THEORY AND PRACTICE OF
MUSICAL SOUND SYNTHESIS

Section 1

Introduction And Overview

1.1 THE USE OF ELECTRONIC SOUND SYNTHESIS:

1.1a INTRODUCTION: THE REALITY OF ELECTRONIC SOUND SYNTHESIS:

There are numerous reasons to consider the use of electronic methods for the generation of musical sounds. Perhaps above all such apparent needs for using electronic methods we find a motivation that is more difficult to describe. We use the electronics to make music because we want to make music and the electronics is available to us. That is, the existence of a technique or piece of hardware, or the potential of such, is enough to start us, (in fact, virtually to compel us), to try out the ideas that come to mind. This is in fact enough motivation, and has always been enough throughout the history of all the arts. A new media for expression will always draw some artists to it.

From a more pragmatic point of view, the employment of electronics may be a virtual necessity for an advancement of the musical arts. It is difficult to imagine the state of the musical arts today if some form of electronics were not employed. In particular, the electronic recording, reproduction, and transmission of musical compositions has brought music in a quality and variety many of us would never have experienced were it not for electronics. Consider which of your favorite compositions you would choose not to have heard at all if it were necessary to experience music only in a live setting. Certainly music is "better" when it is heard live in a concert hall, and there might be more concerts available to us if we had no electronic alternative, but it could not make up in quality for the variety and convenience we get from electronic recordings.

It is difficult to guess at what a world without electronically recorded music would be like, just as it is difficult to suppose what any area of human experience would be like today without modern electronics. The question is of course academic, but the point is to indicate that electronically produced music is only somewhat newer to us than electronic recording. It is upon us at least in part without our making the choice. The incorporation of new technology into art has been with us since the start of human history. Whether the impetus to merge the technology with the art comes from the art side or the technology side is a matter of degree in the individual case, and the result is much the same in either case.

Thus we are faced with the reality of the electronic synthesis of musical sounds. The question is then to consider how best to use this technology. What are our musical needs? What are the problems in applying the technology (practicalities and costs) relative to successes and failures of techniques as judged by whatever objective or subjective standards we may be able to develop? How well does an electronic technique allow us to expand on past practices (imitating traditional acoustic type sounds and/or filling out traditional musical forms)? What sort of promise does the technique have for entirely new musical practices? All these questions should be considered in the evaluation of a potential music synthesis technique.

1.1b GOALS FOR SYNTHESIS SYSTEMS:

In choosing the title for this book, we tried to be careful so that the title did not suggest contents that we did not intend to cover. In the "Theory and Practice" part of the title, we suggest that we are going to study the basic theory behind the techniques, and this is always useful for understanding as much as possible about the techniques so that the practice parts can be better implemented and expanded upon. In the second part of the title, we talk about "Musical Sound Synthesis." This is in our view quite a bit different from "music synthesis." The implication is that we want do discuss how the sounds, the raw material of music, can be synthesized. This is a matter of electronics, physics, math, and psychoacoustics, and at least several other matters, largely technical. The process of "music synthesis" is often used in a manner synonymous with what we call "musical sound synthesis," but we choose to make the distinction. If we were really going to discuss the synthesis of music rather than the synthesis of musical sounds, we would be writing a book on composing. By way of analogy, if this book were being written a few hundred years ago, it would probably have been on "violin making" or something of that nature, not on playing the violin or on composing for the violin.

A bit more on the idea of the word "practice" in the title should be made. We are not going into a lot of detail on the actual circuitry involved in the processes we describe. For one thing, circuit details just change to rapidly as new electronic devices come out, and we would soon be out of date. For another, a frequently updated publication, the Electronotes Builder's Guide and Preferred Circuit Collection (EBG&PCC) is available offering such detailed circuitry. Thus the term "practice" is used in a sense just above the actual circuit level, in a manner somewhat common in recent books on electronics.

In the same sense that we have certain goals for this book, we will be writing this book with certain goals in mind for the synthesis equipment and processes to be described. As suggested above, the systems are capable of producing musical sounds. Whether or not the same type of organization and thinking can lead to the production of actual music (i.e., to composition) is an open question. We will not discuss this except to say that no person who is reading this book for technical information, thus considering himself more engineer than musician, should suppose that it is not possible for him to also compose music. One of the true strengths of electronic synthesis techniques is that they can put a practical instrument on which to compose into the hands of anyone who wants it (just try to get your own orchestra of 100 musicians!).

In saying that we are going to discuss the synthesis of musical sounds, it seems imperative on us to say what a musical sound is. How can we discuss how to produce something we have not defined? Well, the truth is that it is not possible to define musical sounds with any degree of precision that would be likely to satisfy a reader who is probably technically trained. We might say that it is an electronic sound produced in imitation of a traditional acoustic type sound, or that it possesses some property or properties that a traditional acoustic type sound does. This is a useful starting point, to show that our systems at least have a capability of achieving a base set of sounds, but it is not enough. For one thing, we are looking for new sounds, not just an alternative way of making old ones. For another, we do not really understand completely what the properties of traditional acoustic type sounds are. Thus it is encouraging if a synthesis machine makes familiar sounds, but it should also be open ended so that it reaches at least into regions where a sound would not be judged musical by traditional standards. Someone may (someone will) want to expand music into this region.

To summarize the goals of the type of system we have in mind, we say that it should be practical, it should be capable of working in traditional areas of music to a dagree, and it should be capable of producing sounds of a new sort for possible use in musical compositions.

1.1c PRODUCTION OF GENERAL TYPE SOUNDS:

It might be supposed that the most desirable type of music synthesis system would be one that is capable of producing a completely general sound. This is not true, or at least it is not practical to realize such a system. In fact, claims have been made that this or that system is capable of producing any sound ever heard or any sound that ever will be heard.

This claim is almost certainly false for any synthesizer of the voltage-controlled type. Such a synthesizer (the most common type, generically known as the "Moog" synthesizer, but made by dozens of companies now) is composed of functional blocks each of which produces a waveform or processes it in a manner controlled by one or more control voltages. These control voltages are the "parameters" of the system, and a practical number is in the range of five to fifty or so. Through the manipulation of these parameters, such a synthesizer is capable of producing a wide variety of sounds, but the synthesis of a truly arbitrary sound is just something we do not know how to do in the first place, and don't expect to be able to do with a small set of parameters. We can achieve sounds of great variety, of complexity, unexpected sounds, but the exact realization of a desired sound is something we can not always do. This limitation, which is <u>not</u> a serious limitation in many applications, is evidenced by the fact that we produce "trumpet-like" sounds and "clarinet-like" sounds, not sounds that are indistinguishable from the original we have in mind. We can come close, but not <u>arbitrarily</u> close.

When we consider the claim of complete generality for a computer type of synthesis system, the validity of the claim with respect to theoretical limits is clear (the claim is valid), but with respect to practical limits, the validity is quite obscure. By a digital or computer system we have in mind one that is capable of producing a sound from digitally generated "samples" in time. A voltage-controlled synthesizer produces sounds by processing continuous waveforms, but a digital type synthesizer puts the waveform together bit by bit. In fact, this is the only way the digital system works. It can be shown that if we assemble in the digital memory of the device, enough samples of enough accuracy, we can reproduce (and by implication, synthesize) <u>any</u> sound. This demonstrates theoretical capability. However, two additional points cloud this simple picture. First, the number of samples required for any second of sound is about 30,000, and we also need perhaps 10 or 12 bits to represent each digital word (to represent each sample's value). Thus we need about 300,000 bits per second. That's a lot. If digital storage is cheap, we can do this when the samples come to us for little or no effort (when we are recording for example), but if we need to <u>synthesizer</u> these 300,000 bits for each second, we have a problem. It becomes a question of do we <u>want</u> to do all this work for a second of music, and do we <u>know</u> <u>how</u> to do it even if we want to? Probably not. The second point that clouds the picture relates to the rate at which a human being can process information presented (the rate at which the brain can handle information usefully). This rate is something like 40 bits per second maximum. Thus the general digital system is somehow "overspecified" by a factor on the order of 10,000.

