Proceedings of The Institute of Acoustics

THE EXTRACTION OF FM SYNTHESIS PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

Russell G.Payne and Michael Greenhough

Physics Department, University College, Cardiff.

INTRODUCTION

Frequency modulation provides a simple method for generating audio signals with very complex, time-varying spectra. The technique offers extraordinarily good value in the world of electronic and computer music, where the demand is often for a wide variety of musically interesting sounds from minimal circuitry and processing power. For this reason there was widespread implementation of FM synthesis algorithms on the mainframe computer-music systems of the 1970s. With the dramatic fall in the cost of digital hardware in the 1980s, portable commercial synthesisers based on the FM technique have become extremely popular. The very successful Yamaha DX series has brought flexible, high-quality sound synthesis within the reach of everyone.

It is possible to produce a great variety of rich sounds by manipulating just a few of the FM parameters such as the frequencies and amplitudes of the carrier and modulating signals. The amplitudes of the resulting spectral components are related to the modulation index, I, via the set of Bessel functions. The irregular shape of these functions ensures that even simple time variations of I produce spectra rivalling, in complexity, those from real acoustic systems and thus capable of holding the attention of the ear/brain.

However, it is difficult to determine just what kind of setting of the FM parameters will produce a certain desired sound. Though basic guidelines can be given, an element of trial and error is normally required. This can be time consuming and often leads to frustration, particularly amongst musicians with few technical inclinations.

There is clearly a need for a method of deriving the appropriate FM parameters directly from a physical description of the desired sound. Thanks to the cheapness of microcomputers and the establishment of the MIDI convention [1], it is now a relatively routine matter to exchange data between synthesisers and computers. An obvious approach, then, is to develop a computer alogrithm which will accept, as input, a sampled version of a real sound and produce, as output, the parameters required to synthesise that sound using the frequency-modulation technique. That sound and, more importantly, variants of it can then be resynthesised very economically from the FM parameter domain.

EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

IMPLEMENTATION

The system on which these developments are taking place is an IBM PC-AT controlling a Yamaha DX7 synthesiser via a MIDI interface board. The DX7 provides the equivalent of six FM oscillators which can function either as carriers or modulators and may be interconnected in a wide variety of configurations.

The analysis method adopted was proposed by Justice [2] and our initial implementation is described in [3]. In essence, we start with a sampled version of a signal, Sr(n), which is purely real, and derive from it the so-called analytic signal, S(n) = Sr(n) + jSi(n), which is complex. This derivation can be accomplished by applying the discrete Hilbert transform to the sampled signal, though in practice a short-cut using fast Fourier transforms is used.

We can now extract the amplitude envelope,

$$E(n) = \sqrt{[Sr(n)^2 + Si(n)^2]}$$

and the phase, $X(n) = \arctan\,[Si(n) \,/\, Sr(n)]$

and hence express the original signal as: Sr(n) = E(n) cos [(X(n)].

When more than one period of the input waveform is analysed (as is usually the case) the argument of the arctan function, Si(n) / Sr(n), will exhibit infinite discontinuities. However, a continuous phase function can be extracted if X(n) is allowed to evolve accumulatively rather than being confined to its principal values. Inevitably, though, uncertainties in the gradient are introduced in this process of 'phase unwrapping' [4]. We can extract the carrier frequency from the gradient of the phase X(n) by doing a least-squares fit. If this carrier component is then subtracted from X(n) it leaves a residual which we interpret as the modulating component. The analysis process can then be reiterated to reveal the amplitude envelope and frequency of this modulating function. If the sampled sound is sufficiently simple to be represented by a single carrier and modulator then the process will terminate here. However, for more complex sounds it is likely that a vertical chain of several oscillators, each modulating the one below, will be required. The analysis will then require one iteration for each of the levels of modulation. For a finite number of levels the process should converge and, ideally, may be completely automatable. In practice a degree of supervision and intervention may be necessary.

EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

Up to six of the DX7's oscillators may be connected in parallel (pure additive synthesis), in series (pure 'stacked' FM), or a mixture of both. A question which arises is whether all natural musical sounds can be simulated using pure stacked FM. Whilst the answer is as yet unclear, we can say that the analysis scheme described above will at least yield a starting point for resynthesis and an understanding of the structure of a given sound sample. Closer simulations may then be obtained by 'fine-tuning' of the system parameters.

