The electronic music synthesizer and the physics of music

W. M. Hartmann
Department of Physics
Michigan State University
East Lansing, Michigan 48824
(Received 2 December 1974; revised 19 March 1975)

We describe the principal modules of analog electronic music synthesizers and discuss some ways that we have used a synthesizer in demonstrations, psychophysical experiments, and an undergraduate laboratory course in the physics of music and acoustics. We consider the synthesis of both steady and transient auditory phenomena.

I. INTRODUCTION

In this paper we consider the use of an electronic music synthesizer for demonstrations and laboratory experiments in the study of the physics and psychophysics of sound. Our emphasis is on analog creation of sounds, rather than the digital techniques used in computer synthesized music. Hybrid instruments, consisting of an analog synthesizer controlled by a computer, are a natural extension of the analog devices to be discussed.

In Sec. II we describe briefly the functions of the most important modules of a typical analog synthesizer. In Sec. III we discuss the use of the synthesizer in experiments on steady-state auditory phenomena, those in which the relevant auditory properties can be mainly characterized by a time-independent spectrum.

In Sec. IV we consider transient phenomena, defined as nonperiodic or subaudible periodic spectral fluctuations. Transient effects typically involve fluctuations with a time scale from 10 msec to 1 sec (e.g., attack and decay, vibrato, glissandi, and irregular intonation).

II. ANALOG SYNTHESIZERS

Modern electronic music synthesizers are available in a variety of sizes and prices from kits costing $200 to studio behemoths costing $20 000 or more. Most appropriate for an introductory physics course in sound is probably one of the class of small portable synthesizers known as ''minis,'' which cost $2000 or less but which include a keyboard and enough modules to provide a useful laboratory device. Although there are about a dozen different commercial mini analog synthesizers the basic concepts, described by Moog,[1] are common to all. These principles of operation are an interesting combination of musical utility and engineering convenience.

A useful musical synthesizer is a collection of electrical signal generating and processing modules with inputs and outputs that are mutually compatible. The degree to which the individual modules are hardwired together or permit independent inputs and outputs is a compromise between convenience and flexibility of operation and economy of manufacture. The principal feature of the synthesizer modules, however, is that significant parameters of the modules can be controlled by a control voltage input to the module. The basic voltage-controlled modules are as follows.

A. Voltage-controlled oscillator (VCO)

Mini synthesizers have three or four oscillators (function generators), typically generating triangle, square, and sawtooth wave forms. These wave forms have no particular auditory significance but are easily generated electronically and with suitable processing can be converted into interesting auditory signals. The inverse period (fundamental frequency) is variable from about 0.01 Hz through the range of orchestral instruments or higher. The precise tuning of a VCO is done with manual controls; the fundamental frequency f_V is then determined by a control voltage V over any selected (3- to 5-octave) portion of the total range, according to the law

$$f_V = \alpha \exp(V/V_0). \tag{1}$$

Constant α tunes the instrument and V_0 establishes a scale. A typical synthesizer keyboard switches contacts on a chain of resistors providing equal steps of increasing voltage for higher keys depressed at a rate of 1 V per keyboard octave. Therefore, synthesizer keyboard control is equitempered, but the number of notes per octave can be varied by manually changing the constant V_0. For the traditional 12-note scale, $V_0 = (\ln 2)^{-1}$ V. Less commonly, the frequency is linearly controlled:

$$f_V = \beta V. \tag{2}$$

B. Voltage-controlled filter (VCF)

A low-pass filter with at least a 12-dB/octave asymptotic cutoff is musically the most useful, but band-pass and high-pass filters are available on some synthesizers. Typically, the filter cutoff (or band center) frequency is voltage controlled, again according to Eq. (1). Therefore, if the same control voltage (say from the keyboard) is used to control both the VCO and the VCF, then the filter will track the frequency of the VCO. The VCF typically has a manual ''resonance'' control that can introduce a sharp peak at the cutoff or band-pass frequency into the response characteristic of the filter.

C. Voltage-controlled amplifier (VCA)

The voltage gain g $(0 \leq g \leq 2)$ of the VCA is proportional to a control voltage V. On some synthesizers the gain is optionally proportional to $\exp(V/V_1)$. Typically, it is a VCA that is used to determine the starting and stopping of a note played on the synthesizer. Turning on a note with the VCA preserves the spectral shape of the note, unlike the start-up of most conventional musical instruments. If the output of an oscillator running at low (subaudible) frequency is used to control the VCA, then a tremolo (periodic variation in loudness) will result.

The above three modules are the basic building blocks of a voltage-controlled synthesizer. The two fundamental processors are linear devices. Other available synthesizer modules will be discussed in remaining sections as appropriate. There are, for example, direct coupled mixers, to add two or more signal voltages or control voltages, balanced ("ring") modulators to form the algebraic product of two signals, reverberation springs, sample and hold circuits, clock pulse generators, preamplifiers for low-level external input signals, (1 mV, rms), and sequencers, to produce a series of control voltages. Most synthesizers include a source of white noise (constant average spectral power per unit frequency) and, in some cases, pink noise (constant average spectral power per octave). Furthermore, additional parameters may be voltage controlled. The wave form of the VCO or the width of the resonance peak of the VCF, for example, are controlled by a control voltage on some synthesizers. Some VCOs provide a rectangular wave with a pulse width that is voltage controlled.

Despite its keyboard, the synthesizer is basically a monophonic instrument; it plays one note at a time. For musical purposes one often brings all the available oscillators and processing modules to bear on the creation of the one note. Polyphony is accomplished in electronic music studios by successive overdubs on a multitrack tape recorder.

