ELECTRONOTES 220

Newsletter of the Musical Engineering Group

1016 Hanshaw Road, Ithaca, New York 14850

Volume 23, Number 220 January 2014

REVIEWING THE CURRENT STATE OF MUSIC SYNTHESIS

-by Bernie Hutchins

INTRODUCTION:

Electronic music synthesis came along at exactly the right time for many of us who had a strong love of music, but a lesser degree of musical talent, with alternative skills that were more like those of engineers. Many such engineers of my generation have seen a rather impressive gamut of theoretical and practical examples of music synthesis such that our technical skills were required so that the musicians could achieve the product they desired, or at least one they could successfully sell down the line to the next consumer. Many assumptions had to be changed, and some disappointment had to be weathered. We learned a lot, basically more trial-and-error and one step at a time. Rarely (perhaps never?) was any single development truly “revolutionary” even as it became nonetheless lasting.

Accordingly we propose to initially define “success” in terms of commercial success – either grand-scale or niche-filling. So we shall first have less concern with achieving a “universal” musical instrument as long as what we do design and sell proves useful for some applications. In this sense, a standard voltage-controlled synthesizer had to be a judged a resounding musical success. From a more academic perspective, it could just as easily be judged a musical failure. There were things some musicians wanted to do, which they could not do. This is not to suggest that, if they had been able to achieve their desires, they could have gone further (let alone made a buck).

Recently I had occasion to assemble a bunch of "Moog stories" and decided to finally get a long envisioned project of scanning my 1974 Moog interview and republishing it. This was not as hard as I supposed, as the OCR errors in the scan were routine. Further, I had the pleasure of actually again reading and studying what Bob said. It was quite astounding to see how on point most everything he said was. The full interview and other stories are here [1]:

http://electronotes.netfirms.com/ENWN14.pdf

One point which fit in perfectly with what I wanted to discuss here was this**:**

> *" Another shortcoming was the failure to achieve the proper degree of complexity, and to this day, we are still living with this. This brings to mind a lecture I once heard by Peter Kubelka. Kubelka Is an Austrian filmmaker, an incredibly intelligent and cultured guy. It turns out that his hobby is cooking. Someone once asked him if he thought that cooking was an art form. He said that of course cooking is an art form as it has complexity and dynamics. I think this is true of any medium of artistic expression. If it is capable of complexity and capable of dynamics, then it is a valid medium. With the first voltage controlled elements that we made, we certainly realized the potential for dynamics. The potential for complexity is still to be satisfactorily realized."*

It seems clear to me that what Bob meant by dynamics was the ability to produce sounds that were not just correct in terms of pitch and duration, but in amplitude, and very importantly, spectral evolution. In terms of the modular synthesizer, this was the VCA and the VCF, as well as multiple modulation capabilities. True enough, we had a handle on these needs back in 1974. And we understood this all theoretically for the most part.

In terms of complexity, it is less clear what Bob had in mind, but I suspect this was multi-faceted. Bob loved complexity of sound itself (a "fat sound" as he called it) and certainly appreciated note-to-note musical structures of great complexity. In 1974 we had ideas about these things. But they were not immediately for sale. Yes you could get a fat sound with many nominally tracking (with slight errors) VCO's. And we understood that complex scores could be programmed by a computer outputting control voltages. Much progress has been made in the area of complexity – but largely trial and error. And we don't even know today that much about how to give control of this complexity to the musician. There is no overreaching theory here.

TWO STARTING POINTS

After the release of ***Switched On Bach*** about 1969, people (some of us anyway) were very much interested in how the Moog Synthesizer worked. Upon learning that it was constructed from analog oscillators, filters, amplifiers, and such, its limitations were apparent. As were its extremely impressive capabilities – electronically produced, and yet very impressive imitations of traditional acoustic sounds. This was important as this is what a lot of people wanted – not some crazy sci-fi sounds. Thus when I read in some news reports that the synthesizer could produce any sound one could imagine, I knew (1) that this was not so, and (2) that Moog himself never would have said this.

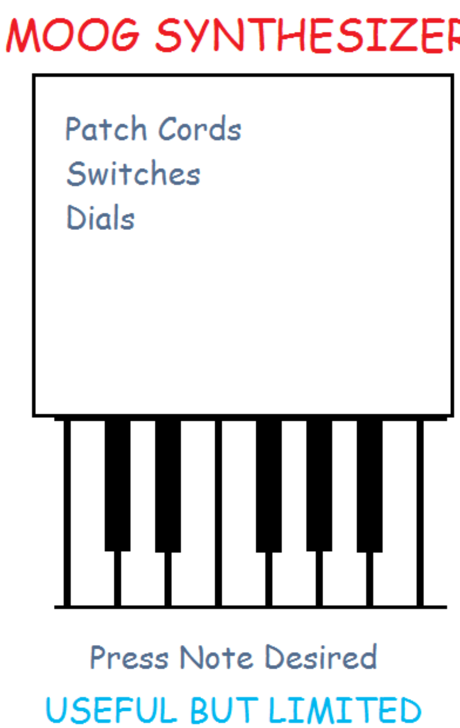
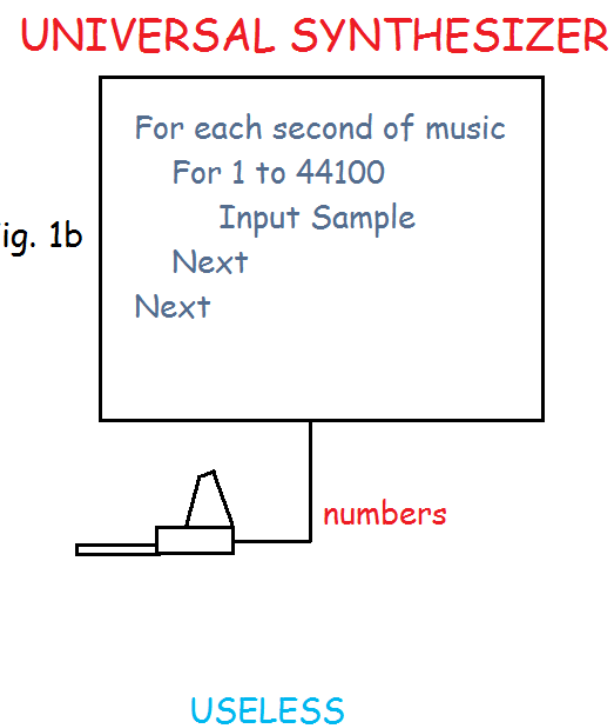


Fig. 1a

Fig. 1b



It is never a surprise that media reports are over-hyped. But by at least the early 70s we had a pretty good idea what the new Moog Synthesizer did, and we also knew about discrete time sampling so that the information content of musical signals was less of a mystery. Accordingly we were in a position to recognize that the Moog Synthesizer was far from universal, and we knew what would be involved with making a universal synthesizer – at least in theory.

The Moog Synthesizer and other similar machines that soon followed were composed of voltage-controlled oscillators, voltage-controlled filters, voltage-controlled amplifiers, envelope generators, and perhaps a few other “modules”. The machine was “programmed” (Fig 1a) by setting switches and plugging in “patch cords” and as Moog tried constantly to emphasize, by initialized switches and pot-settings (dials). Thereafter control was turned over to a human player through a conventional electronic organ keyboard. In this sense, when used this way, it often just became another “voice” in the sense a musician used the term “voice”.

True enough, it could make some strange sounds. For example it produced “clangorous” sounds through the frequency modulation process by plugging one VCO into the control input of another. That was kind of new and a worthy attention-getting demo. But eventually it got down to making the synthesizer sound like a trumpet, a violin, or a string bass. Possibly nothing indicated the capabilities of the analog synthesizers (such as a Mini-Moog) more than having them first synthesize sounds comparable to acoustic instruments of the same physical size as themselves, and then jump to producing sounds comparable to instruments many times their size.

The notion of a “Universal Synthesizer” is suggested in Fig. 1b. If you sat in front of such a machine, it would have a numerical keypad of some sort, and probably a screen prompting the user to “Input Next Sample”. This really is universal as there is no restriction on what the samples are. Readers here immediately see the traps.

First there is the overwhelming task of inputting 44,100 samples for each second of music we want to produce. Secondly, even if you were willing to spend many hours for each second of music, how would you choose the sample values? Well – you would be quick to point out – you don’t enter by hand, but you use a program to assemble the sequence of samples. Something like $\sin(2\pi fnT)$ would produce a sine wave of frequency f at sampling rate (1/T) for a large range of integers n. That would work; but what have you done!