It is clear that the analysis algorithm described above makes some implicit assumptions about the sound data it is likely to be given. For example, since each iteration through the analysis extracts a frequency parameter, then we can state the following:

1. the Nyquist criterion must be obeyed,
2. the waveform must be identifiably periodic ,
3. several periods must be present in the sound sample,
4. amplitude envelopes must be slowly-moving with respect to the signal frequency.

Most 'musical sounds' used as sample inputs to the algorithm will easily satisfy these conditions. However, the scheme is likely to fail in the analysis of, for example, the sound of a glass bottle breaking. It is important to note that this statement does not imply that such sounds could not be simulated using FM, but rather that analysis is likely to be difficult.

EXAMPLE ANALYSES

1. Two-oscillator FM with low Modulation Index.

Fig. 1 shows the results of the analysis of a short-duration, 2-oscillator synthetic sound. The carrier and modulator frequencies are 200 Hz and 100 Hz respectively, and the modulation index is about 1.5. The latter half of the sampled waveform (Fig. 1(a)) is clearly frequency modulated, and both the envelope and frequency paramaters are successfully extracted by the analysis algorithm. This comes as no surprise, since the analysis scheme is ideally suited to this synthesis configuration. Resynthesis of this sound using a DX7 would require matching the extracted envelope shape for each oscillator (actually a continuous function) with the discrete envelope parameters required by the synthesizer, together with the relevant oscillator frequency parameters.

EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

2. Two-oscillator FM with high Modulation Index.

The behaviour of the analysis algorithm with more complex sounds is demonstrated in Fig. 2. Here, the input sound was again synthesized on the DX7, using 2 oscillators, with frequencies as above, but the modulation index now about 10. The resulting frequency spectrum, Fig. 2(a), shows that most of the sound energy was centred at freqencies much higher than the carrier used. Thus we might expect the analysis scheme to produce different resynthesis parameters to those used to create the original sound, and this is indeed the case. The carrier frequency component is not present in the input spectrum, through the cancellation action of reflected 'negative sidebands'. The average phase extracted corresponds to a carrier freq of 692Hz ; further iterations produce two layers of modulators, each of 98Hz. Fig. 2(b) shows the frequency spectrum of the resynthesized sound using this arrangement. As might be expected from a comparison of Figs. 2(a) & (b), the sound of the original and the resynthesised signal are very similar.

3. Analysis of sounds requiring more than one FM stack.

Since the DX7 allows the parallel connection of several vertical oscillator stacks, the possibility of applying our analysis scheme to discrete regions of sampled sound frequency space presents itself. Fig. 3 shows the time-evolution of a synthetic sound suitable for this treatment. The sound is like that of a harp pluck with a flute-like sustain, and is conveniently generated using two chains of three oscillators, summed in parallel. Examination of the first few time frames of Fig. 3 reveals that the spectral regions occupied respectively by the harp and the flute partials do not significantly overlap. Setting a threshold frequency of 650 Hz allows separate processing of the two regions. The extracted amplitude envelope and frequency parameters for the first analysis iteration of these regions (processed over 40 time frames, each of about 110 ms duration) are given in Fig. 4. The oscillatory behaviour of both the frequency and amplitude plots of Fig. 4(a) is due to a 'beating' effect in the flute sound (similar to that produced by two oscillators of slightly different frequencies). The analysis scheme extracts, alternately, the 'true' carrier frequency (220Hz) , and a higher octave (440Hz).

EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

Fig. 4(b) demonstrates a similar effect in the analysis of the upper (harp) frequencies. Here, the analysis algorithm initially appears to latch on to arbitrary frequencies around 1760 Hz. As the higher partials gradually decay, successively lower harmonics are 'favoured' until the 'true' carrier frequency of 880 Hz is observed. Towards the end of the sound sample, where the signal amplitude is very low, complex interactions between the remaining significant partials again yield unpredictable frequency parameters. Further iterations would then be required to complete the analysis of this sound. The implications of this example (for future development of the analysis algorithm) are currently being considered.

CONCLUSIONS

The scheme enabling computer-assisted analysis of musical sounds into FM resynthesis parameters is still under development,although some assessment of its potential has been made. At present we are using contrived, but complex, sound samples in order to establish a physical basis for this development. Whilst it seems likely that the scheme could cope with sounds fulfilling the four criteria listed earlier, completely arbitrary sounds would present considerable difficulties, at least in any automated analysis mode. At present the scheme requires a large amount of processing time even for short sound samples. Methods of reducing the amount of data processing required are being considered to facilitate more efficient development of the algorithm.