III. STEADY-STATE EXPERIMENTS

For physics demonstrations or lab experiments on steady-state signals, there is little that one can do with the synthesizer that one could not, in principle, do with the collection of audio apparatus typically available in physics departments. However, the convenience in patching and complete compatibility of the modules of a synthesizer allows a variety of experimentation that would be prohibitively cumbersome with conventional laboratory gear. Furthermore, the keyboard tuning provides a useful musically oriented framework for frequency control. On the other hand, control calibration on commercial synthesizers is generally inadequate and external instrumentation (frequency counters, voltmeters, and oscilloscopes) is required for quantitative experimentation. The accuracy of the various modules of synthesizers also reflects their musical orientation. Because human ears are sensitive to changes in pitch (± 0.02 semitone), synthesizer VCOs provide accurate frequencies with good short-term stability. Because human ears are less sensitive to audio intensity changes, synthesizer manufacturers can (and do) compromise on the accuracy of VCAs.

A. Modulation

Because its modules are voltage controlled, the synthesizer provides a signal source for demonstrations and experiments in modulation. Both amplitude modulation and frequency modulation are routinely available. A voltage-controlled pulse width rectangular wave generator allows pulse width modulation experiments to be performed. With an oscilloscope and a spectrum analyzer, one can observe the results of modulation.

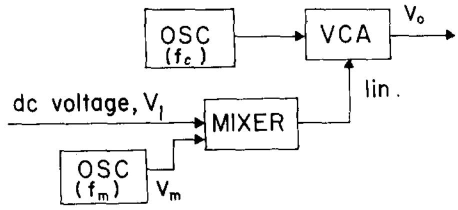

Fig. 1. Block diagram for amplitude modulation. Voltage V_1 must be positive so that the mixer output is nonnegative to avoid distortion. Generally, signal voltages are indicated by horizontal arrows. Control voltage inputs to voltage-controlled modules are indicated by vertical arrows.

To produce an amplitude-modulated output voltage v_0 from a carrier signal v_c with modulating signal v_m,

$$v_0(t) = [V_1 + v_m(t)]v_c(t), \tag{3}$$

the patch of Fig. 1 is made. The modulation percentage is $\max|v_m|/V_1$. When the fundamental frequencies f_c and f_m of v_c and v_m do not differ greatly, it is still easy to tune 100% modulation if v_m is a square wave. Relative to any component of the carrier, the square wave side bands occur at $\pm f_m$ (− 4 DB), $\pm 3f_m$ (− 14 dB), $\pm 5f_m$ (− 18 dB), etc. Note that the first side bands at $\pm f_m$ are actually stronger than the side bands for 100% modulation by a sine wave (− 6 dB).

Balanced modulation is a special case of amplitude modulation in which constant $V_1 = 0$ and the average values of v_m and v_c are also zero. If, for example, both v_m and v_c are sine waves, then the output v_0 will be the sum of two sine waves, one with frequency $f_m + f_c$ and the other with frequency $|f_m - f_c|$. Because the carrier signal is suppressed, balanced modulation can be used to alter drastically the pitch and timbre of an audio signal. The effect is a favorite among electronic musicians and modern composers. Because a VCA is only a two-quadrant multiplier, with zero output for all negative control voltages, the patch of Fig. 1 is not suitable for balanced modulation. Instead separate modules, called ring modulators for historical reasons,[2] are supplied to perform the required four-quadrant multiplication.

Frequency modulation (FM)[3] is obtained with the patch of Fig. 2. To simplify experiments in FM the frequency of VCO 2 should be a linear function of the control voltage [Eq. (2)] rather than exponential. Musically, an advantage of exponential control is that a given vibrato control voltage will result in equal variation in musical interval at all center frequencies.[1] The spectrum of an exponentially controlled VCO for wide-band FM has been calculated by Hutchins.[4] However, in the narrow-band limit, the differences between exponential and linear control can often be neglected.

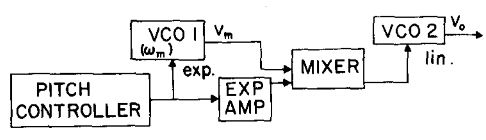

Fig. 2. Block diagram for frequency modulation in which the musical interval between the modulating signal from VCO 1 and the modulated signal from VCO 2 is constant for all notes of the scale.

In the special case that modulating signal v_m is a cosine wave of angular frequency ω_m, the instantaneous angular fundamental frequency of VCO 2 becomes

$$\omega_2 = \omega_c + \Delta\omega \cos\omega_m t. \quad (4)$$

The spectrum of the output v_0 contains an infinite number of side bands with spacing ω_m. The decay of spectral strength of the side bands away from the carrier frequency depends upon the modulation index[3] $\beta \equiv \Delta\omega/\omega_m$; narrow-band and wide-band frequency modulation regimes are characterized by $\beta < \pi/2$ and $\beta > \pi/2$. In electronic music, frequency modulation is often used to generate inharmonic tones simulating bells or gongs.

