Some Mathematical Tools for Music-Making

Miller Puckette [*]

Abstract

Electronic music constantly uses transformations of functions of time. Some frequently-used mathematical operations are described, with an eye to their effects on sound spectra and possible musical applications.

1 Introduction

Music, like any art form, defies scientific analysis: the scientific method seeks traits common to a given species, whereas a piece of music is significant precisely because of its differences from other pieces of music. But even if we aren't able to find many reliable scientific laws about music, we can sometimes create interesting new tools for making it. Doing so requires a combination of intuition about music and mathematical understanding of the underlying medium, sound. Here I'll discuss one point of view on sound as a medium, which emphasizes its dependence on time.

More than any other art form, music concerns itself with time. A piece of music takes us on a path from its beginning to its end. A work of visual art, reflected from right to left, might still be recognizably the same work. But a piece of music turned from back to front is garbage. Time is the independent coordinate of music on which the dependent ones (pitch, loudness, ...) depend. Musical phrases are functions of time in the same way that a gesture in a painting is a function of two spatial dimensions.

This time-orientation may derive partly from the fact that music is usually transmitted using sound.

When we perceive a sound, we can detect extraordinarily small changes in timing, easily down to a millisecond; but our acoustical perception of spatial location, even in the best of conditions, is limited to about one degree of resolution and is usually much worse. Visually, we can sometimes resolve images to about 1/60 of a degree, both horizontally and vertically, but our eyes usually can't perceive time to better than 50 milliseconds or so of accuracy.

This is reflected in the common digital formats for storing sounds and moving images. Sounds are usually stored at between 40,000 and 50,000 frames per second, but with only between 2 and 6 spatial channels. Video normally requires only 30 frames per second but even low-quality video typically needs 500,000 or more spatial channels (pixels).

Since music is made of sound, and since sound, in our perception at least, can approximately be reduced to one or a few real-valued functions of time, we can get some insight into the workings of music by thinking about the real-valued functions of time and our perception of them. This will never lead to anything like a theory of music, but can shed light on the reasons music is what it is in certain respects—and can help us greatly in our attempts to make music.

1.1 Time symmetry

The space of real-valued functions is symmetric with respect to translations in time: $f(t) \mapsto f(t+\tau)$. This is a linear operator whose eigenfunctions are sinusoids:

$$f(t) = Ae^{i\alpha t}$$

where A is an amplitude and α an angular frequency. They behave under translations like this:
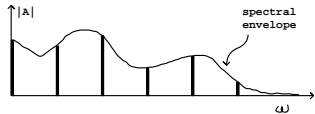
$$f(t + \tau) = e^{i\alpha\tau}f(t)$$

1

Figure 1: Spectrum of a periodic signal.

Figure 2: Interference pattern between two time-shifted signals. Top: spectrum of the original signal. Bottom: the result.

To our great fortune, our ears perform (very approximately) a sort of eigenvalue expansion of an incoming sound. This feature may have evolved in order to help us distinguish individual sound sources from the unruly mixture of sounds that reaches our ears. Our hearing systems appear to search for additive, approximately periodic components in complex sounds.

1.2 Periodicity

Both natural sounds and the sound of the human voice abound with signals (real-valued functions of time) that are approximately periodic. Since periodicity is a time symmetry, it is natural to look at the eigenvalue expansion of a periodic function of time. A signal $f(t)$ with period τ (and hence with frequency $2\pi/\tau$ radians per unit of time) can be expanded as:

$$f(t) = \ldots + A_{-1}e^{-i\alpha t} + A_0 + A_1 e^{i\alpha t} + A_2 e^{2i\alpha t} + \ldots$$

which is the well-known Fourier series for $f(t)$. Since periodic sounds occupy only a discrete subset of all available frequencies, it is possible to imagine confronting a spectrum of unknown origin and analyzing it as a sum of periodic functions.

Spectra of periodic functions (their Fourier series) can be graphed as in Figure 1. The amplitude of the ith harmonic is $|A_i|$. The *timbre* of the periodic sound is thought to depend mostly on a (not well-defined) curve called the *spectral envelope*, shown in the figure; if the fundamental frequency is changed but the spectral envelope kept the same, the resulting sound often has a similar timbre.

Combining two or more periodic signals whose fundamental frequencies have simple ratios (fractions with integers less than about 7 in the numerator and denominator) gives rise to spectra with many shared frequencies; this is the basis for the Helmholtz the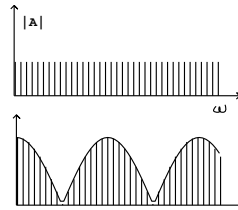ory of harmony [1], which depends on the fact that the Fourier series, considered as a spectrum, occupies frequencies at integer multiples of the fundamental frequency.