REFERENCES

1. 'MIDI Specification 1.0', The International MIDI Association, 1187 Hartshook St., N. Hollywood, CA 91607 USA.

2. J.H. Justice, 'Analytic Signal Processing in Music Computation', IEEE ASSP, Vol.27, no.6, 670-684, (1979).

3. R.G. Payne, 'Synthesizing Complex Audio Sounds With FM ', Proc. IOA, Vol. 8, Part 6, 90-97, (1986).

4. J.M. Tribolet, 'A new Phase Unwrapping Algorithm', IEEE ASSP, Vol. ASSP-25, no.2,170-177, (1977).

EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

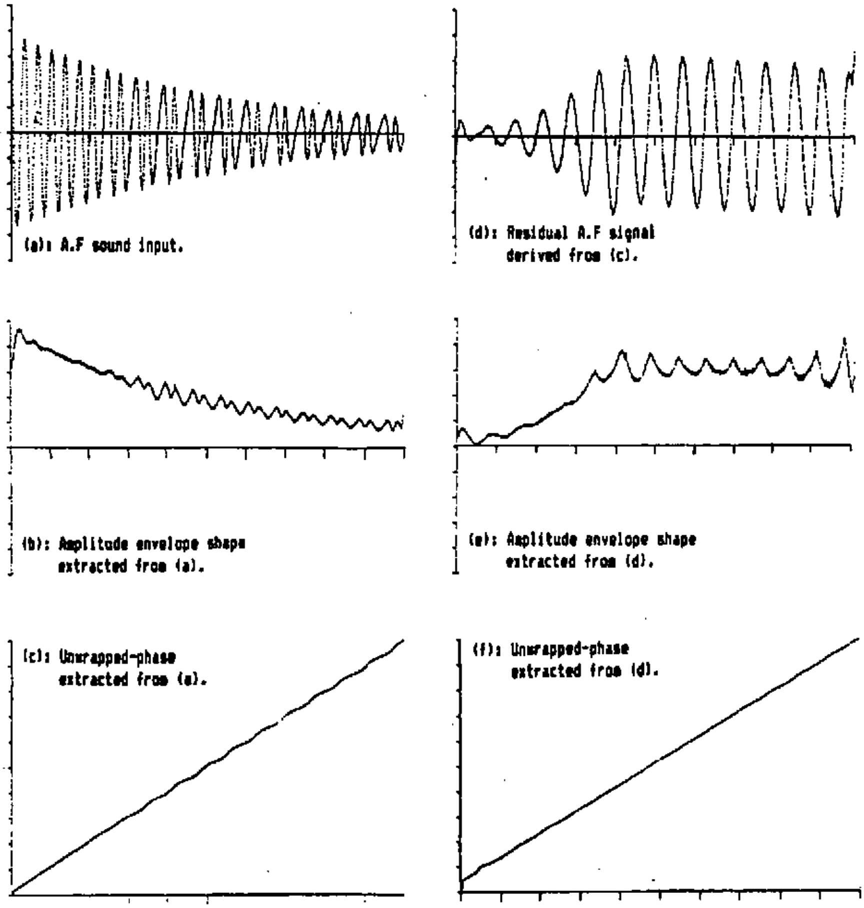


FIG. 1 : Analysis of a simple, two oscillator FM sound with:
Modulation Index = 1.5.
Carrier frequency = 200 Hz.
Modulator frequency = 100 Hz.

EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.

FIG. 2 : Comparison of frequency spectra produced by the FM systems shown.

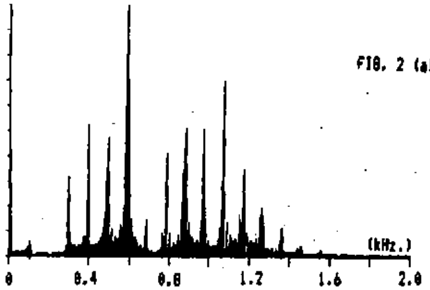
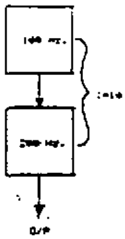


FIG. 2 (a): High modulation index, two oscillator system.