Before we discuss pulse modulation, we need to consider the spectrum of a pulse wave form with period T, which is in a low state for time T/n and in a high state for time $T(1 - 1/n)$. For $n = 2$ the function is a square wave, with all even harmonics missing. As n increases through integer values, making the pulse sharper, the total higher harmonic content increases relative to the fundamental, and those harmonics that are integer multiples of n are absent from the spectrum. For n nearly integral, the nth harmonic is small [cf., Eq. (6)]. The total audio power in the wave is proportional to the variance of the signal voltage v from its mean $\langle v \rangle$:

$$P \approx \langle v^2 \rangle - \langle v \rangle^2 \approx \frac{4(1 - 1/n)}{n}. \quad (5)$$

We therefore have the musically unusual situation that, for fixed amplitude, as the harmonic content increases the total power decreases. The 10-dB decrease of power as n increases from 2 to 40 can be measured with VU meters. For accurate measurements one must use a low fundamental frequency (≤ 100 Hz) to minimize the meter roll-off at high frequencies, which contribute importantly to the total power at large values of n. In Fig. 3 we show the theoretical spectrum of a rectangular wave with $n = 8$. Generally, the contribution of the pth harmonic to the power spectrum of a rectangular wave is

$$p^{-2}[2 - \cos(2\pi p/n)]. \quad (6)$$

and all the harmonics are in phase.

Throughout this paper we shall use the conventional audio intensity units of decibels. This logarithmic scale gives a distorted representation of the sensation of loudness, when the relative importance of high upper partials is compared to that of the lower partials of a tone.

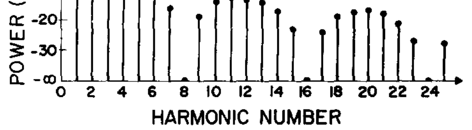

Fig. 3. Power spectrum of a rectangular wave that is low for 1/8 period and high for 7/8 period.

An easy way to make a pulse wave form is to input a triangular wave to a Schmitt trigger. If the frequency of the triangle wave is modulated, then the trigger output exhibits pulse-position modulation (PPM). If the dc central voltage of the triangle wave is modulated, relative to the trigger threshold voltages, then the output is a pulse width-modulated (PWM) signal.

In calculating the spectrum for the pulse signals, it is perhaps helpful to consider instead the time derivative of the pulse train. (The two amplitude spectra are related by a simple factor of ω_n.) Then the calculation of the spectrum of the modulated pulse train is isomorphic with the problem of scattering in a linear diatomic chain with alternating + and − delta function potentials and a periodic displacement of the atoms. Setting up the calculation in this way we find several properties of pulse modulation: (i) As in the case of frequency modulation, the number of side bands depends upon a modulation index that is an expansion parameter. For PPM the natural small parameter is $\Delta\omega/\omega_c$ and for PWM with pulse width τ, $\Delta(\tau^{-1})/\omega_c$. These small parameters are clearly different from the modulation index, $\Delta\omega/\omega_m$, for ordinary frequency modulation. The pulse width modulation technique above is really a modulation of the fraction of a period for which the output is high. Therefore, the parameter $\Delta(\tau^{-1})/\omega_c$ is determined only by voltage levels and is independent of ω_c; that is, the output spectra for different notes are congruent for a given ω_m. If the modulation frequency ω_m is varied, the relative importance of the various side bands is again left unchanged. (ii) To leading order in the expansion parameter (analogous to the phonon expansion for the diatomic chain), there are two side bands, 180° out of phase, for each harmonic of the unmodulated pulse wave described above. However, the power spectrum of these side bands oscillates at a different rate with increasing frequency than does the spectrum for the parent harmonics. (iii) In the case of PPM, the absence of a harmonic in the unmodulated pulse wave implies the absence of the modulation side bands to this harmonic. In the case of PWM, however, side bands of a harmonic may be strong even if the harmonic itself is absent from the spectrum by symmetry.

B. Filters

The synthesizer filter can be used in various qualitative experiments in Fourier series. Effective demonstrations make the output wave both audible, through an accurate audio system, and visible on an oscilloscope. Suppose that the square wave or sawtooth output of an oscillator is sent to a low-pass filter. If the filter cutoff frequency is manually increased, then the filter output provides a demonstration of cumulative Fourier synthesis of a square or sawtooth wave. The demonstration is greatly improved if the low-pass response function includes the sharp resonance for the following reason: If the low-pass response were ideally rectangular or if additive synthesis were used, then the cumulative synthesis of a square or sawtooth would include oscillations with as many peaks and valleys per cycle as indicated by the highest harmonic included. In practice, the usual maximally flat-amplitude-response low-pass filter *does not* lead to Gibbs phenomena of this sort. However, adding the resonance at the cutoff frequency produces oscillations by subtractive synthesis that are similar to those which would be produced by the more exact additive synthesis.

The sharply tuned band-pass filter of a synthesizer can be used in semiquantitative Fourier analysis. By using the keyboard to control the filter passband, one can quickly pull out the various harmonics of a short-lived tone such as a percussive or plucked note. One simply tunes the filter so that the passband is centered on the fundamental frequency of the note to be analyzed when key C1 is depressed. Then depressing keys C2, G2, C3, E3, G3, B*b*3, and C4 will reveal harmonics numbered 2–8. (The filter is typically broad enough that B*b* is an adequate approximation to the seventh harmonic.) With some careful coordination with the instrumentalist, the filter keyboard operator can select specific harmonics of the notes of a tune to transform a tune played by the instrumentalist into quite a different tune.

A low-pass filter which tracks a VCO is useful in creating reference sounds used in pitch-matching experiments. None of the usual wave forms produced by a function generator is truly satisfactory for these experiments. Many musical people have difficulty matching the pitches of sine waves. Square and triangle waves contain no even harmonics, and sawtooth or ramp waves are too rich in harmonics. A low-pass-filtered sawtooth wave is a good choice for critical experiments.

C. Noise bands

Some interesting acoustical experiments involve mainly the white noise source of the synthesizer.