You have drastically reduced the amount of information that is possible. Before you chose to program, you had complete freedom. But if you exercised this freedom, you would produce only white noise. Using the equation, samples are automatically determined by a relatively small number of “parameters”. This you had to do to be practical, and you had to do it to get a desirable result as well. Dealing with music in terms of a set of parameters has gone on for hundreds of years – we call the set of parameters “notes”. A typical note on a score tells us the pitch, the spectrum (at last the instrument to be used), duration, and loudness, and perhaps other things (things like pizzicato perhaps).

Using the Moog synthesizer we are, first of all, probably looking to have the electronic device serve as a substitute for where we might otherwise employ an acoustic instrument and a professional player. Perhaps we are saving money, or want personal control, are anxious to experiment, of mainly having fun. We expect as well that we may come across, probably by experiment, some additional sounds we can use. Still, there will remain a possibility of using a universal (digital) approach, for which we are now resolved to programming.

THINGS THAT WORK

Readers here likely will always think of the voltage-controlled synthesizer as the first electronic musical instrument. We may also of course remember the much older Theremin as a somewhat limited earlier device. Further the voltage-controlled synthesizer was also predated by the "RCA Synthesizer" which was a digital computer from about 1955: http://www.uv201.com/Misc_Pages/Misc_Sounds/RCA%20Synthesizer%20Side%202.mp3

which was impressive (recordings at this site). Likely you all have heard it. In the mid-50's there was a popular recording from the RCA synthesizer that featured, among other works, the singling of ***Bicycle Built for Two****.* In the movie 2001 as one of the astronauts was disassembling the mutinous computer HAL, the computer reverted to its childhood and could only sing this song. Very funny for those who knew the full story. So the RCA Synthesizer was a success. but as much as today's generation could not imagine it, the generation of the mid 1950's could not imagine even seeing, let alone using, let alone owning a computer. So our first thing that worked is the voltage-controlled synthesizer which we will date about 1965.

(1) THE VOLTAGE-CONTROLLED (MOOG) SYNTHESIZER

This we suggested in Fig. 1a. Jim Michmerhuizen who was at the Boston School of Electronic Music perceptively described the voltage-controlled synthesizer as a "collection of gadgets". True enough, and Moog would have agreed, even as both of them recognized that a lot of customers would not care to exploit the full possibilities. Instead, many gravitated toward the "standard patch" (Fig. 2a) which became, essentially, the hard-wired patch of the Mini-Moog and most small synthesizers. It was at least a reference point, a starting point, and often the whole point. The patch has a signal path at the top, a VCO feeding a waveform of high harmonic content (a saw, square, triangle, or very usefully, a variable pulse) to a VCF which altered the spectrum as a function of time, and finally a VCA which shaped the envelope of the tone, defining the duration. In the lower part of the figure we see the control paths. From the keyboard we derive pitch information (what key is down) and timing information (like a gate – some key is down; and a trigger – the key is going down or is changing). The pitch information controls the frequency of the VCO, and is one control of the VCF. The VCF "tracks" the VCO. The VCF also has second control, an "envelope" that causes the filter characteristic to vary during the course of the tone, relative to the keyboard pitch. Eventually, the filtered signal from the VCF is shaped by the VCA in response to an envelope (which may be the same one used for the VCF, but better a separate one). Most small synthesizers had at least two VCOs. Most also had other modules, typically a random noise source, a sample-and-hold, and a "ring modulator" (a multiplier or similar).

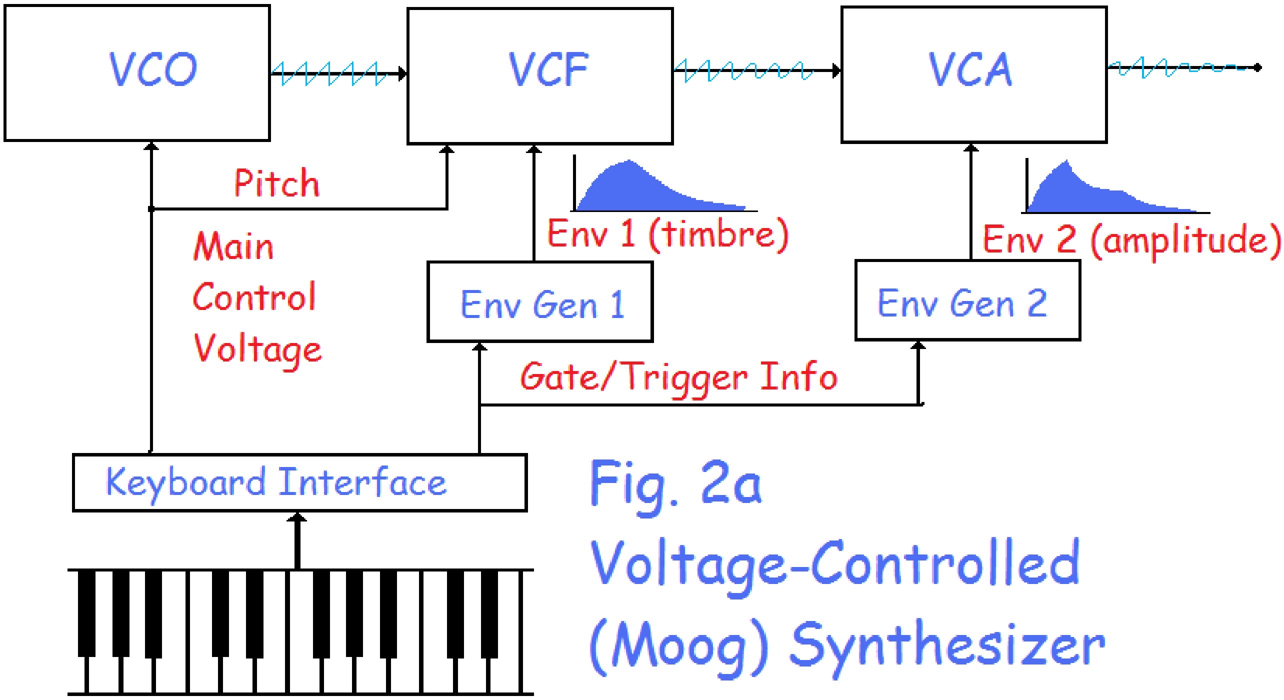


Fig. 2a
Voltage-Controlled
(Moog) Synthesizer

So Fig. 2a represents a good percentage of what was useful. The additional gadgets were handy when showing off. And once you got that much built, you were in a position to experiment. This experimentation was important for the commercial manufacturers, but was probably a lot more available to the individual builders (readers of this newsletter) who weren't facing very many business deadlines.

(2) MODULATION

The second thing that worked was modulation. Modulation in many forms was pretty well understood in terms of radio broadcasting where audio frequencies were transmitted by encoding the audio signal as part of a radio-frequency signal. Typically AM and FM were well understood. Less easy to immediately approach were the notions of what happens if both the signal and the "carrier" are audio frequencies. The mathematics here was automatically available if the engineer could develop the proper perspective!

A VCA has inherently a potential to do AM and a VCO to do FM, and Moog in his classic paper [2] described the use of FM to produce "clangorous sounds". By 1973, Chowning had quite independently developed a digital version of FM [3]. The notion of "dynamic depth" FM was fundamental for the best results, and the exponential control function inherent in keyboard synthesizers complicated this [4]. Dynamic depth was most useful when modulation could go "through zero" and this was developed [5] and recently reviewed in this newsletter [6].

In addition, the voltage-controlled synthesizer had a "Pulse-Width Modulation" (PWM) feature that was trivial to implement (and under-used?), and in many ways, as useful as FM, and this was also recently reviewed in this newsletter [7].