In consideration of the digital system in the paragraph above, we found that complete generality brought with it the responsibility for assembling very large amounts of data, more than is practical, and more than is necessary from a brain rate of processing point of view. Thus we should never see digital systems at the present time, but we do see them. The reason is that digital systems are also capable of being programmed, and we can write programs to assemble portions of data in an automatic manner. Thus the user assembles his sounds by specifying parameters which are then turned over to the program, which then assembles the 300,000 bits per second needed. Is this a way around the original problem? Yes and no. Yes the machine becomes practical, but no we still have an overspecification, and certainly we sacrifice the original generality, because we can now only assemble sounds in a way that the programs allow.

Thus a voltage-controlled synthesizer is inherently limited in its generality, and a digital (computer) type synthesizer is practically limited, and surprisingly, limited to a similar degree. Perhaps this is no real surprise. Both types of machine have evolved to a degree of generality (or lack thereof) that produces sounds of a variety, and for a degree of effort, that is compatible with the needs of the user (composer). There is only a certain degree of generality that an average composer needs, and only a certain degree of effort that this average composer is willing to go to for a second's worth of music. Successful synthesis machines thus are produced to meet these needs.

1.2 TYPES OF ELECTRONIC MUSIC SYSTEMS:

1.2a CLASSIFICATION OF SYNTHESIS SYSTEMS:

Electronic music synthesis systems generally can be classified as "analog" or as "digital" systems, and there is often an associated synthesis technique that goes with them. Generally, analog synthesizers work in a voltage-controlled mode and employ "subtractive synthesis," which is the filtering down of a harmonically rich waveform. Digital systems on the other hand generally work on what would be considered "additive synthesis," the building up of a waveform from component frequencies. Both types of system may also use "modulation synthesis," the enriching of a waveform by modulating it and listening to the result (i.e., not demodulating it). Further, most analog systems are capable of a limited degree of additive synthesis, and many digital systems are capable of some subtractive synthesis by making use of digital filters programmed in. There are no real rules however - most things that can be done in an analog manner can be done in a corresponding digital manner. Most things that can be done in a digital manner can be done in an analog manner, although perhaps with limitations. No system is limited by theory to any one synthesis technique, and there are no rules that say that you can not employ two or more techniques at the same time. For example, we can modulate a voltage-controlled oscillator, producing a complex set of component frequencies, and then filter this subtractively.

To a degree, analog systems are considered to be the standard voltage-controlled or "Moog" synthesizer. By the same degree of grouping, digital systems are considered to be "computer music" systems. Analog systems thus have a starting commercial price in the range of $1000, and perhaps as low as $200 for certain simple kits or for other home-built alternatives. Computer systems on the other hand have an implied set of instructions that begins with "first build a computer system" which is quite a bit more expensive when the various needed peripherals (terminals, CRT's, disk memories, D/A converters) are counted. Probably at least a factor of 10 difference is reasonable to consider. Yet this is not absolute, as the most costly analog system (over $50,000) is quite a bit more than the least expensive digital system (under $30,000). Then when you look at the details of the systems, you will probably find a lot of digital elements in the analog system (particularly with regard to control), and the digital system may well have given up a lot of its original generality so that its features resemble quite closely those of a corresponding analog system. In short, you have to look closely at the individual system for its features, its method of implementation, and its costs.

Another way of classifying systems has to do with the way it is used. There are "studio machines" and "concert machines." Again the distinction is not clear with regard to the actual pieces of machinery gathered together, but rather the distinction is with regard to portability (concert) or lack thereof (studio) and association with quality recording equipment (studio). Some concert setups may involve just one $1000 voltage-controlled synthesizer. Other more affluent performers may take on the road more equipment than some university studios possess. Many academic studios are computer based, although a few computer based systems are capable of going on the road. Some small concert or "performance oriented" machines will be all that is needed in some studios that are mainly aimed at recording bands with many other instruments, or will be all that the studio has invested in so far.

A discussion of analog versus digital should also mention speed of operation at this point. An analog system essentially runs freely, producing its waveforms and processing them in real time with no problem. There is thus no problem of working in real time, and no thought that it could be otherwise. Computer music systems on the other hand have traditionally been out of real time. This is easily understood when you consider the 30,000 samples that must be produced for each second of music. If we had to enter these in by hand with a terminal, even if we could enter them at one per second, it would take over eight hours to enter one second's worth of music. This is another aspect related to the practicality of realization that we discussed above, and is why we needed programming to automatically assemble the sets of samples from a few guiding parameters. Let's assume that we can enter the parameters in real time (thus we essentially "play" the instrument). Now we are limited by the speed with which the computer program can automatically assemble the data sets. For 30,000 samples per second, it has to have a speed to produce one sample every 33 microseconds. Since the generation of a sample would involve the examination of the current parameters, arithmetic calculation of the current values of the sample, various checks of the program's progress, and the outputting of the sample for playback or for storage, we can see that 33 microseconds is not all that long. If the program is sophisticated, we won't have enough time. If the program is not sophisticated, we will find our data sets too trite, and any resulting musical sounds too boring.

Thus we often consider computer music to be out of real time by a substantial amount, perhaps taking 10 to 100 seconds to calculate numbers for one second of music. Such an instrument is not playable. You have to enter parameters, and then wait. Such a system is impossible in the traditional concert, and is limited to studio work. Yet not all digital systems are out of real time nor need they always be out of real time as system speed improves in the future. Either through the use of restrictions, through the use of simpler programming, or through the use of digital structures in *only some places, systems properly termed "digital" can work in real time*. Thus, again you have to consider the individual case.

One particularly interesting system is the hybrid analog/digital system where control is digital, but the actual generation of tones is analog. You can think of it as a computer playing a voltage-controlled synthesizer if you wish. In this case, programs can be very sophisticated, as the computer needs only generate and update a relatively small number of parameters each second. Updating 50 parameters 50 times per second is 2500 numbers each second, or 400 microseconds for each number. If we have fewer parameters, or if we do not recalculate parameters that are not changing, of if parameters only change at certain predetermined times (predetermined by the rhythm pattern of the composition), calculation times of many milliseconds are possible. If there is a maximum of 16 notes per second and 10 parameters per note, calculation time would be 1/160 = 6.25 milliseconds. Thus the hybrid approach can make a lot of sense, and its attractiveness is further enhanced by the recent release of a number of dedicated electronic music integrated circuits which make the analog part of the circuitry less expensive and much easier to build, making parallel multi-voice (polyphonic) systems quite practical.

Thus we have classified systems as analog or digital, as to their main technique of synthesis (additive, subtractive, modulation), by their size and application, and by their speed. *It is also possible and important to classify them according to* their "features" and the way they may be utilized. A voltage-controlled synthesizer with a sample-and-hold feature is different from one without this feature. A system that may be "patched" (interconnected with external "patch cords") is different from one which is "prepatched" or "hard wired" into a factory determined configuration. Many of these features and methods of utilization become involved with the intent of the user, the type of music that he intends to produce, and the availability of other equipment. Thus "user oriented" considerations appear, and the "human engineering" of the system must not be neglected. A system might appear quite proper on paper, but be difficult to actually use. It should not be surprising that such descriptions are not very standard, are warped by advertising jargon, are highly subjective, and a matter of personal taste. Here the individual builder has the advantage of understanding his system much better, and being in a position to change things.

1.2b VOLTAGE-CONTROLLED SYNTHESIZERS

Prior to about 1965, music synthesizers were very rare, and were either based on computer generation or on special purpose expensive hardware. After 1965, a new concept in music synthesizers appeared, the voltage-controlled synthesizer as developed by Robert Moog, Alan Pearlman, Donald Buchla, and others. The concept of voltage-control as originated by Moog was arrived at in an indirect manner in the sense that it was first chosen as a means of control, and then was found to have some very useful extensions. In particular, it not only served as a means of implementing manual control, but by its nature could accept control from any suitable voltage, including voltages too rapidly varying, to complicated, or too precise to be controlled manually.

