

---

# Musical Sound Transformation by *Convolution*

Curtis Roads

Les Ateliers UPIC  
5 allée de Nantes  
91300 Massy (Paris)  
France

## ABSTRACT

Since the dawn of the vacuum tube era, musicians have sought to transform musical sounds by electronic means (Bode 1984). Increasing computer power continues to open up new possibilities toward this end. Previously exotic and computationally-intensive sound transformations can now be realized on personal computers. *Convolution* is one such technique. A fundamental operation in signal processing, it "marries" two signals (Rabiner and Gold 1975). As a form of *cross-synthesis*, convolution is also implicit in operations such as filtering, modulation, excitation/resonance modeling, spatialization, and reverberation. Treating these operations as convolutions, we can extend them in interesting directions.

A few musicians have experimented with arbitrary convolutions. But without an understanding of the musical implications of the technique, the results are may be confusing or disappointing. An analogy with paint is appropriate. When a child smears arbitrary hues of paint together; the result is usually an amorphous mess. Convolution can easily smear sounds. Many convolutions that appear to be interesting musical ideas ("How about crossing a clarinet with a speaking voice?") result in blurred, ringing sonorities. If one is not careful, convolution can easily destroy the time structure and identity of its inputs, emitting an indistinct sonic blob at the output. Thus it is imperative for musicians working with convolution to comprehend its implications so that they can better predict the results of arbitrary convolutions.

This paper summarizes the theory and presents some results of our research into musical applications of convolution, as well as new techniques involving convolution with grains and rhythm input. These results were first presented in scientific conferences in Trieste, Capri, and Paris a year ago (Roads 1992). In lecture form, this paper is accompanied by sound examples.

## STATUS OF CONVOLUTION

Convolution occupies an odd position today. It remains unknown to most musicians, yet to signal processing engineers, it is a basic topic. The mathematical theory of convolution was nailed down long ago, so that signal processing textbooks inevitably present it abstractly in the first few pages, reducing it to a handful of mathematical clichés (just as frequency modulation is presented, by the way). Unfortunately, the musical significance of these equations is not well known or appreciated, either by engineers or musicians. The manifold effects of convolution have rarely been studied in a musical context (an exception is Dolson and Boulanger 1985).

Listeners are familiar with the effects of convolution, even if they are not aware of its theory. Convolution is disguised under more familiar terms such as filtering, modulation, and reverberation. Recent software tools unbundle convolution, offering it as an explicit operation, allowing any two signals to be convolved (Erbe 1992; The MathWorks 1992). These tools provide a stable basis for musical exploration of the technique, and prompt a need for more universal understanding of its

powers. We begin this teaching task in the present paper, sections of which are drawn from our textbook *Computer Music Tutorial* (Roads 1994).

## IMPULSE RESPONSE

The notion of impulse response is central to convolution. Signal processing theory defines the term "filter" broadly—any system that accepts an input signal and emits an output is said to be a filter (Rabiner et al. 1972). So convolution is a filter. A good way to examine the effect of a filter is to see how it reacts to test signals. One of the most important test signals in signal processing is the *unit impulse*—a brief burst of energy at maximum amplitude. In a digital system, the briefest possible signal lasts just one sample. This signal contains energy at all frequencies that can be represented at a given sampling frequency. The output signal generated by a filter that is fed a unit impulse is called the *impulse response* (IR) of the filter. The IR corresponds to the system's *amplitude-versus-frequency response* (or frequency response). The IR and the frequency response contain the same information—the filter's response to the unit impulse—but plotted in different domains. That is, the IR is a time-domain representation, and the frequency response is a frequency-domain representation.

Convolution is the bridge between the time domain and the frequency domain. Any filter, for example, convolves its impulse response with the input signal to produce a filtered output signal. The implications of convolution with an IR are vast. One can start from the measured IR of any audio-frequency system (microphone, loudspeaker, room, distortion, delay effect, filter, modulator, etc.); through convolution, one can impose the properties of this system onto any audio signal.