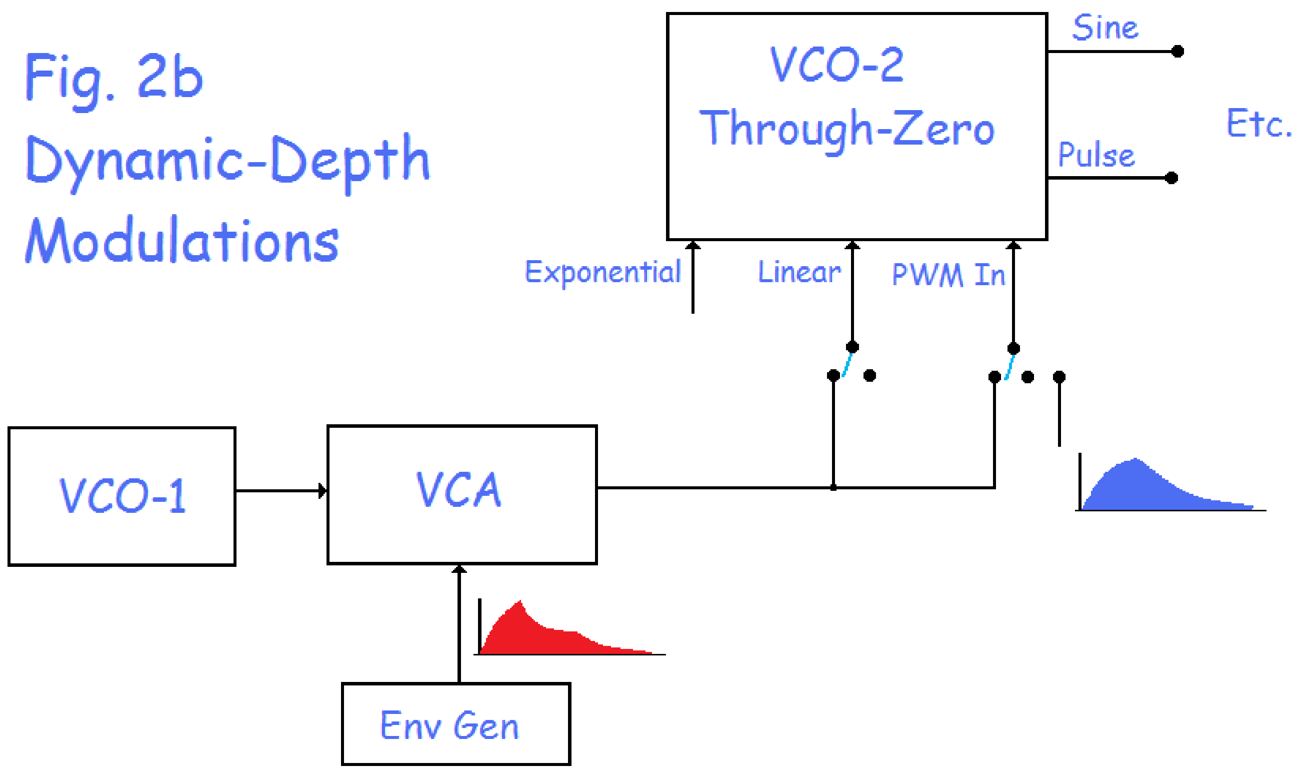


Fig. 2b shows our second success in a form of a scheme that suggests the use of existing modules to implement FM and PWM which we envision here in an analog context. (As noted, digital versions of the FM technique were widely used). The setup is capable of examining FM and PWM separately, although the urge to do both at once need not be resisted.

In Fig. 2b, VCO-1 is an ordinary VCO, quite possibly controlled by a keyboard (or not). VCO-2 on the other hand is a through-zero VCO with a linear, as well as an exponential control. Note that the output of VCO-1 is shaped by the red envelope. When the red envelope is low, the drive to the linear input of VCO-2 is zeroed, and as the red envelope rises, the level of the sinewave (or other modulating waveform) at the linear input increases, increasing the "depth" or "modulation index". Since this is a linear control, it can drive the instantaneous frequency to zero, and below. Having arranged to accommodate negative control levels (through what amounts to a time reversal), the depth does not have to stop at zero. We have mentioned in the past, and reemphasize here, that the audible effect of going through zero is negligible. Whatever we heard occurring as the control approached zero just continues to increasingly be heard. More, and better.

The use of PWM is common, and the implementation as a feature of the VCO is quite trivial. For the most part, we have used this by driving the PWM input with its own envelope (the blue one shown), much as we drive a VCF. In fact, the effect is quite similar to a VCF, a dynamically changing spectrum. For reasons that are not clear there is less use of PWM in response to a periodic waveform. That is, less use of the PWM switch to the left as opposed to the right. Here we suggest using a dynamic depth much as we did with FM, by using the shaped modulating signal out of the VCA. Beyond selecting an output of VCO-2 (only the pulse output responds to the PWM control) the "Etc." suggests a follow-up VCA and perhaps even a VCF – whatever.

These modulation methods are capable of non-harmonic tones (the general case) and harmonic ones with a bit of care (integer multiples for VCO-1 and VCO-2). It is thus more general than Fig. 2a. Fig. 2a is essentially a "low-pass model" for a tone and is inherently best suited to harmonics. But we can always add some modulation to produce non-harmonics, even in the scheme of Fig. 2a. The thing about modulation "sidebands" is that we have less of a direct feel for what will happen relative to what we suppose we are trying to do. Lowering a filter cutoff has a well-established effect. Changing a modulation frequency may be full of surprises. But again, it's good to have lots of possibilities.

Before we leave modulation we mentioned a "ring modulator" which was one of our added modules, in its most basic form (although not usually the cheapest) it is just a four-quadrant analog multiplier. Multiplying waveforms of differing frequencies produces sum and difference frequencies. Generally, this was non-harmonic. The dedicated ring-modulator was often used with live sound from a microphone to modify voice and singing in a "robotic sounding" way.

We mentioned also the availability of noise sources and sample-and-hold modules. The noise source was often raw sound material filtered into quasi-pitched material. The sample-and-hold just grabbed instantaneous value of an input voltage and held them, typically for small fractions of a second up to several seconds. By sampling a noise source, a random series of control voltages (like a random melody when applied to a VCO) was obtained. Alternatively, sampling any waveform or envelope broke it into discrete steps.

(3) PHYSICAL MODELING

Physical modeling means pretty much what you would suppose. Fig. 2a is really the model of a low-pass **signal**, and Fig. 2b a model of a modulation **signal**. Both are characterized by the results they give, or don't give. A physical model is done with reference to a **system**: generally a mechanical (acoustical) one. We can consider two approaches. First, we write up equations and solve them and get an equation that is the answer. Only a few such cases are easily solvable in closed form. Second, we may just iterate (simulate) the physical process.

We have a wide choice of acoustical instruments to model, vibrating strings, vibrating membranes, vibrating air columns, various types of reeds, resonant structures, and that sort of thing. Almost always we have an "ideal" version of the structure. For example, ideal strings are treated in many places (with a harmonic solution), and an ideal vibrating membrane has been treated in this newsletter [8], with non-harmonics resulting. Interesting enough, but not very realistic. For example, the membrane problem is different from a "kettle drum" because the latter was developed to be more harmonic than the isolated membrane.

For a second example, a real vibrating string has not only damping, but a stiffness that shifts overtones sharp to some degree. A plucked string has overtones that go sharp, and yet the higher frequencies damp out much faster than the lower ones. We would envision a model for this plucked string to use a parallel set of bandpass resonators with slightly increasing positions relative to harmonics, and of decreasing Q. Tedious, but it should work.

Perhaps much as practical factors were involved in the Moog 4-pole low-pass, which nonetheless was quite adequate for a low-pass model, a practical attempt at plucked string synthesis resulted in the "Karplus-Strong" (KS) model, which then more-or-less looks like a physical model. At this point we acknowledge an unjustified neglect of KS here in ***Electronotes***. Why?

The launch paper for the KS method was "Digital synthesis of plucked-string and drum timbres" [9] published in ***Computer Music Journal*** Vol. 7, No. 2, pp 43-55, Summer 1983, which is online here:

http://users.soe.ucsc.edu/~karplus/papers/digitar.pdf

I became familiar with the paper about that time as Kevin Karplus began teaching at Cornell Electrical Engineering, and we saw each other frequently, and talked a lot. The methods was clearly digital, and Kevin thought of it as an algorithm (a self-modifying "wavetable") while I thought of it as a digital filter. It is however clear that the authors had a good notion (see "Analysis…" section of their paper) that the resonances of the system were understood to be similar to those which occurs with acoustical string instruments. Note however that they "estimate" the positions of the poles rather than calculate them. This simply reminds us that root-finder programs were hard to access at that time. So a study that would take us an hour today was largely out of the question at that time. As to the question of modified wavetable or physical model, if I recall correctly it was the wavetable that was the driver, with an appreciation that the result fortuitously resembled a physical model. So although the motivation may not be in exactly the right order, the great success (and simplicity) of the method is why I am using it here as a best example of a physical model.

