

# Digital Music Synthesizer

Hanhee Lee

December 4, 2024

## Abstract

A digital music synthesizer (DMS) is a device that create new waveforms (i.e. sound) using digital signal processing techniques (DSPT) [1]. This paper provides a walk-through of using a DMS to create new waveforms. This includes converting the analog signal to digital using an analog to digital converter (ADC), applying various DSPT to modify the waveform, and converting the processed digital signal to analog using a digital to analog converter (DAC).

## 1 Introduction

The development of the DMS began in the mid-1900s, changing the way that sound is created and manipulated [2]. These digital instruments can create an infinite range of sounds, expanding the creative freedom of users in different industries such as gaming, music, movies, etc. In order to understand how the DMS works, it's important to define what a sound is.

A sound is anything that can be heard as a result of vibrations in the air [3]. To differentiate between sounds, there are three components:

1. **Amplitude (Loudness):** Corresponds to the height of the waveform.
2. **Frequency (Pitch):** Corresponds to the number of cycles that occur in a period.
3. **Timbre (Quality):** Corresponds to the number and type of overtones (i.e. harmonics of the fundamental frequency), and their relative strength to one another.

These components can be precisely controlled in the DMS to obtain a desired sound. The DMS process is as follows (see Figure 1) [4]:

1. **Digital Oscillator:** Generates digital audio signals using an analog to digital converter, which form the basic waveforms before being synthesized using different DSPT [3].
2. **Additive Synthesis:** Adding waveforms together to form new sounds [5].
3. **Modulation** Change various sound parameters using low-frequency oscillators (LFOs), which modulate parameters using slow and periodic signals, and frequency modulation (FM), which modulate the frequency of a carrier waveform using another waveform [6].

4. **ADSR Envelope:** The sound are shaped using an envelope generator, which consists of four stages attack, decay, sustain, and release so that it modulates parameters such as amplitude and pitch over time [7].
5. **Filters:** Filters modify the timbre of the waveforms by attenuating or amplifying specific frequency components using subtractive synthesis [7]. Common types include low-pass filters (LPF), high-pass filters (HPF), band-pass filters (BPF), and band-stop filters (BSF).
6. **Digital to Analog Converter:** Convert the processed digital signal back into an analog format by using the Sampling Theorem to prevent aliasing [8, 9].

Ultimately, sound is processed digitally by converting the signal into the discrete time (DT) system to be processed through the digital music synthesizer before getting converted back into analog format using an ADC. Note, some of the following concepts will be described in the continuous time (CT) domain. However, DMS uses digital sequences, therefore, these DSPT will actually be in DT. It is assumed in this paper that the DSPT will result in samples,  $y(nT_s)$ , but the specifics will not be discussed.

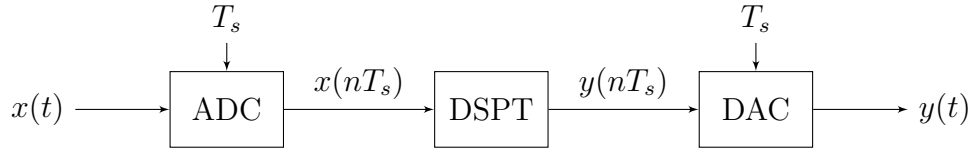


Figure 1: Block diagram of the DMS process [8].

## 2 Digital Oscillator

The DMS uses a digital oscillator to replicate any musical instruments in the 20[Hz] to 20000[Hz] range (i.e. human hearing range) [3]. Specifically, a typical type of digital oscillator is a wave-table oscillator, which performs the wave table lookup operation consisting of two steps [7]:

1. Pre-compute digital audio signals,  $x[n]$ , consisting of a sequence of samples  $n = 0, 1, \dots, N - 1$ , where  $N$  is the length of the sequence. For periodic waveforms, the  $N$  samples of any desired waveform can be precomputed using the Fourier series.
2. Produce an output  $z[n] = x[y[n]]$  such that the input signal  $y[n]$  is the index to get the specific values from  $x[n]$ .