This much is well-known in the engineering community. By generalizing the notion of impulse response, however, one arrives at quite another set of possibilities. Simply stated, let us consider any sequence of samples as the impulse response of a hypothetical system. Now we arrive at a musically potent application of convolution: cross-synthesis (or cross-filtering) by convolution of two arbitrary signals. In musical signal processing, "cross-synthesis" describes a variety of different techniques that in some way combine the properties of two sounds into a single sound (Roads 1994). This usually involves some kind of spectrum shaping of one sound by the other, which is indeed one of the effects of

cross-synthesis by convolution.

What then precisely is convolution? The next sections present, in an intuitive manner, the theory of convolution. The rest of the paper assesses the musical significance of convolution, offering practical guidelines for musical use and new extensions of the technique.

## THE OPERATION OF CONVOLUTION

To understand convolution, let us look at the simplest case: convolution of a signal  $a$  with a unit impulse, which we call  $unit[n]$ . At time  $n = 0$ ,  $unit[n] = 1$ , but for all other values of  $n$ ,  $unit[n] = 0$ . The convolution of  $a[n]$  with  $unit[n]$  can be denoted as follows:

$$output[n] = a[n] * unit[n] = a[n]$$

Here "\*" signifies convolution. As figure 1a shows, this results in a set of values for  $output$  that are the same as the original signal  $a[n]$ . Thus, convolution with the unit impulse is said to be an identity operation with respect to convolution, because any function convolved with  $unit[n]$  leaves that function unchanged.

Two other simple cases of convolution tell us enough to predict what will happen at the sample level with any convolution. If one scales the amplitude of  $unit[n]$  by a constant  $c$ , the operation can be written as follows:

$$output[n] = a[n] * (c \times unit[n])$$

The result is simply:

$$output[n] = c \times a[n]$$

In other words, we obtain the identity of  $a$ , scaled by the constant  $c$ , as shown in figure 1b

In the third case, we convolve signal  $a$  by a unit impulse that has been time-shifted by  $t$  samples. Now the impulse appears at sample  $n - t$  instead of at  $n = 0$ . This can be expressed as follows:

$$output[n] = a[n] * unit[n - t]$$

The result of which is:

$$output[n] = a[n - t]$$

That is,  $output$  is identical to  $a$  except that it is time-shifted by the difference between  $n$  and  $t$ . Figure 1c shows a combination of scaling and time-shifting.

