6. On Predicting a Performance Benchmark

6.1 Overview

MAS is a statistical optimisation based upon certain assumptions about frequency stationarities in note-based music. It is a challenge to provide a definitive performance benchmark in comparison to classical AS because the principle of generality - a characteristic strength of AS - is retained and cost of computation is adapted to the application in question in an attempt to reach a compromise between the conflicting points of section 1.2. For instance, some specialist sounds (e.g. Sheppard Tones) (Moore, 1990) require mapping into the highest level subband $s_{0,1}$ of the proposed hierarchy with poor performance due to the overheads of supporting an unexploited multirate paradigm. In contrast, a classical 'pipeless' organ (Comerford, 1993) with fixed pitch notes may map with great efficiency into the terminal integer subband series $s_{3,1..8}$ for K=3.

In the context of commercially marketed digital music synthesisers, note-based music will remain the pre-eminent application for the foreseeable future whereas parallel fields of music technology research, (e.g. granular synthesis) are oriented towards facilitating the composition of 'avant-garde' electro-acoustic music. The research motive of MAS is in concordance with the former imperative, and hence pieces of music in traditional score notation provide the most appropriate statistical source for benchmarking. Score notation is based upon discrete pitches with occasional expression marks for pitch inflection such as vibrato and glissandi; such accentuation is sparingly used in practice, or restricted to small sections of ensembles. Hence this investigation is confined to fixed-pitch note synthesis, though a limited account is taken of average pitch dynamics.

Eqn. (3.1) in section 3.1.7 forms the basis for quantifying the relative efficiency of MAS relative to classical AS. However, three items of data are required for its evaluation, two of which are u_{mas} and $v(K)$. The third item n_k is application-specific and for benchmarking, the best choice is 'representative' partial frequency distributions derived from actual musical scores; in particular, those composed for large ensembles such as

symphony orchestras. The justification for an ambitious modelling scenario is that a common application of AS is the resynthesis of acoustic instruments, and MAS is designed to facilitate large scale synthesis requiring large *S* in eqn. (1.1). From a statistical perspective, the large quantity of input data represented by a symphonic work improves the integrity of output data from the benchmarking model.

The simulation presented in this chapter models the resource allocation required for ensemble synthesis and performs no oscillator bank / filterbank computation. Score data must be combined with both an orchestration, and a database of analysed instrument timbres, from which it is then feasible to predict the control data set for a multirate oscillator bank as it evolves during the piece. Subjective questions about the purity of sound generated via MAS must therefore remain unanswered, though it is argued - accepting both the well-known properties of AS outlined in section 1.1.1 and the principles of functional transparency adhered to in the development of MAS - that this issue is not of great concern: sound fidelity for MAS, like AS, is dependent principally upon implementation parameters such as f_s, wordlength and S/N tolerances etc.

One proposal for optimising AS is relevant to mention here: the imposition of an auditory model to prune out imperceptible masked partials - 'receiver coding'- founded on the premise that TOB's are expensive to compute (c.f. arguments of section 1.3.2). A discussion of receiver coding and its inherent problems is given in section 2.3. In the MAS paradigm and associated MASC design, it is asserted that the founding premise of the technique is weak because oscillators become far more economic to compute and control on a large scale in a MASC, in comparison to a TOB, and thus techniques exploiting masking are disregarded in this chapter.

6.2 Description of The SHARC Timbre Database

SHARC is a database developed by Sandell (1991, 1994) which catalogues timbres, in the form of Fourier spectra, for most acoustic instruments to be found in a modern symphony orchestra. The source data originated from live recordings which were compiled in a CD library labelled the McGill University Master Samples (MUMS). For each instrument, a professional musician played a chromatic scale in each playing style

(e.g. open, muted for a French horn) such that, after digital recording, individual notes could be isolated for analysis. To facilitate a complete spectral model, SHARC therefore provides a timbre for each note within the compass of each instrument.

As the analysis procedure assumes that the signal is a stationary tone composed of a sum of partials in strict harmony, segmentation is carried out upon the sustain portion of the note and a number of waveforms are extracted to facilitate ensemble averaging and accurate pitch estimation. Hence, information on spectral evolution during attack transients is lacking. However, note sustain is slightly time-varying and a 'representative point' is required such that the analysis segment is maximally stationary according to an objective measure (to minimise spurious spectral components) so that the instantaneous spectral envelope encapsulates the characteristic sustain timbre. A four-step segmentation approach is implemented to preprocess sample data for FFT analysis:

1. Identification of the longest sustain interval for which minimum amplitude was greater than 75% of the maximum,
2. Phase vocoder analysis frames for the interval are averaged and one frame is identified which has least-mean-square error from the average.
3. A segment five periods in length (determined by expected fundamental pitch) is extracted from about the frame mid-point. Estimation of exact pitch is then performed via autocorrelation.
4. With the pitch estimate, four periods of the waveform (starting with a zero crossing) are extracted from this segment and interpolated to the next longer power-of-two for input to a Hamming-windowed FFT.

Use of an oversampled FFT minimises the smearing between frequency bins caused by windowing. As exactly four periods are input into the FFT, decimation of the output by four yields the complex spectra of partials which is stored in terms of magnitude and phase. The source signal is at CD rate and is decimated by two to 22.05kHz prior to analysis meaning that information on partials above 11.025kHz is absent. Sandell (1994) justifies decimation by the fact that experimental results indicate that this upper region contains minimal energy for most of the instruments analysed for SHARC (with some

isolated exceptions such as the muted trumpet). However, this region does have major significance in the perception of transient phenomena and percussion (Masri and Bateman, 1996).

The database structure of SHARC has a simple two-level hierarchy. At the top level are 39 directory names (listed in Appendix A) formed of abbreviations of each instrument and playing style (analysed in separation, hence four entries for the cello). Within each sub-directory is a 'CONTENTS' file which has a row for each analysed note and columns listing comprehensive parameters from each analysis. Each row contains a filename which references an ASCII file in the same subdirectory which describes the spectral envelope of the timbre by a list, starting from the fundamental, of partial magnitudes (in dB's) and phases. Magnitude is relative to the highest energy partial which is normalised to 0dB. The format of the CONTENTS file is standardised and facilitates automated database searching, which is exploited in the next section.