### 2.1 Sampling

While the wave table oscillator is effective for generating a range of periodic waveforms, its limited by having to know what the analytical form of the waveform is before pre-computing

the digital audio signals. As a result, to ensure that the DMS can receive any sound (e.g. instruments, sound effects, etc), sampling can be used. Sampling is a process that involves recording a live audio signal into a wave table (i.e. digital sequence) [7]. These samples will accumulate in the DMS' memory as a stream of numbers, which can then be further processed using DSPT [3].

To ensure that there is no loss of information from turning the analog signal into a digital signal, the conversion process must follow the Sampling theorem [10, 9]:

Let the continuous-time signal  $x(t)$  be band-limited, i.e., the Fourier transform  $X(f)$  satisfies:

$$X(f) = 0 \quad \text{for } |f| > f_{\max}. \quad (1)$$

Then  $x(t)$  can be fully recovered by sampling in intervals of  $T_s$  (sampling period), if  $x(t)$  is sampled at the Nyquist rate:

$$f_s > 2f_{\max} \quad (2)$$

- $f_s$ : Sampling frequency.
- $f_{\max}$ : Maximum frequency of the original continuous-time signal.

It is a reasonable assumption to have  $x(t)$  be bandlimited due to the audio signals being heard between 20[Hz] to 20000[Hz] range for a human. As a result, live waveforms can be sampled at a rate of  $f_s$  such that

$$\{\dots, x(-2T_s), x(-T_s), x(0), x(T_s), x(2T_s), \dots\},$$

which will stay ensure that the signal can be reconstructed later on [9]. The Nyquist rate ensures that aliasing doesn't occur where the sampling frequency  $f_s < 2f_{\max}$ , which would distort the original signal, making it impossible to reconstruct the signal [10]. Even though the minimum requirement is at least  $2f_{\max}$ , DMS will over sample to the standard audio sample rate of 44.1[kHz] or higher [11]. This enables the user to have greater creative freedom in terms of synthesizing the audio for different purposes without losing information in the signal.

In order to convert these signals into a digital sequence, the following process is used (see Figure 2). Firstly,  $x(t)$  is sampled at a rate significantly higher than the Nyquist rate (oversampling), which have a continuous range of possible amplitudes [8]. As a result, a quantizer is used to map these continuous amplitudes to a discrete set of amplitudes with quantization interval  $\Delta$  to produce  $\hat{x}(nT_s)$ . Finally, the quantized signal is passed into an encoder, which produces a digital sequence for further processing in the DMS.

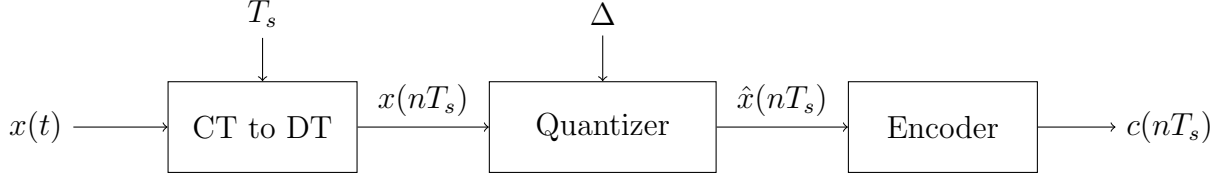


Figure 2: Block diagram of the ADC [8].

### 3 Additive Synthesis

Additive synthesis takes advantage of the fact that any sound can be expressed as a sum of sinusoids [5]. Specifically, by summing over  $N$  sinusoidal components, the user can modulate these components by relatively slowly varying amplitude and frequency envelopes. The equation for additive synthesis is:

$$y(t) = \sum_{i=1}^N A_i(t) \sin(\theta_i(t)) \quad (3)$$

- $A_i(t) \in \mathbb{R}$ : Amplitude of  $i$ th partial over time  $t$
- $\theta_i(t) = \int_0^t \omega_i(t) dt + \phi_i(0) \in \mathbb{R}$ 
  - $\omega_i(t) = \frac{d\theta_i(t)}{dt} \in \mathbb{R}$ : Radian frequency of  $i$ th partial vs. time.
  - $\phi_i(0) \in \mathbb{R}$ : Phase offset of  $i$ th partial at time 0.