At this point, compare KS to FM.   Both were clearly very valuable, and simple additions to the toolbox which started with the voltage-controlled low-pass structure. FM greatly simplified the production of percussive sounds while KS was an astoundingly good plucked string (with drum capabilities as well).   Both were similar in that while they produced excellent individual tones, getting them on pitch and the tones thus generated into a melodic sequence was more difficult than with the low-pass keyboard-controlled structure. And the two were different in that FM could be analog (mostly) while KS seemed restricted to being digital. In the mid 1980's, we were still doing mostly analog stuff at ***Electronotes***.   KS was quite extensively extended (again in a digital context), with much work involving the way of obtaining a desired pitch, which could be complicated.   Quite simply, ***Electronotes*** fell quickly behind as far as KS was concerned.

Fig. 2c summarizes the physical modeling as we want to introduce it here, and shows an example of the KS process. First we will discuss the general notions on the upper portion of Fig. 2c. In the pink box, we indicate that the process is simply a matter of writing down the correct equations, and solving them.   The solution may be (very rarely – such as an ideal string) in "closed form". By closed from we mean that we write down an expression for the synthesized signal x(t) for all t, and then plug in t over the range we need.   As we stated above, this is rare, and generally even when possible, not useful. It is not useful because it is usually an ideal case that is calculated, and it is the non-ideal effects that our ear find interesting. So an ideal string sounds not so much like a real string as it does just a perfectly periodic harmonic waveform (boring).

In the more usual case, we do not get a closed form. Instead, we solve the equations numerically. This is usually the solution of differential equations by iteration. Less we suppose this is a cop-out, we should recognize that similar procedures send space probes to Mars, and so on**;** a closed-form to the "three body problem" (let alone, more than three bodies!) not being available. The mathematics in this case is quite properly a simulation. If you prefer, the calculation IS the machine!

Quite often, different musical tones differ in one or relatively few "parameters" (such as different pitches) and we can adjust one solution to the varied parameter without a full calculation. This is similar to an "interpolation" of solutions.   It is however true that day by day (year by year at least) full calculations have become practical due to hardware advances.

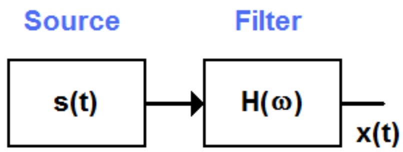
The upper right of Fig. 2c shows a way of thinking about many physical models: the so-called "Source/Filter" model. Here we produce a sound x(t) by constructing a suitable source s(t) and passing it through a suitable filter with impulse response h(t). For example, s(t) could be a bowed string and h(t) is the resonant "box" of a violin. It is not unusual that s(t) would provide the performance determining parameters (pitch, loudness) while h(t) is a fixed filter in the background for all tones (such as the resonant box which only the carpenter changes). Nor is it unusual to have h(t) vary! Famously the human speech mechanism has a time-

Write down the physical equations of mechanism

Solve in closed form

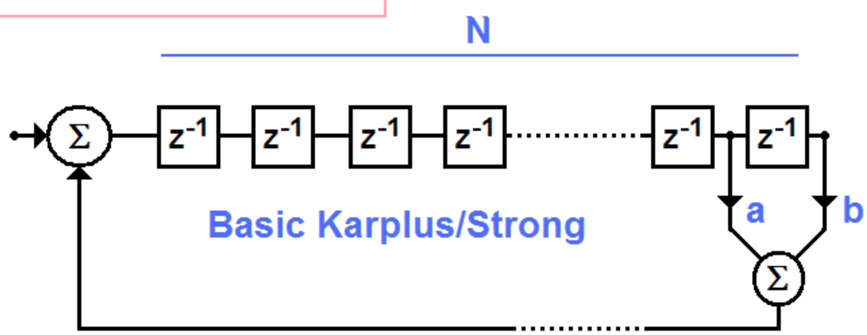
OR

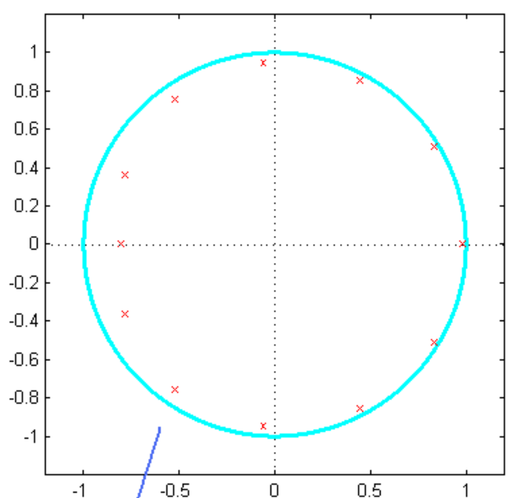
Simulate Iteratively

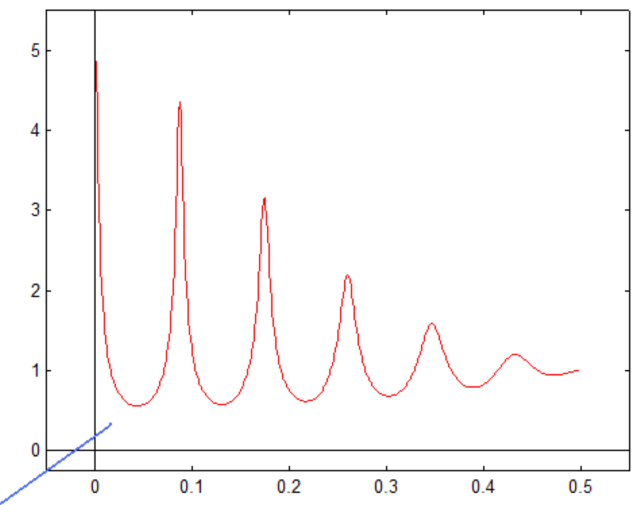


$x(t) = s(t)*h(t)$

$X(\omega) = S(\omega)H(\omega)$







Poles

Freq. Resp.

Imp. Resp.

Example a=0.4 b=0.4

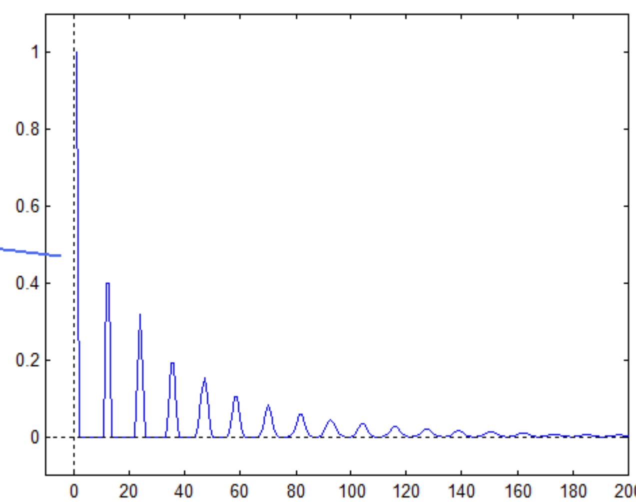


Fig. 2c

varying “formant” filter, the formant being the vocal tract as it articulates different sounds. That is, it is not just the pitch changing against a fixed filter, but mainly the varying filter that provides meaning. In this source/filter view, we at least start with the idea that this is a linear system (which may be good enough) as indicated by the time-domain convolution and frequency-domain equations below the sketch.

Now the KS example, the lower portions of Fig. 2c. In the original KS scheme the feedback coefficients a and b are both 1/2. Consider that if a=0 and b=1 we just have a storage loop of length N, and a boring wavetable read. If N=13 there are 12 poles equally spaced on the unit circle (one at z=1). If a=0 and 0<b<1, we have 12 poles on a circle at radius $b^{(1/12)}$, and so on. With a=1/2 and b=1/2 the poles are no longer on a circle, but one pole is at z=1 again. (This dc pole does not care if the delay is 12 or 13.) I have never cared for this (conditionally stable) pole at z=1 which comes with the original KS. Instead much the same audible result is obtained if a and b sum to less than 1. Thus the example here has a=0.4 and b=0.4, with the poles shown. Now no pole is on the unit circle (at +1) but rather one pole is at +0.9808 and another at -0.8076 with the others skewing to the right. This skew is exactly what we were looking for to lower the “Q” of the higher frequency poles. Also in accord we see the frequency response to have smaller resonances for higher frequencies, as shown. The impulse response starts out with an impulse (delayed) and this smears and rounds to something approaching sinusoidal as it decays. This is what we wanted from a plucked string synthesis.

