Ring Modulation and Structure

WHEN ELECTRONIC MUSIC STUDIOS were first established in the early 1950s various electronic devices were inherited almost without modification from the telecommunications and recording industries. Electrical wave generators are used both to test audio equipment (amplifiers, telephones, etc.) and themselves to carry information (morse signals, and indirectly radar and sonar). Wartime developments had improved substantially the quality and quantity of such equipment, valve-based at the time. Thus sine, square, ramp and impulse generators were available. From radio-electronics, too, came the *ring modulator*. The development of radio frequency electronics during the 1930s and 40s had thrown up various circuits for the *modulation* of one wave with another. Amplitude modulation was, for example, until recently the predominant form of broadcasting method: a *carrier frequency* was changed in amplitude in a form which encoded the broadcast *signal*. The development of frequency modulation, in which the same signal could be encoded by changing the *frequency* of the carrier, led to much greater fidelity and is now common. The ring modulator is a device which demonstrates one form of frequency modulation. There is in musical terms no need to refer to 'carrier frequency' and 'signal', although older textbooks still use these words. We do nonetheless require two electrical input signals to this device which modulate each other in frequency. The arrangement of a ring of diodes gives the circuit its name.

Textbooks, articles and concert programmes tend to simplify an explanation of the way the ring modulator works. The best known examples are the 'Dalek voice' and, perhaps, the bell-like timbres produced by the ring modulated piano in Stockhausen's *Mantra*. The simplest mathematical explanation is to say that the frequencies of the two inputs *add* and *subtract*. If we had two *pure sine waves* of, say, 100 Hz and 500 Hz we would receive on output the addition, 600 Hz, and the difference, 400 Hz, *simultaneously*. (Some versions separate the two resultant products, but those commonly available, used in the works to be discussed, do not. One should not detect either of the original inputs on the output: better versions of the ring modulator have less 'breakthrough'. Both inputs need to be present to produce any output).

We shall see, however, that as soon as either input becomes more complex, the process of addition and subtraction applies to each and every component frequency of the signals. Even the slightly more complicated case of modulation of a ramp wave, in which a note of 100 Hz has overtones of 200 Hz, 300 Hz, 400 Hz, etc. present, an altogether more extensive set of products will be produced. This simplified approach, inherited from telecommunications use, ignores too the very much more complex case of natural sounds picked up through microphones. These rarely behave according to any rulebook. For example, a piano note — an apparently ordered sound in comparison with many other microphone-captured sounds — has numerous overtones, each of which begins and ends and changes in very unpredictable ways and each of which may be modulated.

Ring modulation is perhaps an overused resource in electronic music in general and live electronic music in particular. One reason for this is its use as an arbitrary colouring device unrelated to other structural aspects of the work. Two domains will be examined in an attempt to overcome this. One concerns the knowledge of the full range of possibilities of the generation of pitch modes as electronic products, the second, a fuller knowledge of the colouristic, i.e. timbral aspect: the reasons for the 'bell-like' qualities too readily used.

I THE PRODUCTION OF MODES

The input frequencies, which in this section we shall assume to be harmonic sounds with a strong fundamental (true of most instruments and electronic generators), we shall call A and B. Our simplified theory tells us that the result will be two waves of frequency $A+B$, $A-B$.

Now let the *input interval* be called I. This may be expressed as the ratio of the input signals:

$$A/B=I, \text{ therefore } A=B.I$$

The *output interval* we may call R, and express it as the ratio of the output frequencies, thus:

$$R=\frac{A+B}{A-B}$$

Combining the two equations:

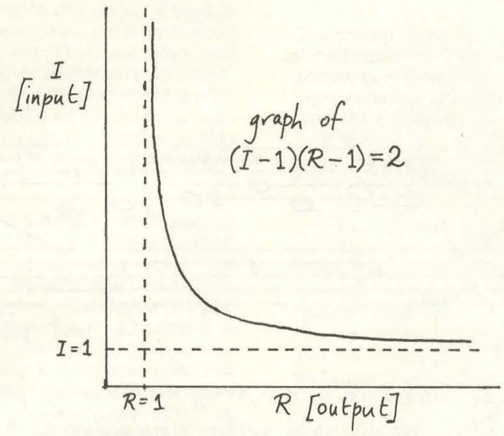
$$R=\frac{B.I+B}{B.I-B}=\frac{I+1}{I-1}=1+\frac{2}{I-1}$$

