APPLYING BALANCED MODEL TRUNCATION TO SOUND ANALYSIS/SYNTHESIS MODELS

J.P. Mackenzie, J. Kale and G.D. Cain
School of Electronic and Manufacturing Systems Engineering, University of Westminster,
115 New Cavendish St., London W1M 8JS, U.K. Tel: +44 071 911 5000 x5642 or 911 5083,
e-mail: mackenzj1@wmin.ac.uk or kale@cmen.wmin.ac.uk

Abstract: Balanced Model Truncation (BMT) is a powerful technique that can be used to greatly reduce the order of certain digital filters with little distortion to their frequency and phase responses. Since digital filters are commonly used in computer music applications, BMT may be a tool of considerable practical use. To demonstrate this potential, BMT is applied to an autoregressive (AR) drum sound model, showing that its order may be reduced considerably with little change to the accuracy of the model.

1. INTRODUCTION

The intention of this paper is to introduce the technique of Balanced Model Truncation (BMT) as a powerful tool for use in computer music applications. BMT is a means by which the order of a digital filter may be significantly reduced while accurately maintaining its input/output response. This is achieved by analysing the filter to quantify and then 'cut' it of system redundancy. In (Belc-Zmyski, B.), it is shown that certain high-order FIR filters may be converted to low-order IIR ones while faithfully reproducing the frequency and phase response of the original.

In the field of computer music, digital filters are widely used in the generation and processing of sound where, typically, the demand is for filters that model complex acoustic systems, but that are of low orders for implementational ease and speed. The application discussed in this paper is of an autoregressive (AR) model of percussive and drum sounds which consists of an all-pole IIR digital filter excited by a white noise source. It has been reported in (Sandler, M. 1999) that for such a model to faithfully model a given drum sound, it must be have an order of several hundred. Such a high order, however, presents significant implementation problems.

In this paper we demonstrate how BMT can be used to derive a much lower order autoregressive moving average (ARMA) model with a performance that is mostly indistinguishable from the AR one. Furthermore, this model is of greater accuracy than an ARMA model of the same lower order derived using a conventional method.

2. AR AND ARMA MODELLING

Autoregressive (AR) models are widely used in computer music as part of powerful sound analysis/synthesis schemes - see for example (Moorer J.A.) and (Laney P.). Using linear prediction to estimate the model parameters it is possible to directly find an all-pole filter with frequency response that is closest, for a given model order in a least-squared-error sense, to the spectrum of a given signal. A sound is typically modelled in segments, each of which is represented by its own AR filter. These filters may then be used to resynthesise the sound and perform powerful modifications to it by altering the excitation to the filters, or the filter parameters themselves. In (Sandler, M. 1999) drum sounds are modelled with the goal being to facilitate the resulting filters into second-order sections so that the individual resonances of the drum may be modified independently. Because of the high orders involved, however, the filters are not easily implemented this way and a solution to this problem is still being sought (Sandler, M. 1992).

It would therefore be preferable to use an autoregressive moving-average (ARMA) model instead as the presence of both poles and zeros allows greater flexibility of the filter's spectrum and consequently lower model orders. The ARMA parameter estimation problem is, however, nonlinear with respect to the parameters and therefore no direct solution can be found as is the case for the AR model. An alternative solution is to use nonlinear optimisation techniques as is the case, for example, in the
software package MATLAB. The result, however, is often computationally costly to obtain and is not guaranteed to be optimal in any sense.

