The Synthesis of Complex Audio Spectra by Means of Discrete Summation Formulas

JAMES A. MOORER

Department of Music, Stanford University, Stanford, CA 94305

A new family of economical and versatile synthesis techniques has been discovered, which provide a means of controlling the spectra of audio signals, that has capabilities and control similar to those of Chowning's frequency modulation technique. The advantages of the current methods over frequency modulation synthesis are that the signal can be exactly limited to a specified number of partials, and that "one-sided" spectra can be conveniently synthesized.

INTRODUCTION: The search for ways to give electronic sounds a "natural" quality has been considered very important among many researchers. Most methods for doing so were quite cumbersome until Chowning's marvelous discovery that natural sounding tones could be produced in a simple manner by the use of frequency modulation to generate complex, time-varying spectra [1].

By way of review, the premise of frequency modulation synthesis is that "the character of the temporal evolution of the spectral components is of critical importance in the determination of timbre" [1]. Through our studies of the physical properties of actual music instrument tones and their perceptual implications [2], it is clear that the perception of timbre is a very complex subject, but that it is critically dependent upon temporal change. This gives us hope that by simulating certain changing properties of a music instrument tone we can give the illusion of the instrument without necessarily having to duplicate the waveform exactly. Our studies with frequency modulation synthesis have given us much encouragement in this direction.

Although it is not necessarily the goal of electronic musicians to be able to synthesize exactly the sounds of orchestral instruments, for there are numerous excellent musicians who can do this quite well, it is of interest to use these tones as starting points for the exploration of timbre. In this way, we begin with a tone that already sounds "natural", so that hopefully our extensions into un-known timbres can preserve some of this natural quality.

In this paper we discuss four members of a family of synthesis techniques that includes frequency modulation synthesis as well as many others. We call this family *discrete summation formulas*, because they are derived as sums, finite or infinite, of trigonometric series which have closed-form sums. The hope is that the closed forms are easier to compute than the entire series itself. In many cases, such as frequency modulation synthesis, this is in fact the case.

We should point out that this method was discovered independently by David Lewin of Harvard University somewhat after our own discovery [3]. Since the formulas are quite old, it is possible that the discovery has been made by still other researchers.

DISCRETE SUMMATION FORMULAS

The equation we will start with has been known for quite a few years. Although we shall not do so here, it can be simply derived as an exponential series:

$$\sum_{k=0}^{N} a^k \sin(\theta + k\beta) =$$

$$\frac{\sin\theta - a\sin(\theta-\beta) - a^{N+1}[\sin\{\theta+(N+1)\beta\} - a\sin(\theta+N\beta)]}{1 + a^2 - 2a\cos\beta}. \tag{1}$$

This can be found in Jolley [4] and many others. In this form it is a mathematical identity, valid for all values of

a, θ, and β. We can make this into a synthesis technique by setting θ to $2\pi f_c t$ and β to $2\pi f_m t$, mimicking some of the notation from frequency modulation synthesis. Accordingly, we will call f_c the center frequency and f_m the modulating frequency. We will call a the ratio, because the amplitude of each partial is related to the amplitudes of adjacent partials by factors of a.

Note that we can make a harmonic series by setting $mf_m = nf_c$, where m and n are integers. Similarly, we can make an inharmonic series by making m or n into irrational numbers.

Fig. 1 shows an example where $f_c = 3000$ Hz, $f_m = 2000$ Hz, $N = 3$, and $a = 0.5$. This plot is the discrete Fourier transform of an actual waveform generated through the use of Eq. (1). As we see, each partial is 6 dB lower than the previous partial, just as would be predicted from setting the ratio to 0.5. Fig. 2 shows an example like the previous one but where $a = 2.0$, producing a spectrum such that each succeeding partial is 6 dB greater than the previous one. This gives us a way to produce an exactly band-limited signal where the ratio of the amplitudes of adjacent harmonics can be controlled as a function of time to produce time-varying spectra.

A more simple version can be produced by allowing the summation to extend to infinity. This produces the following:

$$\sum_{k=0}^{\infty} a^k \sin(\theta + k\beta) = \frac{\sin\theta - a\sin(\theta - \beta)}{1 + a^2 - 2a\cos\beta}, \qquad a < 1. \quad (2)$$

If we use a ratio that is small enough so that the partials become very small as their frequencies approach half the sampling frequency, as is commonly done with frequency modulation synthesis, then there is no objectionable aliasing. This form is, of course, restricted to values of the ratio that are less than 1, for the sum does not converge for $a \geq 1$. We can easily see that if $f_c = f_m$, the numerator can be simplified even further.

These formulas express "one-sided" spectra. The partials extend in positive integral multiples of f_m from f_c. There are several ways to produce "two-sided" spectra. Possibly the cleanest way is through the following formula, which can be easily derived from Eq. (1):

$$\sin\theta + \sum_{k=1}^{N} a^k \{\sin(\theta + k\beta) + \sin(\theta - k\beta)\}$$

$$= \frac{\sin\theta\,(1 - a^2 - 2a^{N+1}\,[\cos\{(N+1)\beta\} - a\cos N\beta])}{1 + a^2 - 2a\cos\beta}. \quad (3)$$

In Fig. 3 we show an example of Eq. (3) for $f_c = 7000$ Hz, $f_m = 2000$ Hz, $a = 0.5$, and $N = 2$. Likewise, Fig. 4 is the same except that $a = 2$. Again, these figures are discrete Fourier transforms of actual waveforms generated by Eq. (3). We can take the sum to infinity, which gives an unusually compact formula:

$$\sin\theta + \sum_{k=1}^{\infty} a^k \{\sin(\theta + k\beta) + \sin(\theta - k\beta)\}$$

$$= \frac{(1 - a^2)\sin\theta}{1 + a^2 - 2a\cos\beta}, \qquad a < 1. \quad (4)$$
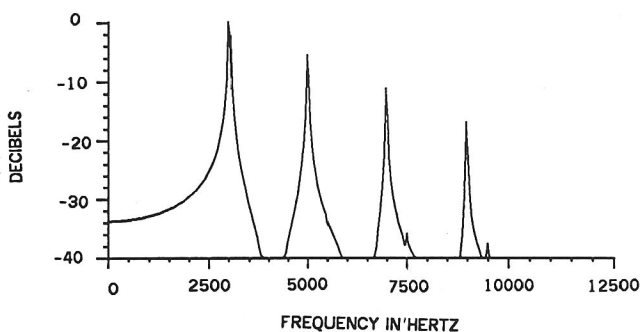

Fig. 1. Spectrum of a waveform produced by Eq. (1) for $F_c = 3000$ Hz, $F_m = 2000$ Hz, $N = 3$, and $a = 0.5$. This is the discrete Fourier transform of an actual synthetic waveform. Since this is a numerical and not a theoretical transform, the peaks have nonzero width, and other peaks (at 7500 and 9500 Hz) are evident due to numerical considerations.
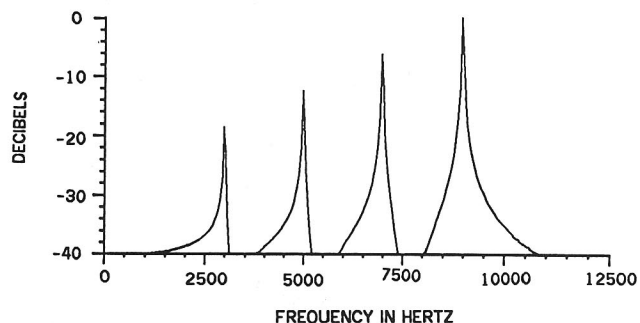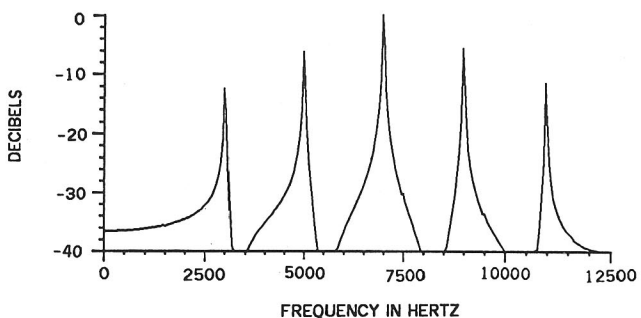

Fig. 3. Spectrum of a waveform produced by Eq. (3) for $F_c = 7000$ Hz, $F_m = 2000$ Hz, $N = 2$, and $a = 0.5$.