It will be useful here to consider an example, the case of a voltage-controlled oscillator (VCO) as shown in Fig. 1-1. The oscillator has a frequency that is a function of the control voltage. Many VCO's are linear, but the ones most useful for electronic music have an exponential function relating the frequency to the control voltage typically as:

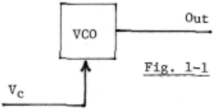


Fig. 1-1

$$f = f_o 2^{V_c} \qquad (1\text{-}1)$$

There are numerous advantages to using the exponential relationship which will be discussed in a later section. Note that when $V_c = 0$, the frequency is not zero but rather $f_o 2^0 = f_o$. If the voltage V_c increases to +1 volt, the frequency becomes $f_o 2^1 = 2f_o$, and if V_c becomes -1 volt, the frequency becomes $f_o 2^{-1} = f_o/2$. It is evident that any change of one volt in V_c will change the frequency by a factor of two, which is a musical octave. Hence the standard electronic music VCO has a one-volt-per-octave control relationship. [Note that octave in music refers to eight notes: do, re, me, fa, so, la, te, do, and that's where the eight in the word octave comes from].

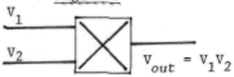
We can assume for now that the control function is ideal and as given by equation (1-1), and now to know what frequency the VCO is producing (thus, the musical pitch), we need only know how V_c varies in time. A typical source of V_c would be from a musical keyboard. A standard musical scale is the 12-tone "chromatic" equally tempered scale, which is 12 notes equally spaced within one octave. For the C Major scale, these notes are C, C#, D, D#, E, F, F#, G, G#, A, A#, and B. Each of these notes is spaced above the other by the same ratio, the twelfth root of two. The exponential function of equation (1-1) does this exactly. Thus if we arrange to have the keyboard output voltages that are 1/12 volt apart between keys of the chromatic scale, we get a VCO tuned to a standard keyboard. The user pushes the key he wants. This is of course manual control. Another type of manual control would be to have a pot connected between two voltage levels, and the wiper of the pot could then be fed to supply V_c. As the knob of the pot is turned, the pitch changes continuously. In fact, all electronic music VCO's have this pot to set an initial pitch for a keyboard, and it can serve as a manual control.

▷ Now, the importance of voltage-control will be understood when we consider that a voltage V_c does not have to come from a manual source. For example, it could be supplied from a D/A output from a computer. This is useful because a computer can come up with a number and feed it through an output port to a D/A, but it can't very well push a key on a keyboard or turn a knob. Another useful control voltage source is to use any voltage available in the voltage-controlled synthesizer. Typically the synthesizer will have three or more VCO's in it as well as other "modules" to be discussed. Thus, for example, one VCO can control another. Thus as the voltage of the controlling VCO goes up and down, the pitch of the controlled VCO goes up and down. Clearly we can turn the frequency of the controlling VCO up to a point where it varies the controlled VCO faster than we could expect to do with manual control. We might be able to turn a knob up and down five times per second, but what if we want a variation of 100 times per second? Thus we see that voltage-control offers an extension to our manual control. We can set processes going automatically. We can achieve control of many parameters of the sound, not just those from our two hands. We can turn control over to devices beyond our reach. We can sum controls, manipulate and process them. We can store controls and recall them. We can "fan out" controls for use in controlling several parameters. We can cascade controls (one voltage controls a module which produces a voltage which controls another module). Any voltage in the system can be a signal (a potential sound) or it can be used as a control. There are no rules. Everything has been put on a more or less equal basis.

While we have looked at the VCO as an example, there are many types of voltage-controlled modules that are common, and still more that are experimental or available home built. Voltage-Controlled Amplifiers (VCA's) and Voltage-Controlled Filters (VCF's) are almost always found in the synthesizer, the former controlling the amplitude of the musical signal (or some control) while the latter performs a filtering, thus achieving the subtractive part of subtractive synthesis. Other modules commonly found include Envelope Generators (EG's) which provide control-voltage contours when properly started. Noise sources and sample-and-hold units are also standard. These and others will be discussed in detail later.

One item which we have not listed yet is the Balanced Modulator (BM). This device is properly called a voltage-controlled device, but it also relates to the modulation method of synthesis, and also, it is of a nature that we want to examine carefully. The BM, also called a "ring modulator" is really nothing more than an analog multiplier. Fig. 1-2 shows a symbol for a BM as a multiplier, and note that the output of the BM is the algebraic product of the input voltages. It is a "four-quadrant" multiplier because it is capable of multiplying the voltages and returning the correct algebraic sign. Thus two positive voltages or two negative ones when multiplied yield a positive output, while a negative and a positive one when multiplied yield a negative output. Which input to the BM is the input and which one is the control? It does not matter - both are essentially the same. The difference is more in the use. In its most typical application, two audio signals are input to the BM, and the product forms a new signal. We will see later that this is a modulation, and like all modulations it produces a structure of frequency "sidebands" which offer us sounds not composed of harmonics of one fundamental. Thus the BM offers us something new, since the outputs of a VCO are periodic waveforms and as such all are composed of harmonics of one fundamental (the fundamental being the frequency of the VCO while harmonics or overtones are due to the fact that the VCO usually offers waveforms other than pure sine waves).

Fig. 1-2



Balanced Modulator

In another application of the BM, we can consider one input as an audio signal and the other one as a slowly varying control, zero or positive let us suppose. While we have not explained much about a VCA, which we said controls amplitude, it *should be clear that the BM used in this way performs amplitude control as well.* In fact, a BM can be used as a VCA, although it is usually better to use an actual VCA which is a "two-quadrant" multiplier, since the VCA has better rejection of the input when total shutoff is desired.

The BM or the VCA are examples of what we can call a "Controlled Gain Block" (CGB). The CGB is in many ways the essence of voltage-control. VCA's and BM's are formed from CGB's (two- and four-quadrant multipliers respectively in this case). VCF's generally have at least two CGB's internally in them, and VCO's may use one as well. The fact that the CGB is so important should be no real surprise. If we are going to control parameters by feeding controlling voltages to a module, then there must be something inside the module that responds to this control voltage, and alters its electrical properties accordingly. The CGB is the ideal device for this purpose. It takes an input voltage, examines the controlling voltage, and changes the level at the output accordingly. CGB's are generally linear. If you change the control voltage, the gain (ratio of output to input) changes in direct proportion. CGB's can all be considered analog multipliers of one form or another. One popular form is the Operational Transconductance Amplifier (OTA) which is actually a two-quadrant multiplier. One input accepts a bipolar voltage, while the other input is actually a unipolar current. The output is a bipolar current. We are not concerned that a device responds to or produces current rather than voltage, as conversions between voltage and current are extremely simple (essentially resistors do this in accordance with Ohm's law!).

The essential things to grasp about voltage-control are that it provides us with new forms of manipulation beyond what we can do manually, and that a voltage-controlled synthesizer is an interconnection of various voltage-controlled blocks or modules, with some voltages serving as signals and others as controls. Later a full section will explore the voltage-controlled synthesizer in more detail.