The AR signal model is of the form,
\[ x(k) = \sum_{i=0}^{\infty} A_i e^{-i \omega k} \]
where \( y(k) \) and \( e(k) \) are c-transforms of the signal \( y(k) \) and a white noise sequence \( e(k) \), both of length \( N \), where \( k \) is the discrete-time index, \( A_i \) is a polynomial of order \( f \), \( A_i = 1 + a_1 z^{-1} + \cdots + a_f z^{-f} \). Given some signal, \( v(k) \), of length \( N \) to model, the goal is to find the set of \( a_i \) such that the squared difference between \( v(k) \) and \( y(m) \), that is,
\[ \sum_{k=0}^{N-1} (v(k) - y(m))^2 \]
is to be minimized. Solution of this problem leads to a set of linear equations in the unknowns (\( a_i \)) which can be solved in a number of ways. For this paper, the MATLAB AR routine is used which uses forward and backward prediction of the time-series to be modelled to derive the (\( a_i \)).

The ARMA parametric model is an AR model with an added moving average part having the form
\[ y_{ma}(z) = \frac{B_{ma}(z)}{A_{ma}(z)} e(z) \]
where \( B_{ma} \) is another polynomial, of order \( g \), independent of \( A_{ma}(z) \), but of the form \( B_{ma}(z) = b_1 z^{-1} + \cdots + b_g z^{-g} \). The MATLAB ARMA routine used for this work minimizes the model error with an iterative Gauss-Newton optimization routine and typically the true minimum is not found, but is only approximated.

3. BALANCED MODEL TRUNCATION (BMT)

BMT provides an analysis of a linear time-invariant system that reveals which of the states are contributing to the input/output response and quantifies the degree of redundancy, if any, that exists. Having determined which states of a system have an insignificant contribution, it is then possible to truncate the system to rid it of the redundancy, with only a very small effect on the system response. In many cases, it is possible to accurately approximate both the magnitude and phase response of the system while significantly reducing the system order.

A system transfer function \( F(z) \) of order \( n \), can be written as a state space difference equation, \( \hat{x}(k+1) = A\hat{x}(k) + B\hat{u}(k) \), \( y(k) = C\hat{x}(k) + D\hat{u}(k) \), where \( \hat{x} \) is the \( n \)-dimensional state vector, \( u \) is the scalar system input, \( y \) its output, and \( A, B, C \) and \( D \) the matrices which define the system. The system transfer function is related to state space by the relationship \( F(z) = C(zI - A)^{-1} B \). Any system transfer function has an infinity of state space realizations that are input/output equivalent. These are related by any nonsingular matrix, \( \Sigma \) such that \( \Sigma^{T}A \Sigma \), \( \Sigma^{T}B \), \( \Sigma^{T}C \) is where \( C(zI - A)^{-1} B = F(z) \).

Two matrices that give important information about the behaviour of states within a particular realization are the controllability and observability grammians \( P \) and \( Q \). A state space realization for which \( P = Q = \Sigma \) is known as a balanced realization. This has the property that the states are ordered according to their contribution to the state responses which are quantified by the values of the (\( c_i \)). These are called the Hankel Singular Values (HSV) of the system. Moreover, the balanced system may be partitioned in state space into two subsystems: \( (A_s, B, C_s) \), called the true-rated system and having order \( m \), and \( (A_n, B_n, C_n) \) called the rejected system having order \( n-m \), where
\[
\begin{bmatrix}
A_s & A_n \\
B_s & B_n \\
C_s & C_n
\end{bmatrix}
\]
The truncated state space system corresponds to a transfer function \( F_c(s) \) with a similarity to the original system \( F(s) \) which is quantified by the Hankel norm \( \left\| F(s) - F_c(s) \right\|_\Sigma = 2 \text{tr} \left( \Sigma_c \right) \), where

\[
\Sigma = \begin{bmatrix}
\Sigma_c & 0 \\
0 & \Sigma
\end{bmatrix}
\]

That is, the difference between the original and approximate system transfer functions is a function of the sum of the Hankel Singular Values of the rejected system. It therefore follows that if the HSVs of the rejected system are chosen to be substantially smaller than those of the truncated system, then the approximation error will be small. In practice, the possibility of reduction is dependent on the range of the HSVs for the balanced form of the original system. The HSVs can be displayed on a graph to allow the truncated system order, and consequently the accuracy of the approximation, to be chosen. An example HSV graph can be seen in Figure 1.

