Proceedings of The Institute of Acoustics

HIGH ORDER LPC ANALYSIS/SYNTHESIS OF ATONAL PERCUSSION

M.B.Sandler

Dept. Electronic & Electrical Engineering, King's College London, UK.

ABSTRACT

In this paper the Linear Predictive Coding (LPC) algorithm is discussed showing how it may be used to formulate a data efficient representation of acoustic phenomena - in particular for drums. Its use in analysis and synthesis of drums is developed and some time domain graphs, showing the effectiveness of the algorithm, are presented.

1. Introduction

The LPC algorithm is now established as the primary means for efficient coding of human speech signals for digital transmission purposes. It has been developed to produce many varieties, but in one of the more basic forms, is capable of data reduction on the order of 10:1 compared to straightforward PCM (8 bits is the norm for speech).

LPC works by making the assumption that the acoustic system producing the phenomenon is resonant. This is in distinction to the Fourier model (as used in the Fourier Transform and the FFT) which assumes a superposition of sinusoids as the generative process.

It is also otherwise known in some branches of the Digital Signal Processing community, and in economics and statistics, as an Autoregressive (AR) modelling algorithm.

1.1. The LPC model

The form of the AR model generated by the LPC procedure is

$$AR_p(z) = \frac{1}{A_p(z)} \qquad (1)$$

where

$$A_p(z) = 1 + \sum_{i=1}^{p} a_i z^{-i} \qquad (2)$$

and $A_p(z)$ is a polynomial in z, known as the inverse filter, p is the order of the model and the polynomial, $\{a_i\}$ are the filter, model or polynomial coefficients, which are determined by the algorithm. The order p is a measure of the accuracy of the model.

Evaluation of $AR_p(z)$ for $z = e^{j\theta}$, ($\theta = 2\pi f / f_s$,$\theta \in [0,\pi]$) results in the complex function describing the Fourier Transform of $AR_p(z)$, and both the magnitude and phase functions may be generated as a consequence.

Those values of z for which $A_p(z) = 0$ or $AR_p(z) = \infty$ are known as the poles of $AR_p(z)$ or zeroes of $A_p(z)$. If p is even and $AR_p(z)$ is stable, then its poles occur as complex conjugate pairs inside the unit circle. In isolation, each pole pair acts as a resonator (or formant in speech parlance).

The way in which the Fourier and LPC models differ is demonstrated by Figures 1 (a) and (b) below.

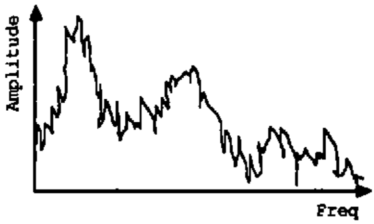

Figure 1 (a)

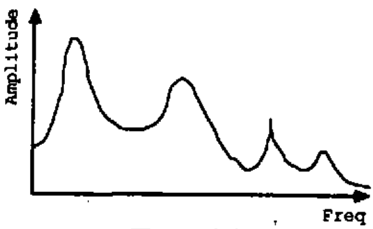

Figure 1 (b)

In these, Figure 1 (a) gives an example of the type of spectrum that might be obtained by Fourier analysis, perhaps by using an order 256 FFT, whilst 1 (b) shows an LPC model spectrum, as generated from the same original data, but using say a 10th order model. It may be said that the LPC spectrum is a smoothed version of the FFT spectrum. Its importance lies in the fact that important features, like the resonances shown, become clearer and that representation is significantly more efficient.

1.2. LPC and Speech

The basic speech production model is shown below in Figure 2. It consists of a time-varying resonant filter, which is the vocal tract, and a choice of inputs, pitched and noise. In speech synthesis, each fundamental utterance, or allophone, is generated by selecting an appropriate input and the resonances of the filter. The sequence of utterances is merely the concatenation in the time domain of the output of this model as its parameters change. For analysis, the time-varying nature of speech is dealt with by assuming that speech is stationary for short time periods and fixing the model over those periods.