6.3 Allocation Simulation for Orchestral Synthesis

6.3.1 Overview of Experiment

The first stage of benchmarking is to derive an estimate of expected allocation statistics from orchestral music in the form of a histogram. Fig. 6.1 illustrates how such an allocation simulation divides into three sections; (1) SHARC pre-processing via FBIN.C, (2) representation of orchestration and score, and (3) the simulation program ALLOCSIM.C itself. These are described, respectively, in sections 6.3.2 to 6.3.4. Results are presented in section 6.3.5. The entire experiment was coded in 'C' under UNIX. The histogram quantifies the mean number of oscillators required per time unit during the piece of music *versus* an ascending series of frequency bins spanning the audio spectrum, numbering n_{bins}. An estimate is thus provided of the computation required for AS as a function of frequency band. Fine frequency resolution is desirable so that the histogram may be accurately superimposed upon a QMF subband hierarchy which includes logically-excluded deadbands. Therefore n_{bins}=256 was chosen as a compromise between computational complexity (as data is analysed in a spreadsheet) and simulation

accuracy. The upper frequency limit of the histogram is determined by SHARC at 11.025kHz.

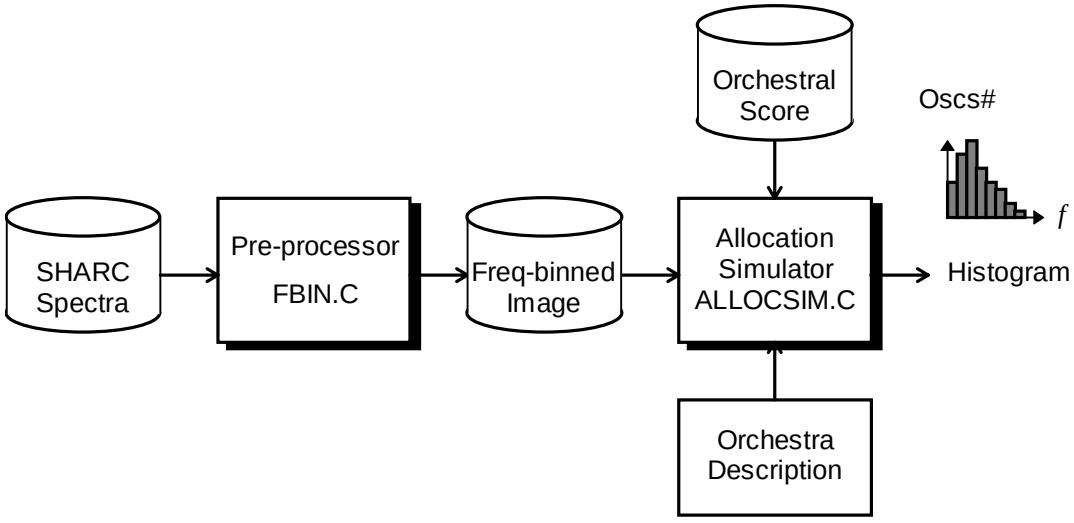

Figure 6.1 Dataflow of Allocation Simulation

6.3.2 Preprocessing of SHARC Data

The data record format of SHARC is not suitable for immediate input into ALLOCSIM.C because spectra are organised on the basis of an equal-tempered musical scale. Therefore a pre-processor FBIN.C was written to translate SHARC into 'image' with an integer-series of n_{bins} bins that matches the format of the required histogram. The image has the same fundamental directory structure as the source database, but with the difference that each note file comprises a list of n_{bins} values denoting the number of oscillators required per bin. A problem is that an arbitrary note has a theoretically infinite partial series and to include all those below the Nyquist limit of $f_s/2$ would generate an equal number of oscillators per bin: computation would be dependent on note pitch rather than timbre rendering the exploitation of SHARC redundant. Some method is required, based upon a perceptual criterion, for optimising the number of partials actually synthesised to the timbre in question.

In traditional AS, the cost of computation grows linearly with the number of partials as summarised in eqn (1.2). However, many will have insufficient power to add to the

subjective tone quality constituting redundant computation. An optimal partial set is generated by sorting partials in terms of perceptual significance and sequentially including them into the set until an acceptable level of synthesis quality is reached. This requires imposition of an auditory model which is a non-trivial task in the context of ensemble synthesis (c.f. section 2.3). A simpler 'pruning' technique was adopted in FBIN.C. Partials are sorted in terms of amplitude and included into the set until the power ratio of included to excluded partials exceeds a quality threshold ε, which may be interpreted as an S/N measure with the excluded residual signal as 'noise'. Given a series of partial amplitudes A_i:{$1 \le i \le N$}, the RMS power P of the resulting time-domain signal is given by eqn. (6.1) which is used as the pruning criterion.

$$P = \frac{\sqrt{\sum_{i=1}^{N} A_i^2}}{\sqrt{2}} \tag{6.1}$$

The resulting image of each SHARC timbre is a function of spectral envelope, note pitch and also ε. As ε is a variable in the simulation, a number of SHARC image databases are compiled over a pertinent range of ε in order to observe how MAS allocation efficiency is affected as a function of synthesis quality. In the simulation a range of $\varepsilon \in \{40,60,80\}$dB was chosen to illustrate this point. As $\varepsilon \to \infty$, few partials are excluded and therefore the impact of timbre information in SHARC has little effect on the histogram as explained previously. Also, there will be redundancy in the computation of partials which, individually, contribute relatively little to the overall quality of the perceived timbre: for example, those with low energy partials in the top octave from 10kHz to 20kHz. However at $\varepsilon \cong 0$dB, few partials are included and synthesis quality will be unacceptably low. These observations imply that there is an optimal value of ε that provides the best histogram for benchmarking. In conclusion, it is pertinent to emphasise that the application of eqn. (6.1) is empirical, but that the experimental results (Figs. 6.3 and 6.4) corroborate with our intuition in that increasing the synthesis quality (ε) causes an upward sweep in the amount of AS resources dedicated to the region above 5kHz which can be interpreted as increasing 'sparkle'.

