Timberlake, Brittany Spears, and Bruno Mars) was created with a synthesizer. Even the sub-bass in rap is just a simple sine wave. Synths have forever influenced and shaped music to what it is today.

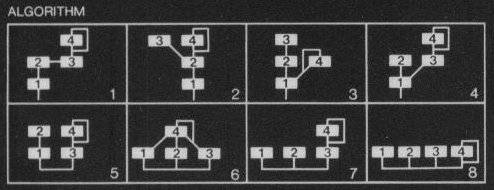
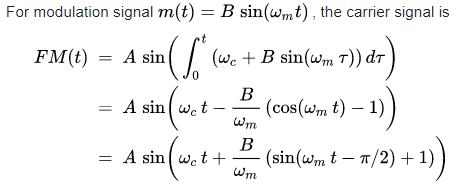
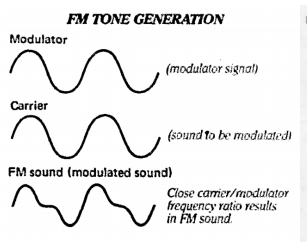
**Waveform Visual Representations**

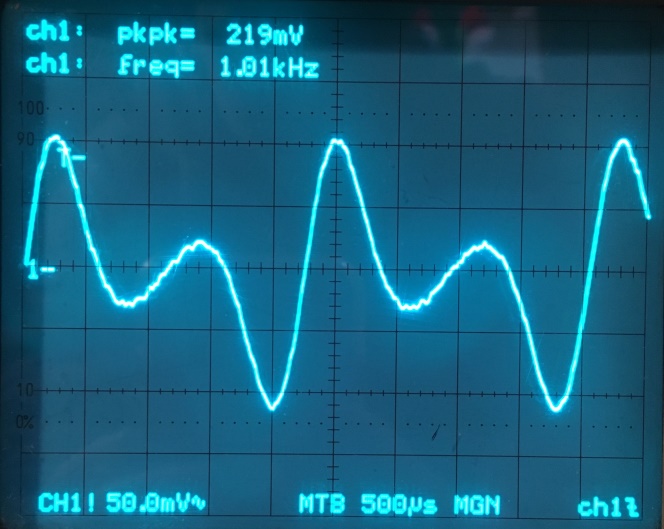
**What is FM Synthesis?**

There are many different types of synthesis. Original synthesizers were fully analog, meaning they used discrete electronic components to generate and manipulate voltages to create sound. They would create certain waveforms such as a saw, square, pulse, triangle, and sine, and then used filters and envelope generators to shape the sound. You essentially would start with a sound, and filter out what you don’t want. The filter would remove harmonics starting from the highest down to the lowest. As technology further emerged and digital began to take over, frequency modulation (FM) was born.