Karplus and Strong viewed this as a self-modifying wavetable. That is, it plays back a stored length-N sequence, but modifies the reloads. The modification is indeed low-pass in the case of tapping the last two taps in the general manner shown. Another view is that there is a filter in the feedback loop (with taps a and b). While we are quite free to input any signal we like, the case where the input is an impulse is fundamental, and Karplus and Strong also discussed the loading of the line with random noise for drum-like effects. Most likely it is the skewing of the poles for unequal Q’s and for unequal spacing that is of the most use to us here.

As I check back issues, I find brief discussions of KS in EN#139 and EN#189. In my teaching work, there was much more, including some implementations with DSP cards and at least a half-dozen student projects with myself, or advising students in other courses. Many of the academic papers on KS involve a fundamental problem with KS of getting the pitch right. It is one thing (an important thing) to say that the basic sound is a very good plucked string. It is a different thing to control the pitch, which starts out looking like about the reciprocal of the total delay (perhaps minus half a delay). Or most likely the pitch was determined by the placement of the poles, the imperfectly (non-harmonically) placed harmonic resonances. This is exactly what occurs with a real string. Variations of the weightings were used. But the methods just did not seem to fit into our analog based world at that time.

(4) SAMPLING

For good, bad, or with judgmental indifference, we note that our successes have gravitated toward methods that produce more-or-less traditional sounds similar to those obtained acoustically. There is a lesson here. It is no surprised that we arrive at another success in our survey of "artificial" music**:** RECORDING.

Much as we may wonder at what it might be like not to have air travel and cell phones, etc., probably few who really love music have not contemplated what the situation must have been like even 100 years ago when music was mostly live. Really – how did those folks survive? I guess they played music at home, went to a few concerts that were available, or did without. Eventually, music was on the radio, and 78 and 33-1/3 RPM records appeared, and so on. At some point, we went from starvation to much more than we could listen to. And it goes forward today with increases in inventory and in accessibility. Really - what's not to like?

One thing we might consider in lieu of having recordings is "hearing the music in our heads". We are familiar with musicians who can apparently look at a score and "hear" the music without it ever being realized. Probably most of us can "re-hear" music we have heard acoustically (perhaps listened to it many times) in our heads. It's not the same – but it's not entirely different either. (For myself, the correct temporal ordering of phrases is largely abandoned.) In discussing this with students and others over the years, comparing it to what is required (sampling-wise), the consensus is always that the music we store "in our heads" is nothing like a digital recording. Everyone feels it is "abstracted" in some manner. I don't have a clue.

If we decide to actually record music, we quickly become impressed with many advantages of a digital storage medium rather than an analog one. One thing we have to bear in mind is that whenever the barrier of "insufficient memory" has come up, it has been just a matter of waiting a year or two before we push that barrier aside and set a new goal. So at some point it might have seemed practical to synthesize trumpet tones by actually recording just a few and manipulating them for different pitches, durations, loudnesses, etc. We considered all sorts of codings, interpolations, models to get a desired tone as derived from a small set on hand (recorded). As memory became much much more available (smaller, cheaper) the starting set could perhaps become dense enough that we had no need to manipulate the available set – we just could find one that was close enough. If not today, perhaps tomorrow.

Surprisingly enough, electronic musical instruments based on actual recordings have a history longer than the Moog synthesizer. In particular, the "Mellotron" was a product of the early 60's, and was very limited (mostly for accompaniments) based on magnetic tapes:

http://en.wikipedia.org/wiki/Mellotron

Conceptually, when you wanted a tone (usually you were producing sustained chords) you pressed keys on a keyboard that initiated a tape and mixed notes from multiple playback heads. Yes – you could run to the end of the recording. After you lifted the key, the tape rapidly rewound. Interested readers can find much material and audio examples. Here we just ask the reader to ponder the mechanism if it involved more and more instruments. It became an issue of the limitations of (analog) storage, with the prospect of extensive, realistic digital storage still many order of magnitude beyond anyone's then current imagination.

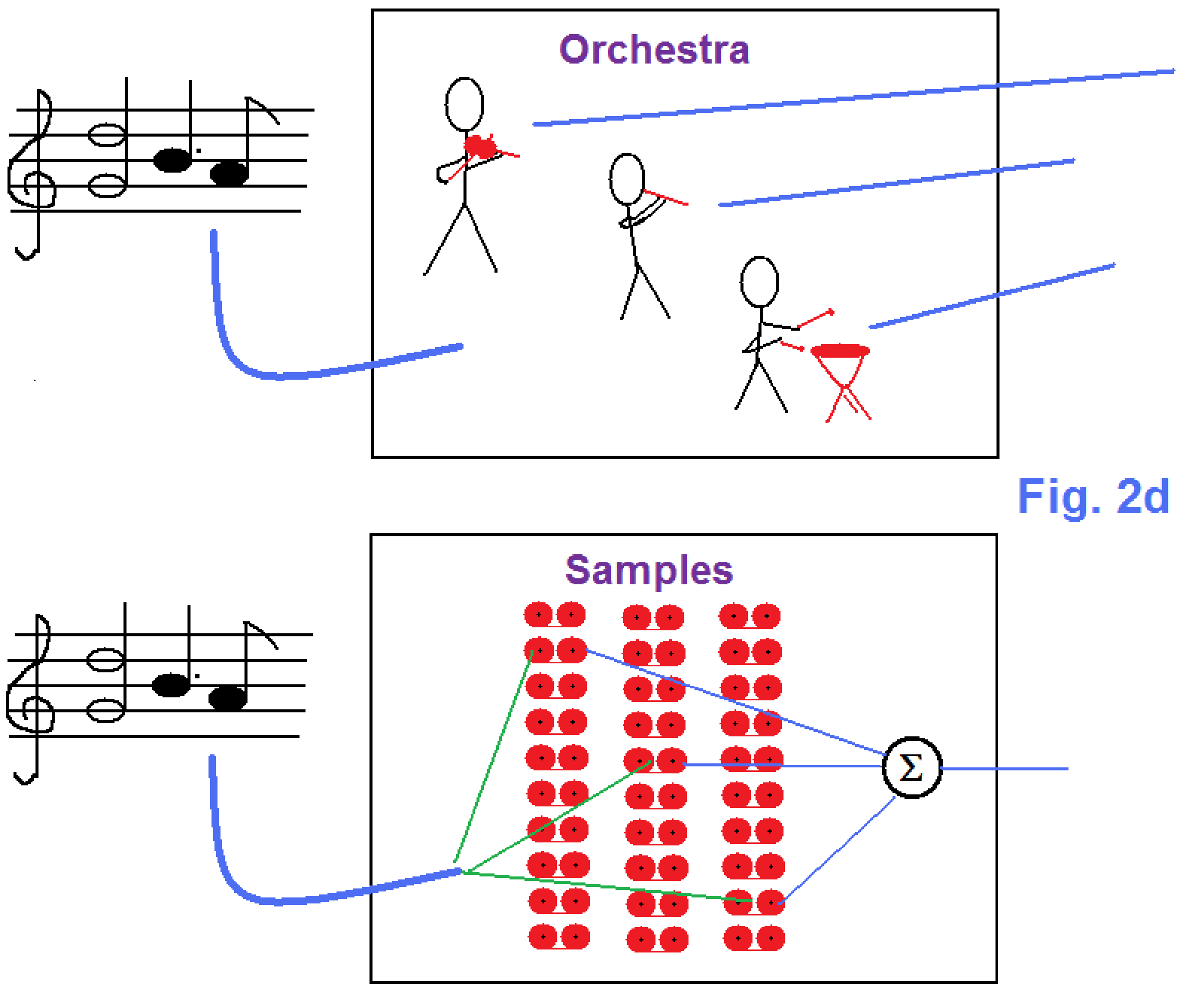


Fig. 2d

Fig. 2d is intended to provide a perspective for the problem. We are talking first of all of using live players (an orchestra), either directly (top) or we will make a nice collection of individual recordings. While it is traditional in signal processing to consider a sample as one number, with sampling instruments, "samples" are more like individual notes. Thus the lower portion of Fig. 2d is intended to represent many many recordings available such as in a Mellotron.

The input in either case is something like a traditional musical score. If we use a live orchestra we put down the individual parts in front of the individual players, and they all cooperate (perhaps with the assistance of a conductor) to produce the music. In the bottom, the score is input by a keyboard or as a computer file. For each note, the machine looks up the best candidate tone. So we have shown our recordings (little red tapes) in an array. Indeed we could envision an array of much higher dimension (pitch, duration, loudness etc.) so there could be a lot of memory. We are not concerned so much with the actual costs (a temporary condition presumably) of hardware memory, as noted above. Rather there is a problem in logistics relating to human players just to get the one-time recordings. That is, obtaining good recordings (in all the parametric dimensions) is likely to be seen as the largest problem here.