6.3.3 Organisation of Score and Orchestration Data

The source chosen for score data were oversize ‘full’ scores from the Music Section of the University of Durham Library. These list the orchestral parts vertically and are read in a single pass across each page, representing the bar-by-bar activity of parts in parallel. This data requires translation into a machine readable format for use by ALLOCSIM.C. MIDI files are a readily available resource but are limited by the functionality of target hardware (e.g. PC sound card) and the range available over the INTERNET was considered unsatisfactory. For instance, MIDI has only 16 channels and ALLOCSIM.C uses a minimum of 26 parts. Manual score entry gave greater control over the origin and complexity of the music input into the allocation simulator.

To this end, a graphical editor program was written in MS QBASIC on a PC that permitted an intuitive, fast and visually verifiable means of entering score data bar-by-bar. The output ‘part’ file is ASCII and consists of a list of triplets describing {*note start time in beats from start of piece, MIDI pitch, note duration in beats*}. Polyphony is represented by permitting notes on consecutive lines to have the same start time; a feature which is required, for instance, when a violin section which is usually monophonic becomes biphonic. Though the rate of score entry was initially slow, after refinements to the editor and accustomisation to the routine, progress accelerated vindicating the idea.

Sir Edward Elgar’s Variations on an Original Theme: Enigma (Op. 36) was chosen as the source score for a number of reasons (Elgar, 1899). Each variation is relatively short and thus the burden of score entry was lightened: a full symphony is a task of questionable utility. Also, there is a wide range of orchestration density between variations, permitting a tentative analysis of the relationship of histogram to style. Therefore two were chosen which use full orchestral resources at some point and represent contrasting extremes of expression; (1) No. 1 ‘C.A.E.’ - ‘light’, and (2) No. 9 ‘Nimrod’ - ‘intense’. Appendix B tabulates the parts required, to which SHARC timbre they relate, and the number of voices chosen to make up each part. The Late Romantic orchestration is typical for the date of composition (Westrup and Harrison, 1959). A match is found for all parts except the timpani.

An important point is raised here point about how MAS is used to build up ensemble textures. For instance, SHARC includes an analysis of a violin ensemble. In reality, individual violins in a section are never precisely in tune and hence the ensemble sound cannot be truly described by a single harmonic partial series. It is the interaction of many voices in imperfect consonance distributed over an acoustic space that generates the characteristic rich texture (Meyer, 1993). MAS facilitates polyphony in the example violin section by reducing the cost of synthesising many instantiations of a single prototype. Also, the control problem cited in section 1.1 is reduced by exploiting hierarchy in the application structure. Instead of *n* independent instantiations, *n* variants of the prototype are initialised and related to the latter by intuitive metaparameters e.g. a detuning control (Jaffe, 1995).

To recreate the spatial distribution of individual voices is more difficult because the computational efficiency of MAS (explained in section 3.1.3) relies on the fact that the number of voices exceeds the number of filterbanks, requiring that voices are grouped and post-processed as a single composite signal. Even if independent voice synthesis were feasible, the overhead of supporting reverberant post-processing for each voice is extravagant. In the context of an orchestra, the instruments of a section are located closely together and the variation in impulse response between players to a distant listener will not vary significantly. Group post-processing of the section is therefore likely to be of acceptable fidelity, though sub-division of a large section with a set of filterbanks is an optional refinement.

6.3.4 ALLOCSIM.C Algorithm

Initialise histogram array h[1..*end of piece*,1..n_{bins}] to zeros

For each part *p*

Read in SHARC image *b*[*min..max register*,1..*nbins*] relating to *p*

Set of currently sounding notes *N*={}

For each time instant *t*=1..*end of piece*

Initialise new notes in *N* from part file of *p* starting at *t*

Remove any notes in N terminating at t

For each $n \in N$

$h[t,1..n_{bins}] = h[t,1..n_{bins}] + (b[n,1..n_{bins}] * (\text{voices in } p))$

Compute required statistics from h

ALLOCSIM.C is based upon a simple sequencing algorithm which is summarised in the above pseudocode. It takes as arguments ε (to address the specified SHARC image) and a directory containing the part files of the specified piece. At each iteration of the main loop ALLOCSIM reads a new part file (determined by the orchestration of Appendix B) and models the set of notes sounding at each time instant (quantised to semiquavers for the Enigma Variations). A histogram associated with each instant is incremented with the predicted frequency bin allocation of oscillators for that part so that histogram evolution throughout the piece is recorded. This enables the post-computation of other statistics than a mean histogram, if so desired. A problem encountered in execution was that some timbres are missing from SHARC resulting in an incomplete compass for some of the parts. This was resolved by locating the nearest note at frequency f_n to the desired note at f_d which has a SHARC timbre and copying its image. To take into account the pitch difference, each bin of the desired note is multiplied by f_n / f_d.

6.3.5 Analysis of the Allocation Histograms

The results of the simulation for the two pieces are plotted in Figs. 6.3 and 6.4 in section 6.6 in three-dimensions to express the fact that the number of oscillators required is a function of both frequency and ε. Though n_{bins} is set to 256 in the simulation, the resulting histograms are 'noisy' and are therefore smoothed in the graphs by using 16 major bins, each comprising the sum of 16 minor bins. This may be compared with the frequency quantisation of a subband hierarchy of depth K=3 which has 4 subbands spanning DC-10kHz, which is too coarse for illustrating the trends involved.

The first observation is that though the two pieces are different in character, they display a marked degree of correlation between their respective histograms. The only significant difference is that the allocation density in 'Nimrod' is higher than in 'C.A.E.' as expected because the former has greater orchestration density than the latter. In the limit at

ε=80dB, ‘Nimrod’ requires ×1.48 the resources of ‘C.A.E.’. The correlation is attributable to the common orchestration, the pruning artefacts of SHARC, and to a much lesser extent, by the shared provenance of the pieces.