Because of room resonances, a white noise source is the best choice of signal to demonstrate the inverse square law for audio intensity. The loudspeaker should be pointed directly at a microphone and the microphone output read on a VU meter. Different microphone–speaker separations can be measured with a tape measure to confirm the inverse square law for separations less than about 10 ft. For larger separations in the average lecture hall, one notes that the constant reverberant sound field is equally as strong as the direct field.[5]

If one runs white noise through a sharply resonant band-pass filter and controls the passband frequency with the keyboard, one has something like the ultimate in subtractive synthesis. By playing the theme from "The Bridge on the River Kwai," one obtains a qualtitative notion of the physics of choral whistling.

In a more serious vein, the patch of the above item provides a well-controlled band of noise to be used in masking experiments. One can easily demonstrate the asymmetry of masking for loud sounds; namely, a band of noise will effectively mask a tone of frequency higher than the noise band but will not effectively mask a tone lower than the noise band.[6]

D. Repetition pitch

A psychophysical effect with implications for the theory of hearing is that of repetition pitch. Bilsen and Ritsma[7] give a charming history of this effect and have reviewed experiments in which a wide-band signal plus the same signal delayed by τ seconds ($1 < \tau < 10$ msec) is presented to one or both ears of a subject. The subject perceives a pitch with an apparent frequency equal to $1/\tau$.

We find that the effect can be easily observed by a subject between two loudspeakers placed near the center of a large room, with both driven (in phase) by a white noise source as shown in Fig. 4.

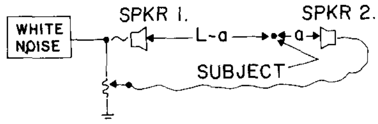

Fig. 4. Experimental arrangement for demonstrating "repetition pitch."

Relative to the noise from the near speaker (2), the noise from speaker 1 is delayed by $\tau = (L - 2a)/c$, where c is the speed of sound. The subject is given a remote control box for tuning a filtered sawtooth wave of small amplitude and for switching this signal in and out of speaker 2. The subject tunes the VCO/filter to match the pitch of the noise, the VCO frequency is read on a counter, and the experimental result is compared with the expected perceived frequency,

$$f = 13\,560/(L - 2a), \quad (7)$$

for room temperature, where L and a are in inches. For our experiments we took L to be 150 in. and tried to make both speakers sound equally loud to the subject. Moving the head between the speakers immediately verifies that the noise is colored—has some tonal character. To identify the pitch accurately requires a little practice. However, once the subject has learned the art, he can identify the pitch of the noise *within the accuracy of the length measurement* a, at least for frequencies between 500 and 150 Hz. The above experimental arrangement prompts the observation that the constant spacing between the spectral peaks is the frequency for which the difference between speaker–subject distances ($L - 2a$) is equal to one wavelength. Although experiments with short pulses of sound and with phase-shifted signals appear to make the effect a complicated one,[7] it is instructive to consider only the power spectrum, $P(\omega)$. For time delay τ and speakers in phase,

$$P(\omega) = \left| \int_{-\infty}^{\infty} [v(t) + v(t-\tau)] \exp(i\omega t)\, dt \right|^2 = 2\,|v(\omega)|^2[1 + \cos(\omega\tau)]. \quad (8)$$

Since $|v(\omega)|^2$ is constant by construction, the spectrum consists of a periodic series of broad lumps with common spacing $\Delta f = \tau^{-1}$. To first approximation, the pitch is unchanged when the plus sign above is replaced by a minus by phase reversal. Therefore, it is not the center frequency of the lowest spectral band that is responsible for the pitch cue. Rather, many of the equally spaced broad lines contribute to the sensation of pitch at frequency τ^{-1}.

E. Periodicity pitch

The effect of periodicity pitch has been important in the evolution of the current theories of pitch perception.[8,9] If one listens to the sum of two sine waves,

$$v(t) = v_0\{\sin[2\pi(2f_1t)] + \gamma \sin[2\pi(3f_1t)]\} \quad (\gamma < 1), \quad (9)$$

then one may perceive a note of fundamental frequency f_1, despite the fact that the spectrum of $v(t)$ contains no component at frequency f_1. A variety of experiments have shown that the effect is not caused by mechanical nonlinear distortion of the ear.[8]

Two VCOs of a synthesizer can be tuned a perfect fifth apart (frequency ratio of 3/2) and phase locked to remain in that (nonequitempered) relationship over the entire keyboard. Therefore, one can play a tune or any portion thereof by exploiting the effect of periodicity pitch.

F. The oboe

Experiments on periodicity pitch with only two principal components reveal the effect to be something of a "Cheshire cat," obvious one moment, elusive the next, and sometimes dependent on musical context. Curiously, the pitch of an oboe, the traditional tuning reference for orchestras, may owe much to the effect of periodicity pitch. We find that our most successful electronic simulation of oboe and other double-reed timbres is a tone with little spectral strength at the fundamental frequency. To obtain such a tone, we use frequency modulation with modulating and carrier frequencies that have an integral ratio. The digital version of this technique has been discussed by Chowning,[10] who also considers transient effects. To create a double-reed timbre, we use the FM configuration of Fig. 5 with a sine wave from VCO 1 tuned to frequency f_1 and a triangle wave from VCO 2 tuned to frequency $3f_1$. With a modulation index $\beta = 0.45$, the double-reed effect occurs over a frequency range $110 \leq f_1 \leq 600$ Hz. The spectrum of $A440$, measured with a spectrum analyzer, appears in Fig. 6 and is in agreement with the measured modulation index. The fundamental frequency component is down by 35 dB relative to the third harmonic and is the weakest of the first ten harmonics. Nevertheless, a pitch of frequency f_1 is unmistakable. The nasal double-reed cue is also clear, and with an appropriate starting envelope the above signal can easily be mistaken, in a crowd, for an oboe. Although our range for the double-reed cue is approximately that for which the third harmonic falls within an alleged formant band of the oboe,[11] the extent to which something useful is learned about oboes by this exercise is still an open question. Wedin and Goude[12] have measured the spectrum of an oboe for $A440$ and also find that the fundamental is weak compared with upper partials. We conjecture that the oboe is used to tune the orchestra for the following reason: Because the pitch of the oboe sound derives from its upper partials, the pitch is not masked as other instruments tune to the oboe's fundamental frequency.