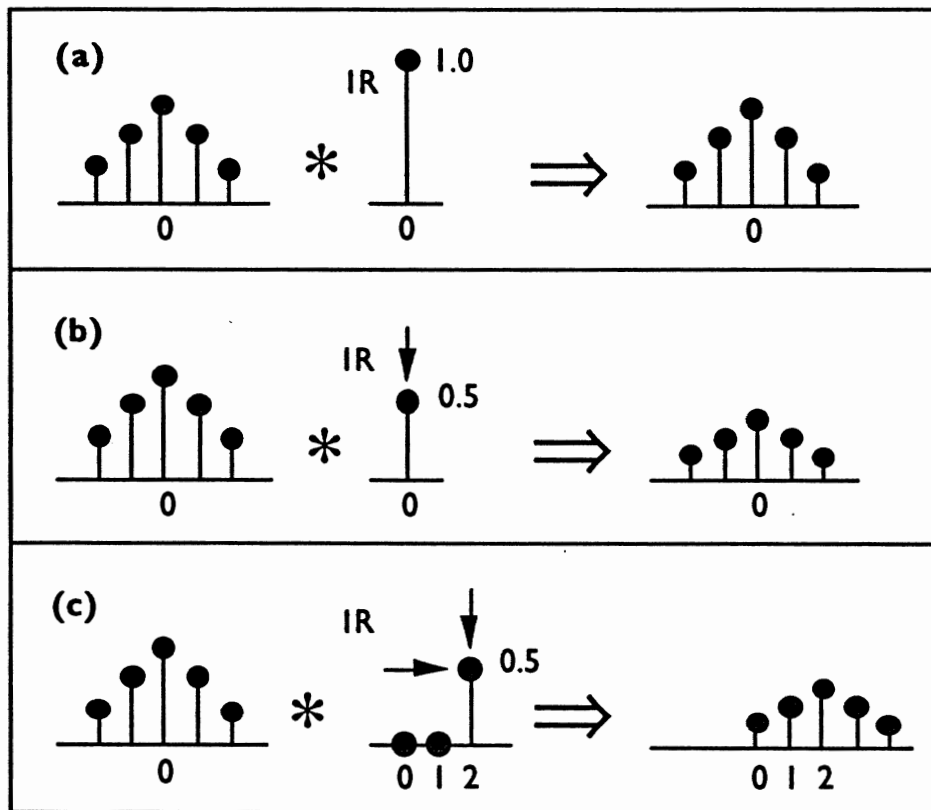


Fig. 1. Convolution by scaled and delayed unit impulses.

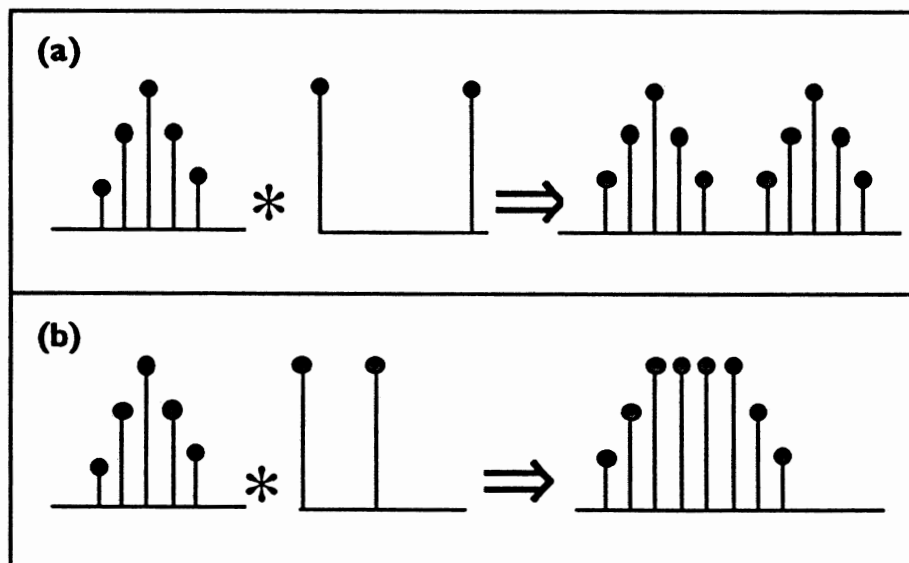


Fig. 2. Echo and time-smearing induced by convolution.

Putting together these three cases, we can view any sampled function as a sequence of scaled and delayed unit impulse functions. They explain the effect of convolution with any IR. For example, the convolution of any signal  $a$

with another signal  $b$  that contains two impulses spaced widely apart results in a repetition or echo of  $a$  starting at the second impulse in  $b$  (figure 2a). When the impulses in

$b$  move closer together, the scaled repetitions of  $b$  start to overlap each other (figure 2b).

Thus, to convolve an input sequence  $a[n]$  with an arbitrary function  $b[n]$ , one places a copy of  $b[n]$  at each point of  $a[n]$ , scaled by the value of  $a[n]$  at that point. The convolution of  $a$  and  $b$  is the sum of these scaled and delayed functions.

## MATHEMATICAL DEFINITION OF CONVOLUTION

Theory defines the convolution of two finite sequences of samples  $a$  and  $b$  as follows:

$$a[n] * b[n] = \text{output}[k] = \sum_{n=0}^{N-1} a[n] \times b[k-n]$$

where  $N$  is the length of the sequence  $a$  in samples and  $k$  ranges over the entire length of  $b$ . In effect, each sample of  $a[n]$  serves as a weighting function for a delayed copy of  $b[n]$ ; these weighted and delayed copies all add together.

The conventional way to calculate this equation is to evaluate the sum for each value of  $k$ . This is *direct convolution*. At the midpoint of the convolution,  $n$  copies are summed, so the result of this method of convolution is usually rescaled (i.e., *normalized*) afterward.