At low values of ε (e.g. 20dB), where only a few high amplitude partials are included in the synthesis set, the related histogram displays a steep roll-off showing that synthesis resources are concentrated in the lower frequencies below major bin 8 (approximately 5kHz). Above this threshold, negligible resources are required. As ε is increased towards 80dB, the rolloff becomes progressively shallower until it is almost flat. This is because more low amplitude high-frequency partials are included into the set, constituting increasing ‘detail’ in the treble region above 5kHz and therefore a higher quality of synthesis, thus confirming the observations of section 6.3.2.

At this point, it is appropriate to discuss two caveats which impinge upon the results subsequently presented. The first is that only stationary spectra are considered and that the synthesis overheads of transients, which are nonstationary and perceptually important during note attack, are absent from the histograms. Essentially, transients are short noise bursts with a flattish spectrum implying increased energy at higher frequencies relative to the ‘low-pass’ nature of SHARC spectra: the requirement for substantial AS resources above 5kHz is indicated. This could be realised by setting $\varepsilon \rightarrow \infty$. However, eqn. (6.1) is based upon assumptions of stationarity. One solution (c.f. section 3.3) is the generation of broad-band energy through AM or FM (Jansen, 1991), (Fitz and Haken, 1995). Hypothetically, the AS resources required are relatively low as only a few nonstationary oscillators, selectively placed, are needed to span the whole audio spectrum in contrast to the scenario where $\varepsilon \rightarrow \infty$.

The second problem concerns the validity of the pruning algorithm of eqn. (6.1): the intuitive assumption that a partial’s energy relates to its perceptual significance is complicated by masking effects. However, the validity of imposing an auditory model upon an individual timbre is relevant only when the note is heard in isolation. For orchestral synthesis, it should be imposed upon the complete oscillator set for the orchestra and this philosophy is eschewed for the reasons set out in section 6.1. In these circumstances therefore, the chosen algorithm, despite its empiricism, appears quite

justifiable and is supported by the evidence of histogram behaviour as a function of ε in Figs. 6.3 and 6.4. It should be emphasised that the experimental methodology of this chapter represents a 'first-cut' investigation into AS resource allocation, for which there is a lack of published work and yet a good potential for substantive research, and it is to be hoped that more advanced analytical methods could be developed.

6.4 Benchmark Calculation

6.4.1 Quantifying MAS Overheads

A problem with eqn. (3.2) for predicting a benchmark is that values must be attributed to u_{mas}, u_{as} and $v(K)$ in the absence of a full MAS implementation. How are reliable values to be estimated and according to what standard are they to be measured? The commonest measure for quantifying the cost of numerical algorithms is the number of multiplies required for a given task. For a linear signal processing system such as a filterbank, the frequency of multiply / accumulate (MAC) operations required is a highly suitable measure. However, the overheads of oscillator bank computation are dominated by the computation of $\sin(\Phi[n])$ - a non-linear calculation - and in managing the control data bandwidth of PWL envelope uncompression. A single MAC alone is required per output sample for imposition of $A_i[n]$ and for the accumulation of S sinusoids in eqn. (1.1).

The cost of transforming an oscillator bank from an AS to MAS paradigm is minimised by the binary-tree subband decomposition of the filterbank proposals of Chapters 4 and 5 which have a small set of K+1 sample rates related by powers-of-two integer ratios. As explained in Chapter 9, the imposition of MAS requires multirate scheduling to support multiple oscillator sample rates, but the complexity increase over that which is required for a TOB is shown to be of not overwhelming significance: a latency penalty is the most important side-effect. As proposed in section 1.4.3, MAS seeks to preserve the optimality inherent in the TOB model and therefore it is considered reasonable to assume that $u_{mas} \cong u_{as}$. The principal computational overhead of MAS is represented by filterbank computation.

$$E1 = \frac{\text{net number of AS oscillator updates per second}}{\text{net number of MAS oscillator updates per second}} \tag{6.2}$$

$$E2 = \frac{\text{net number of MAC's in AS per second}}{\text{net number of MAC's in MAS per second}} \tag{6.3}$$

To avoid the eventuality of incommensurate cost measures leading to an unreliable value of E, it was decided to decompose eqn. (3.1) into two benchmarks, $E1$ and $E2$, summarised in eqns. (6.2) and (6.3). $E1$ measures the performance increase caused by the net reduction of oscillator sample rates in MAS compared to AS based on the equality $u_{mas} \cong u_{as}$, but excludes filterbank computation. In contrast, $E2$ measures the performance increase caused by the net reduction in the number of MAC's required per second by MAS compared to AS: the denominator includes those required by filterbanks and a MAS oscillator bank. The justification for abstract benchmarks arises from the fact that the real cost measure, as discussed in section 3.1.7, is implementation-dependent (e.g. CPU time or VLSI area): $E1$ and $E2$ provide useful figures which will be a dominant in such a measure and can be evaluated without implementation knowledge. $E1$ is chosen deliberately to illustrate a strong optimisation facet of MAS whereas $E2$ places a "devil's advocate" emphasis on the overheads of filterbank computation in MAS in terms of a standard measure of DSP complexity.

6.4.2 Filterbanks: Quantity, Topology and Cost

Before $E2$ can be evaluated, values must be attributed to the number of filterbanks n_{fb}, and $v(K)$. Appendix B describes a hypothetical allocation of the orchestral instruments required by Enigma Variations to a total of n_{fb}=8 filterbanks. Families of instruments (e.g. woodwind) and sections of identical instruments are allocated to individual filterbanks. This illustrates how MAS exploits hierarchy in the synthesis application by factoring the large number of outputs from S individual oscillators in AS into a small number of logically related streams. In the case of an orchestra, an intuitive grouping is by locality in an acoustic volume. A separate 'effects' process is envisaged in Chapter 8 which locates each filterbank stream within a virtual acoustic space in relation to the listener who requires a multichannel signal for perception of the synthetic music.