As far as a project along these lines for the home hobbies - well you would have a difficult time getting the cooperation of all the necessary players. You could however possibly get a fun electronic version of <u>your own playing</u> by making a limited set of recordings at your own leisure. One way to do this would be to hire a studio. Another way to make anechoic recordings (which would be the best approach) is to do them outside, like in the middle of a large, flat, well-vegetated farm field. A large, flat, snow-covered field is even better – at least acoustically.

Computer-aided composition and simulated performance of compositions are certainly attractive products for composers who would have a very difficult time getting the services of even a string quartet, let alone a symphony orchestra. Much as most of us use word-processors for text to write banal memos (as well as formal publications), many take advantage of music score-writing and editing software. This includes the separate rendering of individual instruments parts. A logical extension would be to at least roughly assemble an audio rendition. In this case, the controller of the "synthesizer" is an appropriately formatted score. Certainly not a substitute for real performers, but well worth having available.

Nothing prevents us from having an approach that involves both sampling and synthesis, and points in between. It would seem in fact that this would be essential. For example, some composers might choose to use, for example, a sackbut, which is an antique trombone, and which can be played by many trombone players with a little practice, upon request. Such an offering would be unlikely to be part of a collection of sampled recordings of an orchestra. So we would think of synthesizing a sackbut sound, most likely doing it by modifying trombone recordings It would be an extension of what might be a continuum of related instruments, and/or the production of a tone from close neighbors (perhaps modifying a tone of D to be instead $D^{\#}$).

Sampling can in the most general view be seen as a procedure of working with digitized live sounds, and then processing, editing, and organizing them to a desired end product, integrated with other techniques using any tools available. The test for its success is whether or not the end product is useful.

(5) ANIMATION

Moog spoke of the need for dynamics and complexity. To me, complexity involves, among other things, some animation of a musical voice. Failure to achieve this animation constitutes a failure (see Things that Don't Work below). As far as synthesizers go, most simply, this means don't even think about trying extended soloist tones unless you address this issue.

My standard of badness in this regard has long been a "me too" album that followed the success of ***Switched-On Bach***. I don't think I still have this album, and have forgotten who the artist was, but it had as one of its selections the ***Vocalise*** of Rachmaninoff – one of my best loved pieces. This was, incidentally, not the Tomita synthesizer version that I have in mind here – about which a comment later.

The ***Vocalise*** has many arrangements and performances on many different musical platforms. At its very best it is sung (in my humble opinion), and anyone can sing it, as long as it's Anna Moffo! If you don't know the piece at all, start with her version. It is on YouTube but I will let you find it if you wish – the one I found did not run very well. Much more to the point you can get the MP3 from Amazon for 99¢ and hear it pure and uninterrupted. Notice it is a work (a song) for a soloist, a singer, a soprano, and it needs to be styled individually, with nuanced pitch and articulations. It requires a great deal of art. The synthesizer does not do this easily – at least not from a keyboard.

It seems necessary to consider animation in two senses, depending on the role of the synthesized sounds. It is more or less understood that a simple periodic waveform is dull per se. We have taken on a number of studies of breaking up the regularity to achieve what Moog called a "fat sound" which he thought (correctly) could be achieved with imperfect tuning of VCO's nominally in unison. Our methods, the most successful of which has been the Multi-Phase Waveform Animator (MPWA) [10, and at the link: http://electronotes.netfirms.com/AES4.PDF] had the advantage of using a single VCO and relatively simple and cheap parallel phase shifters. This was effective for generally improving the sound material in making it far less boring.

I'm not at all certain that it is enough to animate a solo voice in this way. That is, if we add MPWA type animation, it is not enough to make a musical voice capable of holding a solo voice line. Such a voice needs to be a controlled complexity. It has to be under artistic control the way the human voice and almost certainly, many other acoustic instruments are. I would characterize the situation as the addition of "styling" to the sound. I certainly do not think this is something an engineer can build in. Accordingly, pro-forma animation is very useful, but can hit a brick wall in what it can eventually achieve.

Returning to ***Vocalise***, Clara Rockmore's Therenin version proves the point that electronic instrument can "sing", given the close control of the Theremin. Tomita avoided the solo issue. His is a realization of an overall piece – he does not even attempt to make it a solo (song).

THINGS THAT DON'T WORK

(1) PERIODIC WAVEFORMS

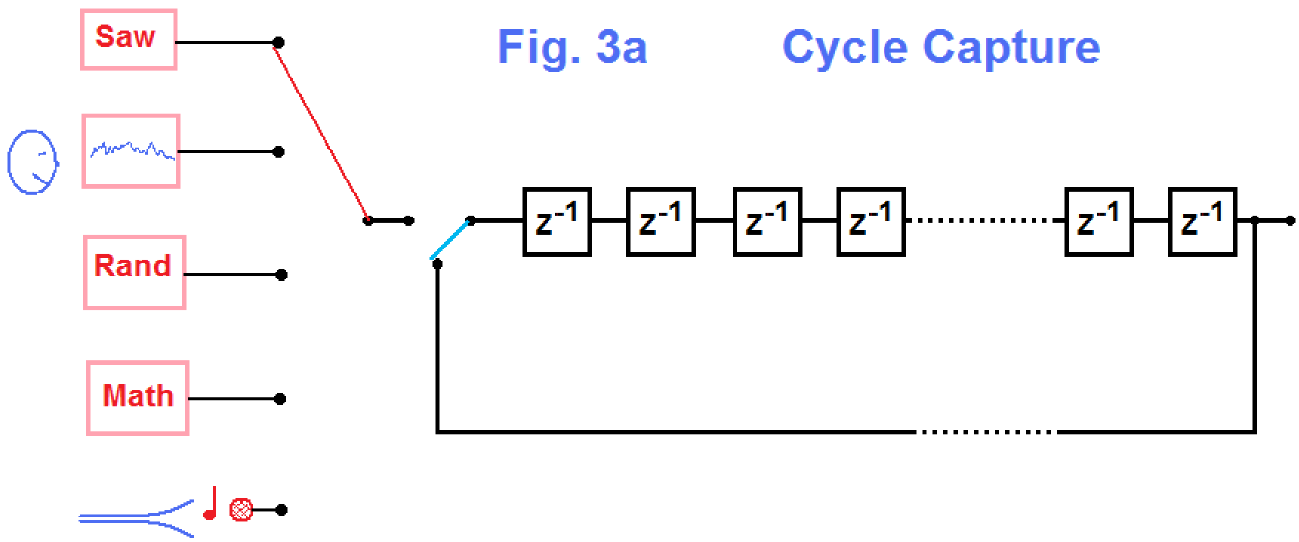
Many of us started out with the notion that musical tones (those with pitch) were periodic waveforms. Experiments with laboratory equipment (signal generators and oscilloscopes) more or less enforced this idea. But, connecting a microphone to the scope and playing an acoustic instrument did not work too well, as we could never "lock in" on a waveform. The scope just would not trigger well enough for a stable display. This was, I guess, seen as the proverbial "bug" (failure of the measuring equipment) rather than a "feature" of acoustic instruments (the periodicity is imperfect). The bug was trying to tell us something!

At the same time, we recognized that signals from signal generators (later function generators) were not exactly "musical". Today we recognize numerous reasons for this. For one thing, we could not conveniently turn the volume up and down, except manually. If we manually changed the frequency (turning the frequency knob) we were embarrassed by the glissandi, and musical notes were supposed to be discrete. The Moog synthesizer tackled these problems (by keyboard conrol), and was unquestionably immediately successful as long as one stuck to fast-moving multi-layered short-tone material. A partial breakthrough.

In as much as a Moog synthesizer sounded like (but not exactly like) an intended musical instrument, we probably weren't surprised. After all, we started with a sawtooth wave (for example) when we should have started with a "trumpet wave". That problem triggering the scope again. But this was clearly just a human limitation was it not? The silly trumpet player just could not hold the tone steady enough. Once we have captured the right trumpet waveform, we will have an artificial player of extraordinary skill. Well, that didn't seem quite right.