Fig. 2. Spectrum of a waveform produced by Eq. (1) for $F_c = 3000$ Hz, $F_m = 2000$ Hz, $N = 3$, and $a = 2.0$.
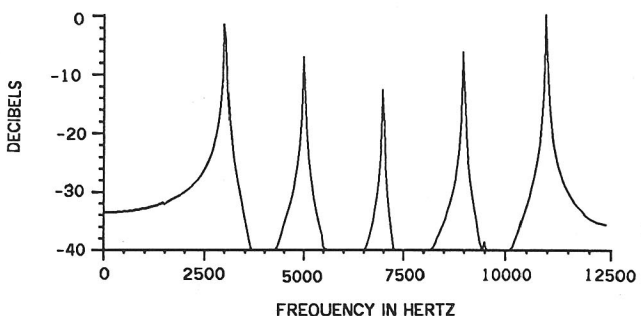

Fig. 4. Spectrum of a waveform produced by Eq. (3) for $F_c = 7000$ Hz, $F_m = 2000$ Hz, $N = 2$, and $a = 2.0$.

These are but a few of the possible formulas that can be derived from Eq. (1). There are a tremendous number of variations. For instance, if we desire only odd harmonics, we can set $f_m = 2f_c$ and get quite a bit of simplification. We hope that this exposition will spur the imagination of the reader and result in a whole range of new and useful formulas.

NORMALIZATION

The formulas presented in Eqs. (1) through (4) have the objectionable quality that the amplitude of the signal changes as the ratio a changes. As in frequency modulation synthesis, it is important that we be able to maintain the same amplitude throughout the whole range of values of a. The question is, what do we mean by amplitude? There are, in fact, several different normalization criteria. Ideally, what we would like to normalize is the loudness of the signal for a wide range of ratios. Unfortunately, this would require a detailed model of loudness perception, which tends to be quite complicated (see, for instance, the model of Zwicker and Scharf [5]). Given that we cannot easily normalize for loudness, there are two other choices that will be discussed here, normalizing for peak amplitude and normalizing for constant power.

If we choose to synthesize a harmonic spectrum, and we use a cosine series rather than a sine series, then all the spectral components will achieve their maximum value at the beginning of the period, so that we can compute the maximum amplitude of the waveform simply by taking the sum of the amplitudes of the harmonics. This is easily done and provides a reasonable approximate normalization in the more general case. This method has a slight difference in perceived loudness between the smallest and largest values of the ratio.

It could be argued that peak amplitude is the *only* important consideration, since when working in integers, the overflow problem is quite difficult to deal with. Our feeling is that this is placing the concern more on the *technical* aspects of scaling rather than on the *musical* aspects. By expending more computer time and money, these technical limitations can be dissolved, so that there is no reason to expect that the diminishing costs of technology would not eventually remove the technical limitations, in which case it is important to deal with the musical aspects without extensive regard to technical aspects. Of course, some attention to economy is given in that one of the factors to recommend this synthesis technique is its efficiency, but we will deal mostly with power normalization here because it does not imply a substantial increase in computer time.

To normalize the power in the signal, we sum the squares of the amplitudes of the partials. The reciprocal of the square root of the sum is then the normalization factor. The normalization factors for the sums in Eqs. (1) through (4) are the following:

$$\sqrt{\frac{1-a^2}{1-a^{2N+2}}} \tag{5}$$

$$\sqrt{1-a^2} \tag{6}$$

$$\sqrt{\frac{1-a^2}{1+a^2-2a^{2N+2}}} \tag{7}$$

$$\sqrt{\frac{1-a^2}{1+a^2}} \tag{8}$$

Since these normalization formulas are somewhat difficult to compute in real time, we suggest computing the normalization factors for the range of a of interest, and storing the factors in a table. This way, we only have to look up the appropriate value rather than compute the square root each time. This can also be done when computing a^{N+1} for Eq. (1) and (3). A table of a^{N+1} could be prepared so as to avoid doing this computation at each sample.

Unfortunately, if any of the harmonics fall directly upon any other harmonics' formulas (5) through (8) are not correct, because the overlapping harmonics may either cancel or reinforce, thus altering the total power in the signal. For this special case we must use a different normalization function. For the one-sided functions, overlap occurs only in the harmonic case when f_m is negative, thus reflecting about zero. In two-sided functions, unless the function is band-limited properly, there will be reflections that in the harmonic case will overlap and possibly cancel other harmonics. We can compute the normalization functions in this case by setting $f_c = nf_m$. For generality, we also add a phase term ϕ into all the sines in Eqs. (1) through (4). This will allow us to determine the normalization functions regardless of whether we synthesize the signal using sine functions, cosine functions, or any arbitrary phase angle.

$$\frac{1}{\sqrt{\frac{1-a^{2N+2}}{1-a^2}-a^{2n}\{1+2n\cos 2\phi\}}}, \quad N \geqslant 2n$$

$$\frac{1}{\sqrt{\frac{1-a^{2N+2}}{1-a^2}-a^{2n}\{1+2(N-n)\cos 2\phi\}}}, \quad N < 2n \tag{9}$$

$$\frac{1}{\sqrt{\frac{1}{1-a^2}-a^{2n}\{1+2n\cos 2\phi\}}} \tag{10}$$

$$\frac{1}{\sqrt{\frac{1+a^2-2a^{2N+2}-2\{a^{2n}-a^{2(N-n+1)}\}\cos 2\phi}{1-a^2}-a^{2n}\{1+2(n-1)\cos 2\phi\}}}, \quad N \geqslant 2n$$

$$\frac{1}{\sqrt{\frac{1+a^2-a^{2N+2}-a^{2(N+n+1)}}{1-a^2}-a^{2n}\{1+2N\cos 2\phi\}}}, \quad N < 2n \tag{11}$$

$$\frac{1}{\sqrt{\dfrac{1+a^2-2a^{2n}\cos 2\phi}{1-a^2}-a^{2n}\{1+2\,(n-1)\cos 2\phi\}}}.$$

$$(12)$$

The formulas for the finite sums [formulas (9) and (11)] have two different cases, depending on how far the reflected sidebands extend. The normalization is different, depending on whether the sidebands extend up to or beyond the center frequency.

Fig. 5 shows the case where $\phi=0$. In this case the reflected sidebands are inverted in phase and thus destructively interfere. The lower sidebands are thus quite attenuated. Fig. 6 shows the case where $\phi=\pi/2$. Here the reflected sidebands add and thus raise the amplitudes of the lower sidebands. In both these figures the infinite sum shown in Eq. (4) was used. The center frequency was $f_c=2000$ Hz, and the difference frequency was $f_m=400$ Hz. Again, these figures are the discrete Fourier transforms of actual waveforms generated by Eq. (4).