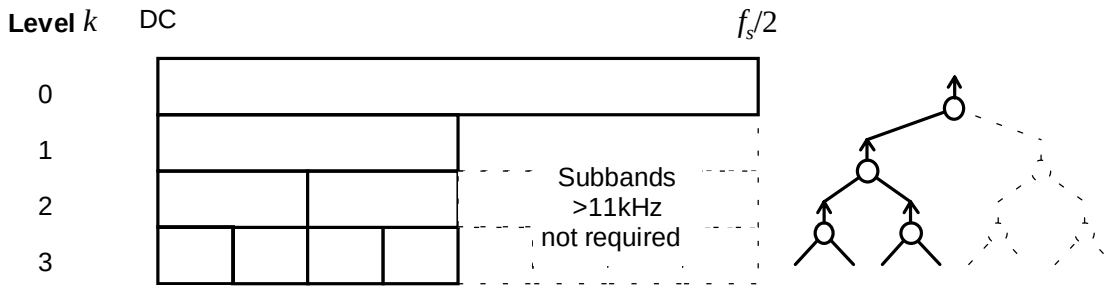

Figure 6.2 Subband Hierarchy and Filterbank Topology Used in Simulation

Each filterbank is assumed to be an incomplete binary tree of the form of Fig. 6.2. This is because SHARC has no information above 11kHz, or $f_s/4$ for a QMF filterbank executing at a nominal CD sample rate of 44.1kHz, and therefore interpolation resources for this upper octave are redundant. Note that it is still feasible to allocate partials with frequencies up to the Nyquist limit of $f_s/2$ to the computationally most expensive subband $s_{0,1}$ for e.g. the synthesis of broadband transients. Such a scenario illustrates the utility of configuring filterbank topology to the partial frequency distribution, rather than imposing complete binary trees. In the PEF filterbank simulation, partials above 10kHz are allocated to $s_{0,1}$ because the upper bound of the subband hierarchy is normalised to 20kHz (see section 5.3.2).

Benchmarking is complicated by the assumption that fixed pitch notes are employed. If this were strictly the case, then all oscillators could be allocated into the terminal integer-series $s_{3,1..8}$ of the proposed subband hierarchy. Players in a real orchestra introduce - as a matter of course - note pitch dynamics, to add expression and texture to the music. Vibrato is the commonest phenomena. However, to quantify the amount required by each instrument in the simulation is highly subjective. A compromise is to introduce a pitch modulation constant of $\pm\tau$ semitones that is applied to all instruments of the orchestra that reflects their mean requirement for pitch dynamics, as documented theoretically in section 3.2.3. Instruments in a real orchestra will span a range of pitch stationarities and τ is thus a rather coarse modelling parameter. Nevertheless, τ allows us to investigate $E1$ and $E2$ as a function of subband allocation pattern.

$$\nu = \nu_s \sum_{k=0}^{K} 2^{-k} t_k \qquad (6.4)$$

Filterbank cost v is evaluated in eqn. (6.4) where v_s is the number of MAC's required per output sample in a prototype stage and t_k:$\{0 \le k \le K\}$ represents the number of filterbank stages required to generate the subband set at each level k in the specified topology. The 2^{-k} term represents the relative decimation between filterbank output and level k. For Fig. 6.2, $t_{0..K}$=(0,1,1,2) for a QMF and $t_{0..K}$=(0.5,0.5,1,2) for a PEF. t_0=0 for a QMF signifies that $s_{0,1}$ is the output fullband and no interpolation is required and t_0=0.5 for a PEF signifies that complex-to-real conversion is necessary which is constituted by half of a PEF stage (in practice, as only a real output is needed, the final $+\omega_h$ frequency shift resolves to two MAC's rather than four). Also, t_1=1 for a QMF signifies that a complete stage is necessary to generate $s_{1,1}$ ($s_{1,2}$ is available by default but unused) and t_1=0.5 for a PEF signifies that a half stage can be used because interpolation of the $H0$ and $H1$ subbands is independent (see section 5.3.2).

Therefore, all three of the filterbanks discussed in Chapters 3 and 4 are benchmarked to evaluate their relative performance. $v_s=4c+8$ for the PEF stage of section 3.2 where c is the cost of a constituent FIR: using the suggested PM-FIR stage design of section 5.3.3. and then the optimal FIR structure of section 3.2.1, c=6 and therefore v_s=32. For a PM-FIR QMF stage, v_s is calculated via eqns. (4.4) to (4.6) and is a function Δ_f: a range of Δ_f must therefore be benchmarked to determine an optimum value where filterbank computation is balanced with the effect of logical deadband exclusion. For the PA-IIR stage discussed in section 4.4, v_s=6.5 with a negligible $\Delta_f \cong 0.01$. From these values of v_s, Table 6-1 records v as determined via eqn. (6.4) with $\delta_s=\delta_p$=-80dB for the PM-FIR.

Filterbank	***v***
PM-FIR QMF	$\cong 1.152/\Delta_f$
PA-IIR QMF	11
PEF	38

Table 6-1 Filterbank Cost v

6.4.3 Applying Allocation Simulation Data

Interpreting the histograms of Figs. 6.3 and 6.4 reveals that ε=20dB and ε=80dB represent, respectively, poor quality synthesis and an asymptotic level of fidelity with computational overkill. The best compromise lies between, but given that the 'optimal' value of ε will remain obscure without the evidence of listening tests, the histograms of ε=40dB and ε=60dB are selected to provide realistic upper and lower bounds to the region where the optimal histogram is likely to be found. A convenient framework for processing the histogram data is as a spreadsheet as integration of all other simulation parameters, such as logical deadband exclusion, is facilitated.