G. Scale studies

Because the synthesizer keyboard is equitempered (though possibly microtonal) and basically monophonic, the synthesizer is not the ideal device for the study of musical scales. However, the collection of synthesizer oscillators, tunable by manual controls and by a common

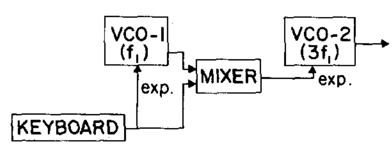

Fig. 5. Patch for generating a steady-state double-reed-like timbre by using frequency modulation. The harmonics of VCO 2 without modulation occur at $3f_1$, $9f_1$, $15f_1$, etc.

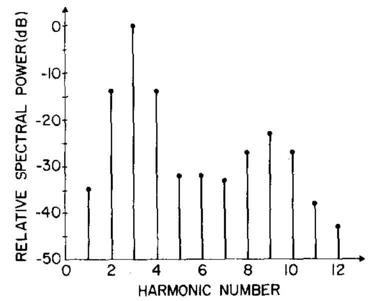

Fig. 6. Power spectrum of the steady-state synthetic oboe tone for $A440$, obtained by using the patch of Fig. 5.

keyboard control voltage, in conjunction with a strobe tuner with a vernier calibrated in cents, allows one to compare individual equitempered intervals with corresponding intervals in other temperaments. It is difficult, if not impossible, to perform synthesizer scale studies in a musical context, as one could do by retuning a piano, but this deficiency is partly compensated by the accuracy and stability of tones created electroncially. One approach to scale studies, mentioned by Hall,[13] is to control the oscillators of the analog synthesizer with control voltages from a digital computer. Since only relative pitches are of primary interest in scale studies, this experimental arrangement does not depend upon long-term stability of the VCOs. However, with minicomputer control some compromise is required between range of control and accuracy due to digitizing errors. A 12-bit word provides only 4096 possible values. On the average, this set of discrete control voltages is optimally used if the digitizing interval corresponds to a constant number of cents (not hertz). Therefore, an exponentially (not linearly) controlled VCO is indicated.

In general, if a maximum pitch error of n cents will be tolerated and the computer word has N bits, then the exponential VCO will have a range of ϕ_e octaves, where

$$\phi_e = 2^N n/1200 \tag{10}$$

and the errors will be of constant size over the entire range. If, on the other hand, a linear VCO is used, then the range is ϕ_l octaves, where

$$\phi_l = \frac{\ln(1 + \phi_e \ln 2)}{\ln 2} \quad (n \ll 1200) \tag{11}$$

and the errors diminish as frequency increases. For example, if we require 1-cent accuracy from a 12-bit word, then $\phi_e = 3.4$ octaves but ϕ_l is only 1.6 octaves.[14]

IV. TRANSIENT PHENOMENA

It is in the area of transient auditory phenomena that an electronic music synthesizer gains its full potential for both music making and acoustical experiments. One of the most useful things that can be learned from experiments with a synthesizer is a notion of the degree to which musical timbre, including instrument identification, is associated with transient behavior and not with the steady-state spectrum of a tone. The importance of starting transients to timbre has been proven by numerous experiments with spliced tape recordings. Several students of musical instrument timbres have analyzed the transients

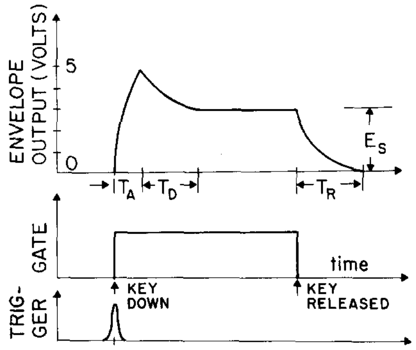

Fig. 7. Function of an envelope generator, showing attack (A), decay (D), sustain (S), and release (R) parameters, plotted with gate and trigger signals.

and recreated realistic tones by subsequent synthesis. Nevertheless, the point requires continued emphasis, especially within physics circles. For example, the oboe timbre discussed above can be transformed into an *even better* banjo by a suitable choice of starting transients!

The transient response of instruments is difficult to measure. Because of the expense of real-time spectrum analyzers and the complexity of a fast Fourier analysis analog/digital computer system it may be that electronic synthesis is the only way to gain some quantitative impressions of the effect of musical transients with a modest investment of funds and time.

The electronic music synthesizer introduces transients into voltage-controlled modules by special one-shot function generators called "envelope generators." While a key is depressed on the synthesizer keyboard, a gate signal is high. The positive derivative of the gate when the key is first depressed is a trigger pulse that is typically used to initiate an envelope as shown in Fig. 7. The attack (A), decay (D), and release (R) times are manually variable from a lower limit, as low as 2 msec, to an upper limit, as large as 15 sec, depending upon the particular synthesizer. The sustain voltage E_s is manually variable from 0 to the maximum envelope voltage, typically 5–8 V. As expected, the curving portions of the envelope obey exponential laws, characteristic of discharging capacitors and of room reverberations. The attack transients of conventional instruments are not exponential, but an exponential is a useful simple approximation. The most obvious use of the envelope generator is to envelop the amplitude of notes played on the keyboard by controlling a VCA which amplifies the audio signal. We begin our discussion of transient effects with a few straightforward experiments on short notes.