As a result, by controlling these parameters, additive synthesis provides a means to generate sounds by assigning each sinusoid a separate envelope that can control the development of timbre and loudness with time [6].

## 4 Modulation

Modulation enables variations in loudness, pitch, and timbre. Two common techniques is using LFOs and FM.

### 4.1 Low-Frequency Oscillator

An LFO is a signal generator that produces slow, periodic waveforms, typically within the range of 0.1[Hz] to 20[Hz] (i.e. below human hearing levels) [6]. The purpose of an LFO is not to generate sound directly, but to modulate various parameters of a waveform periodically. There are many parameters that the LFO can affect [12]:

1. **Waveform:** The shape of the oscillation that the LFO generates (e.g. sine, triangle, square, and sawtooth).

2. **Frequency:** The speed of the modulation, where a low frequency will cause slow changes, while a high frequency will create rapid changes.
3. **Depth:** Controls the extent of the modulation's influence on the parameter of interest.
4. **Mode:** Determines how the LFO's cycle behaves during playback. In trigger mode, the LFO restarts its waveform cycle every time a new signal is played. In envelope mode, the LFO acts as an envelope (completing the period and stopping once the waveform reaches its end).

With these parameters defined, we can now examine how they apply to the most common waveform shapes. The following variables will be used to describe the most common waveforms in CT:

- $A$ : Amplitude of the waveform, determining its loudness.
- $f$  [Hz]: Frequency of the waveform, determining its pitch.
- $t$  [s]: Time.
- $\phi$  [rad]: Phase offset.
- $n$ : Harmonic number, corresponding to integer multiples of the fundamental frequency.
- $T$ : Period of the waveform.

Each of these waveforms offers unique characteristics that influence the modulation that can be produced by the LFO [13]:

1. **Sine Wave:** The sine wave contains only the fundamental frequency with no partial tones, represented as:

$$x(t) = A \sin(2\pi ft + \phi) \quad (4)$$

The sine wave sounds like a pure tone due to only having one frequency, which can be used to create every other waveform.

2. **Square Wave:** The square wave comprises a fundamental frequency and its odd harmonics, represented as [14]:

$$x(t) = \frac{4}{\pi} \sum_{n=1,3,5,\dots}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi t}{L}\right) \quad (5)$$

The square wave sounds richer and buzzier, which is more harsher than the sine wave [13, 15]. The sharp transitions between high and low values give the square wave a characteristic "on-off" sound.

3. **Sawtooth Wave:** The sawtooth wave contains a fundamental frequency and all harmonics of the fundamental, represented as [16]:

$$x(t) = \frac{1}{2} - \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi t}{L}\right) \quad (6)$$

The sawtooth wave is the buzziest and richest in terms of harmonics [15].

4. **Triangle Wave:** The triangle wave contains a fundamental frequency and its odd harmonics, represented as [17]:

$$x(t) = \sum_{k=1,3,5,\dots}^{\infty} \frac{8A}{\pi^2 k^2} (-1)^{(k-1)/2} \cos\left(2\pi \frac{k}{T} t\right) \quad (7)$$

The triangle wave has a harsher sound than sine waves, but a softer sound than square waves [15].

As a result, LFOs have many applications to create different effects. Two main effects are

1. **Vibrato Effect:** Modulates the pitch of a waveform by periodically varying the pitch using an LFO, in which the intensity is determined by the amplitude, and the frequency determines the speed of the fluctuations [18].
2. **Tremelo Effect:** Modulates the amplitude of a waveform by changing the volume of the signal periodically, which are usually made with sine or triangle waveforms [19].