There comes a bit of a problem if we wish to change smoothly from a harmonic case, where the sidebands overlap and thus require special normalization functions, to the inharmonic case, where there is no sideband overlap. If we let $\phi=\pi/6$, then the normalization functions are similar enough to allow normalization by a single function as we change from harmonic to inharmonic. This is only an approximation to true power normalization, but it seems to be adequate for most purposes.

REFLECTED SIDEBANDS

A sinusoid at a negative frequency is equivalent to a sinusoid at the absolute value of that same frequency with a phase reversal, that is to say, $\sin(-\theta)=-\sin\theta$. To see how this applies to the summation formulas, let us take the left-hand side of Eq. (4) and write it out, adding an arbitrary phase angle ϕ:

$$\sin(\theta+\phi)$$
$$+\,a\sin(\theta+\phi-\beta)+a\sin(\theta+\phi+\beta)$$
$$+\,a^2\sin(\theta+\phi-2\beta)+a^2\sin(\theta+\phi+2\beta)$$
$$+\,a^3\sin(\theta+\phi-3\beta)+a^3\sin(\theta+\phi+3\beta)$$
$$\cdots.$$
$$(17)$$

To be specific, let us take the case where $\theta=3\beta$. This can be written as follows:

$$\sin(3\beta+\phi)$$
$$+\,a\sin(2\beta+\phi)+a\sin(4\beta+\phi)$$
$$+\,a^2\sin(\beta+\phi)+a^2\sin(5\beta+\phi)$$
$$+\,a^3\sin\phi+a^3\sin(6\beta+\phi)$$
$$+\,a^4\sin(-\beta+\phi)+a^4\sin(7\beta+\phi)$$
$$+\,a^5\sin(-2\beta+\phi)+a^5\sin(8\beta+\phi)$$
$$\cdots.$$
$$(18)$$

Now let us generalize. We shall set $\theta=n\beta$, which represents one form of the harmonic case. We shall group the harmonics together, using the trigonometric relation $\sin(-x)=-\sin x$, and starting at zero frequency.

$$a^n\sin\phi$$
$$+\,a^{n-1}\sin(\beta+\phi)-a^{n+1}\sin(\beta-\phi)$$
$$+\,a^{n-2}\sin(2\beta+\phi)-a^{n+2}\sin(2\beta-\phi)$$
$$+\,a^{n-3}\sin(3\beta+\phi)-a^{n+3}\sin(3\beta-\phi)$$
$$\cdots$$
$$+\,a\sin\{(n-1)\beta+\phi\}-a^{2n-1}\sin\{(n-1)\beta-\phi\}$$
$$+\,\sin(n\beta+\phi)-a^{2n}\sin(n\beta-\phi)$$
$$+\,a\sin\{(n+1)\beta-\phi\}-a^{2n+1}\sin\{(n+1)\beta-\phi\}$$
$$+\,a^2\sin\{(n+2)\beta-\phi\}-a^{2n+2}\sin\{(n+2)\beta-\phi\}$$
$$+\,a^3\sin\{(n+3)\beta+\phi\}-a^{2n+3}\sin\{(n+3)\beta-\phi\}$$
$$\cdots.$$
$$(19)$$

Thus we can see that by rotating ϕ, we can get the harmonics to add or cancel to any degree we desire. If $a<1.0$, the amplitudes of the reflected sidebands will fall off and soon become negligible. If we use the finite form as shown in Eq. (3), then the reflected sidebands will not become negligible. Some most interesting spectra can be produced in this manner.

If the relation between θ and β is not rational, then there is no overlap, and the reflected sidebands are interleaved with the unreflected sidebands.

The one-sided form of the formula as shown in Eqs. (1) and (2) can also have reflected sidebands. They occur when $\beta<0$. If we set $\theta=-n\beta$, we can write down the harmonics in order as we did above:
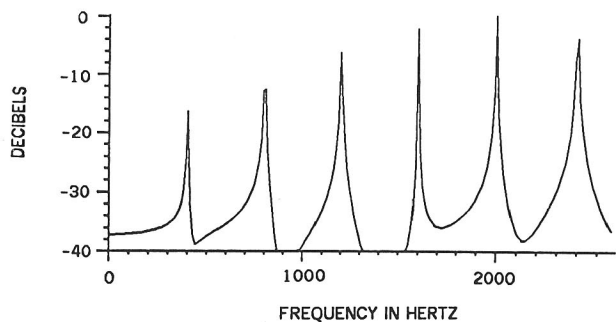

Fig. 5. Spectrum of a waveform produced by Eq. (4) for $F_c=2000$ Hz, $F_m=400$ Hz, $\phi=0$, and $a=0.8$. The reflected lower sidebands have reduced the amplitude of the partial at 400 Hz because of destructive interference.
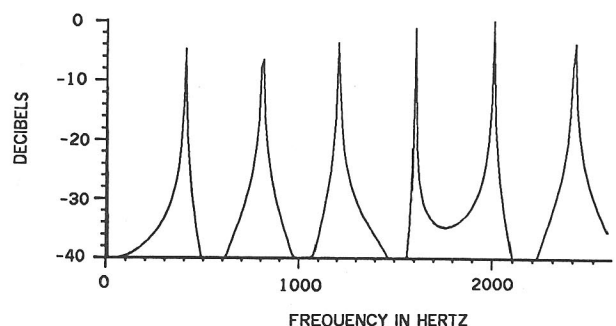

Fig. 6. Spectrum of a waveform produced by Eq. (4) for $F_c=2000$ Hz, $F_m=400$ Hz, $\phi=\pi/2$, and $a=0.8$. This is the same as Fig. 5, except that the phase has been changed. The reflected lower sidebands have increased the amplitude of the partial at 400 Hz because of constructive interference.

$$a^n \sin \phi$$
$$+ a^{n-1} \sin (\beta + \phi) - a^{n+1} \sin (\beta - \phi)$$
$$+ a^{n-2} \sin (2\beta - \phi) - a^{n+2} \sin (2\beta - \phi)$$
$$+ a^{n-3} \sin (3\beta + \phi) - a^{n+3} \sin (3\beta - \phi)$$
$$\cdots$$
$$+ a \sin \{(n-1)\beta + \phi\} - a^{2n-1} \sin \{(n-1)\beta - \phi\}$$
$$+ \sin (n\beta + \phi) - a^{2n} \sin(n\beta - \phi)$$
$$- a^{2n+1} \sin \{(n+1)\beta - \phi\}$$
$$- a^{2n+2} \sin \{(n+2)\beta - \phi\}$$
$$- a^{2n+3} \sin \{(n+3)\beta - \phi\}$$
$$\cdots . \qquad (20)$$

This is similar to the two-sided form, as shown in Eq. (19), except that there are no more overlapping terms above the center frequency term at $n\beta$. For $a < 0$, this produces a very prominent harmonic at $n\beta$. Quite often a prominent harmonic like this will cause the tone to "de-fuse," so that two tones are perceived: a tone at the fundamental frequency and a tone at $n\beta$.

Fig. 7a shows the one-sided form with $f_m = -500$ Hz, $f_c = 2000$ Hz, $a = 0.8$, and $\phi = 0$. By rotating the phase angle ϕ through $\pi/2$, we can get the maximal effect on the overlapping sidebands. This is shown in Fig. 7b.