2 Operations on signals

With this simple spectral model of sounds in mind, we can now develop some of the fundamental techniques for operating on sounds. These techniques recur constantly in efforts to synthesize and process musical sounds electronically, for example with a computer.

2.1 Interference patterns (filtering)

Adding a sound to a time-delayed copy of itself sets up an interference pattern in the spectrum of the sound. This is the acoustic analog of a diffraction grating in optics. If the delay between two copies is τ, the two will interfere constructively at frequencies $0, 2\pi/\tau, 4\pi/\tau, \ldots$ and destructively halfway between these points. Figure 2 shows a spectrum of an input signal, and the resulting spectrum from adding two copies, one delayed in time compared to the other.

Linear combinations of many differently delayed copies of a signal give rise to more complicated interference patterns in the spectrum. A huge field of study is concerned with choosing particular linear combinations so that the interference pattern has desired properties, such as enhancing one frequency range compared to another [2]. Electrical engineers and electronic musicians call this technique *filtering*.

2

Filters arise in the natural world whenever sound encounters a cavity or barrier; for instance, the human vocal tract can be thought of as filtering the raw output of the glottis (vocal fold). To see why this is so, imagine the sound of the glottis scattering, separately, off each point of the surface of the vocal cavity. At the output (the mouth and nose) you get the superposition of all the scattered (and hence delayed) copies: a filter. This is essentially the same picture as Feynman used to describe quantum scattering as a superposition of all possible paths of the particles in a system.

2.2 Frequency shifting (modulation)

Returning to our complex sinusoid, $f(t) = A \cdot \exp(i\alpha t)$, we try multiplying it by another one, say $g(t) = \exp(i\beta t)$. We get a third one,

$$f(t)g(t) = Ae^{i(\alpha+\beta)t}$$

of frequency $\alpha + \beta$. Since in the real world we usually have access only to the real part of an incoming signal and can only send real-valued signals to our speakers, a more frequently encountered scenario is:

$$f(t) = A\cos(\alpha t), \quad g(t) = \cos(\beta t)$$

$$f(g)g(t) = \frac{A}{2}\left[\cos((\alpha+\beta)t) + \cos((\alpha-\beta)t)\right]$$

If the function $f(t)$ has many sinusoidal components, by the distributive law, multiplying by $g(t) = \cos(\beta t)$ acts individually on each one. Figure 3 shows a possible spectrum of a periodic function $f(t)$ and the result of multiplying it by a real-valued sinusoid. Engineers and electronic musicians call this *modulation.*

The resulting spectrum can again be that of a periodic signal (if the ratio of the frequency of the modulating signal $g(t)$ to the original fundamental frequency is a fraction with small numerator and denominator), or otherwise it might have no audible fundamental. Both possibilities can be musically useful, depending on the context.

The spectral envelope of the result resembles that of the original, unmodulated signal as long as the modulating frequency is small compared to the original fundamental; otherwise it can be greatly distorted. So the one operation can affect both tuning
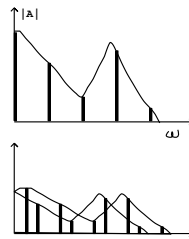

Figure 3: Modulating a periodic signal. Top: Spectrum of the original signal; Bottom: the result of multiplying it by a real-valued sinusoid.

and spectral envelope with some degree of independent control.

2.3 Nonlinear transfer functions (waveshaping)

A technique familiarized in Rock and Roll music of the sixties, but with antecedents in electronic music, is simply to distort sounds to change their timbres. If $f(t)$ is an incoming signal, we compose it with a nonlinear *transfer function* $h(t)$, and listen to the result, $h(f(t))$. In the R&R tradition, $f(t)$ is the electric guitar signal and $h(t)$ is the transfer function of the overdriven amplifier.

For example, let $f(t)$ be a real-valued "sinusoid" with time-varying amplitude:

$$f(t) = A(t)\cos(\alpha t)$$

and choose $h(t) = t^2$ as a transfer function, giving:

$$h(f(t)) = \frac{A^2(t)}{2}\left[\cos(2\alpha t) + 1\right]$$

Possible input and output functions are graphed in Figure 4. If $h(t)$ is chosen to be a polynomial or a convergent power series, the actions of the monomials in $h(t)$ will be mixed in ratios depending on $A(t)$: changing amplitude in the input changes timbre in the output. Skillfully chosen input and transfer functions can give rise to "nice" spectra; for instance, choosing

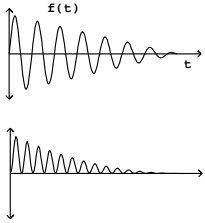$$h(t) = \frac{1}{1 + t^2}$$

3

Figure 4: Waveshaping a sinusoid. Top: the incoming sound as a function of time. Bottom: the result of applying a simple non-linear function.

turns a sinusoid into a spectrum with exponentially dropping partials, with the rate of rolloff determined by A [3].