Convolution lengthens its inputs. The length of the output sequence generated by direct convolution is:

$$\text{length}(\text{output}) = \text{length}(a) + \text{length}(b) - 1$$

In the typical case of an audio filter (lowpass, highpass, bandpass, bandreject),  $a$  is an IR that is very short compared to the length of the  $b$  signal. For a broad smooth lowpass or highpass filter, for example, the IR lasts less than a millisecond.

## THE LAW OF CONVOLUTION

A fundamental law of signal processing is that the convolution of two waveforms is equivalent to the multiplication of their spectra. The inverse also holds. That is, the multiplication of two waveforms is equal to the convolution of their spectra. Another way of stating this is as follows:

*Convolution in the time domain is equal to multiplication in the frequency domain and vice versa.*

The law of convolution has profound acoustical implications. In particular, convolution of two sounds is equivalent to filtering the spectrum of one by the spectrum of the other. Inversely, multiplying two audio signals (i.e., performing *amplitude modulation* or *ring modulation*), is equal to convolving their spectra. Convolution of spectra means that each point in the discrete frequency spectrum of input  $a$  is convolved with every point in the spectrum of  $b$ . Convolution does not distinguish whether its input sequences represent samples or spectra. To the convolution algorithm they are both just discrete sequences. (The spectrum sequence generated by Fourier analysis is in the form of complex numbers, however; see the section on modulation later.)

Another implication of the law of convolution is that every time one reshapes the envelope of a sound, one is convolving the spectrum of the envelope with the spectrum of the reshaped sound. In other words, every time-domain transformation results in a corresponding frequency-domain transformation, and vice versa.

## RELATIONSHIP OF CONVOLUTION TO FILTERING

The equation of a general *finite-impulse-response* (FIR) filter is as follows:

$$y[n] = (a \times x[n]) \pm (b \times x[n-1]) \pm \dots (i \times x[n-j])$$

One can think of the coefficients  $a, b, \dots, i$  as elements in an array  $h(i)$ , where each element in  $h(i)$  is multiplied times the corresponding element in array  $x[j]$ . With this in mind, the general equation of an FIR filter presented earlier can be restated as a convolution:

$$y[n] = \sum_{m=0}^{N-1} h[m] \times x[n-m]$$

where  $N$  is the length of the sequence  $h$  in samples and  $n$  ranges over the entire length of  $x$ . Notice that the coefficients  $h$  play the role of the impulse response in the convolution equation. And indeed, the impulse response of an FIR filter can be taken directly from the value of its coefficients. Thus any FIR filter

can be expressed as a convolution, and vice versa.

Since an *infinite-impulse response* (IIR) filter also convolves, it is reasonable to ask whether there is also a direct relation between its coefficients and its IR. In a word, the answer is no. There are, however, approximation techniques that design an IIR starting from a given IR (Rabiner and Gold 1975, p. 265).

## FAST CONVOLUTION

Direct convolution is notoriously intensive computationally, requiring on the order of  $N^2$  operations, where  $N$  is the length of the longest input sequence. Many practical applications of convolution use a method called *fast convolution* (Stockham 1969). Fast convolution for long sequences takes advantage of the fact that the product of two  $N$ -point *discrete Fourier transforms* (DFTs) is equal to the DFT of the convolution of two  $N$ -point sequences. Since the DFT can be computed rapidly using the *fast Fourier transform* (FFT) algorithm, this leads to a great speedup for convolution.

Before the FFTs are taken, both sequences are lengthened by appending zeros until they are both equal to the convolution output length (as discussed previously). The results of the convolution can be resynthesized from the inverse FFT.

Fast convolution takes on the order of  $N \times \log_2(N)$  operations. Consider the direct convolution of two 2-second sounds sampled at 48 KHz. This requires on the order of 9,216,000,000 operations. Fast convolution with the same two sounds takes less than 1,500,000 operations, a speedup by a factor of 6100. Put another way, a fast convolution that takes one second to calculate on a given microprocessor requires 101 minutes to calculate via direct convolution.

For real-time applications where immediate output is needed, it is possible to implement convolution in sections, that is, a few samples at a time. Sectioned and nonsectioned convolution generate equivalent results, but sectioned convolution begins to produce output after only a short delay. See Rabiner and Gold (1975) for explanations of standard techniques for sectioned convolution.