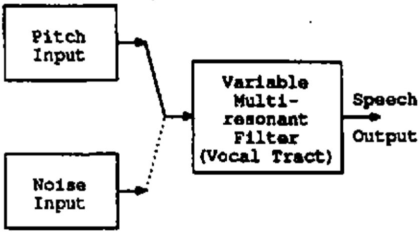

Figure 2

1.3. LPC and Music

In LPC analysis and synthesis of musical instruments, this same model is used, though with some modifications. The pitched input might be the vibrations of the lips of a brass player or the reed of a woodwind instrument. The noise input may be assumed for a percussive instrument such as a drum. The resonant filter may be a pipe, a string, or a skin in the case of a drum.

For most instruments, the resonant filter has fixed properties throughout the duration of a single note, though for most, the filter will change from note to note. This is not the case for a drum beat.

1.4. LPC and Drums

A drum consists of a skin (or membrane) which, conventionally, is circular, and is stretched over a supporting metal frame and connected to a cylindrical wooden cavity. In many cases the cavity will be made airtight at the bottom, by means of a second skin. This is depicted in Figure 3.

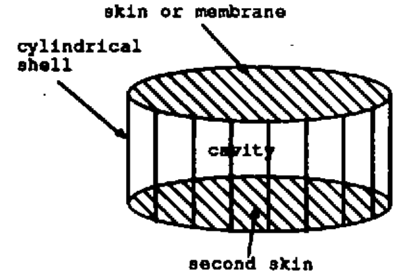

Figure 3

The top skin is struck with a wooden stick (the use of brushes and padded beaters will be ignored for the purposes of this discussion) which causes it to resonate. The physics of resonant circular membranes is well documented, having been discussed as long ago as the last century, by Rayleigh [1]. Essentially, the resonant modes or frequencies are related to one another as the zeroes of first kind integer order Bessel functions, with the fundamental being determined by factors such as the tension in the membrane (skin tightness), its diameter and mass. Since the zeroes of $J_n(x)$ are not evenly distributed for variation in x, the overtones of a drum are not harmonically related. It is this last important point that gives rise to the characteristic atonal sound of a drum.

The resonances of the upper membrane are affected by the characteristics of the drum shell and the second skin, but not in any way that significantly modifies the discussion above. Besides, both these and the parameters above are static, in that they will not be expected to change either during a single drum beat or indeed during a performance.

There are however two dynamic parameters influencing a drum's sound: these are the position and force of the strike with the stick. The strike position will determine how strongly the various modes of resonance are stimulated (i.e. the relative amplitudes of the resonances). The more important influence, however, is the strength of the strike. This causes the skin to be distorted, imparting further tension and causing the simple stretched membrane model above to be, at least partially, invalidated.

The sonic effect is that the fundamental (and all overtones) reduce in frequency as the drum beat progresses, this being caused by the added tension gradually diminishing. This means that in comparison to say a single note from a flute (ignoring for convenience the attack and decay periods) a drum beat must be analysed not as a single sound record, but as a concatenated sequence of records, just as in the case of speech. For each record a new model, as in equation (1), is computed and the complete set of models represents the drum beat.

The differences between the model for speech and for an atonal percussion instrument such as a drum may thus be summarized as follows:-

- the shift in resonant frequencies from one time segment to the next is small, predictable and smooth for drums, but may be large, discontinuous and unpredictable for speech
- speech synthesis requires only a few resonances to be identified for reasonable perceptual quality, whereas drum synthesis will require many resonances for a high quality rendition
- the input to a drum model need only be a short burst of noise, but for speech it

must be either a sequence of pulses of a given pitch or a noise sequence lasting the duration of the utterance

2. Theory

The LPC spectral model of equations (1) and (2) may be rewritten as

$$AR_p(z) = \frac{1}{1 + F_p(z)} \tag{3}$$

where

$$F_p(z) = \sum_{i=1}^{p} a_i z^{-i} \tag{3a}$$

(for Digital Signal Processing afficianados $F_p(z)$ is an FIR filter)

In block diagram form, the system implied for sound synthesis is shown as Figure 4 below.

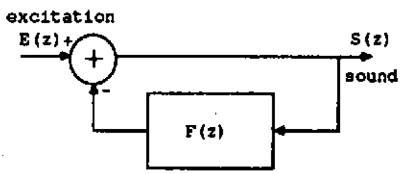

Figure 4

In this diagram, the output S(z) is given as

$$S(z) = E(z) \cdot AR_p(z) \tag{4}$$

$$= E(z) \cdot \frac{1}{1 + F_p(z)} \tag{4a}$$

Thus the LPC algorithm finds the coefficients of the 'inverse' filter F_p needed to produce sound S given an excitation E. The algorithm attempts to minimize the energy in the unknown excitation, E, which if the analysis model fits exactly with the physical process (i.e. S is produced entirely by resonance) would be a single impulse or continuous white noise.