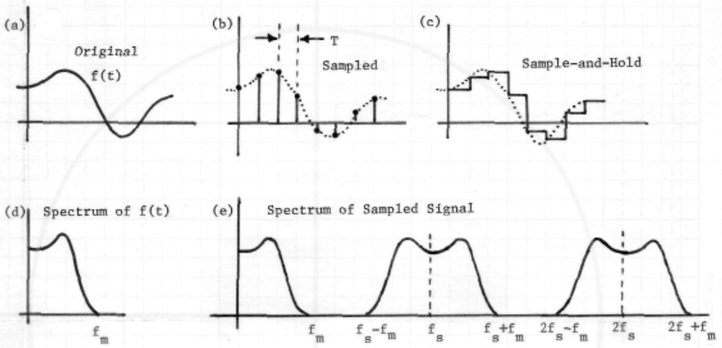
1.2c DIGITAL GENERATION OF SOUNDS:

The generation of waveforms by a digital means is radically different from the generation by analog means, although the end result may be indistinguishable for most practical purposes. Perhaps it is best to make the true distinction between the methods in the way they treat time. Analog methods treat time in the ordinary way as going on continually. Thus an analog oscillator for example will give you a waveform that has a value at the present time, and for all later times. Digital systems on the other hand treat time as existing only for certain points, usually equally spaced in regular time. This is a radical difference, since these points are infinitesimally small. Essentially a waveform then becomes a series of samples - numbers that represent the value of the waveform. We can understand that this is necessary from the nature of a digital system. In a digital system, we have only logic zeros and logic ones, and these internal zero/one states are organized in certain ways to form logic conditions, multi-bit binary numbers, etc. Such a digital system is defined by the existence of "states" which are determined by the internal pattern of ones and zeros. If these states are not stationary at any one instant (as when one or more zeros are changing to ones and vice versa), then the digital machine is not well defined. To be useful, we must obtain information from the machine only when it has had a chance to settle firmly into a well-defined state. Thus we can remove information only at a certain rate, and in between, we must assume that the state is a constant. Thus there must be a time between the outputs, and effectively the value of various outputs exists only at the readout times.

Fig. 1-3 shows in (a) an analog waveform f(t), while in (b) we see the result of sampling the waveform at time intervals T. Note that we show sampling as point samples - time exists only at discrete points, and notin between. In an actual recovery process, the waveform is actually sample-and-held as in (c), and this is mathematically different, so we must be aware of what the actual case is.

The reader probably knows that a signal can be described in terms of its time variation, or in terms of its spectrum or frequency content. For example, the spectrum of the continuous signal f(t) in Fig. 1-3 (a) might be as shown in Fig. 1-3 (d). We then can ask what the effect of sampling is as viewed from a spectral point of view. We will discuss this more later, but the effect of sampling is to cause the spectrum to repeat, being reflected about multiples of the sampling frequency $f_s = 1/T$. Thus the spectrum of Fig. 1-3 (b) would be as shown in (e). If we suppose that the signal f(t) has no frequency components above f_m, then we can arrive at a simple formulation of the "sampling theorem." In Fig. 1-3e we do not want the original part of the spectrum (the "baseband") to overlap with any upper part of the sampled spectrum. If it does, when we try to recover the original signal f(t) from the sampled version (which we need to do to actually hear it) using a low pass filter, we will get some interfering signal components, generally appearing in new positions of the original spectrum. These can be totally destructive to a good audio signal. How do we avoid the overlap? From Fig. 1-3 (e) we see that

Fig. 1-3



there is no overlap in the spectrum if $(f_s-f_m) > f_m$, which can also be written as:

$$f_s > 2f_m \tag{1-2}$$

This important result tells us that if we sample at more than twice the frequency of the maximum component of the input, we will not have any overlap. The reader can then understand how we could recover the original spectrum by using a sharp low-pass filter, filtering Fig. 1-3 (e) with a low-pass cutoff at $f_s/2$ yields Fig. 1-3 (d). Before going on, we need to mention how the use of sample-and-hold Fig. 1-3 (c) rather than sampling as in Fig. 1-3 (b) changes the picture. Basically we have to sample-and-hold to recover the signal. To see why consider that the point samples of Fig. 1-3 (b), while representing the correct mathematical picture, are infinitesimally narrow, and thus contain no useful energy. If we low-pass filter, we get nothing out. Thus we must sample-and-hold (or use some better interpolation between point samples). The effect of the holding action is to give a spectrum similar to that in Fig. 1-3 (e), except there is also a 1/f term that multiplies the spectrum. Thus exact reconstruction can be a bit trickier than we might suppose in the simplest case. Yet this does help us to understand the repeating spectrum of Fig. 1-3 (e). We tend to think not of sampling, but of sample-and-holding, and intuitively we do not expect a waveform such as that in Fig. 1-3 (c) to have a spectrum that does not fall off with frequency. Thus the somewhat artificial appearance of Fig. 1-3 (e) can be understood not as a realizable spectrum, but as the *mathematical equivalence of a somewhat artificial sampling process of Fig. 1-3 (b).*

So far we have talked about sampling of an existing analog waveform f(t). What we really need to consider is how to digitally synthesize an appropriate set of samples and then recover an analog waveform. As discussed earlier, the synthesis of the set of samples is a matter of allowing various programs to organize them guided by some input parameters. Thus we consider that the samples exist, stored inside the computer memory. Now, the spectrum of these is periodic, as in Fig. 1-3 (e), if for no better reason because the computer does not "know" that it did not get these samples by sampling a live input f(t). However, there is no problem with sampling fast enough, the sampling rate is to be determined by the rate at which these will be read out. What we do know is that if we read out at f_s, then we must filter at $f_s/2$, and thus f_s must be at least twice the frequency of the maximum frequency component we wish to synthesize. Note that a sample-and-hold will normally be used, or will be present as a latching of a D/A converter.

Before going on further, we can say a few words about how a computer program that assembles digital data might work. *First, the program would have a list of the* parameters to produce for a given "note." We say "note" because it may be a more or less conventional musical note, or it may be a somewhat more arbitrary type of sound. In any event, the computer has a list of specifications for a sound event that is to be produced. This might include things like the pitch level, the waveform type, and the type of amplitude envelope to be used, including the duration of the sound. These parameters will direct the main control program to go to appropriate subsections which contain the appropriate program statements. For example, the waveform that is desired might be a sawtooth, or the digital equivalent which is a staircase wave. To generate this, the program would start with a number and then store it. Before going to the output, the program would then look up the current value of the amplitude envelope, and multiply this step in the staircase by this number. The sample is then ready for output. Next the program returns to the value of the sawtooth (staircase) and increments it by one step, goes on to the next value of the envelope, and so on. At a certain point, a number value is reached for the staircase that indicates its peak, and the staircase is then reset to its starting value. Thus there is a check at each step to see if the staircase must be reset. Other checks are needed or may be needed to check if the note is supposed to be finished. When it is, control goes back to the control program to see what the next note is to be.

The above example is probably oversimplified, but at the same time, the reader can see that even for this simple case, a fair amount of programming is required. As the program gets more and more sophisticated, it becomes longer to run. Thus it is not uncommon to find a program with relatively few options, and computer music often sounds lacking in variety, and listeners familiar with the various programs are able to tell which program was used just by listening. One solution is of course to leave more of the programming to the user, so that variety is available. The reader may also note that since the program is controlled by parameters, we can often compare this parameter set with the controllable parameters of voltage-controlled synthesizers. We often find systems quite similar in scope.

We have probably given the impression that computer music is inefficient relative to analog methods, and that users will find it more difficult to compose with. While this is probably true in a general sense, certainly there are composers who prefer the computer, and there can be little doubt that the digital technique will continue to improve. Already machines of attractive capabilities and acceptable size exist, being limited mainly by high cost.

1.3 TECHNIQUES OF MUSICAL SOUND SYNTHESIS:

1.3a ADDITIVE SYNTHESIS

While we are reluctant to try to define a musical sound, we need to have something to begin with when we consider techniques of sound synthesis. The way that immediately comes to mind is to first choose a sound we agree is musical, and then analyze it to see what its structure is like. Analysis could logically be either in the time domain (as we might look at the waveform on an oscilloscope) or it could be in the frequency domain (we would use a spectrum analyzer). Suppose we look at the waveform on a scope. What do we learn? First, we would obviously learn something about the waveform itself, and have some idea how to resynthesize the sound. We would just duplicate the waveform (this might not be at all easy). We would also learn some things about the general nature of musical sound waveforms - mainly that they are very complicated and difficult to pin down in the time domain. In short, we would learn mainly that time analysis probably is not the way to go.