Therefore $R\text{-}1 = \frac{2}{I\text{-}1}$

or $(I\text{-}1) \cdot (R\text{-}1) = 2$ ...(1)

Thus if we plot $(I\text{-}1)$ against $(R\text{-}1)$ we obtain a hyperbola, or more simply, if we plot I against R we have a hyperbola to the R=1 and I=1 lines. The reason for this is that a *frequency difference* of zero is a unison of *interval ratio* unity. Although the hyperbola has a negative part, we may ignore this as we do not distinguish between positive and negative intervals. This produces Fig. 1 which may be used to compute any input and output interval in any mode or scale system.

Figure 1



The input interval need not be continuous. In the works I shall examine, both inputs are constrained to the familiar equal-tempered system of twelve divisions to the octave. This effectively breaks down the curve of Fig. 1 to a series of points: this is best expressed as a table (see Table 1). This is calculated as follows: if the input interval I and the output interval I are equal to x and y semitones respectively, and we

Table 1 (intervals reckoned in semitones)

I	R	I	R	I	R	I	R
1	61.5	19	12.15	37	4.12	55	1.45
2	49.28	20	11.33	38	3.87	56	1.37
3	42.7	21	10.68	39	3.66	57	1.28
4	37.5	22	10.12	40	3.45	58	1.21
5	33.7	23	9.45	41	3.32	59	1.14
6	30.6	24	8.84	42	3.13	60	1.09
7	28.1	25	8.42	43	2.95	61	1.025
8	25.63	26	7.88	44	2.76	62	0.96
9	23.9	27	7.4	45	2.59	63	0.92
10	21.95	28	7.01	46	2.47	64	0.86
11	20.6	29	6.64	47	2.30	65	0.81
12	19	30	6.20	48	2.16	66	0.76
13	17.74	31	5.87	49	2.04	67	0.75
14	16.6	32	5.54	50	1.93	68	0.68
15	15.6	33	5.20	51	1.83	69	0.64
16	14.7	34	4.94	52	1.72	70	0.61
17	13.7	35	4.62	53	1.63	71	0.59
18	12.76	36	4.36	54	1.53	72	0.54

use the symbol a to represent $\sqrt[12]{2}$ (the frequency ratio of a semitone), then:

$I = a^x$ and $R = a^y$

therefore, from (1) $(a^x - 1)(a^y - 1) = 2$

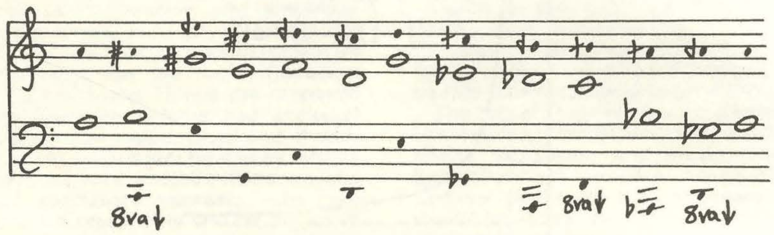
Redistributing and taking logarithms:

$$y = \frac{\log(a^x + 1) - \log(a^x - 1)}{\log(a)} \quad \ldots (2)$$

This leads to Table 1.

I shall examine three examples of works where a well-tempered instrument (in all three a piano) is modulated against a sine wave also constrained to the well-tempered system; the works are Stockhausen's *Mantra*, Roger Smalley's *Monody* and my own *Piano Piece III*. We can now see more clearly the various types of mode generation at work in such pieces. In *Mantra*, the 13 notes of the mantra (see Ex. 1) are used

Example 1



O – Mantra , • – ring modulation product
Ring modulated with sine wave of 220 Hz.
ɖ – quarter flat, + – quarter sharp.