## 4.2 Frequency Modulation

The Yamaha DX7 was invented by John Chowning in 1983, which was the first commercially available DMS that used FM synthesis [6]. In order to understand how FM synthesis works, we will use the Yamaha DX7's methodology of computing FM. Specifically, FM synthesis modulates the frequency of a carrier waveform using a modulating waveform inside the argument. The frequency modulated signal can be represented as

$$\sin(\omega_c t + I \sin \omega_m t) \quad (8)$$

- $\omega_c = 2\pi f_c$ : Carrier angular frequency.
- $\omega_m = 2\pi f_m$ : Modulating angular frequency.
- $I$ : Modulation index, which can determine the extent of frequency deviation  $d$  as  $d = I f_m$

where the values are calculated using digital lookup tables for the function [6].

Furthermore, the FM synthesis equation can be represented in terms of its Fourier Series using the Bessel functions  $J_n(I)$ :

$$\sin(\omega_c t + I \sin \omega_m t) = \sum_{n=-\infty}^{\infty} J_n(I) \sin(\omega_c + n\omega_m)t \quad (9)$$

where  $J_n(I)$  represents the amplitude of the  $n$ th sideband [6]. The Yamaha DX7 implements this Fourier series using two operators:

1. **Operator 1:** Generates the carrier signal, whose amplitude envelope dictates how the final signal's amplitude changes over time [6].

2. **Operator 2:** Produces the modulating signal, and its amplitude envelope controls the modulation index  $I$ , which influences the harmonic content of operator 1's output using the frequency deviation [6].

For small values of  $I$ , the frequency spectrum is dominated by  $\omega_c$ , resulting in a pure sine wave [6]. As  $I$  increases, the frequency spectrum has more harmonic components, with higher-order sidebands contributing more significantly. Consequently, operator 1 defines the amplitude of the final output signal, while operator 2 shapes the timbre by controlling the modulation index  $I$ . As a result, FM synthesis can create sounds such as bell timbres, metallic tones, tine tones of electric pianos, punchy bass, and synthetic brass sounds [20].

## 5 ADSR Envelope

The ADSR (attack, decay, sustain, and release) envelope generator (EG) produces an audio signal that rises and falls smoothly, controlling the loudness of the sound [7]. The four stages of the ADSR envelope are as follows (see Figure 3) [6]:

1. **Attack:** Time it takes for the initial rise to the maximum amplitude when the envelope is triggered on.
2. **Decay:** Time it takes for the transition from its maximum amplitude to the sustain amplitude level.
3. **Sustain:** The level at which the sound is held at steady state.
4. **Release:** The time it takes for the sound to go to 0 amplitude after the note is released.

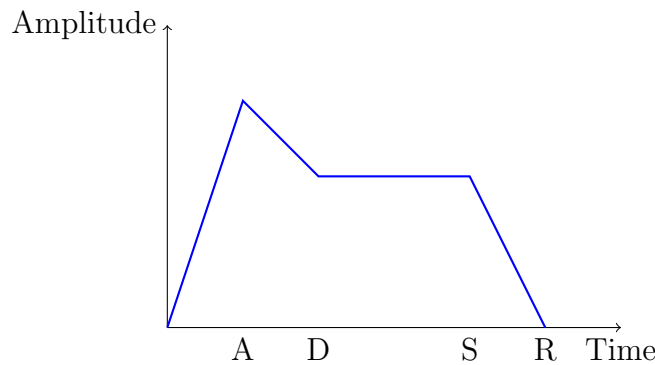


Figure 3: ADSR envelope applied to a sound signal [6].

The ADSR envelope is activated by a trigger control stream that turns the envelope on and off [7]. When the generator is triggered on, the envelope goes through the attack, decay, and sustain segments, while triggering it off causes it to start the release segment.

The ADSR EG has 5 main parameters [7]. Firstly, the level parameter determines the envelope's maximum amplitude. Secondly, the attack and decay parameters set the duration

of these segments, respectively. Thirdly, the sustain parameter specifies the amplitude at which the sound is held during the sustain segment. Fourthly, the release parameter controls the duration of the release segment. Together, these parameters, along with the timing of the on and off triggers can determine the ADSR envelope’s output.