In the time domain, E(z) and S(z) are denoted as the discrete time sequences e[n] and s[n].

In terms of a drum, E(z) is the spectrum of the energy put into the skin/cavity combination and $F_p(z)$ (or equivalently $AR_p(z)$) represents the resonant modes. E(z) is reasonably approximated by a short burst of noise.

Once the model, as the set of values $\{a_k\}$, has been found for a sound fragment, the effective excitation may likewise be computed by inverse filtering. Consider equations (4), these may be rewritten as

$$E(z) = S(z) \cdot \frac{1}{AR_p(z)} \tag{5}$$

$$= S(z) \cdot A_p(z) \tag{5a}$$

$$= S(z) \cdot (1 + F_p(z)) \tag{5b}$$

which means that the excitation sequence may be obtained by filtering the acoustical output, S(z), with the filter block of Figure 4 operating in open loop mode as in Figure 5 below.

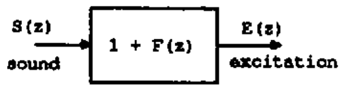

Figure 5

3. Analysis

The LPC analysis of a drumbeat proceeds as follows. The beat might last about 2 seconds and will be sampled at 20kHz or higher. Thus a typical soundfile consists of 64kwords of information, which in all the analyses described herein are 16 bit 2's complement integers, which are converted in all programs to floating point representation. The soundfile is read into the LPC program in frames (or segments or fragments) of between 200 and 2000 samples, depending on the order of the LPC model requested. It is normal to use a frame size up to about ten times greater than the model order as this will give an adequately accurate representation and provide a 10:1 data reduction. The analysis results are written to a file as a sequence of coefficient sets, $\{a_k\}$, one set for each time frame in the soundfile.

The coefficient file is then used for a number of tasks. The most important of these is as input to a program which will identify the individual resonances (see Discussion). It is also used to re-generate the excitation sequence assumed by the modelling procedure, and for direct resynthesis as per equation (4). It is this last use which will receive most attention in this paper.

4. Synthesis

As in analysis, the sound is created in frames, which are concatenated together to form the overall new soundfile. Each frame corresponds to a single set of coefficients. The input to the LPC synthesizer of Figure 4 is either an excitation as extracted by inverse filtering, or it may be a burst of random excitation, generated within the program. For flexibility in resynthesis, the recreated sound segments may be the same size as those used to generate the corresponding coefficient set, or indeed may be longer or shorter. Changing the duration of each segment will not only increase or reduce the overall duration of the synthesized soundfile, it will also slow or speed the pitch shift which occurs during a drumbeat, as caused by the deformation of the skin and explained earlier.

5. Experiments

The experimental set-up is as shown in Figure 6, which depicts the three main programs used, together with the files produced and used by these. This has largely been explained in the two proceeding sections. Important parameters effecting the behaviour of the program are also shown.

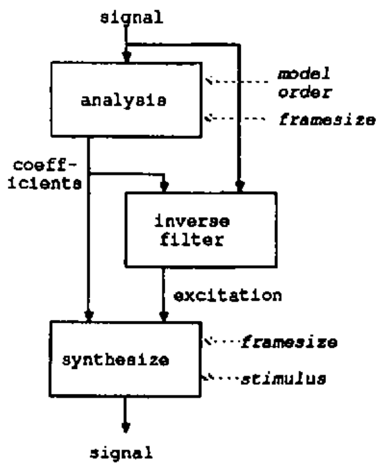

Figure 6

The original soundfile used in this work is shown as Figure 7. It was recorded from a 14" tom-tom drum in an acoustically dead environment, and ensuring no signal clipping and other visibly gross non-linearities in the recording chain, onto a Betamax VCR/PCM recorder. It was transferred to a personal computer using a 16 bit analogue-to-digital converter operating at 20kHz. With computer noise etc, it is estimated that the final resolution is about 14 bits.