A histogram is represented as a row vector **h** comprising n_{bins} elements spanning DC to 11.025kHz. This is mapped on to an allocation pattern **a** (using the specified filterbank topology **t** of Fig. 6.2) which is also a row vector of size n_{bins} with identical frequency span denoting which level k:{$0 \le k \le K$} each bin of **h** is mapped to in the subband hierarchy. Each bin of **a** has, rather, the weighting 2^{-k} denoting the decimation caused by allocation to level k in comparison to classical AS in subband $s_{0,1}$. For a PM-FIR QMF implementation **a** is a function of Δ_f, which determines the impact of logical deadband exclusion. τ determines the Q of a partial series and whether **a** tends towards the terminal integer-spaced series $s_{3,1..8}$ (high τ and Q) or the fully-overlapping octave-spaced series $s_{0..K,1}$ (low τ and Q as illustrated theoretically in section 3.2.3).

$$E1 = \frac{f_{as} \sum_{i=1}^{n_{bins}} \mathbf{h}_i}{f_{mas} \mathbf{a}(n_{bins}, \mathbf{t}, \Delta f, \tau, f_h) \mathbf{h}^T} \tag{6.5}$$

$$E2 = \frac{f_{as} \sum_{i=1}^{n_{bins}} \mathbf{h}_i}{f_{mas} (n_{fb} v + \mathbf{a}(n_{bins}, \mathbf{t}, \Delta f, \tau, f_h) \mathbf{h}^T)} \tag{6.6}$$

$E1$ is expressed in eqn. (6.5) where f_h denotes the upper frequency limit of **a** (f_h =22.05kHz for a QMF, f_h =20kHz for a PEF) and is used to normalise bin widths in **a** to

those of **h**. Account must be taken of output sample rates. f_{as} relates to classical AS at f_{as}=44.1kHz whereas f_{mas} relates to MAS and is set at f_{mas}=44.1kHz for critically-sampled QMF filterbanks and f_{mas}=50kHz for the oversampled PEF (see section 5.3.3) and thus the cost increase of oversampling in a PEF is incorporated. The assumption that $u_{mas} \cong u_{as}$ holds for complex signals in a PEF because, though oscillator sample rate is halved relative to single-phase representation, the cost of an update is approximately doubled (see Chapter 7). $E2$ as expressed in eqn. (6.6) is based on eqn. (6.5), but includes filterbank costs in the denominator where n_{fb}=8 from section 6.4.2 and v is dependent on the specific filterbank from Table 6-1.

6.4.4 Analysis of Benchmarking Results

The benchmark results are plotted in section 6.6. Figs. 6.5 and 6.6 display the $E1$ benchmarks for 'C.A.E.' and 'Nimrod' respectively, versus filterbank type and τ (the pitch variation tolerance) for both the ε=40dB and ε=60dB histograms: 'PA' refers to the PA-IIR QMF, and PM-FIR QMF Δ_f is shown for four representative values $\Delta f \in \{0.05, 0.1, 0.15, 0.2\}$. Both graphs are approximately identical because (i) their respective histograms are highly correlated and (ii) filterbank computation is excluded. The PA-IIR gives the best results as the filterbank is critically sampled and provides a near-ideal subband hierarchy. The PEF is a close second because it has an ideal subband hierarchy but is oversampled. For the PM-FIR, it can be seen how $E1$ falls off rapidly with increasing Δ_f due to logical deadband exclusion. The deleterious influence of deadbands increases towards the middle of the hierarchy (on the frequency axis). This why $E1$ for the PM-FIR is lower for the ε=60dB histogram (which has a higher density of partials in this region) than that for ε=40dB, particularly at high values of Δ_f. The effect of τ is to reduce $E1$: the falloff is quite steep for filterbanks close to the ideal subband hierarchy, but shallow for the PM-FIR with large Δ_f because logical deadband exclusion already promotes many partials into higher subbands and thus the extra proportion promoted by large τ is small.

Figs. 6.7 and 6.8 display the $E2$ benchmarks in the same format as those for $E1$. The inclusion of filterbank computation into $E2$ illustrates the economy of scale exploited by MAS in that $E2$ for 'Nimrod' is consistently higher than 'C.A.E.' which has less

orchestration density. The PA-IIR has pre-eminent performance due to its low-cost and near-ideal subband hierarchy, but also has the steepest rolloff of $E2$ against τ for the reasons previously discussed. The PEF has worst performance due to the expense of oversampling and complex-signal representation. However, $E2$ is greater than unity for the PEF, even at high values of τ, indicating that net savings in MAC computation are made. The PM-FIR has a performance intermediate between the PA-IIR and PEF and displays the interesting characteristic of a global maximum of $E2$ versus Δ_f. An explanation is that at low values of Δ_f, filterbank cost (and latency) are excessively high, though a near-ideal subband hierarchy is obtained. Conversely, at high values of Δf, filterbank cost is low but the subband hierarchy suffers severe deterioration due to logical deadband exclusion. The value of Δ_f for which $E2$ is a maximum increases with τ, for the reason that a quasi-ideal subband hierarchy is not required for high τ where most partials are promoted anyway. A value in the region $0.1 \leq \Delta_f \leq 0.15$ appears to bound max($E2$). This factor, in combination with the observations of section 5.2.2 are strong evidence for the optimality of K=3.

6.5 Review

The theory of MAS economics advanced in section 3.1.7 is supported. If taken *cum grano salis*, then the benchmarks predicted in this chapter may be expected to reflect those gained by a prototype MAS implementation and extensive listening tests, which is outside the scope of this thesis. $E1$ is an 'optimistic' measure concerned solely with throughput optimisation in a TOB. $E2$ is a 'pessimistic' measure which is weighted towards filterbank computation and includes only the single output MAC operation of a TOB. However, both display consistent values significantly greater than unity supporting the superior optimality of MAS over AS. Performance increase is largely determined by filterbank choice and can be ranked in ascending order as (1) PEF, (2) PM-FIR and (3) PA-IIR which, in descending order, reflects functional transparency as determined in Chapters 3 and 4, thus illustrating the tradeoffs inherent to MAS.