## MUSICAL SIGNIFICANCE OF CONVOLUTION

A veritable catalog of sonic transformations emerge out of convolution: cross-filters, spatialization and modulation, models of excitation/resonance interaction, and time-domain effects. Indeed, some of the most dramatic effects induced by convolution involve temporal transformations: attack smoothing, multiple echoes, room simulation, time smearing, and reverberation (Dolson and Boulanger 1985; Roads 1992).

It is important to realize that the type of effect achieved depends entirely on the nature of the input signals. Pure convolution has no control parameters.

The following sections glance at each type of transformation. A bullet in front of an indented section indicates a practical guideline.

### Cross-filtering

Any filter can be realized by convolving an input signal with the IR of the desired filter.

- If both signals are long in duration and one of the input signals has a smooth attack, the dominant effect of convolution is a spectrum alteration.

Let us call two sources  $a$  and  $b$  and their corresponding analyzed spectra  $A$  and  $B$ . If we multiply each point in  $A$  with each corresponding point in  $B$  and then resynthesize, we obtain a waveform that is the convolution of  $a$  with  $b$ .

- If both sources are long duration and each has a strong pitch and one or both of the sources has a smooth attack, the result contains both pitches and the intersection of their spectra.

For example, the convolution of two saxophone tones, each with a smooth attack, sounds like the two tones are being played simultaneously. Unlike simple mixing, however, the filtering effect accentuates metallic resonances common to the tones.

Convolution is particularly sensitive to the attack of its inputs.

- If either source has a smooth attack, the output will have a smooth attack.

Listening to the results of cross-filtering, one sometimes wishes to increase the presence of

one signal at the expense of the other. Unfortunately, there is no straightforward way to adjust the "balance" of the two sources or to lessen the convolution effect. Attenuating one of the sources simply lowers the amplitude of the output; it does not change the balance.

### Echoes

Echo effects can be explained as follows. Any unit impulse in one of the inputs to the convolution results in a copy of the other signal. Thus if we convolve any brief sound with an IR consisting of two unit impulses spaced 1 second apart, the result is a clear echo of the first sound (figure 2a).

- To create a multiple echo effect, convolve any sound with a series of impulses spaced at the desired delay times.

### Room Simulation

The IR of a room may have dozens of strong impulses, corresponding to reflections off interior surfaces—its echo pattern.

- To simulate the effect of a sound playing in an arbitrary room, record the impulse response of the room (i.e., its response to a loud impulse). If such an IR is convolved with an arbitrary sound, the result is as though that sound had been played in that room.

If the peaks in the IR are closely spaced, however, the repetitions are time-smeared (figure 2b). Time-smearing smooths out transients and blurs the precise onset of events.

### Spatial Positioning

This technique is closely related to the previous one, but requires that the convolved sounds are recorded in a multichannel format.

- To position a sound at a precise point in a space, generate an impulse at that point and convolve the sound with a multichannel recording of the impulse response of the space.

### Reverberation

The combination of time-smearing and echo explains why noise signals, which contain thousands of sharp peaks, result in reverberation effects when convolved.

- To generate a basic reverberation, convolve any signal with white noise, where the noise has a sharp attack and an

exponential decay. The length of the noise determines the reverberation time.

- To color this reverberation, filter the noise before convolving it.
- If the noise has a logarithmic—not exponential—decay, the input sound appears suspended in time before the decay.

### Importance of Mixing

The raw output of convolution is not always usable as a musical result. In many instances, and especially for realistic spatial effects, it is essential to blend the output of the convolution with the original signal.

- For realistic spatial effects, mix the convolved signal down at least -15 dB with respect to the unprocessed signal.

### Modulation as Convolution

Amplitude and ring modulation (AM and RM) both call for multiplication of time-domain waveforms. The law of convolution states that multiplication of two waveforms convolves their spectra. Hence, convolution accounts for the sidebands that result. Consider the examples in figure 1, and imagine that instead of impulses in the time domain, convolution is working on lines in the frequency domain. The same rules apply—with the important difference that the arithmetic of complex numbers applies. The FFT, for example, generates a complex number for each spectrum component. Here the main point is that this representation is symmetric about 0 Hz, with a replica of each component (halved in amplitude) in the negative frequency domain. This negative spectrum is rarely plotted, since it only has significance inside the FFT. But it helps explain the positive and negative sidebands generated by AM and RM.