In general, we prefer the two-sided form, even though it requires a complicated normalization function, because it produces more reflected sidebands. In the inharmonic case this gives a much more dense spectrum. In the harmonic case the phase angle ϕ gives us another degree of freedom in controlling the amplitudes of the lower harmonics. With overlapping reflected sidebands we can produce a tremulant much like that found with frequency modulation synthesis by simply setting $nf_c = mf_m + \delta$. The quantity δ we call

the skew frequency. It causes the reflected sidebands to occur at frequencies slightly different from the sidebands they will overlap. This difference causes the harmonics to beat, with the result that the effective amplitude of these overlapped harmonics will rise and fall periodically, at a rate determined by δ. This will cause a periodic change in loudness, a change in timbre, and a periodic change in the perceived pitch of the tone. The pitch appears to change because the frequency of each overlapped harmonic will vary between the frequencies of the two sinusoids that are beating. These reasons make the two-sided form quite attractive for general-purpose synthesis.

DYNAMIC SPECTRA

To use the summation formulas effectively, we must gain insight into how the spectrum varies as the various parameters are changed. A convenient way to do this is through the use of three-dimensional plots of harmonic amplitudes as functions of some parameter. The parameters we have to deal with are the number of sidebands N, the ratio between the amplitudes of adjacent harmonics a, the center and difference frequencies f_c and f_m, and the phase of the harmonics ϕ.

Fig. 8 shows three such plots for the two-sided formula [Eq. (3)] with $N = 5$, $f_c = f_m = 1$, and the phase ϕ set to 0, $\pi/6$, and $\pi/2$, respectively. These plots show a perspective representation of a three-dimensional function. We are plotting, in this case, amplitude on the vertical axis, frequency on the rear-to-front axis, and ratio on the horizontal axis. We see the effect on the harmonic amplitudes of changing the ratio continuously from 0 to 1. The smaller plots in the upper part of each graph show how the ratio changes with distance along the horizontal axis, what the power normalization function [as computed by Eq. (11)] is as a function of the ratio, and how the power (actually, the square root of the power) of the entire signal changes with distance along the horizontal axis. We can see that with $\phi = 0$, the reflected sidebands cancel out the lower harmonics, whereas with $\phi = \pi/2$, they reinforce. A perfect balance is achieved at $\phi = \pi/6$, where they approach a smooth spectrum as the ratio approaches 1.

Fig. 9 shows similar plots for the one-sided formula [Eq. (1)] with Eq. (9) as the power normalization function. In these figures, $f_c = 3$, $f_m = -1$, and $N = 8$. Again, the phase angle ϕ has a similar effect on the reflected sidebands.

By use of this same program for display of time-variant spectra, we can show the change of the spectrum with time for an actual musical case. In fact, this program is very useful for gaining insight into how the spectrum evolves with time. For instance, Fig. 10 shows the display for a brass-like tone. Here the horizontal axis denotes time, thus showing the evolution of the harmonic amplitudes with time. We have chosen $\phi = \pi/6$ for spectral continuity, and are using the two-sided form for generality. In this case, $f_m = f_c = 1$, and the ratio function is similar to the amplitude function. This causes the fundamental to enter first and the higher harmonics to enter later.

Fig. 11 shows the display for a gong-like tone. Since this tone is inharmonic, the partials are not at integer fre-
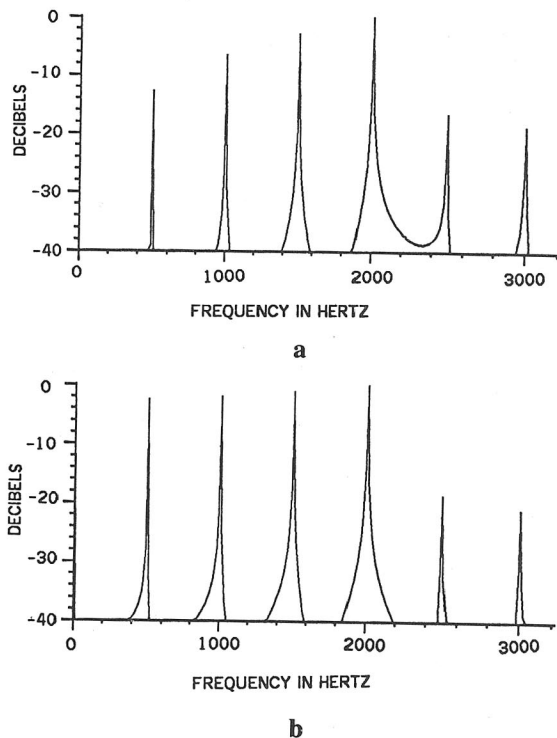

Fig. 7. Spectrum of a waveform produced by Eq. (2) for $F_c = 2000$ Hz, $F_m = 500$ Hz, and $a = 0.8$. **a.** $\phi = 0$, causing the amplitude of the lowest partials to be reduced. **b.** $\phi = \pi/2$, causing the amplitude of the lowest partials to be increased.

quencies. This plot was done with $f_c = 1$, $f_m = 1.414$, and $N = 9$. The function is two-sided, and the phase ϕ is irrelevant because there is no harmonic overlap with the reflected sidebands. Only frequencies up to 5 are shown in this figure. On this tone, the ratio was tapered down with the amplitude function so that the inharmonicity would gradually die away, leaving only the fundamental.

IMPLEMENTATION

The implementation of the sine summation formula as a synthesis technique will be described in terms of the MUSIC V program [6]. Although this is not necessarily the best way to implement the technique, it seems a good way to communicate music synthesis algorithms, much as FORTRAN and ALGOL are used as languages for communicating useful programming algorithms. To do this in MUSIC V, however, requires two important additions to the language. We need a divider and provision for a table lookup. Let us define these in roughly the way as in Mathews [6].
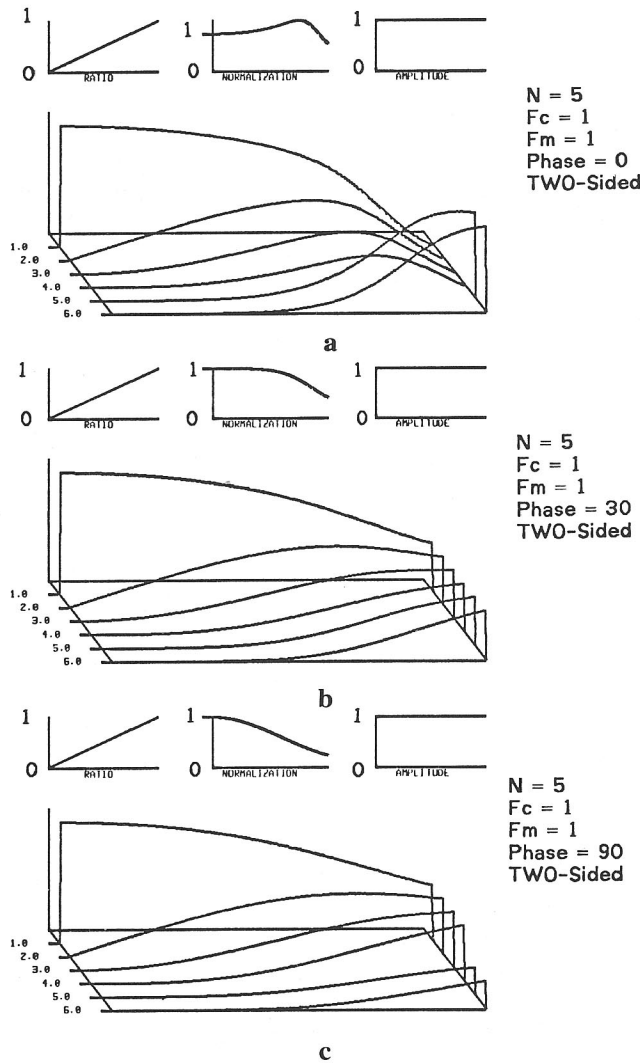
DIV Generator

The DIV generator divides two numbers. It is diagrammed in Fig. 12. The equation of operation is

$$O_i = I1_i/I2_i$$

where I1 and I2 are the two inputs and O is the output. In the score,

DIV B2 B3 B4

associates I1 with B2, I2 with B3, and O with B4. Division by zero is, of course, a problem that we will not attempt to deal with here. Let us just assume that the divisor is always nonzero for the purposes of illustration.