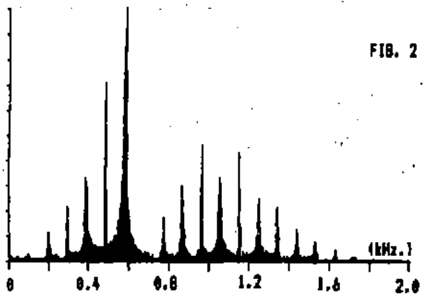
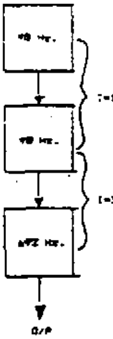


FIG. 2 (b): 'Alternative' system as determined by analysis scheme.



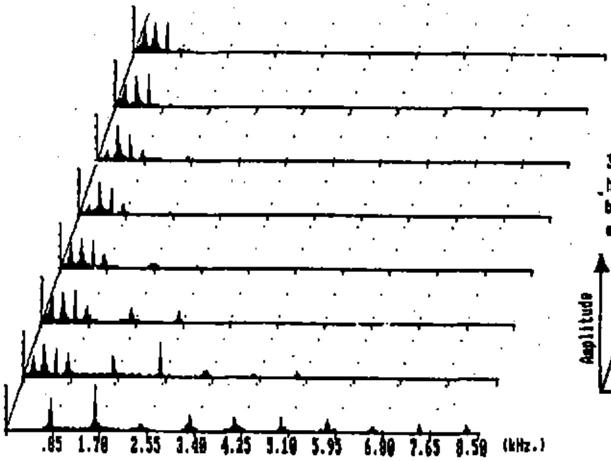
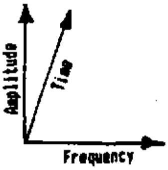
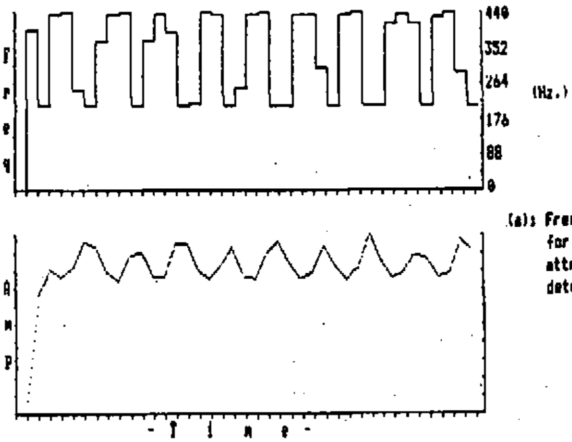


FIG. 3 (left):

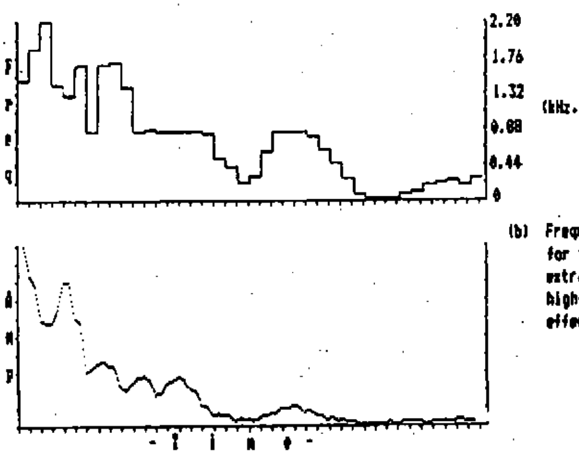
3-D spectral plot of synthetic harp and flute sound. The spacing between time frames is 440 ms, and each frame is 110ms in duration.



EXTRACTION OF FM PARAMETERS FROM SAMPLED MUSICAL SOUNDS.



(a): Frequency and amplitude envelope plots for the flute sound. Oscillatory behaviour attributed to frequency 'beating' between detuned oscillators in original sample.



(b) Frequency and amplitude envelope plots for the harp sound. Analysis algorithm extracts correct frequency after initial higher partials have decayed. 'Beating' effect also evident.

FIG. 4 : Plots of the amplitude envelope and frequency parameters extracted from the processing of the combined 'harp and flute' sound of Fig. 3. The total sample time is 4.4 s.