Using primarily these three fundamental techniques, electronic musicians operate on a starting palette of sinusoids, white noise, and/or recorded sounds, to produce a huge variety of electronic sounds for use as raw materials in making new forms of music.

3 Analysis

It is frequently desirable, in dealing with sounds electronically, to analyze the frequency content of a sound. In the simplest situation we assume the sound is a finite sum of sinusoids and we would like to know their frequencies and (complex) amplitudes. This would be easy except for the fact that the frequencies and amplitudes in question are usually changing, sometimes quite rapidly. Very few sounds in nature or of electronic origin are well modeled as a sum of eternally unchanging sinusoids.

The most frequently taken approach to this problem is to extract a short segment of the signal to be analyzed, hoping that whatever components are present haven't had time to vary much within the segment. For example, if the function to analyze is $f(t)$, one could first multiply it by a windowing function such as:

$$w(t) = \begin{cases} \frac{1}{2}\left[\cos(2\pi t/S) + 1\right] & |t| < S/2 \\ 0 & \text{otherwise} \end{cases}$$
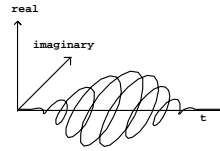

Figure 5: A (complex-valued) sinusoidal wave packet as a function of time.

where S is the length of the segment to analyze. Suppose that f is a complex sinusoid:

$$f(t) = Ae^{i\alpha t}$$

so that the product $w(t)f(t)$ is as shown in Figure 5. The Fourier transform of the product (as an L^2 function [4, p. 168]) is:

$$\mathcal{FT}\left\{w(t)f(t)\right\}(\omega) = \mathcal{FT}\left\{w(t)\right\}(\omega - \alpha)$$

or, in words, it is just the Fourier transform of the windowing function $w(t)$ shifted in frequency by α, as shown in Figure 6.

The main peak of Figure 6 is $8\pi/S$ in width, centered about α. The shorter we make the time segment S, the more spread-out the frequency-domain peak will appear. This is the Heisenberg uncertainty principle in action [5, p. 126].

Since the Fourier transform is linear, a superposition of sinusoids would give a superposition of peaks on the frequency axis. To fully resolve them we would need the peaks to be separated by the peak width, $8\pi/S$. If the sound is periodic, the analysis should be done over a length of time containing at least four periods of the sound. (In practice we can often allow some overlap, reducing this to about three).

Looking at a sound's Fourier transform, we can determine the frequencies and amplitudes of its sinusoidal components by fitting the observed peaks with their known theoretical shape. The spectral envelope can be estimated as well.

Furthermore, a given audio signal can be modified in interesting ways by taking its Fourier transform (using a sequence of overlapping analysis segments called *windows*), performing some operation,
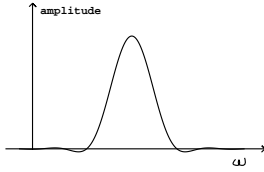
Figure 6: Fourier transform of the wave packet.

and then taking the inverse Fourier transforms to reconstruct a modified signal. For example, the spectral envelope of one signal can be "stamped" on another by modifying the magnitude of the latter's Fourier transform non-uniformly.

Many details have been glossed over in this very brief introduction; moreover, only that part of electronic music which deals with realization of pieces of music has been treated. Present-day research also touches on the design of real-time software systems and human-friendly controls for music making; computer understanding of musical form and computer-aided composition; music perception; and music in multimedia applications. Within the narrower field described here, many problems remain open and improvements are constantly sought in our existing repertoire of techniques. Mathematicians can find excellent work here.

References

[1] Hermann Helmholtz. *On the Sensations of Tone.* Dover, New York, fourth edition, 1954. Translation, A.J. Ellis.

[2] Kenneth Steiglitz. *A Digital Signal Processing Primer.* Addison-Wesley, Menlo Park, California, 1996.

[3] Miller S. Puckette. Formant-based audio synthesis using nonlinear distortion. *Journal of the Audio Engineering Society*, 43(1):224–227, 1995.

[4] Walter Rudin. *Functional Analysis.* McGraw-Hill, New York, 1973.

[5] Charles F. Stevens. *The Six Core Theories of Modern Physics.* MIT Press, Cambridge, Massachusetts, 1995.