TAB Generator

The TAB generator does a simple table lookup. It is diagrammed in Fig. 13. To operate, it needs a table which
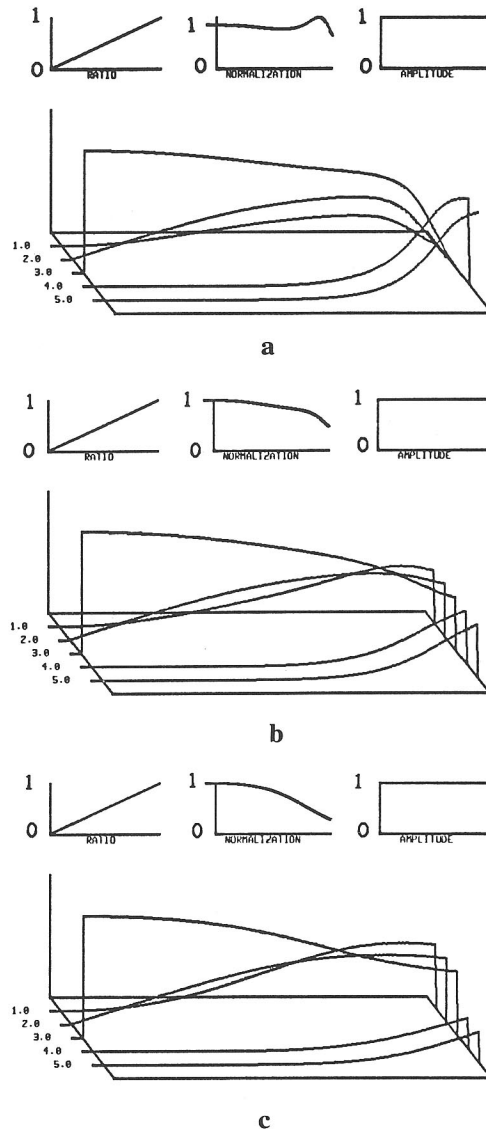

N = 5
Fc = 1
Fm = 1
Phase = 0
TWO-Sided

N = 5
Fc = 1
Fm = 1
Phase = 30
TWO-Sided

N = 5
Fc = 1
Fm = 1
Phase = 90
TWO-Sided

a

b

c

Fig. 8. Perspective plots of the partials of a harmonic tone produced by Eq. (3) with $F_c = 1$ Hz, $F_m = 1$ Hz, and $N = 5$. **a.** $\phi = 0$. **b.** $\phi = \pi/6$. **c.** $\phi = \pi/2$. The ratio is swept from 0 to 1.0 as the graph proceeds from left to right. The frequency axis comes out of the page with the lowest frequency at the rear and the largest frequency displayed in the foreground.

a

b

c

Fig. 9. Perspective plots of the partials of a harmonic tone produced by Eq. (1) with $F_c = 3$ Hz, $F_m = -1$ Hz, and $N = 8$. **a.** $\phi = 0$. **b.** $\phi = \pi/6$. **c.** $\phi = \pi/2$. The ratio is swept from 0 to 1.0 as the graph proceeds from left to right. This plot is analogous to that of Fig. 8.

stores some function (the *range*), and a scale factor that expresses what values of the input (the *domain*) get mapped into what table locations. The equation of operation is

$$O_i = F_i([I1_i * SCALE]_{Mod\ 511})$$

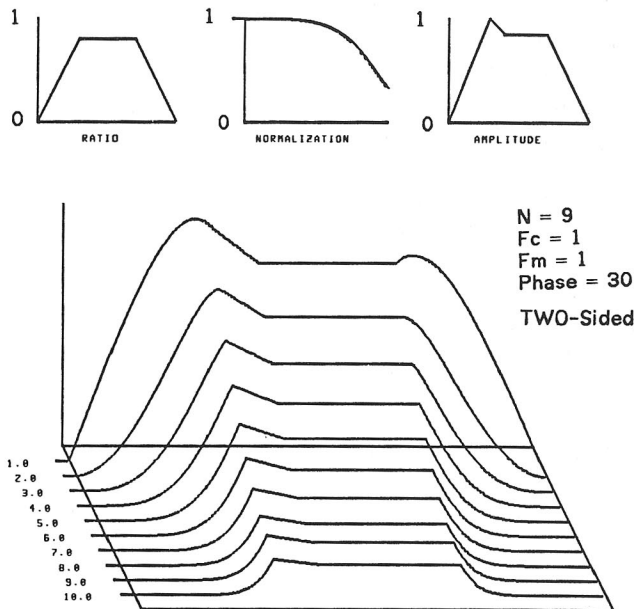where i is the index of acoustic output samples and

Fig. 10. Perspective plot of the dynamic spectrum of a brass-like tone. Time proceeds from left to right. In this case the small plot in the upper right shows the overall amplitude of the signal with time. The small plot in the upper left shows the ratio as a function of time. The amplitude and ratio curves are designed to allow the fundamental to enter first and the higher harmonics to enter subsequently. This tone would take place in about 0.5 second. The horizontal axis in the lower plot thus represents 0.5 second of time. The upper plots of the ratio and the overall amplitude also represent 0.5 second.
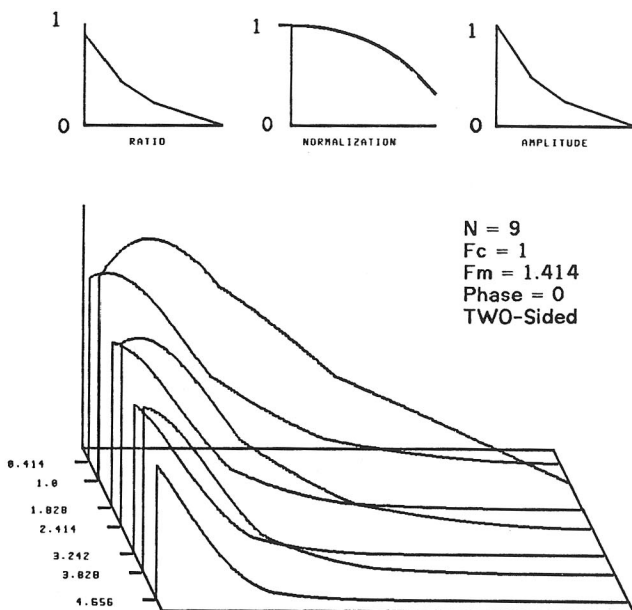

Fig. 11. Perspective plot of the dynamic spectrum of a bell-like sound. In this case the signal is inharmonic. Both the amplitude envelope and the ratio have exponential curves. The horizontal axis in this case represents several seconds.