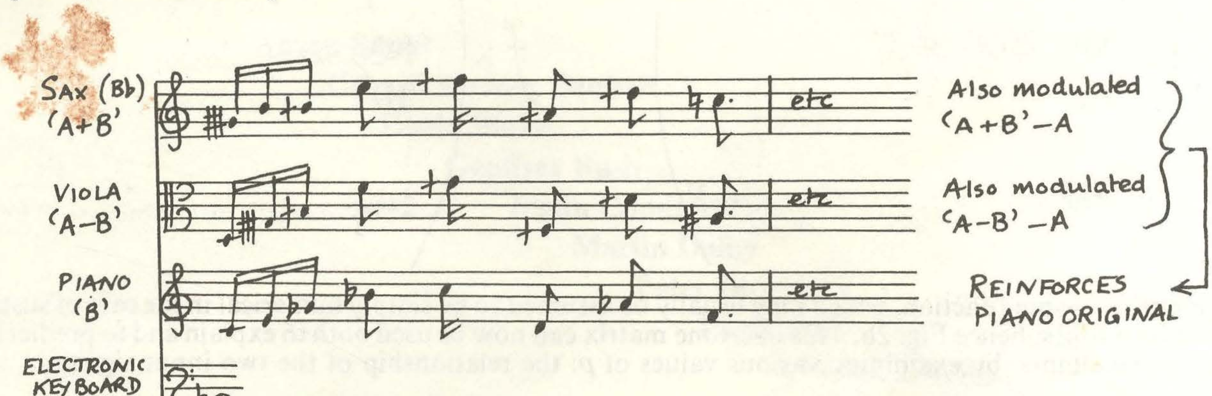
not only of course to generate, through expansion and contraction, the pitch material of the whole work,[1] but also *untransposed*, to generate the fundamental pitches of the sine waves used for modulation in each of the 13 sections of the work (see the introduction to the score). Admittedly, wider-ranging glissandi are used, but the confinement of the electronic wave to within an octave of frequency for most of the work has the double-edged effect of reducing interest in the area of timbral variety, but also thereby increasing the possibilities of appreciating the resulting consonance and dissonance structuring, for the composer claims: 'The intervals of the mantra itself are composed such that they move away from the central note, produce increasingly more deviations, micro-intervals . . . and then return.'[2] We may easily read off the resulting tones when the mantra is ring modulated with its first pitch, A=220 Hz (see Ex. 1). This does not entirely fulfil the composer's intention, and with the further confusion of the mantric expansions, at only certain points does this idea of 'cadential' function[3] become apparent. Indeed it probably demands the simplification of line effected in Roger Smalley's *Monody* to make such assertions about the structural functioning of ring modulation apparent. In this work three (or four if available) octaves of sine tone are used, and the various combinations of high tone/low piano, etc. are used in alternating sections. As the piano line is entirely monodic (no pedal and no chords) the modes are clearly created, though the rhythmic energy of the piece may accidentally detract from this. Although on paper Ex. 2 shows a great similarity to the example from *Mantra*, it has an entirely different function in this work. The opening statement is built from a mode: the whole-tone scale on C, plus G (and a passing B natural). The phrase is repeated eight times, being modulated with each note of the mode in turn. The resultant tones will be simple permutations and transpositions of one another: the principle of consistency on which the whole piece is based. On the second page of the work an extended version of the phrase is modulated with the other five chromatic notes

Example 2

not present in the first mode. This results in more complex relationships — yet again the absolute positioning of the sine tones is important: no notes may be arbitrarily shifted an octave either way, and the fifth relationship to the original mode guarantees a considerable degree of consistency.

