Review TNM054 - Sound Technology I

Nathalie Ek, Albert Cervin

May 2010

Abstract

This little paper will try to summarize important parts of the course TNM054 given at Linkping University spring 2010. This is in no sense complete and has no intention to be.

1 Synthesis

1.1 Periodic Signals

Periodic signals can be seen as the sum of sinusoids at frequencies that are integer multiples (harmonics) of the fundamental frequency $f_0 = \frac{1}{T}$ (Fourier Series). If we for example take the functions $sin(x)$ and $sin(3x)/3$ and sum the two together we get another signal, still with period 2π but with a different look and sound.

1.2 Musical tones

The human perception of pitch (frequency) is not uniform and because of this the scale used in most western music has non-uniform spaced musical tones. The distance of tones increases with increasing frequency. An octave is 8 whole tones and is represented by doubling the frequency.

$$f_{octave} = 2^n f_0, \quad n \geq 0 \tag{1}$$

In equation 1 every octave f_o following the fundamental frequency f_0 can easily be obtained. In addition to this, harmonics are integer multiples of the fundamental frequency. The coincidence with musical notes are not exact though. This is due to our tempered scale, which in practice means that all tones in our scale are a little out of tune. This is done to reduce the number of tones in the scale.

1.3 Synthesis

Sound synthesis aims at reproducing existing or non-existing sources by means of simple algorithms. Efficient synthesis structures are controlled by a minimal set of parameters, such as pitch, amplitude, duration, force, velocity etc... The aim of synthesis is to fool our "bad" perception by creating a replica of the original sound. To be useful, these algorithms should be able to perform synthesis in real time.

1.3.1 Table Look-up Oscillator

The table look-up oscillator is probably the simplest and most naive synthesis method. It works by holding a table of signal values in a memory table. These values are then accessed with a step size and the output is created. This gives a frequency directly proportional to the step size used. Vibrato (frequency modulation effects) can be introduced by adding a LFO (Low Frequency Oscillator) to the step calculation.

1.3.2 ADSR Envelope Generators

ADSR stands for **A**ttack **D**ecay **S**ustain and **R**elease and represents the four piecewise linear phases of amplitude envelope in a sound. Attack is the initial phase, decay is the first decrease in amplitude, sustain is the length of the sound and release is the “die-off” of the sound.

The ADSR is used in order to include the slow variations of amplitude described above, that is missing in the table look-up oscillator.

1.3.3 Additive synthesis

Fourier additive synthesis is done by superposing several sinusoidal oscillators. Amplitude and frequency (phase) modulation is then controlled separately.

1.3.4 FM synthesis

Cascading two or more sinusoidal oscillators with arbitrary center frequencies results in a simple FM structure where the signal is given by

$$y(t) = Asin(2\pi f_c t + Isin(2\pi f_m t)) \tag{2}$$

where I is the modulating index which controls the depth of the modulation, f_c the carrier frequency, f_m the modulation frequency and A is the amplitude.

The frequency spectrum is harmonic **iff** the ratio is rational. FM synthesis is not based on a complete representation which gives artifacts. FM synths are often considered good at producing bell-like sounds.

1.3.5 Wavetable samplers

Wavetable sampling is a technique inspired by table look-up method that stores isolated tones of instruments in lookup tables. The table is subdivided into three phases: the initial transient, the sustain and the release. In order to avoid aliasing, interpolation and sample rate conversion is used. The sounds stored in wavetables can be time-stretched or compressed by means of granular interpolation/decimation.

1.3.6 PSOLA

PSOLA stands for **P**itch **S**ynchronous **O**ver**L**ap **A**dd ad is a method for synthesizing where the sound is modeled as a collection of overlapping grains, each two or more periods long. Different combiniations, repeating or discarding grains, can be used to lengthen or shorten the produced sound. If we take the famous sound of the trumpet, we have a period of roughly 128 samples. The signal is then split into “chunks” of twice the length of the period (256 samples) with 128 samples overlap. By repeating every chunk twice and then overlap adding them together again, a sound with double length is obtained.

1.3.7 Subtractive synthesis

Subtractive synthesis is obtained by filtering a richer signal (noise or pulse train) by means of a filter matching the spectral characteristics of the signal that we want to obtain. This model of synthesis is often used in speech synthesis.

1.3.8 Physical models

Synthesis by physical models relies on wave propagation equations within the structures and scattering at the interfaces. By space and time sampling (or by discretization of the equation) one can generally emulate wave propagation by means of a digital structure (digital waveguide/mesh). Together with boundary condition this structure provides the synthesis algorithm.

2 Time-frequency Representations

2.1 Frequency representation

A signal is a superposition of an infinite number (a continuum) of sinusoids at distinct frequencies ω and complex amplitudes $F(\omega)$ called the Fourier transform. There are four types of Fourier representations:

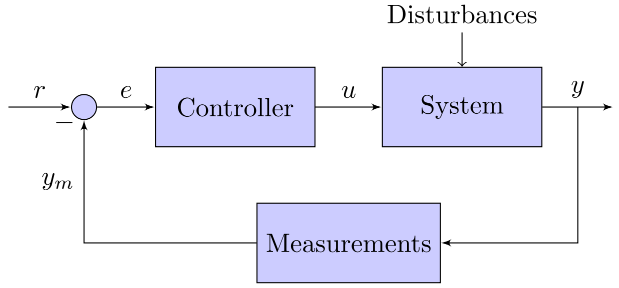
- Continous time Fourier Transforn (continous time, continous frequency)
- Discrete-Time FT (discrete time, continous frequency)
- Fourier Series (continous time, discrete frequency)
- DFT (discrete time, discrete frequency)

2.2 Time vs. Frequency Representations

The Fourier transform averages the signal over the entire time axis. The representative elements are sinusoids (narrowband in frequency, infinite duration in time).In time representation the representative elements are narrow pulses (localized in time, very spread out in frequency). These are the two extreme cases of each representation. By choosing something in between this, we are closing in on the Short Time Fourier Transform.

2.3 Short Time Fourier Transform

The Short Time Fourier Transform (hereafter STFT), is a variation of the Fourier transform which instead of transforming the whole signal, operates on windowed segments of the signal. The result of the STFT is a collection of the Fourier transforms of the signal frames, which are averages of the frequency content over the window length. The shorter the signal (in time) the wider is the Fourier spectrum (in frequency).



The STFT allow us to localize signals in the time-frequency plane within the limits of the uncertainty principle.

$$\Delta t \, \Delta\omega \geq \frac{1}{2} \tag{3}$$

The uncertainty principle described by equation 3 tells us that we can not have infinately good resolution in both time and frequency. In fact, when the time resolution is increased, by lowering the window size, the frequency resolution is decreased.