We would then probably try a spectrum analyzer, and might come out with something like that shown in Fig. 1-4. In this view, we show a three dimensional

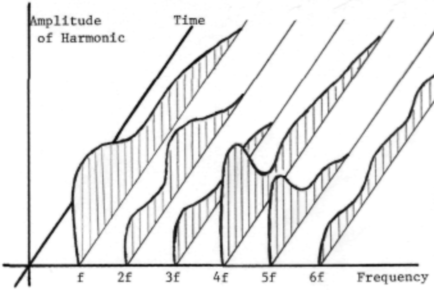


Fig. 1-4

plot showing frequency, amplitude, and time. We see several things from our drawing, which might represent a standard type of note (a pitched sound from a conventional acoustic instrument). First, the sound consists of discrete frequencies that are harmonics of a frequency f. Secondly, each harmonic varies in time, and each with a different amplitude-time contour. We can also relate this a bit better to what we hear with the ear. We recognize the lowest or "fundamental" frequency as giving the musical pitch, and we recognize the presence of harmonics as supplying "overtones" to the sound, resulting in a brighter tone color.

At this point, a word should be said about spectra. The one such as we have indicated in Fig. 1-4 is really an amplitude spectrum. It gives the amplitude of the supposed sinusoidal components that make up the waveform. Yet the amplitude *spectrum is not a complete description - we could not reconstruct the waveform from* the amplitude spectrum. We would also need the phase part of the spectrum. That is, we need to know how the sinusoidal components are lined up in phase. Now, here is a key point which has been established by studies of hearing. The phase does not matter. At least not much. As far as the ear is concerned, the sound is determined by the amplitude spectrum alone. Two points should be noted here. First, the fact that the spectrum (meaning amplitude spectrum here, and usually later on) is an incomplete description mathematically, but one that is adequate for the ear, makes it ideal for our purposes. Secondly, it is because the details of the waveform do depend on phase that a description in terms of waveform is very complicated.

Having more or less determined that we want somehow to synthesize the general form of an time dependent amplitude spectrum, we need to think about how best to do this. One way to attempt to synthesize the spectrum of Fig. 1-4 would be to start with six oscillators at frequencies f, 2f, 3f, 4f, 5f, and 6f, and input each of these to its own voltage-controlled amplifier. We would then go to some trouble to generate the contours that represent the amplitudes, and feed these to the VCA's. The outputs of the VCA's would be summed (see Fig. 1-5).

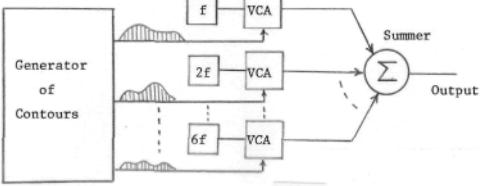


Fig. 1-5

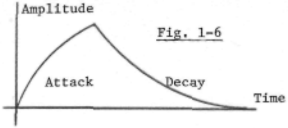


Fig. 1-6

Thus we arrive at what appears to be a reasonable approach to additive synthesis, at least as viewed from the possibility of achieving a successful sound. What are the practicalities? First, there is the problem of tuning the oscillators f, 2f,...6f, and changing them all in parallel when we want to achieve a note of a different pitch. Thus these oscillators must really be VCO's, so for n harmonics, we need n VCO's and n VCA's to start with. A more serious limitation has to do with the generation of the contours or "envelopes" for each harmonic. The common type of envelope that we have available in voltage-controlled synthesizers is of the form shown in Fig. 1-6, a simple "AD" (attack-decay) type formed from the exponential charging of a capacitor. In many cases, the shape of Fig. 1-6 will be a reasonable approximation to the contours of Fig. 1-4, especially the decaying exponential part. [Note that the attack and decay times can be varied independently over a wide range, and need not be exactly as in Fig. 1-6.] Thus, the general form of Fig. 1-5 makes an interesting instrument when the generator of contours is a bank of six envelope generators. The lack of precise tuning between the six oscillators can result in a more natural sounding tone, unlike the case of subtractive synthesis where a single waveform (all harmonics precisely set) is filtered dynamically.

However, it is probably also evident that there is a fair amount of hardware involved if we do this with analog elements. Thus it is probably the case that a more useful realization strategy for additive synthesis would involve digital generation. In such a digital scheme, the summation of the various harmonics as controlled by their contours is just a matter of mathematics. The number of harmonics can be doubled easily if desired, with only a doubling of computation time as a penalty (as compared to buying twice as many analog modules). Digital methods will also offer much more sophistication in accurate generation of the contours. We can conclude that additive synthesis with analog elements is a useful idea on a small scale, but if a relatively involved system is contemplated, probably a digital computer music system should be employed.

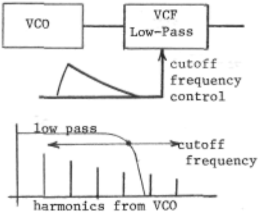
1.3b SUBTRACTIVE SYNTHESIS:

We have a hard time not thinking of subtractive synthesis as the more fundamental process as compared to additive synthesis. This is due to its extensive use, and the fact that many of us became involved in electronic music at the time it was becoming practical for the experimenter in its voltage-controlled version. In later sections of this book, we will be doing subtractive synthesis first.

Two things must be understood in order to understand the operation of subtractive synthesis. First, it must be understood that all "complex" waveforms (any waveform other than a pure sine wave) are composed of frequency components that are harmonics of a fundamental frequency. The components are determined by the "Fourier Series" (FS) expansion of the waveform. Secondly it is necessary to understand the way filters work, and to understand that it is possible to understand the filtering of a complex waveform by considering the way each and every component is filtered individually, as though it were the only one being processed by the filter. This is the principle of superposition, which is possible because filtering is a linear operation. Thus you must understand Fourier analysis and filtering.

It is useful to say why subtractive synthesis may be a general sort of process of the type that may result in musical sounds. One reason is that one necessary aspect of a musical sound seems to be a dynamically changing spectrum (see Fig. 1-4). In subtractive synthesis, instead of generating and controlling each component separately, we will be starting with a complex waveform that already has a lot of components, and will then let a filter move over these components to get a dynamic spectrum. Certainly we are free to experiment with this, choosing waveforms and filter types in an arbitrary manner, and this is often done. We can at the same time find some justification for certain successful choices however, based on a *physical idea about conventional acoustic musical instruments. This idea is that* the higher frequency components of a sound will die out more rapidly than the low frequency components. We know this from our physical intuition. If we strike a large bell we get a low frequency tone that rings for many seconds. A very small bell however will ring at a very high pitch, and only for an instant. The same general result is available with a piano. Now, in a mechanical system, which an acoustic instrument is, we can expect that various harmonics will be radiated from different vibrational modes. A mechanical system is not likely to be truly linear, but the dissipation of energy from the various modes is at least roughly independent of dissipation from other modes. Thus we can understand the fact that higher *harmonics will die out rapidly, while lower ones will sound for longer periods of* time. This suggests that the combination of a waveform with a high harmonic content (such as a sawtooth or a sharp pulse) which is then filtered by a low-pass filter, might be used to synthesize a sound somewhat like that of an acoustic instrument (see Fig. 1-7). According to the discussion above, we would want the filter cutoff to move up rapidly at the beginning of the tone, and then fall back as the tone progresses, cutting off upper harmonics one at a time. There are many subtle points and refinements to this picture, but this is the general picture, and it is no accident that many simple synthesizers aimed at the general imitation of acoustic instruments use only a low-pass VCF as their filter.

Fig. 1-7



Filtering has an obvious effect on the spectrum. In fact, the output of the filter is the spectrum of the input as multiplied by the frequency response function of the filter. The filter also has an observable effect on the waveform itself. The effect is not as simple as just breaking the waveform into Fourier components, multiplying each of these by the frequency response, and then reassembling (adding) the sine wave components together. The reason for this is as we have discussed before - phase matters for the waveform shape, but not for the spectrum. All real filters have in addition to their alteration on the amplitude of components, a phase shift that also depends on the frequency. Thus for an accurate reassembly of the filtered components, we would have to take account of the phase as well as the amplitude. However, the majority of the phase shift expected occurs in the same region where the frequency response is changing rapidly. Thus for a very sharp cutoff low-pass filter, those components that are strongly phase shifted are also strongly attenuated, and their effect is minor. We can therefore get a good picture of the effect of filtering on waveshape by looking at the effect of simply removing some components completely from the FS of the waveform. Fig. 1-8 shows the waveforms for various combinations of harmonics for a sawtooth waveform. Similar waveforms can be observed using a sharp low-pass filter and observing the waveform on an oscilloscope.