As a result, the ADSR EG can control the amplitude, tones, and timbre of the sound [7]. Some examples include [21]:

- **Rich, lush tones:** Long attack, long release.
- **Percussive, staccato sounds:** Short attack, short release.
- **Snares or Hi-hats Instruments:** Long release time.
- **Focus on peaks:** Short sustain, long decay.

## 6 Subtractive Synthesis

Subtractive synthesis uses filters to shape the spectral envelope of a sound [7]. Specifically, the resulting sound is the product of the envelope of the original sound with the frequency response of the filter. Starting from a waveform, the user can remove specific frequency components using filters to shape the final sound signal. The process can be represented in the frequency domain as:

$$Y(f) = X(f)H(f) \tag{10}$$

- $X(f)$  : Frequency response of the original sound.
- $H(f)$  : Frequency response of filter.
- $Y(f)$  : Frequency response of the output signal.

By manipulating the harmonic content of waveforms, filters significantly influence the brightness of the sound [6]. The operation of a filter is typically described in terms of [7]:

1. **Pass-band:** The range of frequencies that the filter allows to pass with minimal attenuation.
2. **Stop-band:** The range of frequencies that the filter attenuates significantly.
3. **Transition-band:** Due to physical limitations, ideal filters with an abrupt transition are not realizable in practical implementations. Therefore, the transition band denotes the intermediate region between the pass-band and stop-band.

### 6.1 Types of Filters

In the DMS, four main filter types are used: low-pass, high-pass, band-pass, and band-stop filters. Each type has a specific role in subtractive synthesis, and their functionality is described below with the following variables:



- $H(f)$ : Frequency response of the filter.
  - $f_c$ : Center frequency of the pass-band.
  - $W$ : Bandwidth of the pass-band.
1. **LPF**: Allows low frequency components to pass through while attenuating higher frequencies above a cutoff frequency  $f_c$  [7]. The ideal frequency response of an LPF is expressed as:

$$H(f) = \begin{cases} 1 & \text{if } |f| \leq f_c, \\ 0 & \text{if } |f| > f_c. \end{cases} \quad (11)$$

2. **HPF**: Allows high-frequency components to pass through while attenuating frequencies below a specified cutoff  $f_c$  [7]. The ideal frequency response of an HPF is:

$$H(f) = \begin{cases} 1 & \text{if } |f| > f_c, \\ 0 & \text{if } |f| \leq f_c. \end{cases} \quad (12)$$

3. **BPF**: Allows frequencies within a specified range, or *bandwidth*, to pass through while attenuating frequencies outside this range [7]. Band-pass filters are useful for emphasizing specific harmonics or resonances. The ideal frequency response is:

$$H(f) = \begin{cases} 1 & \text{if } |f| \leq f_c + W, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

4. **BSF**: Attenuates frequencies within a specified range (notch) while allowing frequencies outside this range to pass through [7]. The ideal frequency response is:

$$H(f) = \begin{cases} 1 & \text{if } |f| \geq f_c + W, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

## 7 Digital to Analog Conversion

DAC is the process of converting a DT signal into a CT signal so that analog devices (e.g. speakers, headphones, etc) can play the sound.

The DAC process (see Figure 4) involves two steps. First, interpolating the samples after applying DSPT

$$\{\dots, y(-2T_s), y(-T_s), y(0), y(T_s), y(2T_s), \dots\}$$

to fully reconstruct a synthesized CT signal without loss of information [8, 9]. This can also be represented mathematically by multiplying the samples by an impulse train, along with its corresponding Fourier transform using the sifting property as:

$$y_s(t) = \sum_{n \in \mathbb{Z}} y(nT_s) \delta(t - nT_s) \xleftrightarrow{\mathcal{F}} \frac{1}{T_s} \sum_{k \in \mathbb{Z}} Y \left( f - \frac{k}{T_s} \right) \quad (15)$$

Second, in an ideal situation, the samples are passed through an ideal LPF, which removes the frequency components. The ideal LPF  $h(t)$  and its Fourier transform  $H(f)$  is:

$$h(t) = \text{sinc}\left(\frac{t}{T_s}\right) \quad H(f) = T_s \text{rect}(fT_s) \quad (16)$$