A. Loudness of short tones

The duration of notes shorter than 0.5 sec is an important psychophysical variable. For example, the sensation of loudness of a short tone does not depend upon amplitude alone but also depends strongly on the duration of the tone. For imprecise but musically realistic demonstration of this effect, the envelope of Fig. 7 can be used with $E_s = 0$ so that the trigger initiates an automatic and reproducible envelope, attack–decay. For more precise experiments, the envelope may be replaced with the gate of a timer and the gate time measured with digital counter. In this way the VCA approximates a tone burst generator. The VCA is a balanced differential amplifier so that no component of the control voltage appears in the signal output; the turnon and turnoff of notes is free of noise. As the duration of the tone is increased from zero, the sensation of loudness increases from zero to the asymptotic value for sustained tones. The lower the frequency of the tone, the longer the tone duration must be to perceive a certain fraction of the asymptotic loudness. For example, a 100-Hz tone is perceived at half its asymptotic value if it is 300 msec long, whereas a 10 000-Hz tone need be only 125 msec long for its loudness to be half the asymptotic value.[9] Various criteria for loudness may be used experimentally. The amplitude threshold for hearing short tones or for hearing short tones masked by white noise has been used by Plomp and Bouman.[15]

B. Pitch of short tones

Tone duration also has an effect on our perception of pitch. A tone which lasts for less than 10 msec sounds like a click, and a subject finds it difficult to match this tone in any pitch-matching experiment. Such an experiment on short tones can be done with the apparatus as described above and with a frequency counter to count both the frequency of the tone which is caused to burst, either once or repeatedly, by the VCA and the frequency of an oscillator tuned by the subject who tries to match the pitch.

It has been conjectured that a psychoacoustical uncertainty principle applies to this experiment[16]:

$$\Delta f \Delta t = K, \tag{12}$$

where Δt is the duration of the tone burst and Δf is the width of the distribution of tones matched by the subject. Either the responses of different subjects to the same tone or the same subject to different tones could be used to provide the distribution. The power spectrum of a sine wave of frequency ω_0 that is instantly started and instantly stopped after Δt has elapsed is

$$P(\omega) \approx \frac{\sin^2(\omega - \omega_0)\Delta t/2}{(\omega - \omega_0)^2}. \tag{13}$$

The fullwidth of the principal maximum of the familiar distribution in Eq. (13) is such that $K = 2$.

If the human ear–brain system merely Fourier analyzed an input signal, in the manner of a spectrum analyzer, then we would expect that the results of the above experiment would follow Eq. (12) with $K = 2$ or $K = 1$. Instead it appears experimentally that, if subjects are required to determine which of two short tones is the higher, then the data[17] are best fitted to Eq. (12) with $K = 0.2$; that is, the human ear seems to beat the uncertainty principle by nearly an order of magnitude. If the subject hears only a single tone burst and is required to tune an oscillator to match the pitch, we find considerable spread in the data. Nevertheless, if we discard data in which the subject jumped octaves or other large musical intervals, then we find that some subjects, especially musicians, can still beat the uncertainty principle convincingly. Our results are quite tentative and further work should be done in this area.

C. Speed of sound

A repeated tone burst achieved by triggering an envelope generator with a low-frequency oscillator can be used in a simple speed of sound experiment. The regular series of tone bursts triggers a dual-trace oscilloscope. One compares the input to a loudspeaker on one scope channel with the output of a distant microphone on the other channel, using the calibrated sweep to measure the delay time. Several prominent echos are typically observable in the microphone output as well.

D. Backwards music

The sound of music or speech played backwards on a tape recorder is a familiar effect. It requires only a small amount of experimenting with the envelope generators of a synthesizer to conclude that most of the effect of backwards music is caused simply by a reversal of the usual transient times, short attack times, and relatively longer decay times. Using only the attack–decay envelopes and a reversible (or fullwidth) tape recorder, one can create the four envelopes of Fig. 8. Envelope (a) is characteristic of a percussive tone, like a toy piano. To make envelope (b), the attack and decay times are interchanged. Envelopes (c) and (d) are (a) and (b) played backwards. The point to be made is that envelopes (a) and (d) sound conventional to our ears, but tones made with envelopes (b) and (c) sound like music played backwards. The curvature of the envelopes *does* have an effect on the timbre of the sound; for example, (a) sounds more natural than (d). Nevertheless, the actual values of attack and decay times are still more important than the curvature. Listening to notes with long attacks and short decays reminds one of several other sounds. An accordian can be played so as to produce the effect, and certain words in the Russian language (e.g., *nyet*) seem to be of this nature.

E. Vibrato–tremolo

Somewhat arbitrarily we include vibrato (frequency modulation with subaudible f_m) and tremolo (amplitude modulation with subaudible f_m) as transient effects. An instrumentalist playing an acoustic instrument probably cannot employ one of these expressive techniques without some of the other. Electronic synthesis is therefore useful to try to sort out the two effects. As long as the period and depth of the modulations are within musically conservative ranges, it is quite common for the listener to be unsure of which effect is being presented.