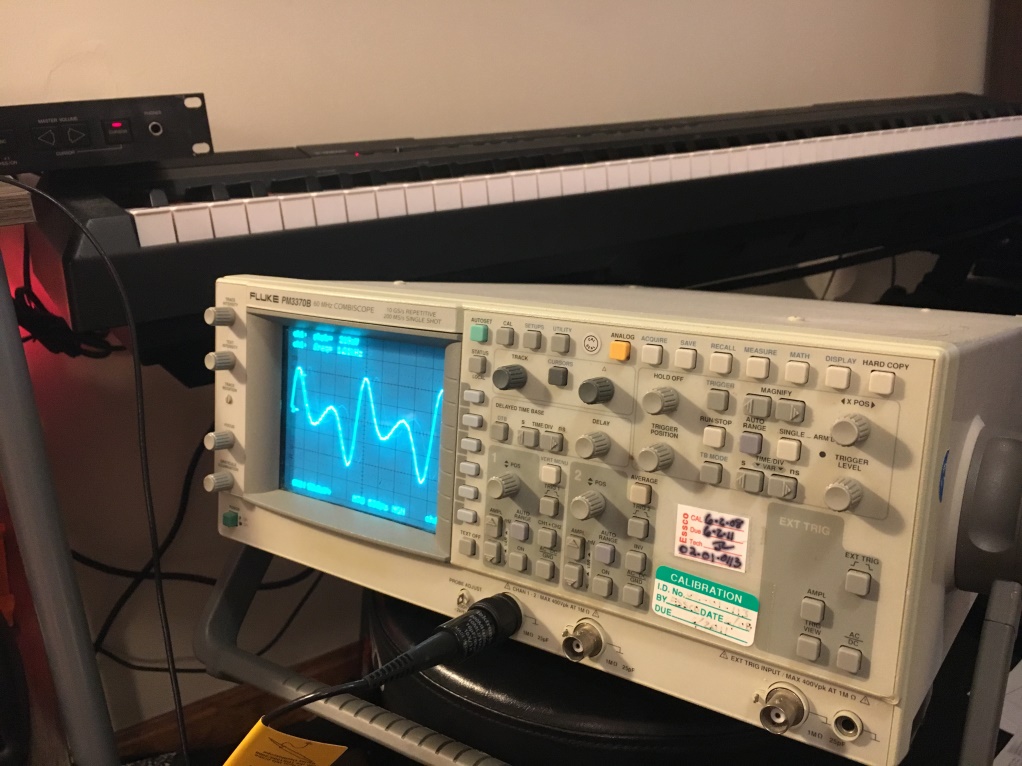
In audio and music, frequency modulation synthesis (or FM synthesis) is a form of audio synthesis where the timbre of a simple waveform (most commonly a sine wave) is changed by modulating its frequency with a modulator frequency that is also in the same or similar audio range (within the range of human hearing: 20hz to 20khz), resulting in a more complex waveform and a different-sounding tone that can also be described as "gritty" if it is a thick and dark timbre. The frequency of an oscillator is altered or distorted in accordance with the amplitude of a modulating signal. In FM synthesis, instead of using analog oscillators, digital operators are used to produce carrier and modulator waves. The particular synthesizer I will be relating to is a Yamaha Tx81z from 1987, which has 4 operators, and 8 algorithms with a feedback loop that can be arranged to program a vast platter of sounds.

In FM synthesis, an operator is essentially a digital oscillator that produces a sine wave. (This synth includes other digital waveforms but to keep the explanation simple we will stick to sine). Frequency modulation is the change of instantaneous frequency of one signal (carrier) in accordance with another signal (modulator). The equation Y = Ac sin(2πfct - Δf/fm cos(2πfmt)) represents the mathematical function where Ac is the Peak carrier amplitude, fc is the Carrier frequency, fm is the Modulator frequency, and Δf is the Peak carrier frequency deviation.

To give a visual representation of this, we will only use operators 1 and 2 of algorithm 1. If we press a440 on the keyboard and set the frequency ratios of both operator 1 and 2 to 1.00, the carrier wave (a 440hz sine wave) will be modulated by another 440hz sine wave. Although the modulator cannot be heard, increasing its amplitude effectively modulates the carrier more, producing more harmonic content. This can be seen by following diagram:



Example of a modulator of 440hz modulating a 440hz   
carrier wave with 68.75% of the amplitude the carrier.   
This produces a rather nasily metallic sound.

In this case you can see the results of FM produced by an FM Matrix, in which the following algorithm was constructed by modifying modulation indices and output levels. Matrices are used to reflect every possible frequency division and operation that can be used with the algorithms. In addition to these operations, the synth also includes a modulation matrix that consists of several low frequency oscillators and multiple destinations. Any modulation source can be mapped to any parameter using this matrix, sending the outputs of the appropriate functions and equations to their destinations where they are additively computed with each other. This creates a diverse range of sonic content which makes this form of synthesis so flexible and diverse.

**Tools Used**

* Synthesizer: Yamaha Tx81Z
* Oscilloscope: Fluke PM3370B
* Music Production Software: Cubase 8