1.3c MODULATION SYNTHESIS:

Soon after the introduction of voltage-controlled synthesizer modules, it became apparent that modulation, an inherent capability of many of these modules, was going to become an important extension of the subtractive synthesis capabilities. For example, since a VCO produces a frequency in response to a controlling voltage as in equation (1-1), it is correct in all cases to say that the control voltage modulates the VCO, and in the particular case where the control voltage is periodic, we have a more or less standard modulation calculation to do (things are complicated by the exponential nature of the function in this case, but we will get to those details in

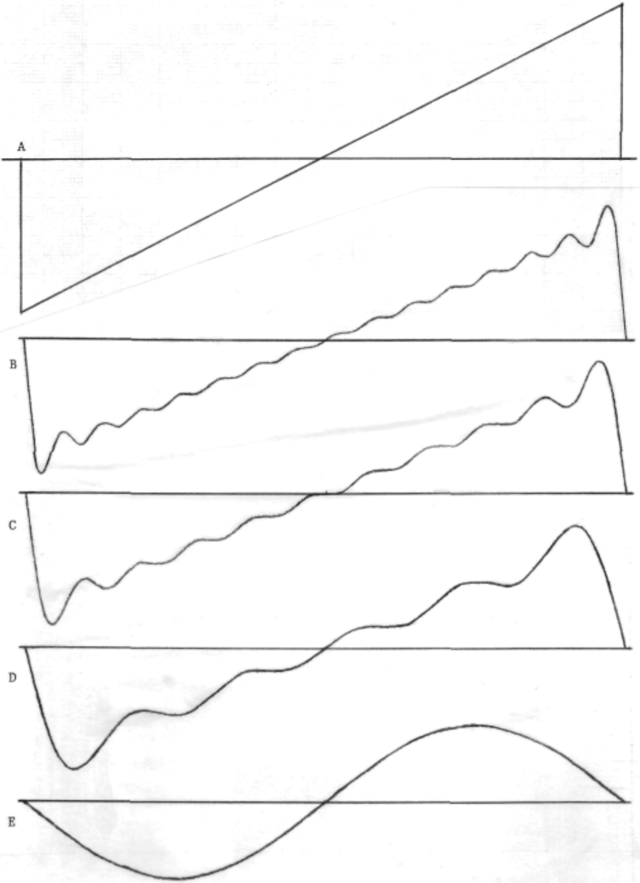


Fig. 1-8 A: Full sawtooth, all harmonics falling off as 1/n. B: First 15 harmonics of sawtooth only. C: First 10 harmonics of sawtooth only. D: First five harmonics of sawtooth only. E: First harmonic of sawtooth only.

when we reach the full section on modulation synthesis. For now we will be looking at the most general results of modulation processes with the assumption made that the modulation process is linear. When it is not, the alteration will be in an unbalancing of sideband patterns. Spacing of sidebands remains the same.

Modulation starts with an original (unmodulated) signal, which is usually called the "carrier." Some parameter of this signal (amplitude, phase, etc.) is then caused to vary in response to another signal called the modulating signal. If the frequency of the carrier is f_c and the frequency of the modulating signal is f_m, then sidebands space themselves at intervals of f_m about f_c. This is true of all signals, be they radio signals, or audio frequency signals. The same mathematical formulation applies. The reader familiar with radio communications theory will basically understand modulation synthesis, but several special points should be made clear. First, we intend to modulate, add information to the signal, but we do not intend to demodulate. We will just listen to the modulated signal directly. Secondly, in the radio case, the carrier frequency is high, in the radio range while the modulation frequency is in the much lower audio range. The result is that we get a grouping of sidebands about the carrier, and thus they too are in the radio range. In the modulation synthesis case, both carrier and modulation frequencies are in the general audio range. Thus we can have sidebands spread over a substantial part of the audio spectrum.

Fig. 1-9 shows a general view of a modulation process where we have one oscillator supplying a "carrier" and the other modulating it. The module doing the processing serves as the modulator. In the case where the module is a VCA, we get amplitude modulation. When it is a balanced modulator, we get balanced (double sideband) modulation, and f_c and f_m are interchangeable. If the module is a VCO, the carrier is generated internally (there is no signal input terminal on a VCO) and we arrive at frequency modulation. Modulation with VCF's and by modulating the pulse width from a VCO pulse output are also common.

Fig. 1-9

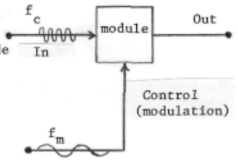


Fig. 1-10 shows the general features of sideband distributions in comparison to a periodic waveform (sawtooth). These sideband distributions are spectra, and we have discussed above that it is spectra that we are really interested in, because this is what the ear really hears. Note that there are several cases of interest. When f_c is not an integer multiple of f_m, the sidebands do not fall in positions that are harmonics of a single fundamental. The result is an "inharmonic" sound, useful for percussive and bell-like sounds, which is not available from a periodic waveform. This result is probably the strongest feature of modulation synthesis as far as an extension of existing technique is concerned. It is also possible to achieve harmonic spectra by making f_c an integer multiple of f_m, and this can be useful for producing more spectral intensity in the middle harmonics, and in ways to be discussed.

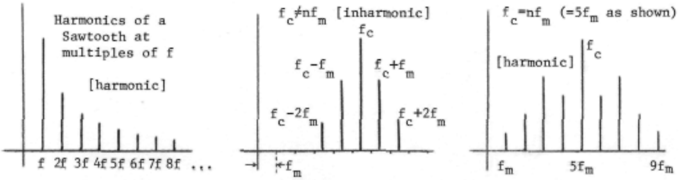


Fig. 1-10 Spectra of Sawtooth and Typical FM Cases

We note that since all periodic waveforms have a spectrum (a Fourier series) that consists of only harmonics of a fundamental, it is reasonable to suppose that a spectrum that consists of non-harmonics will not be periodic. This is in fact the case, and we can think of the non-periodic waveform as the time domain equivalent of the non-harmonic spectrum. We are concerned with the fact that the waveform is not periodic. We are not concerned with the details of the waveform which are only a matter of phase. Note however that while the waveform itself is non-periodic, it still consists of components which are individually periodic, and thus we might expect it to have some sort of a residual musical pitch, or at least result in a sense of pitch motion when the components are all multiplied by the same constant. The sounds may be quite musical in fact, especially in the case where components are vary close to harmonics. In fact, such a case may be more natural sounding to the ear trained on traditional musical instruments. An example of a non-periodic waveform is shown in Fig. 1-11, which consists of a component of unity amplitude at frequency f, and two components of amplitude 0.3, at frequencies 3.1f and 4.1f. [Actually, these all have a fundamental of 0.1f, but this may be too low to be heard as a pitch, or we can easily make the frequencies irrational numbers - the graph would be much the same]. The reader can observe from Fig. 1-11 that while the waveform is definitely not periodic, there are some regularities to it which would certainly suggest a musical pitch. [In fact, musical pitch can result from bandpass filtered white noise].