This LPF will remove all frequencies above the cutoff frequency  $\frac{f_s}{2}$  to ensure that the only frequencies that remain are associated with the synthesized CT signal  $y(t)$  (i.e. within the Nyquist range  $|f| \leq \frac{f_s}{2}$ ) [8]. In the time domain, the fully reconstructed CT signal  $y(t)$  is

$$y(t) = \sum_{n \in \mathbb{Z}} y(nT_s) \text{sinc}\left(\frac{t - nT_s}{T_s}\right) \quad (17)$$

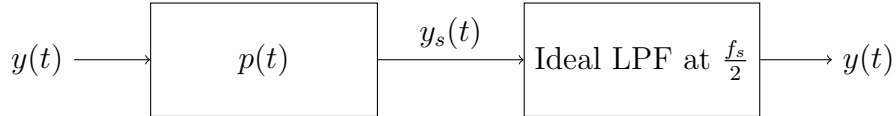


Figure 4: Block diagram of the DAC [8].

Realistically, many DAC use a zero-order hold for the reconstruction filter, which produces a staircase approximation to the original signal  $y(t)$  [8]. The process shown above with the ideal LPF is the same, but with the filter now being [9]:

$$H(f) = e^{-j2\pi f \frac{T_s}{2}} \cdot T_s \text{sinc}(fT_s) \quad (18)$$

## 8 Conclusion

Ultimately, this paper presented an overview of the DMS. Using a digital oscillator, analog signal are converted to a digital form via an ADC. The resulting digital signal can then be modified through many DSPT such as additive synthesis, LFOs, FM, ADSR envelop generators, and subtractive synthesis. Finally, after modifying the digital waveform, it can be converted back to an analog form using a DAC.

For readers interested in further exploration of the DMS, please read the *The Theory and Technique of Electronic Music* [7], *Music: A Mathematical Offering* (Chapters 7 and 8) [6], and *Spectral Audio Signal Processing* [5], along with the references therein.

## References

- [1] “Digital synthesizer,” Wikipedia, [Link](#)
- [2] History and practice of Digital Sound Synthesis, [Link](#)
- [3] H. Massey, “Subtractive 101, part one: The basics,” Synth, [Link](#).
- [4] “Auto-tune - the best vocal plug-ins available,” Auto-Tune - The Best Vocal Plug-Ins Available, [Link](#)
- [5] “Spectral Audio Signal Processing,” CCRMA, [Link](#)
- [6] Music: A mathematical offering Dave Benson, [Link](#)
- [7] The theory and technique of electronic music, [Link](#)
- [8] Schaum’s outlines of digital signal processing, [Link](#)
- [9] F. R. Kschischang, “Notes on the Sampling Theorem”
- [10] GeeksforGeeks, “Nyquist sampling theorem - statement, working, aliasing, applications,” GeeksforGeeks, [Link](#)
- [11] Sample rates and audio sampling: A guide for beginners — adobe, [Link](#)
- [12] S. Haven, “LFO like a Boss: The Complete Beginner’s Guide for 2024,” EDMProd, [Link](#)
- [13] N. Robehmed, “Sine, square, Triangle, saw,” Perfect Circuit, [Link](#)
- [14] “Fourier series–Square Wave,” from Wolfram MathWorld, [Link](#)
- [15] R. Smith, “The basic waveforms in synthesis - on the track,” On The Track - Music theory, beats, and more, [Link](#)
- [16] “Fourier series–sawtooth wave,” from Wolfram MathWorld, [Link](#)
- [17] “Fourier series–Triangle Wave,” from Wolfram MathWorld, [Link](#)
- [18] Complete guide the vibrato audio modulation effect - fox music production, [Link](#)
- [19] K. Pearsall, “Effects guide: Get to know tremolo,” What is Tremolo?, [Link](#)
- [20] “Frequency modulation (FM) synthesis,” Apple Support, [Link](#)
- [21] “ADSR envelopes explained: 4 stages of an ADSR envelope - 2024,” MasterClass, [Link](#)