In the mean time, more evidence was amassing. We learned a lot about the ways musical sounds were produced by acoustical instruments; likely noting that they were imperfect (like not being strictly harmonic). Nonetheless they did what they did and people liked them! Further, we found that music synthesized with "proper" electronic oscillators was inferior if it had extended duration tones (half a second up to several seconds). But the most important experiment was a type as suggested by the "Cycle Capture" scheme of Fig. 3a. What we show in Fig. 3a represents several approaches. The heart is the recycling delay line which can be loaded by five proposed schemes as shown (red switch) and then the samples are recycled by connecting the blue switch to the output feedback position. Note that this is a "wavetable" method, of which Karplus-Strong is an improvement.

Fig. 3a Cycle Capture



With the red switch in the top position (Saw) we suggest that the samples in the cycle are simply set to those of a sawtooth waveform (or any other such familiar choice). Thus it represents a lot of what was already being done with the voltage-controlled ideas. Many times synthesizer users expressed what seemed to be an astounding notion: that they only needed a few waveforms (like sine, sawtooth, and pulse) and weren't that interested in more choices for their musical adventures. Useful information here.

The traditional voltage-controlled waveforms (sine, square, triangle, saw, and pulse) were selected as those available directly from an oscillator, or were made available by simple waveshapers. They were all "simple" in that anyone could draw them in paper, for example. One knew that viewed on a scope, they did not challenge the human visual system. It was clearly possible to "design" waveforms that were visually far more attractive to the eye. It just made sense that these would also be more attractive to the ear. In the second position of the red switch from the top we indicate our designer who has arranged the visually-pretty time sequence to be loaded and cycled. What wonders must surely await?

The middle red switch position shows sample choices chosen at random. This is interesting as we recognize immediately that if the capture line is short, that we probably have a pitched sound. By chance for example, we might expect to sooner or later get samples fairly close to those of a sine wave if the line is perhaps only length 6. If the length is 10^6 however, even the recycling noise might be noise [11, see AN-402]. For the moment, assume the length is only 10 to 100 – something like that.

The red switch position to the Math box suggests that the samples are set by calculation, and that the calculation may be intricate (or simple). For examples, it might be a frequency modulation solution that is calculated and stored (involved) or it might be 4/3 of a cycle of a sine wave (unusual but simple).

The bottom red switch position is the most revealing. Here we assume we have a recording of an acoustical instrument, and with some care, we have isolated and made available a single cycle. This is the realization of a dream to know and use the "waveshape of a trumpet", for example. That is, the waveshape we could not quite trigger and display on the scope face. We finally beat the scope. We can now capture and reliably display a single cycle. And if somehow we didn't quite get it perfectly – well we can edit the file for near perfection. Can't we?

Here is the finding – NO IT DOESN'T WORK !

ALL FIVE SWITCH POSITIONS ARE ALMOST EQUALLY BORING

This is perhaps unexpected. However, going back to what Moog said about Kubelka, we note that the cycle capture lacks dynamics. It's just another unchanging periodic waveform. The periodic waveforms are all boring, and there was nothing special (except engineering considerations) about the easy ones we first used. The first circuit I used that was capable of general waveforms involved Walsh Functions. This was a case of a Math setting of the red switch. I remember setting the controls and watching as I assembled what looked like a pine tree on its side. It looked quite lovely, and I recall being in a hurry to hear just how lovely it would, correspondingly, sound. NOTHING! I don't recall being disappointed; nor did I suspect that I had implemented something wrong. After all, any good lesson is priceless. Hearing and seeing are different. The dynamics had to be specifically implemented (by envelopes) in the Walsh Function case [12] :

http://electronotes.netfirms.com/AES1.PDF

By far the most surprising result is the capture of a single cycle of a recorded instrument (bottom position of the red switch). I have done this experiment – but only so that I could say I had done it, and many years after the Walsh work (using TMS320 DSP chips [13]).

(2) Random Control of Pitches

Here we are talking about what does not work with regard to devices which control melody as opposed to tone. This is tricky because whether it works or not depends on how much control the composer (or performer) retains, and how much selection and editing you are allowed to do. This goes way back to a classic: Don Lancaster's "Psych-Tone".

http://www.swtpc.com/mholley/PopularElectronics/Feb1971/PE_Feb1971.htm

This was a project in ***Popular Electronics*** [14] and realistically, a magazine description of a kit that was for sale. Somehow I managed to afford one (about $50). The device could in fact produce sequences of 63 equally spaced random pitches in a variety of tonalities. Pretty much as advertized it could drive one crazy. This is not to suggest that it could not be modified to be useful. I managed to modify mine so that it made much shorter sequences. [I regret that the resulting device was loaned out, and never returned.] It certainly did inspire a lot of us to look at kit and home-brew electronic music circuitry. And it certainly was an inspiration to use RTL logic IC's in sequencers. In fact, that is what the first four issues of ***Electronotes*** concerned.

In going first to rigging the shorter random sequencer, from which phrases of perhaps no more than 8 to 10 notes at most could be used, and to the perfect control of the RTL-based sequencer in the first four ***Electronotes***, I think we were acknowledging our desire to take back control. Random melodies didn't work first because the brain gave up on them as anything other than a tease, and there were just far far too many options. Something was clearly missing, and almost totally missing. The rate of information flow to the brain was not under control. True, a real composer could fix this by selecting material, and one knew what was acceptable by the end result. It varied some with the individual listener as one would suppose. But complete randomness (exercising complete freedom) didn't work.

It is however important to note that the technique of selection (noticing useful acoustical results, editing down, mixing to a single composition, etc), the result of fighting off random tunes, was quite important, as it put composition and performance (even if for personal pleasure) in the hands of those of us who lacked significance agility with traditional instruments. That is, those of us who lacked performance talent. The approach of using a sequencer graduated to a general appreciation for an utilization of what was really just programming (although we didn't immediately call it that exactly). This too put a tool with musical capabilities in the fumble-fingered hands of those with less talent. Of course, if you had more talent (could actually play a keyboard, for example) that was a bonus.

A couple of related things. One of the problems with random melodies was perhaps the correlation properties of the notes. Random and pseudo-random tones were usually "white noise" and accordingly were uncorrelated, unless otherwise processed. This we could do [15]. Possibly the most important academic notions were offered by Voss and Clarke in January 1978 [16]

http://physics.bu.edu/~redner/542/refs/voss-clarke.pdf

which was quickly picked up by Martin Gardner for ***Scientific American*** for April 1978 [17]

http://www.informatics.indiana.edu/donbyrd/teach/PapersEtcByOthers/SciAmMathGames77_FractalMusic.pdf

These results indicated that a major problem with random tunes was that “white” tumes were too “jumpy” while integrated white noise (now called red noise [18] was too connected (like playing scales up and down. What Voss and Clarke found was that the parameters of musical melodies were not supposed to be white (a flat power spectrum - correlated) nor red ($1/f^2$ power spectrum – too correlated) but more like 1/f (a medium degree of correlation). <u>This explained a lot</u>, and Martin Gardner, as he always could, explained it to perfection and even offered a clever 1/f algorithm. In fact, at that time, 1/f filters (power), or “pink noise” filters were not that easy to do.

This leads me to one last thing: Carl Frederick’s fruit fly! What Carl did was translate DNA code of a fruit fly into a melody. If you have not heard this by visiting our Webnote 2, take a listen to the “Little March of the Fruit Fly”:

http://www.darkzoo.net/clfsite/march.wav.mp3

This is remarkable. Carl assures me that this is really the product of the fruit fly DNA information and not of his own skills at selecting (composing). It’s quite a good tune – and probably an ear-worm. What if anything does this tell us? I haven’t figured out what to make of it. Does it mean that information that has the <u>meaning of the **making fruit flies**</u> is structured or patterned as well to lend itself to reformulation as melodies (or as perceptible patterns) for humans?

(3) <u>RANDOM FOR ANIMATION</u>

One fundamental notion in electrical engineering is the distinction between signal and noise. A signal is often thought of as a relatively simple <u>information bearing</u> signal while noise is thought of as a much more random, <u>meaningless</u> added component . A radio program that you want to hear, and “static” that gets in the way is a prime example. We often have formal engineering descriptions (such as in terms of a spectrum) and accompanying notions of how to separate signal from noise.

Human beings famously separate signal from noise, and often this suggestion is more of a metaphor than a mathematical equation. Acoustically, we are able to bring in powerful processing tools involving the brain as well as the ear, and we describe this as “listening”. Listening is the first step in extracting “meaning” from the overall audio signal. Finding any meaning, or at least not giving up the sound as too boring or too complicated, is one step in separating the signal from the noise. Another consideration is what is signal, and what is noise. A person trying to listen to someone speaking to him/her in a gusting wind may be seeking narrowband components against broadband noise. A person trying to hear if there is the sound of air escaping from a tire (a classical example of white noise), while someone is speaking to him/her, is doing the opposite. A person who has two people talking to him/her at the same time has to decide which person is the signal and which is the noise.