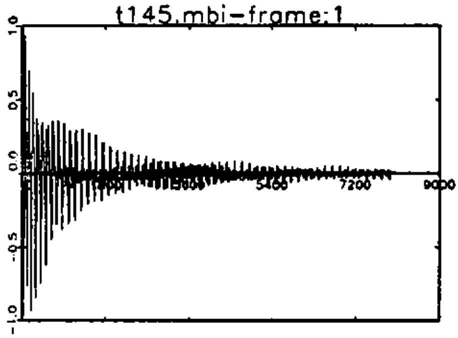

Figure 7

From this, two 200 order models were generated, each using the autocorrelation method of LPC analysis [2] and with a Hamming window applied to the data. Model a) is generated over contiguous blocks of 2048 data samples and model b) over blocks of 1024 samples.

Figure 8 (a & b) shows the two resynthesized versions of the original (Figure 7) in which both were driven with 256 samples of uniformly distributed white noise. In a) each coefficient set regenerated 2048 output samples (corresponding with the manner in which the model was obtained) and b) computed 1024 output values from each set of coefficients.

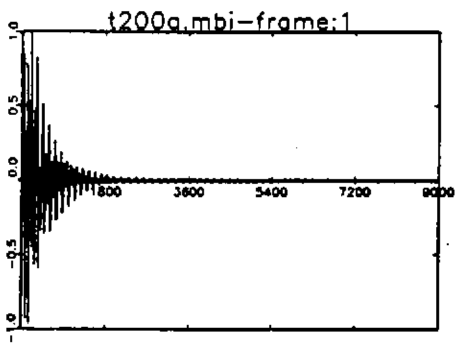

Figure 8 (a)

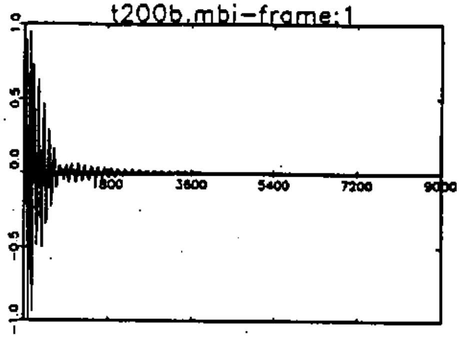

Figure 8 (b)

It can clearly be seen that these two waveforms have the same general characteristics as the original, but differ significantly in detail. Furthermore, both signals decay more rapidly than the original.

Thus further experiments were carried out. In the first of these, the two models were driven with the excitation signal as derived by the inverse filtering operation described above. The results are shown as Figure 9 (a & b). Here a dc offset lasting the duration of the input signal in each case (2048 samples for model a and 1024 samples for model b) can clearly be seen: no explanation has yet been found for this behaviour.

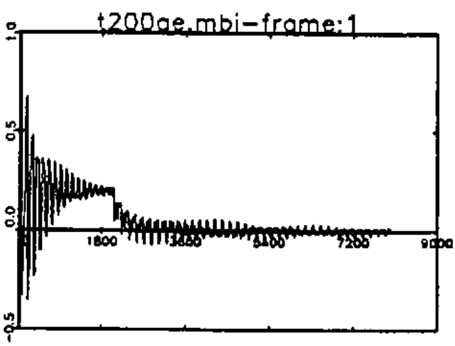

Figure 9 (a)

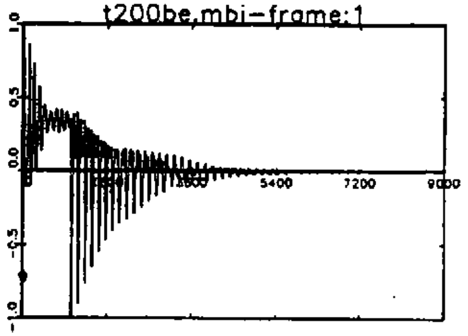

Figure 9 (b)

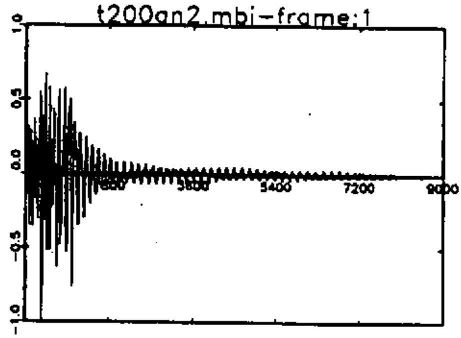

Figure 11

To investigate the effect of the duration of an random noise input, the experiments of Figure 8 were modified in the following ways. Firstly, model a) was driven by 2048 samples and then by 1024 samples of noise, and then model b) was likewise stimulated. The results are shown as Figures 10 to 13, respectively. Here the waveforms look significantly more like the original, particularly in the way that the oscillations die away slowly, though there are still significant differences.