### Excitation/resonance Modeling

Many vocal and instrumental sounds can be simulated by a two-part model: an *excitation* signal that is filtered by a *resonance*. The excitation is a nonlinear switching action, like the pluck of a string, the buzz of a reed, or a jet of air into a tube. The resonance is the filtering response of the body of an instrument. Convolution lets us explore a virtual world in which one sound excites the resonances of another.

By a careful choice of input signals, convolution can simulate improbable or impossible performance situations—as if one

instrument is somehow playing another. In some cases (e.g., a chain of bells striking a gong), the interaction could be realized in the physical world. Other cases (e.g., a harpsichord playing a gong), can only be realized in the virtual reality of convolution.

- To achieve a plausible simulation, the excitation must be a brief, impulse-like signal, (typically percussive), with a sharp attack (or multiple sharp attacks). The resonance can be any sound.

## CONVOLUTIONS WITH GRAINS

A new category of sound transformations involves convolutions with sonic *grains* (Xenakis 1960; Roads 1991, 1992). A grain is a momentary acoustical event. The technique of *asynchronous granular synthesis* (AGS) scatters grains uniformly within the bounds of a *cloud*—a region inscribed on the time/frequency plane (figure 3).

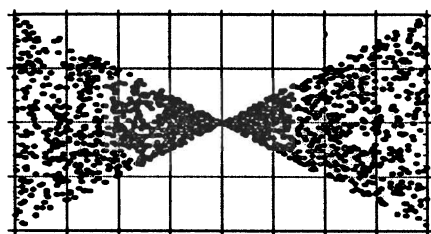


Figure 3. Cloud of sonic grains. Vertical axis is frequency, horizontal axis is time.

In this application, the grains can be thought of as the IR of an unusual filter or virtual space (Roads 1992). The results of convolution with grains vary greatly, depending on the properties of the granular cloud and the input signal.

- For a sharp-attacked input, convolution with a sparse cloud of a few dozen short grains causes a statistical distribution of echoes of the input sound (figure 4).
- Long grains accentuate time-smearing and round off sharp attacks.
- When the input sound has a smooth attack, the result is a time-varying filtering, depending on the spectrum of the grains.
- The denser the cloud, the more the echoes fuse into a reverberation effect (see the next section).

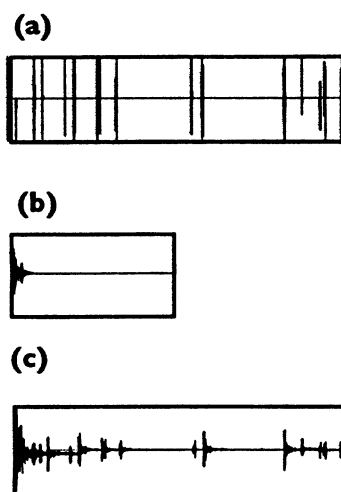


Figure 4. Convolution of grain pattern (a) with a tambourine (b) maps the tambourine to the time and amplitude pattern of the grains (c).

## Atmospheric Reverberation

*The rolling of thunder has been attributed to echoes among the clouds; and if it is to be considered that a cloud is a collection of particles of water...and therefore each capable of reflecting sound, there is no reason why very [loud] sounds should not be reverberated ... from a cloud.*—  
Sir John Herschel, in Tyndall (1875).

Thunderhead clouds act as atmospheric reverberators. The 19th-century acoustical scientists Arago, Mathieu, and Prony, in experiments on the velocity of sound, observed that under a clear sky, cannon shots were singular and sharp. Whereas, when the sky was overcast or a large cloud filled part of the sky, shots were often accompanied by a long rippling "roll," akin to thunder (Tyndall 1875).