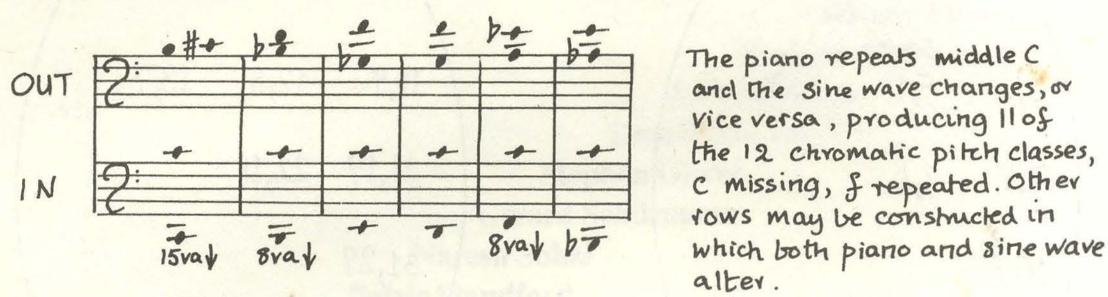
Smalley has extended this modality in the first section of his *Zeitebenen* which adds modulated viola and soprano saxophone (plus unmodulated percussion) to the music of *Monody*. At several points these are used to reinforce the resultant tones produced by the piano modulation, i.e. they actually play the pitches produced by the ring modulation of the piano. Now as they themselves are modulated with an identical pitch, a resultant of *their* modulation is the piano original: a double reinforcement! (See Ex. 3.)

Example 3 (trills excluded)



My own *Piano Piece III* extends this idea of modality by a further constraint: only those intervals between piano note and sine wave that produce a well-tempered interval are used, accepting an error of up to one quarter of a semitone. From this, twelve-note modes are constructed based on the two-note chords produced (see Ex. 4). Suffice it to say that the construction of modes is only the first step toward the real

Example 4



work of composition! Development of these ideas into more microtonal domains is very fruitful research and may be used retrospectively to explain certain harmonic results in works not consciously composed with these views in mind (e.g. sections of *Mantra*).

II FROM COLOURING TO KLANGFARBENMELODIE

This heading indicates what is perhaps an ideal; it is, however, quite possible to move away from the oversimple use of ring modulation as a colouring effect. The simple '*A+B, A–B*' resultant tone calculations become inadequate, reinforcing the tendencies of classical music theory to reduce 'pitch' (a subjective phenomenon) to 'frequency' (objectively measurable), and to reduce a spectrum of frequencies to its fundamental. I do not wish, however, to complicate the matter to the extent of presenting the full general solution to the process: ring modulation is the multiplication of two wave functions. I want to suggest a compromise both simple enough to be grasped for its musical possibilities and more complex than the oversimplification noted above. Let us consider the simplified Fourier transform of two periodic waves:

$$f(n) = \sum a_n \sin(2\pi nft)$$

$$F(m) = \sum a_m \sin(2\pi mFt)$$

where t is time, a_n is the amplitude of the nth partial of frequency nf, and a_m is the amplitude of the mth partial of frequency mF.

The ring modulated product is $f(n).\ F(m)$

$$= \tfrac{1}{2} a_n . a_m \cos(2\pi(nf \pm mF)t) \qquad \ldots(3)$$

This result bears a resemblance to the simplified result often quoted, but it includes the interactions of the other partials in the two input signals. It may be rewritten as an $n \times m$ matrix given each of the possible combinations of overtone interaction (Fig. 2a).[4] This may be simplified further by letting $F = p.f$, where p is

Figure 2a

$$\begin{pmatrix} f \pm F & 2f \pm F & 3f \pm F & \cdots\cdots & nf \pm F \\ & & & \text{etc.} & \vdots \\ f \pm 2F & 2f \pm 2F & 3f \pm 2F & & \vdots \\ f \pm 3F & 2f \pm 3F & 3f \pm 3F & & \vdots \\ \text{etc.} & & & & \vdots \\ \vdots & & & & \vdots \\ f \pm mF & \cdots\cdots & \cdots\cdots & \cdots\cdots & nf \pm mF \end{pmatrix}$$

2b

$$f \times \begin{pmatrix} 1 \pm p & 2 \pm p & 3 \pm p & \cdots\cdots & n \pm p \\ & & & \text{etc.} & \vdots \\ 1 \pm 2p & 2 \pm 2p & 3 \pm 2p & & \vdots \\ 1 \pm 3p & 2 \pm 3p & 3 \pm 3p & & \vdots \\ \text{etc.} & & & & \vdots \\ \vdots & & & & \vdots \\ 1 \pm mp & \cdots\cdots & \cdots\cdots & \cdots\cdots & n \pm mp \end{pmatrix}$$

some relationship function, which may usually be assumed to be simply numerical in the case of sustained harmonic tones, hence Fig. 2b. This overtone matrix can now be used both to explain and to predict many timbre resultants, by examining various values of p: the relationship of the two input signals.