Vibrato has been studied a great deal, particularly with respect to vocalists.[18] The modulation frequency is from 5 to 8 Hz and the depth is about ¼ tone. The vibrato wave form is close to sinusoidal. With a VCO, interesting experiments can be done on vibrato. There is a strong correlation between the perceived depth of a vibrato and the vibrato frequency; namely, the greater the vibrato frequency the smaller is the sensation of vibrato depth. Another practical question is that of the perceived pitch of a tone with vibrato. Although most physicists would instinctively guess that a listener hears the average frequency, players of bowed string instruments often claim that the perceived pitch is close to the highest instantaneous frequency. Synthesizer experiments done at the University of South Carolina by F. X. Byrne, L. Cathey, D. Haines, and the author,[19] presenting varied vibrato rates, depths, wave forms, and central pitches to a dozen subjects who were asked to make nonsimultaneous pitch-matching tunings, suggest that on the average listeners perceive the average frequency. It appears therefore that some physical property other than the frequency is responsible for the perceived rise in the pitch of a bowed string tone when vibrato is introduced.

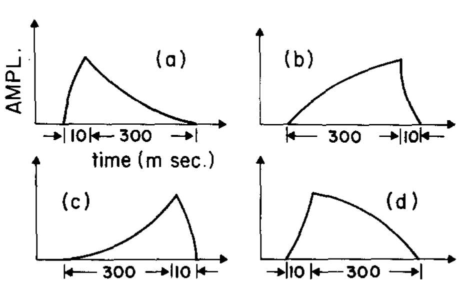

Fig. 8. Four envelopes used to study backwards music. Envelopes (c) and (d) are envelopes (a) and (b) played backwards.

F. Synthetic strings

Matthews and Kohut[20] have pointed out that a significant aspect of violin tones is a frequency modulation (vibrato)-induced amplitude modulation (AM). The resonances of a violin box are considered to be narrow, numerous, and unequally spaced in frequency. Therefore, as the frequency varies with vibrato, the various harmonics of the tone are enhanced or attenuated as they momentarily coincide with peaks or dips of the box response curve. The effect can be simulated with a synthesizer by beginning with a frequency-modulated tone of high harmonic content and passing through a fixed filter bank with alternate channels set high and low. Typically, one might use a narrow pulse wave[21] and pass through the two sections of a ± 12-dB octave-band stereo equalizer connected in series. This patch does not result in a plausible violin sound, probably because the high-frequency resonances of the fixed filter are too broad.

Table I. Spectra of the upper and lower limits of an *A*110 synthetic cello tone to estimate the extent of amplitude modulation induced by vibrato. The strengths of the second harmonic were the same for both upper and lower tones, coincidentally.

Harmonic	Upper note (dB)	Lower note (dB)
1	−28	−30
2	0	0
3	−29	−26
4	−25	−23
5	−29	−29
6	−19	−21
7	−12	−12
8	−14	−12
9	−20	−20
10	−26	−27
11	−23	−23
12	−28	−26
13	−30	−28
14	−31	−30
15	−34	−33

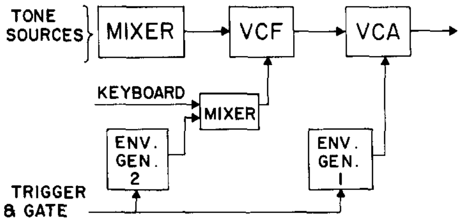

Fig. 9. Block diagram of the canonical signal processing path for a mini synthesizer emphasizing the control of transients.

However, we find that a plausible *cello* timbre is possible with this patch, though only within a range of less than one octave. We examined a successful synthetic cello note, $A110$, with a vibrato depth of slightly more than ¼ tone. We made a spectral analysis of the top and bottom notes of the vibrato to determine the extent of the amplitude modulation. The spectral strengths are given in Table I. None of the differences is very large, but the ear can easily hear out strong vibrato fluctuations in the third harmonic. It did seem that the FM-induced AM, made a positive contribution to the cello cue, though we believe that sharper resonances than those of our maximally spiked ± 24-dB octave-band filter would have been preferable. The list of relative strengths of the various harmonics in Table I is not an attempt to characterize a cello tone with precision. Between the half dozen adjacent synthetic tones which sounded like a cello there were significant relative spectral differences. A tendency for a bump in the region of the seventh harmonic, a relatively weak fundamental, and significant spectral strength in higher partials were common features of these tones.

G. Spectral evolution

An important feature of sustained sounds generated by mechanical means, such as musical instruments, is that the spectral composition of the onset and (less importantly) the fall of the sound is different from that in the steady state. For musical instruments the harmonic content changes as the attack of the note progresses. Spectral studies of this effect have been made by real-time spectral analysis.[22–25] Physically, the effect results from the nonlinear coupling of the various resonant modes of an instrument via the regeneration mechanism, so that the oscillation involves different amounts of these resonant modes at different oscillation amplitudes.[26] The degree of participation of the resonances of higher frequency increases as the tone becomes louder, leading to a change in relative harmonic content, typically an average increase in the strength of upper partials.

This effect is so important to musical timbre that it is essentially responsible for the canonical arrangement of modules in mini synthesizers as shown in Fig. 9.

As the amplitude of the note rises and falls, determined by envelope generator 1, the harmonic content changes because envelope generator 2 is moving the passband of the filter. Essentially, the *nonlinear* mechanism of acoustic instruments that leads to an amplitude-dependent spectrum is electronically simulated by a patch of *linear* components. With this standard patch it is possible to explore somewhat the role of starting transients. For example one can proceed smoothly from a tuba-like sound, with upper partials which build slowly, to a plucked string bass by using a low-pass VCF. In Table II typical envelope parameters are given for the two limits, though the exact values depend upon musical context. The critical timbre times are the 150-m sec filter delay in the tuba attack and the 300-msec fall time for the filter in the plucked bass decay.

Table II. Envelope parameters for demonstrating the effect of transients on timbre. As in Fig. 9, envelope 1 controls the VCA and envelope 2 controls a low-pass VCF.