6.6 Benchmark Results

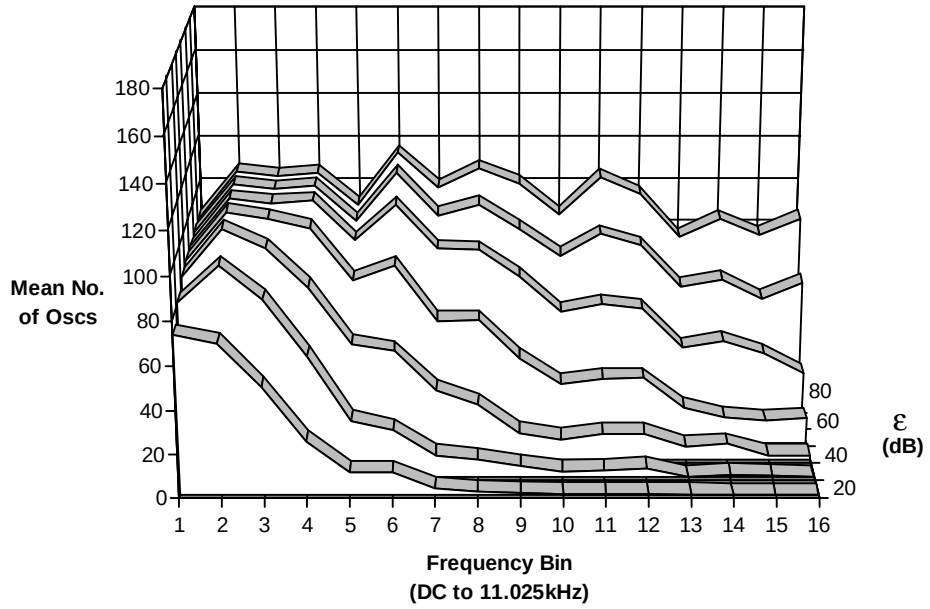

Figure 6.3 'C.A.E.' Allocation Histograms

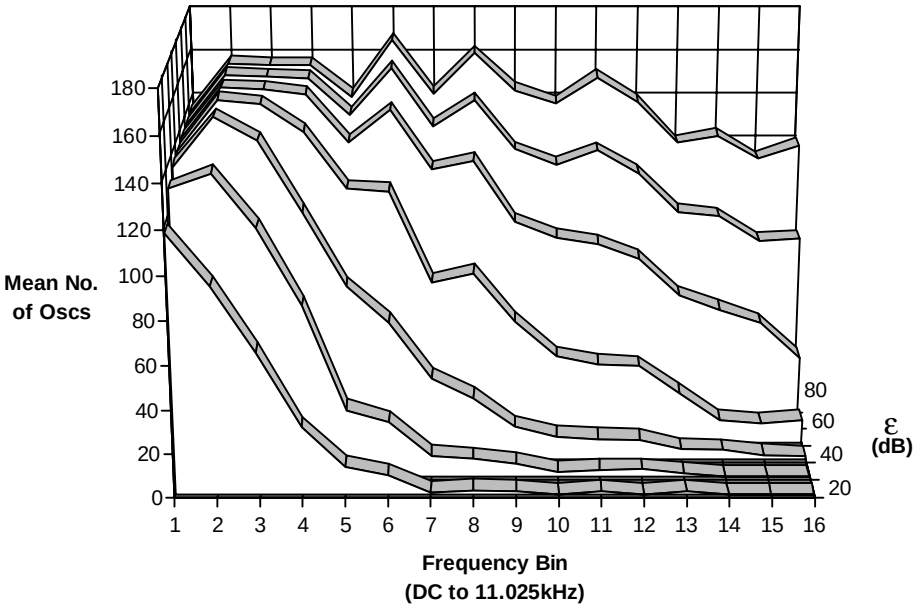

Figure 6.4 'Nimrod' Allocation Histograms

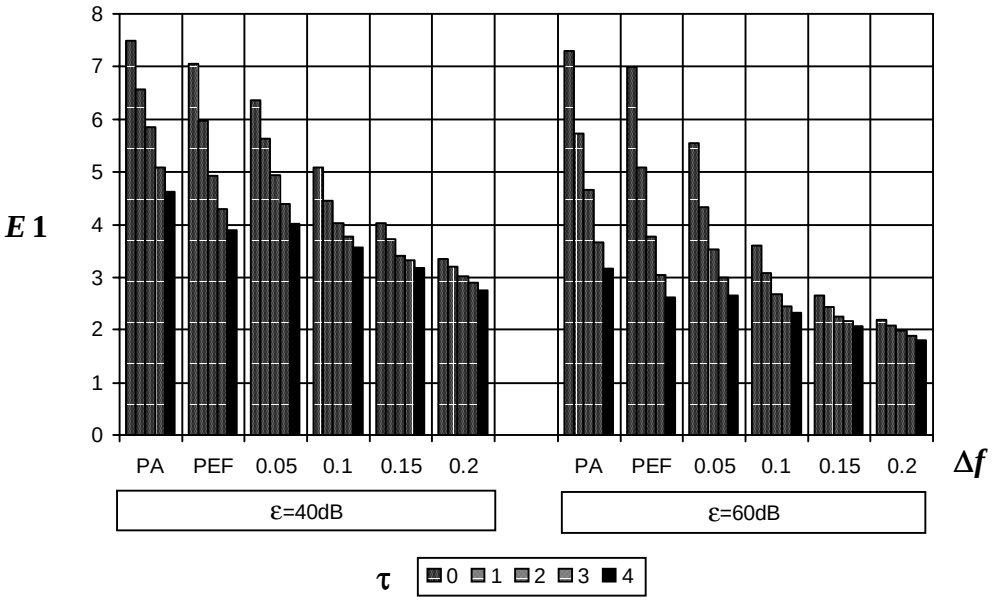

Figure 6.5 E1 for 'C.A.E.'