(1) *When p is an integer,* i.e. any whole number, then one wave is on the other's harmonic series and results in a related harmonic web. The fundamental may be obscured by the formation of very strong formant regions, the existence of which may be predicted from the amplitude term of equation (3): $\frac{1}{2}a_n . a_m$. If p is relatively low the harmonic relationship will be clear: this case has been covered in the first part of this article, albeit simply in terms of the fundamental. For larger values of p, the overtone perception is relatively less clear, as we are dealing with very high, closely-spaced frequencies which our ears are less able to distinguish, e.g. Figs. 3a and 3b show the matrices for $p=3$ and $p=10$ (intervals between

Figure 3a

$$\begin{pmatrix} 4,2 & 5,1 & 6,0 \cdots\cdots \\ & & \text{etc.} \\ 7,5 & 8,4 & \vdots \\ 10,8 & & \vdots \\ \text{etc.} \cdots & \cdots\cdots & \cdots \end{pmatrix}$$

3b

$$\begin{pmatrix} 11,9 & 12,8 & 13,7 \cdots\cdots \\ & & \text{etc.} \\ 21,19 & 22,18 & \vdots \\ 31,29 & & \vdots \\ \text{etc.} \cdots & \cdots\cdots & \cdots \end{pmatrix}$$

the two inputs of an octave and a fifth, and two octaves and a major third respectively). N.B. The sign (+ or –) may be ignored in the resulting number. The matrices indicate only which overtones will be present, a number repeated twice may not indicate any particular strength as phase considerations may complicate matters, but in general strong overtones produce strong modulation products and tend to appear in the top left area of the matrix. The $n \times m$ matrix may then be constructed up to values of n and m that are known to be 'reasonably' strong. The above examples indicate that when $p=3$ the lower eight overtones are present, when $p=10$, higher (11-20s) overtones. As indicated, the former will be nearer a fused timbre, the latter a nonharmonic chord (with timbre).

(2) *When p is a non-integer, rational number,* i.e. expressible as a ratio. This may not be so different from (1) as each input may be considered to be on the overtone series of a third — much lower — note, i.e. if $F = p.f$, $p = 15/7$, say. $F = 15f/7$, or $F/15 = f/7$, i.e. f is the 7th, F the 15th harmonic of the 'ghost fundamental', which may of course be present in the product (see Fig. 4). But because the new fundamental has been inevitably shifted down in the frequency range we tend to be dealing again with high overtone components. The more complex the ratio the more true this is (as above); for ratios such as ½ or ¾, i.e. simple intervals, the less true.

(3) *When p is irrational,* i.e. 'π' etc., such situations may arise in two distinct circumstances: firstly as mistunings of more harmonic cases, but more often in transitional cases between two defined situations. In both instances beats may sometimes be heard. But with poor differentiation by the ear these are often indistinguishable from other high overtone cases.

Figure 4

$$\underset{\text{'ghost fundamental'}}{\frac{f}{7}} \times \begin{pmatrix} 7\pm15 & 14\pm15 & \cdots \\ 7\pm30 & & \text{etc.} \\ 7\pm45 & & \vdots \\ \text{etc.} & \cdots & \cdots \end{pmatrix}$$

Apart from these three cases of varying p we may see more clearly the effects of other constraints and systems.

(1) The special case of one input signal being a pure sine wave (as in section 1 above) may be seen as $a_m=0$, for $m>1$, i.e. $F(m)=a.\sin(2\pi ft)$. The matrix then reduces to a one-dimensional array, row or column, as each overtone of the complex sound is altered and modulated into the two resulting sidebands.