$[I1_i*SCALE]_{Mod\ 511}$ is $[I1_i*SCALE\ -j*511]$, j being selected so that the result always falls between 0 and 510 (inclusive). F_i is the function stored in this table; SCALE is the scale factor that converts the input variable into an index into the table; $I1_i$ is the location controlling input; and O is the output.

Our table is exactly 512 locations long. We might store the function represented by formula (9) in this table for 512 different values of a from $a = 0$ to $a = 1.0$ at equally spaced points. In this case we would set SCALE to 512 so that the value $a = 0$ accesses the first element of the table, and values approaching $a = 1$ approach the end of the table.

In fact, the Mod 511 in the equation of operation is not really meaningful here as it is in the OSC generator, because the function does not necessarily mean anything if the input variable is out of range. It was added here just as an error condition, so that at least values of a out of range would be well defined, although not necessarily meaningful.

We use the TAB generator to store functions of a, such as power normalization functions or values of a^{N+1} which are too costly to compute directly in MUSIC V.

Figs. 14–17 show MUSIC V-like instruments for realizing the functions shown in Eqs. (1), (2), (3), and (4), respectively. Of course, Fig. 17 [Eq. (4)] is the most compact. We have not expanded the diagrams to show the calculations of, for instance, quantities like $(N+1)F_m$, but we just placed these expressions in the figures directly. If one were to actually implement these instruments in MUSIC V, one would have to add further unit generators to compute these quantities, and assignment of parameter numbers to controlling signals would have to be done. These diagrams just show schematically how the flow of data could proceed. The variable Amp is used to denote the overall amplitude of the signal.

For Figs. 14–17 to make sense, we must describe what is in each of the function tables (identified as F1 through F6). To correspond to Eqs. (1) through (4), F1 must be a sine, F2 must be a cosine, and F3 is the amplitude nor-
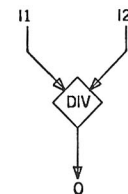

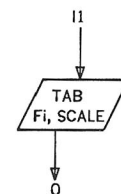Fig. 12. Diagram of the DIV unit generator. The output is just I1/I2.

Fig. 13. Diagram of the TAB unit generator. This multiplies I1 by SCALE and uses the result as an index into the table representing the function F_i.

malization function which, as discussed before, must be somewhat different for different cases. Function F4 is $(1+a^2)$, function F5 is $-a^{N+1}$, and function F6 is $(1-a^2)$. In Fig. 17 the amplitude normalization table F3 should also contain the term $(1-a^2)$ so as to reduce the computation further.

As with frequency modulation, the oscillators must be capable of synthesizing negative frequencies. Terms like $F_c - F_m$ can easily give negative numbers for the frequencies.

We have not shown in these figures two more unit generators which would be desired for dynamic control of the signal. These missing generators should contain the overall amplitude envelope and the ratio a as a function of time.

USAGE

In the following we shall discuss the implementation of a small but useful set of instrument tones, giving specific parameter values as guides (not inflexible rules) to users. We shall use the simplest form of the technique, which is shown in Fig. 17 and in Eq. (4).

Although we shall use the names of orchestral instruments in describing these sounds, we must emphasize that these lables should not be taken too seriously. None of these instruments are indistinguishable from their orchestral counterparts. They are given as guides, capturing some of

the flavor of the orchestral instrument, but without duplicating the rich level of detail found in the waveforms of natural music instruments.

Brass-Like Tones

To produce brass-like tones, we rely on the well-known fact that as the tone begins, only the fundamental sounds at first, with the higher harmonics entering later. Unmuted brass instruments show no signs of formants and have harmonically related partials. Thus we can produce a brass-like tone by setting $F_c = F_m =$ the frequency of the note desired. Fig. 10 shows many things concerning the brass-like tone. The small plot in the upper right of the figure shows the overall amplitude envelope. This is used to control the variable labeled Amp in Fig. 17. As we see it rises to a peak, maintains a steady state, then descends. The entire note length in this case is about 0.5 second. The ratio function rises to a value of 0.78 and maintains it throughout the steady-state region. It is shown in the small plot in the upper left corner of Fig. 10. The ratio function could be arranged to look like the amplitude function with a peak at the beginning, but we must be very careful not to allow the ratio to get too great. The bandwidth increases so rapidly as a approaches 1.0 that great care must be taken with larger values of the ratio. Driving the ratio to 1.0 in the non-band-limited cases [Eqs. (2) and (4)] is analogous to driving the modulation index to infinity in frequency modulation.

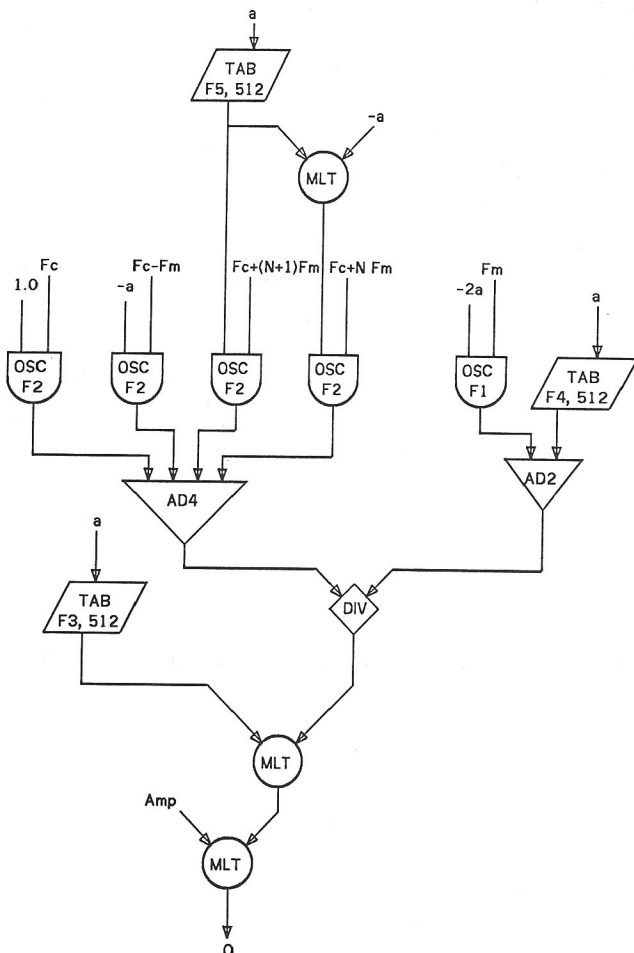In the foreground of Fig. 10, as explained before, we see

Fig. 14. MUSIC V instrument which realizes Eq. (1). F1 is a cosine, F2 is a sine. F3 is the amplitude normalization function. F4 contains the values of $(1+a^2)$, and F5 is $-a^{N+1}$.
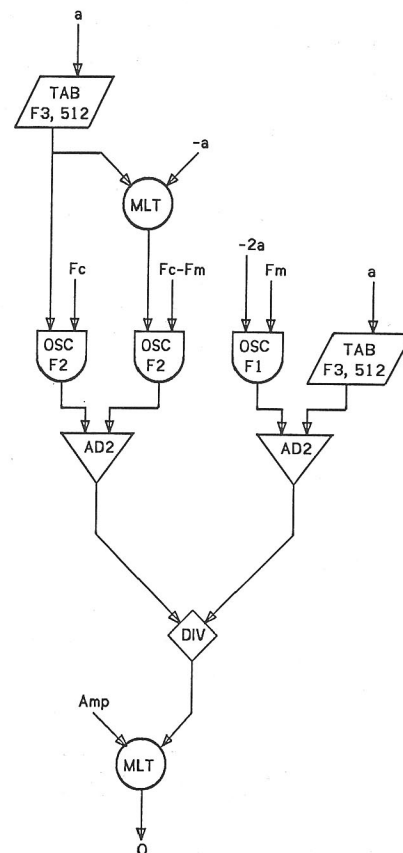

Fig. 15. MUSIC V instrument which realizes Eq. (2). The functions are the same as in Fig. 14.

a diagram of exactly how each of the first few harmonics changes with time. We can see that in fact, this combination of controlling functions causes the fundamental to come in first, with the higher harmonics gradually entering.