Not surprisingly, the convolution of a sound with a cloud of sound particles creates a scattershot "time-splattered" effect analogous to atmospheric reverberation. The term "splattered" conveys the nonuniformity of the effects. These reverberations often undulate irregularly with odd peaks and valleys of intensity.

Atmospheric reverberation begins with a dense cloud of grains generated by AGS. This mass of grains can be thought of as the IR of a cumulus cloud-like enclosure. The virtual reflection contributed by each grain time-splatters the input sound, adding multiple irregularly-spaced delays. If each grain was a single-sample pulse, then the echoes would be

grain may contain hundreds of samples, however, each echo is time-smeared as well as splattered. Time-splattering effects can be divided into two basic categories, which depend mainly on the attack of the input sound.

- If the input begins with a sharp attack, each grain generates an echo of that attack. If the cloud of grains is not continuous, these echoes are irregularly spaced in time.
- If the input has a smooth attack, the time-splattering smooths into a strange colored reverberation. The "color" of the reverberation is determined by the spectrum of the grains—a factor of their duration, envelope, and waveform.

## RHYTHM INPUT

Another new application of convolution is input of performed rhythms. We have seen that a series of impulses convolved with a brief sound maps that sound into the time pattern of the impulses. To enter a performed rhythm, one need only tap with drumsticks on a hard surface, and then convolve those taps with other sounds.

- The convolution of a tapped rhythmic pattern with any sound having a sharp attack causes each tap to be replaced by a copy of the input sound.

This is a fast and precise method of mapping performed rhythms to arbitrary sounds. One can also layer convolutions using different patterns and input sounds. If one positions each tap in stereo space, convolution automatically distributes them spatially.

## CONCLUSION

The vast territory of sound transformations through convolution are beginning to be mapped. Understanding its musical implications will save much time searching for interesting results. Hence, in this paper we have consistently tried to separate the musically useful from the merely possible.

## ACKNOWLEDGEMENTS

I would like to thank my colleagues Prof. Aldo Piccialli of the Department of Physics at the Università di Napoli and Dr. Marie-Hélène

Serra of IRCAM and for stimulating discussions on signal processing. Tom Erbe of the Center for Contemporary Music at Mills College was very helpful in answering questions about the internal operation of his convolution algorithms in the excellent SoundHack program. Thanks also to Gerard Pape and Les Ateliers UPIC for their support.

## REFERENCES

- Bode, H. 1984. "History of electronic sound modification." *Journal of the Audio Engineering Society* 32(10): 730-739.
- Dolson, M., and R. Boulanger. 1985. New directions in the musical use of resonators." Unpublished manuscript.
- Erbe, T. 1992. *SoundHack User's Manual*. Oakland: Mills College.
- The MathWorks. 1992. *Matlab*. Natick: The MathWorks.
- Rabiner, L., and B. Gold. 1975. *Theory and Application of Digital Signal Processing*. Englewood Cliffs: Prentice-Hall.
- Rabiner, L., J. Cooley, H. Helms, L. Jackson, J. Kaiser, C. Rader, R. Schafer, K. Steiglitz, and C. Weinstein. 1972. "Terminology in digital signal processing." *IEEE Transactions on Audio and Electroacoustics* AU-20: 322-7.
- Roads, C. 1991. "Asynchronous granular synthesis." In G. De Poli, A. Piccialli, and C. Roads, eds. 1991. *Representations of Musical Signals*. Cambridge, Mass.: The MIT Press.
- Roads, C. 1992. "Musical Applications of Advanced Signal Transformations." In A. Piccialli, ed. 1992. *Proceedings of the Capri Workshop on Models and Representations of Musical Signals*. Naples: University of Naples Federico II, Department of Physics.
- Roads, C. 1994. *Computer Music Tutorial*. Cambridge, Mass.: The MIT Press.
- Stockham, T. 1969. "High-speed convolution and convolution with applications to digital filtering." In B. Gold and C. Rader. 1969. *Digital Processing of Signals*. New York: McGraw-Hill. pp. 203-232.
- Tyndall, J. 1875. *Sound*. Akron: Werner.
- Xenakis, I. 1960. "Elements of stochastic music." *Gravesaner Blätter* 18: 84-105.