	Tuba		Plucked string bass	
	Envelope 1	Envelope 2	Envelope 1	Envelope 2
T_A	15 msec	150 msec	5 msec	5 msec
T_D	450 msec	450 msec	400 msec	300 msec
E_s	20%	40%	0	0
T_R	450 msec	450 msec	—	—

V. CONCLUSION

In the above sections we have indicated some of the ways in which we have used an electronic music synthesizer in the teaching of the physics of music and acoustics. However, the potential of electronic sound synthesis is so vast that a brief presentation such as this one is bound to be incomplete and the choice of experiments somewhat arbitrary.

As expected, the synthesizer has proved to be a very popular laboratory device among students. Arts students who instinctively balk at trigonometry will gleefully learn about the product $\sin A \; \sin B$ when ring modulation is the context. An essential point of all demonstrations and experiments in musical physics is that the principal measuring instrument is the human ear–brain system. The results of an experiment are directly perceived, with much of the emotional excitement characteristic of human response to musical sounds. Mechanics and the physical principles of waves acquire an obvious relevance when they are demonstrated in a musical context.

ACKNOWLEDGMENTS

I am grateful to Professor D. L. Wessel, Professor T. D. Rossing, Professor J. G. Roederer, Professor J. W. Beauchamp, Professor A. H. Benade, Professor R. T. Schumacher, and Dr. R. A. Moog, for helpful discussions. Michigan State physics students Brent Blumenstock, Ted Green, and Paul Newman contributed importantly to some of the experiments discussed. My sense of musical proportion has profited greatly from extended synthesizer recording sessions with P. J. Hartmann. Equipment for our physics of music course was bought with a grant from the Instructional Scientific Equipment Program of the National Science Foundation.

[1]R. A. Moog, J. Audio Eng. Soc. **13**, 200 (1965).
[2]N. H. Crowhurst, *Electronic Musical Instruments* (Tab, 1971).
[3]M. Schwartz, *Information Transmission Modulation and Noise* (McGraw-Hill, New York, 1959).
[4]B. A. Hutchins, Electronotes **5** (38), 4 (1974); **5** (39), 7 (1974).
[5]H. Kuttruff, *Room Acoustics* (Wiley, New York, 1973), p. 118.
[6]J. Backus, *The Acoustical Foundations of Music* (Norton, New York, 1969).

[7]F. A. Bilsen and R. J. Ritsma, Acustica **22,** 63 (1969/70); see also F. A. Bilsen and J. L. Goldstein, J. Acoust. Soc. Am. **55,** 292 (1974).
[8]F. L. Wightman and D. M. Green, Am. Sci. **62,** 208 (1974).
[9]J. G. Roederer, *Introduction to the Physics and Psychophysics of Music* (Springer-Verlag, New York, 1973).
[10]J. M. Chowning, J. Audio Eng. Soc. **21,** 526 (1973).
[11]H. F. Olson, *Music, Physics and Engineering* (McGraw-Hill, New York, 1952; Dover, New York, 1967), 2nd ed.
[12]L. Wedin and G. Goude, Scand. J. Psychol. **13,** 228 (1972).
[13]D. E. Hall, Am. J. Phys. **42,** 543 (1974).
[14]These formulas, of course, make no allowance for possible errors in digital-to-analog conversion and in the VCO pitch control.
[15]R. Plomp and M. A. Bouman, J. Acoust. Soc. Am. **31,** 749 (1959).
[16]J. C. R. Licklider, in *Handbook of Experimental Psychology,* edited by S. S. Stevens (Wiley, New York, 1951), p. 985 ff.
[17]D. A. Ronken, J. Acoust. Soc. Am. **49,** 1232 (1971); **51,** 1947 (1972).
[18]C. E. Seashore, *Psychology of Music* (McGraw-Hill, New York, 1938; Dover, New York, 1967).
[19]*Experiments on Physics in the Arts, 1974 NSF Workshop,* edited by R. D. Edge (University of South Carolina, Columbia, SC, 1974).
[20]M. V. Matthews and J. Kohut, J. Acoust. Soc. Am. **53,** 1620 (1973).
[21]J. C. Schelling, Sci. Am. **230** (1), 87 (1974). A narrow pulse is an approximation to the velocity of a bowed string. Six-decibels/octave low-pass filtering produces the displacement.
[22]D. Luce and M. Clark, J. Acoust. Soc. Am. **42,** 1232 (1967).
[23]J. W. Beauchamp, *Music by Computers* (Wiley, New York, 1969).
[24]J. W. Beauchamp, J. Acoust. Soc. Am. **56,** 995 (1974).
[25]J. C. Risset and M. V. Matthews, Phys. Today **22,** 23 (1969).
[26]A. H. Benade, Sci. Am. **229** (1), 24 (1973).

SOCIAL STRATIFICATION IN SCIENCE

. . . the institution of science is a highly stratified one. There is at least as much inequality in science as there is in other social institutions. At the same time, . . . the social processes that produce stratification in science may approximate the ideal of a meritocracy based on the application of universalistic principles, and may do this far more than most, perhaps more than all, institutions in our society. Implicitly we have suggested the existence of a significant paradox: that a relatively stable system which is full of sharp stratification of rewards is at its core fundamentally committed to change—intellectual change and "progress" as well as social change. Elites hold their positions only as long as they are active scientists. The notion of passing on "privilege" by anything short of scientific competence is foreign to this institution. Leading scientists can come from any social background if they have the stuff to produce good science. It appears possible that in science, from one generation to another, "the last may be first."

—J.R. Cole and S. Cole, *Social Stratification in Science*
(University of Chicago, Chicago, 1973).