If the extra computer time is available, we recommend the use of Eq. (1) with $N = 8$ for this instrument, rather than the simple form in Eq. (4). This allows the highest harmonic to be specified explicitly, so that some of the "buzzy" quality inflicted upon the listener by the highest harmonics may be eliminated. This produces a much smoother sound.

Woodwind-Like Tones

There are three classes of woodwind tones we would like to discuss here. The most dramatic class is that of the double-reed instruments: bassoon, English horn, and oboe. These instruments have a characteristic resonance that seems to be a very strong cue. We can do this with the sine summation formula by setting F_c to the frequency corresponding to one of the harmonics higher than the fundamental and F_m to the frequency corresponding to the pitch desired. For instance, for a bassoon-like tone we can set $F_c = 5F_m$. For an English-horn-like tone we can use $F_c = 3F_m$. This is shown in Fig. 18. As with the brass-like tones, we use a ratio function that brings in the center frequency first, then brings in the rest of the harmonics. In this case, the center frequency is not set to the fundamental. Fig. 18 shows a perspective plot for this English-horn-
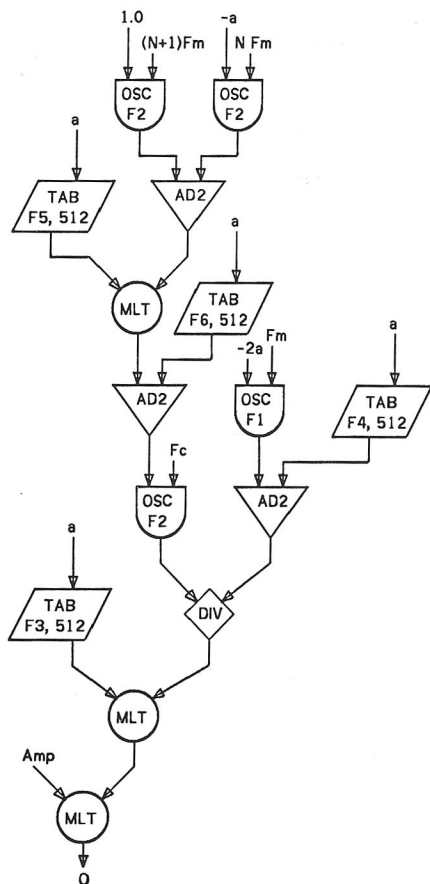
like tone. The existence of a prominent harmonic simulates a resonance and thus delivers a cue that is representative of the double-reed instruments. For a bassoon-like tone the maximum ratio should be reduced to about 0.65.

For saxophone-like tones we must begin the tone with a large number of partials to simulate the noise burst at the beginning of the tone. Fig. 19 shows the functions for such a tone. The ratio begins high (0.92 in this case), then descends to a steady-state value (0.8 is the value shown). For clarinet-like tones the only change is to set $F_c = 2F_m$ so that only odd harmonics are produced.

For these tones, as well as many others, it is often useful to combine two instruments together; one inharmonic instrument to supply the "chiff" at the beginning of the tone, and the other instrument to supply the harmonic portion of the tone. Attempting to generate the "chiff" by using a large initial ratio as was done above is only moderately successful.

Percussive Sounds

As with frequency modulation synthesis, the inharmonic tones seem to be the most fertile ground for realization. Generally, these tones begin with a large number of partials and decay quickly to a pure tone, sometimes called the "hum" tone. A simple way to make inharmonic tones is to set $F_m = 1.414F_c$. This inhibits all overlap, so that there is no common divisor of the two frequencies, except at pitches so low that they cannot be heard.

A bell or gong-like sound can be made by making the overall amplitude and the ratio both exponentials. The exponential decay of the ratio causes the partials to die out faster than the center frequency, so that the tone ends with a single sinusoid. Fig. 11 shows an example of such a tone. Notice that the partials do not fall conveniently upon a harmonic series.

Fig. 16. MUSIC V instrument which realizes Eq. (3). Function F6 contains the values of $(1 - a^2)$.
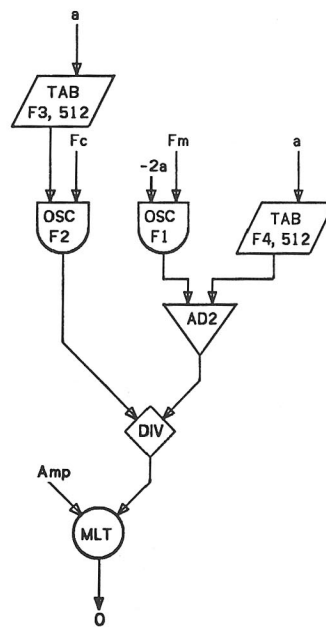

Fig. 17. MUSIC V instrument which realizes Eq. (4). This is the simplest form of the sine summation instrument. Function F3 must be modified in this instrument to include the factor $(1 - a^2)$.

By changing the amplitude function to one that lasts longer before decaying, and by reducing the initial (maximum) ratio to about 0.6, we can produce a drum-like sound that is shown in Fig. 20. A wood-drum-like sound can be made by accelerating the decay of the ratio and by using an extremely high initial ratio. In Fig. 21, we see such a case with an initial ratio of 0.98.

This completes a minimal catalog of sounds possible with the sine summation formulas. These are the same sounds that were described in Chowning's frequency modulation article, and demonstrate the similarity of the techniques.

NUMERICAL CONSIDERATIONS

Numerical problems arise in the use of the sine summation formulas when the ratio approaches 1. In the normalization functions the term $(1 - a^2)$ appears in the denominator of several functions. In some cases the normalization function really does go to infinity, and in some cases it has a finite limit. In the later case one must substitute the limit in the table without attempting to divide by zero.