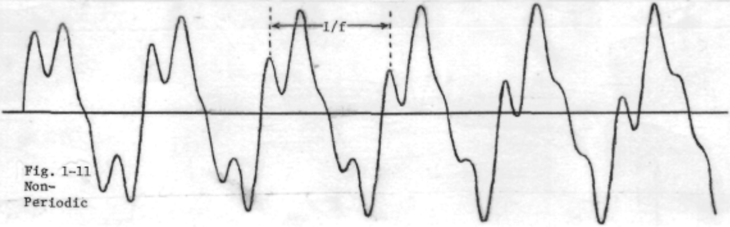


Fig. 1-11
Non-
Periodic

The importance of modulation methods goes further. Two modifications are found to be useful, or at least worth investigating in all types of modulation synthesis. First, the depth of the modulation can be varied dynamically as a tone progresses. This will change the distribution (not the spacing) of sideband energy as the tone varies, resulting in a dynamically changing spectrum, exactly the type of thing we are looking for. The exact implementation of dynamic depth modulation will depend of the type of modulation, and some care must be taken to avoid undesirable secondary effects. Secondly, because of the relative closeness of f_c and f_m, it is entirely possible that sidebands will result that will have negative frequencies (a mathematical result, not an entirely new type of frequency). These appear "reflected" into the positive spectrum, and because their distance from zero, the reflection point, varies, they can appear in strange places in the positive spectrum, resulting in even stranger spectral line distributions. In cases where negative sidebands are not present or present only at insignificant amplitude, it is sometimes possible to modify the modulator so that the modulation process itself passes through the zero point (even though this may seem absurd at first thought). If the modulation process passes through zero, we can think of this as making possible a much wider depth, with correspondingly more complex sideband distributions.

In summary, modulation processes offer us a chance to get spectra where the components are not all harmonics of one fundamental. With the addition of dynamic modulation depth capabilities, and if possible, through zero capabilities, the process can be even further enhanced. Modulation is a somewhat natural process for voltage-controlled synthesizer modules, but digital methods are often found to benefit by solving modulation formulas.

1.4 Additional And Combined Methods:

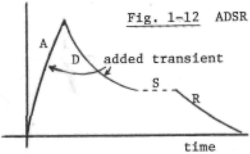
1.4a GENERATION OF TRANSIENT FEATURES

An interesting experiment can be performed in which one first assembles a group of musicians with their traditional acoustic instruments along with a suitable tape recorder. The musician is then asked to hold a long tone on his particular instrument. The person operating the tape recorder will leave the input record level all the way down until after the tone begins. He will then turn the level up, record a portion of the "steady state", and then turn back down before the musician ends the tone. [For some instruments like a piano, there is no steady state since the strings are struck once and their energy then immediately begins to decay. The experiment is still performed in the same manner though.] As a control, the full tones can also be recorded if desired. Now, the tape is played back to a group of persons who are used to telling the sounds of instruments apart, so that normally they would be able to tell a trumpet from a clarinet from a violin, and so on. How well do they perform with the initial and final parts of the tone artificially removed? Well, not too well at all. The task seems very difficult, and some seemingly impossible errors occur (such as mistaking a piano tone for a clarinet) while some perhaps more plausible errors occur quite regularly (mistaking an oboe for a trumpet).

The importance of the experiment is that it shows the important role the so-called "transients" play in our actual perception of the identity of the source of a particular musical sound. This is particularly true of the initial transient (the "attack") and true to a lesser degree for the final transient (the "decay"). There are two points here relevant to our ideas about synthesis. First, if we are going to synthesize a sound in imitation of a traditional acoustic type instrument, we had better pay a good deal of attention to the way we achieve an attack. Secondly, if we are rather looking to just synthesize new musical sounds, we might do well to look at different methods of enriching the amount of information in the attack portion of the tone, as these seems to be of particular interest to the ear-brain.

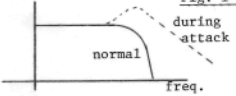
An understanding of musical transients of traditional instruments is not fully developed, and therefore electronic synthesis of transients is imperfect. While many findings from studies of traditional instruments can be carried over to synthesis, it has been mainly through digital synthesis that they have been investigated and even here only superficially. Nonetheless a certain number of successful tricks have been developed which at least attack the problem in the right area. One of the simplest is to modify our envelope generators so that they provide an enhanced feature at the very beginning. Thus we commonly find that envelope generators are of the "ADSR" type for Attack-Decay-Sustain-Release (see Fig. 1-12, as compared to Fig. 1-6). Thus the control from the envelope effectively can be made to go to an extreme initially. This is not a a direct attack on the porblem of course, but a way of using existing processors to add at least something unusual at the beginning. For example, if the ADSR controls a VCA, we at least get a sudden burst of energy at the start. If it controls a VCF low-pass, we get a short burst of extra high frequency harmonics. It works to make sounds more interesting in many cases, although it is not a general way of achieving a given transient effect.

Fig. 1-12 ADSR



Another approach suggested by David Luce is to use a different type of VCF motion. Unlike that of Fig. 1-7 where the response moves up and down, Luce suggests that the response should peak slightly and its cutoff slope should bend upward during the attack portion of the tone. This

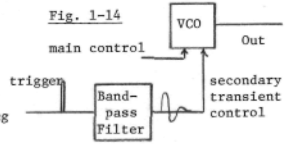
Fig. 1-13



suggestion was based on experimental studies of traditional instruments. Various experimental filters (see Fig. 1-13) have been built and show that the method is in fact useful.

Other methods tend to be in the direction of adding some extra signal, either by direct summation, or indirectly to the control of a parameter, during an attack part of a tone. For example, some transients are quite noisy (such as the sound of air initially blowing across the air hole of a flute or the plucking of a guitar string), so there is some basis for the addition of a burst of noise to the total sound during the attack. A consideration of other instruments shows that the attack involves the setting up of a standing wave in the instrument due to a reflection from the end which returns to the source of vibration. Initially the excitation source may be somewhat off from the final frequency, and it is not until the wave travels to the end and returns that the necessary correction is made (actually the correction is forced due to mechanical considerations). Thus we might consider adding some uncertainty to the initial excitation by putting an appropriate one-time feature on a parameter that is being controlled. This might be a ripple from a decaying rung filter (see Fig. 1-14), a short tone burst, a section of low frequency noise, a short envelope blip, or any one of a number of other possibilities. Most of these methods will offer some improvement or at least interesting sounds for consideration elsewhere.

Fig. 1-14



1.4b ANIMATION OF THE STEADY STATE:

While things like transients are important in adding identity to an electronically produced sound, there is still the steady state portion of the tone to worry about. When people say that a musical tone is electronic, usually in a pejorative manner, what they probably mean is that the sound is too regular and even sounding. Sounds from acoustic instruments tend to have subtle variations even if the performer is trying to hold the tone rock steady. Whether it is just not possible to hold a tone completely steady by manual means, or whether the performer allows or in some unconscious manner causes the variations, is not known. In any event, it is often necessary to take steps to assure that some variations occur in tones of extended duration. As an aside, we can note that many of the most successful synthesizer pieces are those with rapidly moving short notes, not those with long singing-like extended notes.

Certain steps are simple and part of the basic processes we normally use. To keep dynamic variations in the spectrum, we may be able to slow the envelope down so that even for longer notes some changes are going on. Another simple technique is to add a "vibrato" or some similar feature. A vibrato is a frequency modulation of a few percent deviation and at a modulation frequency of from 5 to 10 Hz. The same sort of control can be used to amplitude modulate during the steady state, or a filter can be swept thus assuring a changing harmonic structure no matter how long a note is. This sort of thing is a step in the right direction, and may solve the problem in some cases. However, it should be noted that a vibrato waveform derived from an electronic oscillator is also regular. Thus while we may sense an improvement resulting from variations in the tone during steady state, we soon become aware of an equally bothersome regularity in the vibrato process.

Over the years, many devices have been explored in an attempt to enrich the synthesized tone so that notes of long duration are still interesting to the ear. Such devices as timbre modulators, choral or ensemble simulators, animators, and other devices are employed. Many of these have as a basic operating principle the idea that a regular input can be processed by many separate circuits and then recombined into a single output again. A view of such a processor is shown in

Fig. 1-15, where n processing units P1, P2, ... Pn are controlled by individual low-frequency oscillators v_1, v_2, ... v_n, where all frequencies are different. A typical number for n would be eight. Suitable processors would be phase shifters, delay lines, VCF's, or combinations of these and other units. The general effect is perhaps most like *one of achieving ensemble from a* single source. Instead of one instrument, it sounds more like there are many instruments playing the same note, in the manner in which it is done in an orchestra. Yet it is hard to make any exact statements about the effects achieved, because it is strongly dependent of the nature of the processor.