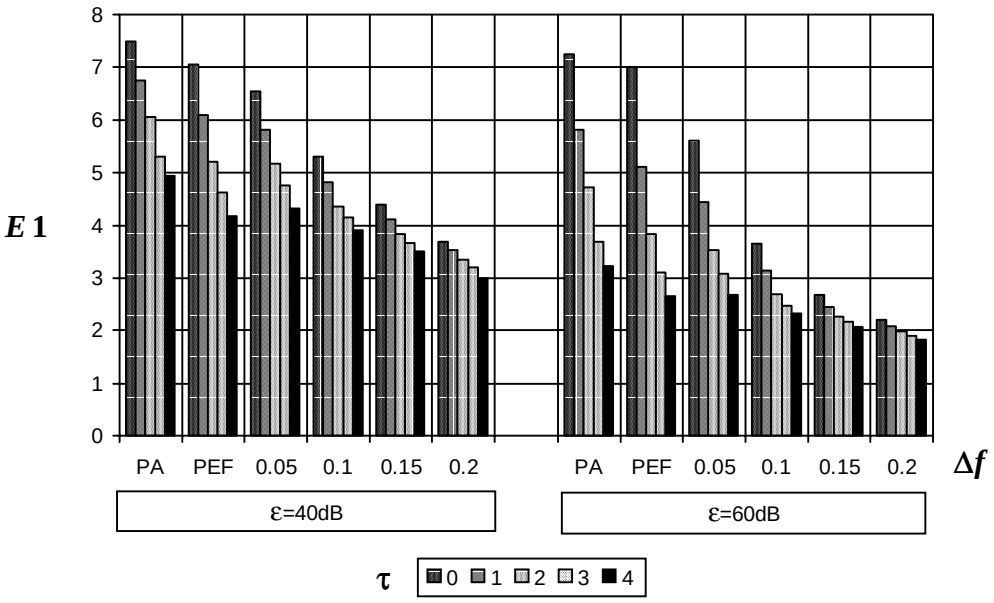

Figure 6.6 E1 for 'Nimrod'

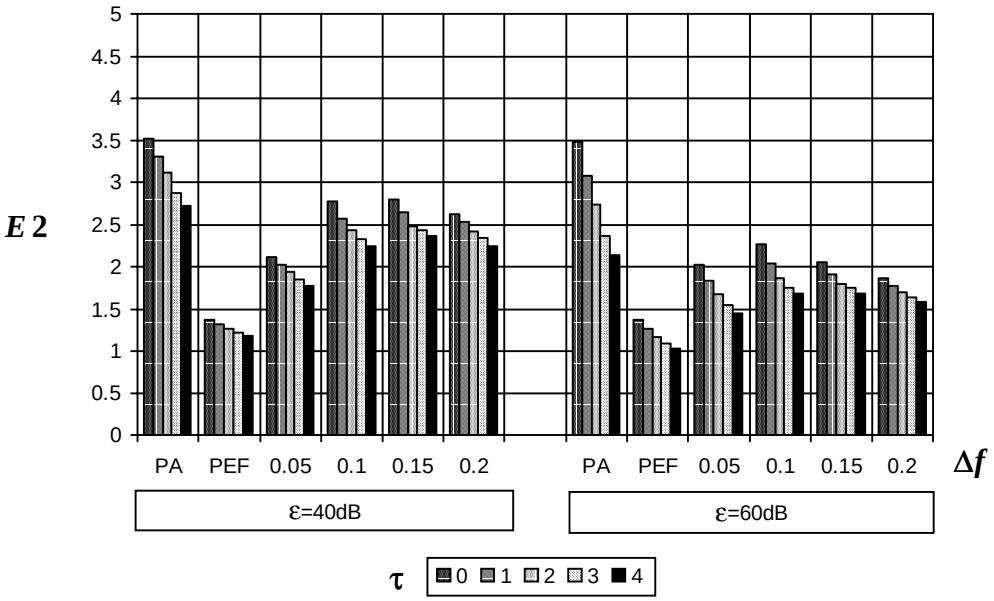

Figure 6.7 E2 for 'C.A.E.'

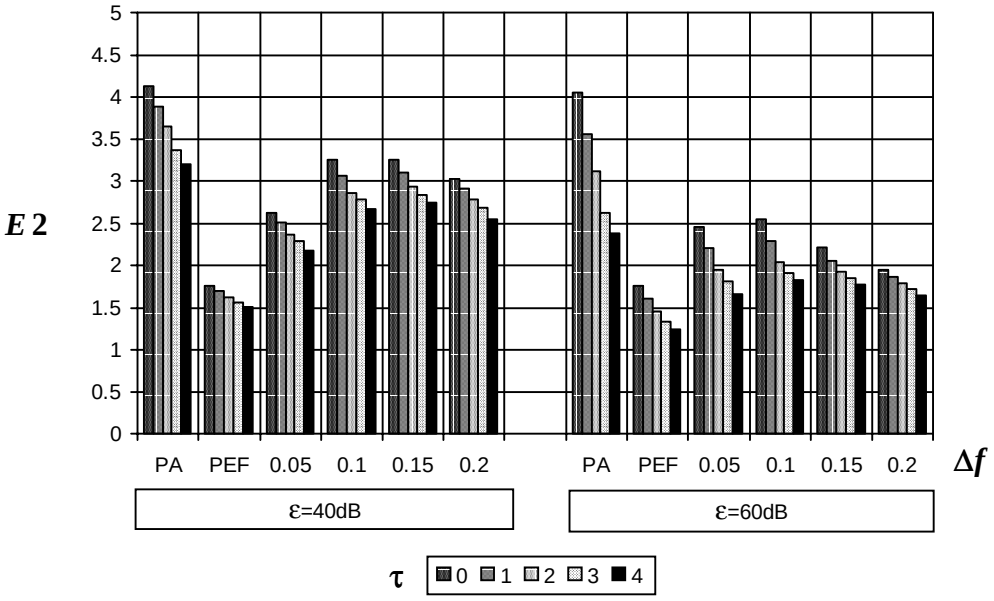

Figure 6.8 E2 for 'Nimrod'