4. EXPERIMENTS AND RESULTS

The experiment to test this technique consists of a comparison of the three modelling techniques, AR, ARMA and BMT, of AR, when applied to the analysis and resynthesis of a woodblock sound.

The sound time series is segmented into a number of shorter signals, of length 1024 samples, each of which is analysed with the three models. In the first instance, the signal is analysed with an AR model of sufficiently high order to give a perceptually good approximation of the original sound. The analysis generates an all-pole IIR filter which is then subjected to BMT to significantly reduce its order. This generates another IIR, but containing an equal number of poles and zeros. Finally, the same signal is analysed with a conventional ARMA method to give an IIR filter of the same order as the reduced version for comparison. The result of the analysis stage, then, is three sets of filters each corresponding to the set of segments of the original sound time series.

Each of the three sets of filters is then used to resynthesise a version of the original sound by exciting each filter with a uniformly distributed white noise sequence of length equal to that of the original signal segment. The resulting signal is then amplitude scaled so that it has the same energy as that of the original and the synthetic segments are simply concatenated to form the resulting sound.

To quantify the accuracy of the models for comparison, all the signals are converted into the frequency domain via the FFT. A difference measure is then calculated as the sum of the squared difference between the synthetic signal magnitude spectrum and that of the original for each of the signal segments.

Table 1 summarises the analysis parameters and resulting accuracies while Figures 2 and 3 show time and magnitude spectra plots for a single segment of the original and each of the synthetic versions. Note that the AR version does not maintain the phase information of the original because the model has been excited with white noise. Therefore the corresponding magnitude spectra plots are similar, but their time series are different. This loss of phase information was not found to be perceptually significant. The reduced AR version, however, does maintain the phase information of the AR model and so both time and frequency magnitude plots are similar.

Figure 1 shows an example of one of the HSV plots that arose from the balanced model realisation of an AR model. The model used is that of the first signal segment. It can clearly be seen that there is a considerable difference between those values associated with the first twenty or so states and those of the rest which therefore dictates the choice of reduction to order 20.

The sounds from the models compare well: the AR model produces a good likeness of the original while that from the reduced version is perceptually indistinguishable from this even though the order of the system has been reduced by a factor of five.

5. CONCLUSION

The spectral difference measures quoted in Table 1 are very encouraging as they indicate that the reduced AR route to an ARMA signal model is more accurate than the conventional nonlinear
optimisation approach. From the time and frequency plots of Figures 2 and 3 it can be seen that the reduced AR model is virtually indistinguishable from the AR version, even though it has only one fifth the order. This is of enormous relevance to the problem of drum sound modelling as it suggests that although high-order AR models are required to capture the essential characteristics of a sound, implementation of much lower order equivalent filters is possible via the BMT technique.

REFERENCES


Table 1 Summary of analysis parameters and resulting spectral difference for a struck wood block sound.

<table>
<thead>
<tr>
<th>Sound</th>
<th>wood block</th>
<th>sample rate: 44,100Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>length:</td>
<td>2490 samples</td>
<td>quantisation: 16 bits</td>
</tr>
<tr>
<td>number of segments:</td>
<td>10</td>
<td>segment length: 1024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Model Order</th>
<th>Spectral Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>100</td>
<td>1.441</td>
</tr>
<tr>
<td>Reduced AR</td>
<td>20</td>
<td>1.417</td>
</tr>
<tr>
<td>ABMA</td>
<td>20</td>
<td>3.312</td>
</tr>
</tbody>
</table>

Figure 1 Hankel Singular Values for each of the 190 system states of an AR model.

Figure 2 Amplitude against discrete time plots of a single segment of the original and modelled wood block sound.

Figure 3 Log magnitude spectrum plots for one segment of the wood block sound. Only the first 100 frequency bins out of 512 are shown for clarity.