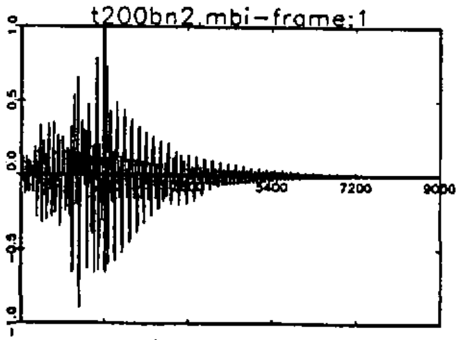

Figure 12

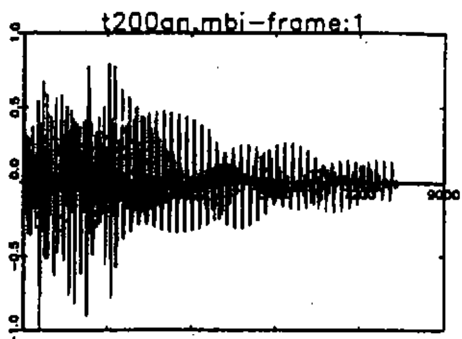

Figure 10

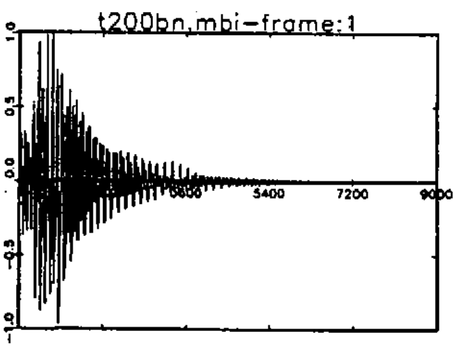

Figure 13

6. Discussion

As shown above, it is possible to create a drum-like waveform using LPC analysis and synthesis techniques. However, it is also clear that it is difficult to exactly re-create the original waveform. It is also clear that a great deal of flexibility is possible with such a synthesis scheme, as evidenced by the waveforms above, the variety having been produced from only two similar models, driven in slightly differing ways. Thus with careful selection of an analysis methodology and later of a synthesis methodology, a plethora of drum-like sounds may be created simply. In the music synthesis field, this is of importance, as it is not necessarily great fidelity to the original that is paramount, rather it is a flexibility of recreation that is desirable.

This work was originally started as an investigation into new means of synthesizing drum sounds, and that reported here is only the initial phases. Later stages are intended to address different synthesis techniques in which each of the resonances in a drum beat are individually created.

In order to do this, techniques have been developed for factorizing polynomials of high order (e.g. 200), such as occur in equation (2), into individual second order (resonant) sections. These sections are then used to produce a synthesizer that has much in common with the formant synthesizers used in speech.

The analysis of AR models as obtained by LPC analysis, into separate complex conjugate pole pairs (resonators) has been reported in [3], and the work continues, presently investigating ways of implementing such high order filters.

It is also hoped in the future, to apply these same techniques, perhaps modified slightly, to other, tonal, instruments.

7. Conclusion

This paper has described the LPC technique for analysis of acoustic phenomena, such as drum beats, and shown how the process can be reversed to create new drum sounds. The experiments reported have shown that some care must be taken in the analysis procedure as well as in the synthesis, but that, in general, LPC does simultaneously offer a degree of flexibility, fidelity to the original and variety, so that it may well become an accepted technique for music synthesis.

8. References

[1] Lord Rayleigh: "The Theory of Sound", 1877.

[2] I.Witten: "Principles of Computer Speech", Academic Press, 1982.

[3] M.B.Sandler: "High Precision Formant Estimation from High Order AR Models", Proceedings of 5th International Conference on Digital Processing of Signals in Communications, 20-23 September 1988, Loughborough.