(2) A fallacy of the simplified approach corrected: if we modulate two sounds of 'the same frequency', our first approximation suggests mere octave doubling:

$A+A=2A,\ A-A=0$

This covers three cases:

(i) in which one input is complex, the other a sine wave, i.e. piano plays A=220 Hz, against a sine wave of 220 Hz (as in the *Mantra* example above). Here the complex part is $n.f$ where f=220, so the product becomes $(n+1)f$. Now for n=2, the first overtone, we see the original fundamental is still present in the modulated product. Obviously the original approximation remains adequate for sounds the fundamental of which is substantially greater in amplitude than the first harmonic, but this would be much less true for an oboe, say!

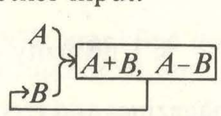
(ii) self-modulation, in which the two inputs are identical. Here too the position is far more complex than octave transposition. Each overtone effectively modulates with every other, and the full array still exists with every integer of the original (untransposed) series: the note does not appear to be transposed, merely to alter in timbre (see Fig. 5).

(iii) two complex sounds of the same fundamental frequency. Here the array will be identical to that for (ii) — Fig. 5, but the amplitude components will, of course, be entirely different. The transient phenomena, even in the steady state wave, will have more pronounced an effect in this case, generating noise components.

Figure 5

$$f \times \begin{pmatrix} 1\pm1 & 1\pm2 & 1\pm3 \cdots \\ 2\pm1 & & \text{etc.} \\ 3\pm1 & & \vdots \\ \text{etc.} & \cdots & \cdots \end{pmatrix}$$

(3) Another system that may be tried for steady state waves (the attack and decay characteristics are very noisy and uncontrollable) is octave division by modulation feedback. The result is empirically proven and the explanation here is far from a proof. A signal A is fed into one input of the ring modulator while the output from the modulator forms the other input:



therefore 'B'$=A+B,\ A-B$. This is not easily soluble, but we may explain the empirical result using the matrix.

Using the relationship p above, the matrix produces Fig. 6 when $p=\frac{1}{2}$ (i.e. we assume the second recycled input to be based on half the frequency with all its overtones). This yields a wave in which $\frac{1}{2}.f$ is also the fundamental and all the overtones are present. In other words (amplitude terms aside) the output side can be matched with our assumption about the nature of the second input. It is interesting to note that in practice the amplitude of the signal is important (high gain is needed).

Figure 6 (all values of $n.\frac{1}{2}f$ present)

$$f \times \begin{pmatrix} 1\pm\frac{1}{2} & 2\pm\frac{1}{2} & 3\pm\frac{1}{2} & \ldots\ldots \text{etc.} \\ 1\pm1 & 2\pm1 & 3\pm1 & \\ 1\pm\frac{3}{2} & 2\pm\frac{3}{2} & 3\pm\frac{3}{2} & \\ 1\pm2 & & & \\ \text{etc.} & & & \end{pmatrix}$$

*   *   *

Perhaps there should be a moratorium on use of the ring modulator in electronic music (especially live-electronics). The failure of any listening audience to perceive beyond the surface sound is not entirely their own fault! Better modulators, more care with compression and equalisation in practical set-ups may improve matters, but the nagging question cannot be avoided that digital frequency shifters would do the job better,[5] and open up greater possibilities of *structural* use of timbre synthesis and alteration. I have examined the ring modulator in a way that I hope is applicable to other analogue devices, and that may be applied retroactively to works already in existence.

NOTES:

[1]See Jonathan Cott, *Stockhausen: Conversations with the Composer* (London: Robson Books, 1974), pp.202-224.

[2]Ibid., p.221.

[3]Ibid.

[4]Each of the following matrices is in effect two superimposed matrices: for addition and subtraction. In Fig. 3, to avoid proliferation, the two results are placed side by side: e.g. 4,2 indicates that 4 results from addition, 2 from subtraction.

[5]See John Schneider, 'New Instruments through Frequency Division', *Contact 15* (Winter 1976-77, pp. 18-21.

CONTACT 18

This issue will include:

a feature on La Monte Young by Dave Smith

the continuation of the 'Music and Society' series with an article by Trevor Wishart

'IRCAM — Paris's new boutique?': an on-the-spot investigation into the workings of Boulez' new brainchild by Richard Witts

'Electronic Music Studios in Britain — 8: University of Surrey' by Robin Maconie

many reviews of new scores, books, magazines and records, and reports of events both in Britain and abroad