We are all aware of the accusatory phrase "You call that music!" People's opinions vary widely. What we have in mind to look at here is not related directly to the style of music but whether or not a sound, per se, can <u>hold</u> our attention and offer itself as a musical component. A violin tone likely will do this. A beeping alarm, a car horn, or a screaming child can get our attention, but not be musical. A gentle wind through trees or ocean surf may not get our attention, may be soothing, but not be musical either. Blaring "music" on a store's PA system may just leave us wondering about the manager's taste, his/her intended effect, if not about the direction to the door. Soft background music in a restaurant may go completely ignored. So we are talking here of music as an intentional acoustical event, largely from a willing producer to a willing listener.

And we are talking mainly about what we would call "timbre" or tone quality. And again with reference to Moog's quoting of Kubelka – we want dynamics and complexity of tone. Above, under things that do work, we spoke of animation (as described) as being a useful step. We added complexity to a simple, purely repetitive tone. Perhaps that was in a direction which might have been suggested just be observing an orchestra. Why are there so many violins? Why only one bassoon? Why do a dozen violinists who can't possibly be playing 100% coherently sound different from one violinist who could perhaps be amplified electronically – saving money! And why do we sometimes have a solo violin and call a piece a concerto?

Going back to the idea above, the rejection of perfect periodicity as holding our interest, we might suppose that we could use a random signal to somehow break this up. Most fundamentally we would just add in the random noise. The ear/brain simply separates the two from the sum. We hear a perfect periodicity, and we hear noise. Two things. When we try to be more subtle and use the random signal to modulate the periodic one, it still sounds much the same as just adding the random signal. Two different things.

A couple of related things can be mentioned. If we take a noise generator and bandpass filter it to a low frequency (say 3-10 Hz) we get what looks like a sinusoidal frequency with amplitude variations, and this can be used with modulation to provide tremolo (AM) or vibrato (FM) and this is not a complete failure! The problem is that the amplitude variations in the bandpass control signal can be too extreme. In the past, we have termed this as "spooky vibrato" in honor of the canonic haunted-house wavering tone. A sharper bandpass (with accompanying slower time response as in the uncertainty calculations) will change things a bit – perhaps an improvement.

The second thing also relates to filtering a noise source. This involves the filtering of the noise with a VCF controlled by a keyboard. This does at the very least "carry a tune" and may be quite nicely pitched. That is – colored noise. Here the random noise is not added, but certain frequencies already in the noise are enhanced. Well worth considering – but not to be used everywhere.

SUMMARY:

With 40+ years of publishing ***Electronotes*** and following other publications, commercial developments, and just talking to people, it is not easy to remember and thereafter relate the most important things we learned, often of course, the hard way. Most likely we could find similar stories in other endeavors. Yet our work in synthesized music is at times relatively unique in that it involves not just engineering, and not just an interaction with human beings, but a substantial measure of esthetics. We are playing around with art. A lot of things that seem important from an engineering point of view just aren't, and a lot of things that would seem technically uncomfortable (for example, allowing clipping [19], that sort of thing) are perfectly OK. Hence a review like this - from time to time.

REFERENCES:

[1] The original publications of the Moog interview are ***Electronotes*** Vol. 7, No. 45 October 14, 1974 - Part 1; ***Electronotes*** Vol. 7, No. 47 December 15, 1974 - Part 2; ***Electronotes*** Vol. 8, No. 50 February 1975 - Part 3.
http://electronotes.netfirms.com/ENWN14.pdf
This is also a good place to mention a very nice video by our friend Trevor Pinch:
http://www.cornell.edu/video/trevor-pinch-history-of-moog-synthesizer

[2] Moog, R. A., "Voltage-Controlled Electronic Music Modules," J. Aud. Eng. Soc., Vol. 12, No. 3, pp 200-206, July 1965
http://www.moogmusic.com/legacy/voltage-controlled-electronic-music-modules
http://moogarchives.com/aes01.htm

[3] John Chowning , "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation, ***" J. Audio Eng. Soc***., Vol. 21, No. 7, Sept 1973, pp 526-534
https://ccrma.stanford.edu/sites/default/files/user/jc/fmsynthesispaperfinal_1.pdf
http://people.ece.cornell.edu/land/courses/ece4760/Math/GCC644/FM_synth/Chowning.pdf
Also see the interesting YouTube: http://www.youtube.com/watch?v=w4g92vX1YF4

[4] B. Hutchins, "The Frequency Modulation Spectrum of an Exponential Voltage-Controlled Oscillator", ***J. Audio Eng. Soc***., Vol. 23, No. 3, April 1975, pp 200-206
http://electronotes.netfirms.com/AES3.PDF

[5] A through-zero FM VCO was offered as an option for the ENS-76 system published in ***Electronotes*** EN#75. A better design is offered as: B. Hutchins, A Voltage-Controlled Oscillator with Through-Zero FM Capability," ***Electronotes,*** Vol. 13, No. 129, Sept 1981 available here: http://electronotes.netfirms.com/EN129.pdf and a very brief summary discussion (to give proper credit to Doug Kraul and Bob Moog) is online here: http://electronotes.netfirms.com/TZFMNote.pdf

[6] Hutchins, B., "What About: Self-Frequency-Modulation," ***Electronotes***, Vol.23, No. 217, Sept 2013
http://electronotes.netfirms.com/EN217.pdf

[7] Hutchins, B., "Revisiting Pulse Width Modulation," ***Electronotes***, Vol. 23, No. 216, July 2013
http://electronotes.netfirms.com/EN216.pdf

[8] Hutchins, B., "The Drumhead Problem – The Vibrating Membrane," ***Electronotes***, Vol. 13, No. 131, Nov. 1981, pp 5-14

[9] Karplus, K., and A. Strong, "Digital synthesis of plucked-string and drum timbres", ***Computer Music Journal*** Vol. 7, No. 2, pp 43-55, Summer 1983
http://users.soe.ucsc.edu/~karplus/papers/digitar.pdf

[10] Hutchins, B., "A Sawtooth-Driven Multi-Phase Waveform Animator," ***Electronotes***, Vol. 10, No. 87, pp. 3-11 (Mar.1978)
See also:
http://electronotes.netfirms.com/AES4.PDF

[11] Hutchins, B.,"Anomalies in Pseudo-Random Generators" Electronotes Application Note No. 402, Jan 10, 2014
http://electronotes.netfirms.com/AN402.pdf

[12] Hutchins, B., "Experimental Electronic Music Devices Employing Walsh Functions", ***J. Audio Eng. Soc***., Vol. 21, No. 8, October 1973, pp 640-645
http://electronotes.netfirms.com/AES1.PDF

[13] Hutchins, B.A. and T. W. Parks, ***Digital Signal Processing Laboratory Using the TMS320C25***, Prentice Hall (1990)

[14] Lancaster, D., "Build the Psych-Tone: Melody Synthesizer with 28 Controls and 63-Note Melody", ***Popular Electronics***, Vol. 34, No. 2, Feb 1971, pg 25
http://www.swtpc.com/mholley/PopularElectronics/Feb1971/PE_Feb1971.htm

[15] Hutchins, "DACs and Psuedo-Random Noise Sequences", ***Electronotes***, Vol 22, No. 212, Dec. 2012; and Hutchins, B., "The ENS-76 Home-Built Synthesizer System – Part 8, Random Sources" ***Electronotes***, Vol. 9, No. 76, April 1977, pp 3-16

[16] Voss, Richard F. and John Clarke. "'1/f Noise in Music: Music From 1/f Noise." ***J. Acoust. Soc. Am***. Vol. 63, No. 1 (1978) pp 258-261.
http://physics.bu.edu/~redner/542/refs/voss-clarke.pdf

[17] Gardner, Martin. "Mathematical Games", ***Scientific American*** April 1978: 16-32. As he almost always did, Gardner had his own innovative way of doing something - generating sequences with 1/f properties in this case.
http://www.informatics.indiana.edu/donbyrd/teach/PapersEtcByOthers/SciAmMathGames77_FractalMusic.pdf

[18] Hutchins, B., "Fun with Red Noise" Electronotes Application Note No. 384, September 1, 2012
http://electronotes.netfirms.com/AN384.pdf

[19] Hutchins, B., "Clipping: Friend or Foe", Electronotes Application Note No. 397, June 15, 2013
http://electronotes.netfirms.com/AN397.pdf