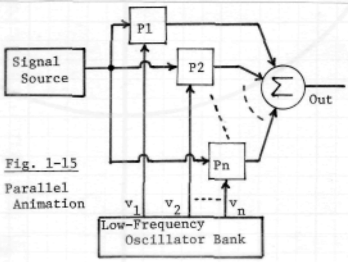


Fig. 1-15

Parallel Animation

The exact way in which this type of parallel animator works is not well understood. It is clear however that if there is only one processing channel, the ear will note the processing, but will easily follow the regularity of the processing pattern (which is due to the low-frequency control oscillator attached to it). If there are two, three, and perhaps as many as four or five processors, the ear may still have a sense that some regularity is present. When you get to eight parallel processors, the pattern is pretty well disguised, and the expected degree of variation is strong enough to be interesting. It is interesting to note however that it is also possible to have too many processing units in parallel. If there are too many, the expected degree of variation is less, as averaging is greater, and the net effect is a step back toward regularity.

1-4c COMBINING METHODS

Nothing prevents us from combining different methods of synthesis in any overall scheme. For example, we can generate a set of non-harmonic sidebands using modulation and then subtractively filter this modulated signal. Some types of animation are something like additive synthesis (see Fig. 1-15 above). Signals produced additively can be amplitude or phase modulated. And so on. Many of these combinations come to the synthesizer user as a natural result of intuition and experimentation. The user with a patchable synthesizer probably forms many of them never having heard of additive or modulation synthesis, and probably not of subtractive synthesis for that matter. In fact, many if not most useful patches that are used on voltage-controlled synthesizers will actually have elements that relate to modulation and/or additive synthesis. Thus combined methods are the usual case. Isolated methods are rare.

1.5 CONTROL:

1.5a MANUAL CONTROL:

For the most part, synthesis systems will not produce usable material unless some control effort is exercised, even if this option is the release of control to a random process. The most fundamental control means is manual, because it is the means used with conventional instruments, and because it could be argued that manual control, or at least human control, is really operative in all synthesizer systems. Manual control as we shall consider it is a matter of employing some sort of physical or mechanical device which is capable of responding to human hands, fingers, feet, lip pressure, blowing of air, etc. These are direct forms of manual control.

Devices such as standard electronic organ keyboards offer a simple means of selecting a given note and determining its duration (i.e., pushing the appropriate key at and for the appropriate time). This is accomplished by first deriving a control voltage by having the key's associated switches select the value appropriate to the keys position on the musical scale, and at the same time deriving "gate" and "trigger" signals which control the timing. Other manual controllers are such things as knobs, which can be turned to provide tones of continuously changing pitch, etc., and two dimensional controllers ("joy sticks") which are really like two knobs with one motion of a stick (up-down say) controlling one voltage, and the other motion (left-right say) controlling a second voltage. Discrete (switched) controllers are in addition to the keyboard, such things as pushbuttons and arrangements of finger switches along a bar or tube in imitation of levers on such instruments as clarinets or flutes, etc. In imitation of wind instruments, we can also devise devices that convert lip pressure and/or air pressure to control voltages, and thus achieve control with the mouth.

An indirect from of manual control is also available through the use of some type of parameter converter. With such a device, a musician is theoretically free to play whatever instrument he pleases. The signal from his instrument is picked up, and parameters (such as pitch, amplitude, timing) are extracted from this signal, converted to control voltage, and used to control the electronics. This sounds good of course, but because of the subtleties already present in the signals from acoustical instruments, the extraction of these fundamental parameters may be a difficult and unreliable task. Thus a universal parameter converter just does not exist. A few pitch-to-voltage converters, useful for a single type of musical instrument do exist and work reasonably well.

1.5b PROGRAMMED CONTROL:

Manual control offers the synthesizer user a means of enlarging his vocabulary of tone colors and sounds. However, the capabilities of the electronics that can be associated with the synthesizer make possible an enlargement in the complexities and structures that can be achieved. This is achieved through programmed control. That is, control voltage sequences, envelope shapes, mixes, etc. can all be set up to occur automatically once set in motion.

In fact, we can think of much of the activity involved with using even a conventional voltage-controlled synthesizer as being programming. Before actually using the synthesizer, the user will set up initial control voltage levels, envelope durations, modulation depths, and so on. This can be considered to be programming since it sets up the device for manual control during the actual playing.

Another type of programmed control is available from the "sequencer" type of device which is available in many forms. Early forms of sequencers had at their heart a digital counter which went through a number of internal states, perhaps in the range of eight to fifty or so. Once set in motion, this sequencer would respond to an internal clock, and output a series of control voltages which the user had preset. Later types of sequencers became more and more sophisticated, offering recording and editing features, and began to look more and more like computers. Thus it was inevitable that the ultimate sequencer would become a digital computer attached to a D/A converter, and also supplying other control and timing signals. Sophisticated sequencers or dedicated controlling computers thus offer the synthesizer user a means of achieving structures of a complexity beyond his own manual abilities, and if desired, beyond those of the most proficient artists or group of artists (a structure of a harmonic or rhythmic complexity that would be unplayable by an individual musician, or impossible to coordinate for a group of musicians, is not difficult to achieve - its musical value may be another matter).

Programmed control may also be found in certain "ornamentation" processes in which a tone is not simply produced, but may have certain ornaments such as trills built into it. Such devices may be associated with transient generation.

1-5c RANDOM CONTROL:

As we see it, there are three types of control: manual, programmed, and random. Manual and programmed controls are things we do intentionally. Random control involves the turning over of part or all of the control to a process over which we have no involvement, or for which we are unsure as to the exact outcome. The reader should be aware of two points concerning random control. The first is that if the control were truly random in all respects, we would get nothing but white noise back. If we modulate an oscillator with noise, we get noise back. However, we may select a voltage sample from a noise signal, and let that control the oscillator over a period corresponding to a single tone, and this achieve what we can call a random pitch. Secondly, the reader should be aware that the use of random elements in music and the other arts is a matter of some controversy, so he himself should be aware of the extent to which control is really turned over to chance.

An example of a random pitch device is the sample-and-hold attached between a noise source and a VCO as shown in Fig. 1-16. As noted above, if the VCO were attached directly to the noise source, we would get noise out of the VCO (unless the modulation depth is relatively shallow). This is because the frequency is trying to change on a basis faster than the expected period of a cycle. By adding the sample-and-hold, we can hold a certain value of the noise for an extended period of time, and we get enough cycles out at a fixed frequency to produce a sense of pitch.

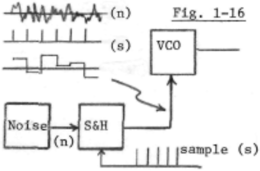


Fig. 1-16

We can ask just how random the series of pitches achieved in the patch of Fig. 1-16 is likely to be. To understand this, we would have to know a fair amount about the statistics of the noise source. What is the distribution of levels like? What is the spectrum of the noise like? Some noise sources may offer all levels between say -10 volts and +10 volts with equal probability for any level (uniform distribution), but it is more likely that a simple device will produce a distribution favoring values around zero (a "Gaussian" or "bell shaped curve" distribution). Thus we might expect pitches about some center value to be fairly common, while extreme pitches would be rare. As to the spectrum, this tells us something about the expected level of correlation between consecutive samples. Thus, given the value of the present sample, we can say that a value around zero is expected next based on the distribution alone, but can we say anything more? If we can, then there is some correlation between components. If the spectrum is truly flat (white noise), we can say nothing more, as the noise is uncorrelated. If the spectrum is not flat, then it favors some frequency components over others. The influence of these favored components has a correlating influence on consecutive samples in many cases. The reader should be aware that often some degree of correlation is highly desirable, and can be added by modifying the sample-and-hold or the noise source spectrum. Samples of truly uncorrelated noise are often subjectively judges to produce unmusical melodies. It is best to have some correlating option that can be added.