In the evaluation of any of Eqs. (1) through (4), the denominator term goes to zero at $\beta = 0$ as $a \to 1.0$. For values near unity, a modulation can occur due to poor
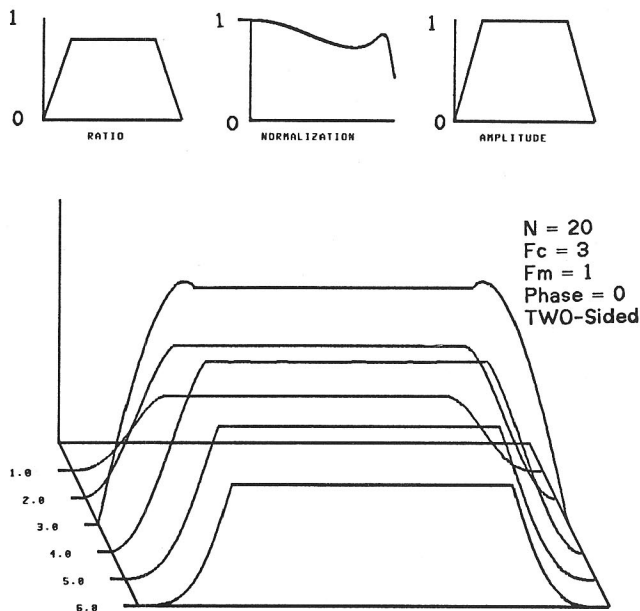
Fig. 18. Perspective plot of double-reed-like tone. In this case $F_c = 3F_m$. This simulates a formant frequency. The horizontal axis represents about 0.5 second.
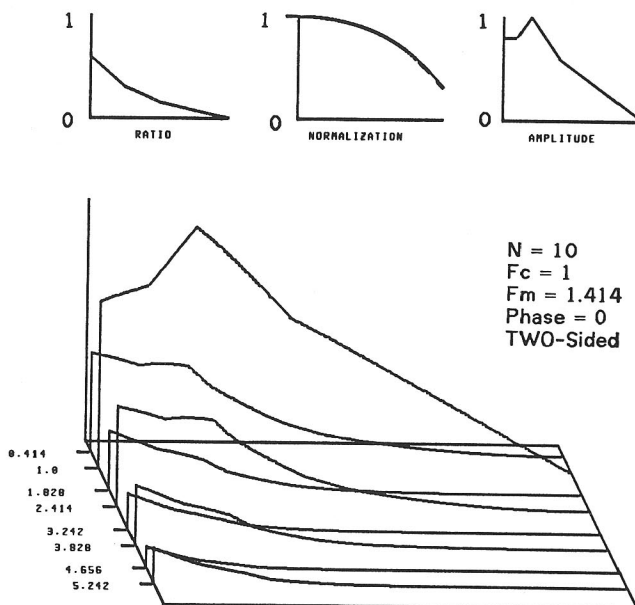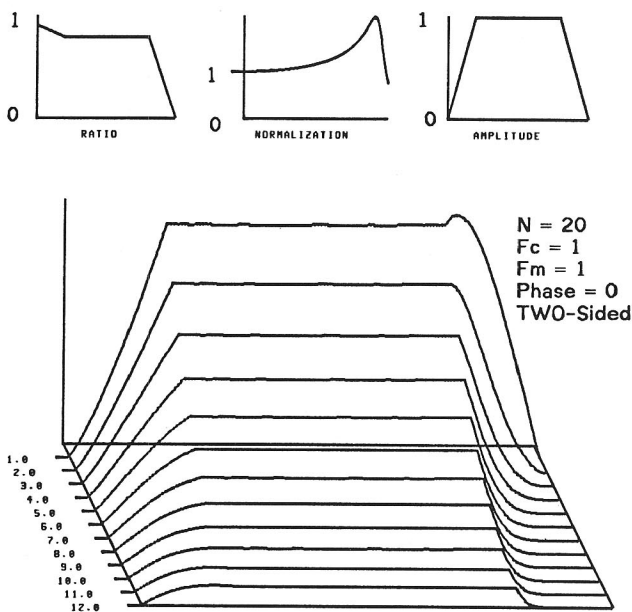
Fig. 19. Perspective plot of saxophone-like tone. The ratio starts high in an attempt to simulate the "chiff" at the beginning of the tone. The horizontal axis represents about 0.5 second.

Fig. 20. Perspective plot of drum-like tone. This tone is inharmonic with $F_m = 1.414 F_c$. The horizontal axis in this case could be about 0.2 second.
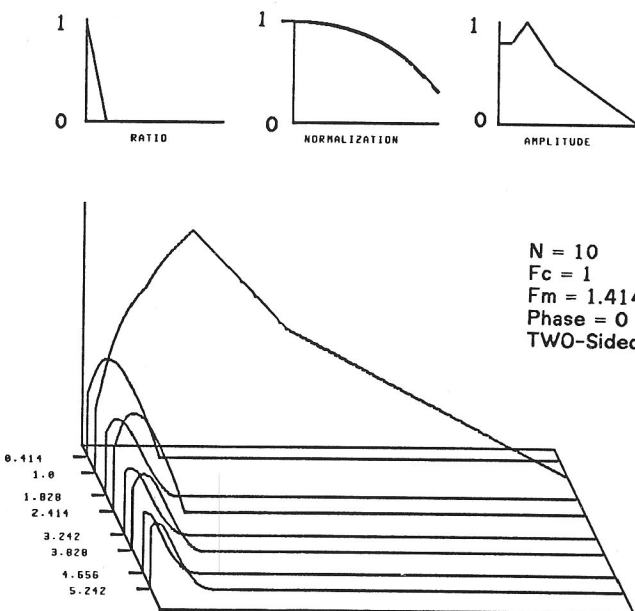
Fig. 21. Perspective plot of wood-drum-like tone. This is similar to the drum tone shown in Fig. 20, except that the ratio decays to zero very quickly. The horizontal axis represents about 0.2 second.

resolution in evaluating the cosine. Our solution to this has been to use a large table for storing the waveforms. A table of 2000 words or more is adequate for even the lowest frequency of use. A method pioneered by Barry Vercoe at M.I.T. uses two 256-word tables. The first is the standard wave table, and the second represents only 1/32 of a period. When the angle is within a 32nd of a period of zero, the second table is used. This gives us 32 times the angular resolution around zero, which is exactly where the trouble arises. The advantage of this method over the use of a large table is that only 512 words of storage are used, rather than the 2000 or more required above. Another technique would be to use a small table (256 words is adequate) and use linear interpolation to give the effect of higher angular resolution.

Even for small values of the ratio, low-frequency tones will have "fuzz" on them due to insufficient angular resolution, unless one of the above methods of increasing the effective table length is used.

SUMMARY AND CONCLUSIONS

The discrete sine summation formulas provide another method for the synthesis of complex time-variant tones that is very similar to Chowning's frequency modulation synthesis. The advantage of these formulas is that the signal can be explicitly band-limited to a specific number of components, and that one-sided spectra can be synthesized. The disadvantage is that even the simplest form is more complex than frequency modulation, and the use of clumsy power normalization functions is required.

The important feature of both sine summation formulas and frequency modulation synthesis is that a large number of perceptual cues for different musical sounds can be realized by dynamic control of simple variables, namely, the amplitude envelope and the spectral envelope of the signal.

There are surely a large number of discrete summation formulas which are useful to composers that have not been discovered yet. It is hoped that this paper will help bring them out of the summation tables and into the studio.

ACKNOWLEDGMENT

REFERENCES

[1] J. M. Chowning, "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation," *J. Audio Eng. Soc.*, vol. 21, pp. 526-534 (Sept. 1973).

[2] J. Grey, "An Exploration of Musical Timbre," PhD Dissertation, Dept. of Psychology, Stanford University (Feb. 1975); distributed through the Dept. Music, Tech. Rep. STAN-M-2.

[3] D. Lewin, personal communication (Apr. 1975).

[4] L. B. W. Jolley, *Summation of Series* (Dover, New York, 1961).

[5] E. Zwicker and B. Scharf, "A Model of Loudness Summation," *Psychol. Rev.*, vol. 1, pp. 3-26 (1965).

[6] M. V. Mathews, *The Technology of Computer Music* (M.I.T. Press, Boston, 1969).

THE AUTHOR

James Anderson Moorer was born November 25, 1945 in Hollywood, Florida. He received his SB degree in electrical engineering from the Massachusetts Institute of Technology in 1967 and an SB degree in applied mathematics in 1968. Dr. Moorer later graduated from Stanford University with a PhD in computer science in 1975.

He worked for four years as systems programmer at the Stanford Artificial Intelligence Laboratory. He is a Research Associate with the newly-formed Stanford Center for Computer Research in Music and Acoustics.