**Room Impulse Response Mixing Time using Fractal Analysis**

Thomas J. Conatya and Dermot J. Furlong

*Music and Media Technologies, Department of Electronic and Electrical Engineering, Trinity College, Dublin, Ireland*

Hybrid artificial reverberation relies on an effective combination of early reflections realized through convolution with a room impulse response, and a reverberation tail created using a filterbank algorithm. In order to optimize the use of both types of reverberation realization, the point at which a room impulse response transitions from early to late reflections (the mixing time) must be identified. Fractal dimension analysis is used to produce fractal dimension profiles for eight room impulse responses. The resulting fractal dimension profiles were analyzed in terms of the length of time taken for the profile to reach a threshold determined by four different criteria, and the resulting values were compared to results from a study of the perceptual evaluation of mixing time. Based on this analysis an equation calculating the appropriate mixing time for use in hybrid artificial reverberation was established. This equation produced an explained variance of 93.49% for the results from a perceptual evaluation of mixing time, where previous model predictors of mixing time had only achieved 81.5%. The results achieved in this study suggest that analysis of the fractal dimension profile of room impulse responses offers a method for calculating the appropriate mixing time for hybrid artificial reverberation that is both simple and accurate.

**FINAL version**

**D.F.**

1. **INTRODUCTION**

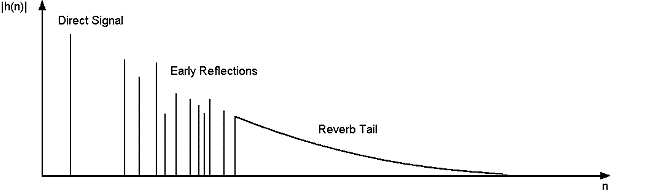
Any room may be characterized by its impulse response between a chosen source and receiver location. From the perspective of perceptual significance, such impulse responses may be divided into three physical stages – direct sound, early reflections, and subsequent reverberation. In general terms, these may be defined as follows: The direct sound constitutes those signal components which arrive at the receiver location from the source without any intervening reflection. The early reflections are indirect sound components which arrive at the receiver location after reflection within a short time of the direct sound. The subsequent reverberation is characterized as the statistically diffuse reflections which arrive with sufficient density as to be individually irresolvable, which condition arises after what is referred to as the mixing time duration. This time, at which reflected components transition from early reflections to reverberation, depends on many room acoustical parameters. However, from a perceptual perspective, the mixing time is an important feature of any room response. Early reflections and reverberation have identifiable, but different, perceptual characteristics, such that it is often important to be able to precisely define the mixing time. Combined with this is the fact that realization of reverberation using a convolution approach is not typically sufficiently efficient to allow real-time computation of complete reverberant responses. Consequently, accurate specification of mixing time would allow convolution realisation of early reflections, with an accompanying synthetic generation of the subsequent reverberation tail.

This has become increasingly important for a variety of research concerns, as well as for applications relating to special effects in the film industry, and in the ever more sophisticated video games industry. The design of Virtual Acoustic Environments (VAE) of increasing sophistication is thus currently being implemented in conjunction with 3D visual environments. As these 3D environments increase in visual sophistication, there is an attendant demand to provide an equally compelling acoustic environment. When constructing a VAE it is thus important to consider the minimisation of the processing power required to create it. One way this can be achieved is through the use of hybrid artificial reverberation (HAR) [Primavera *et al*, 2010]. HAR seeks to exploit knowledge of room impulse response (RIR), and of the threshold of human recognition of distinct reflections within a particular acoustic environment, in order to optimize artificial reverberation computation. Generally speaking, the three parts of an RIR previously identified provide the listener with different information. The direct sound specifies information relating to the source of sound. The early reflections provide the listener with information about the geometry of the physical space in which the source is situated, and the subsequent reverberation arises after the point when the listener can no longer recognise spatial characteristics which depend on inter-reflection detail, but only registers an overall auditory smearing. The point at which the early reflections transition into reverberation can either be described in terms of the acoustical behaviour of the system, known as the mixing time (tmix), or in terms of how the auditory system perceives the behaviour of the system, which is referred to as the perceptual mixing time (tmp). HAR algorithms use convolution of a recorded or simulated RIR with an input signal to generate the early reflections, and then generates the reverberation tail using filter algorithms [Primavera *et al*, 2010; Murphy & Stewart, 2007]. The convolution of an input sound with a complete RIR produces a reproduction of a room’s reflection profile at the point of RIR measurement, but suffers from being computationally expensive. In comparison, algorithmically generated reverberation is computationally inexpensive but can produce unnatural audible artefacts which distort the perceived spatial impression. In that much environmentally specific information about room size and shape is provided by early reflections, any distortion of the early reflection pattern is to be avoided. Hybrid Artificial Reverberation therefore seeks to maximise sound quality while minimising computational load by separately addressing early reflection and subsequent reverberation simulation. To do so, a method is required such that the perceptual mixing time, tmp, can be accurately calculated.

**2. ROOM IMPULSE RESPONSE**

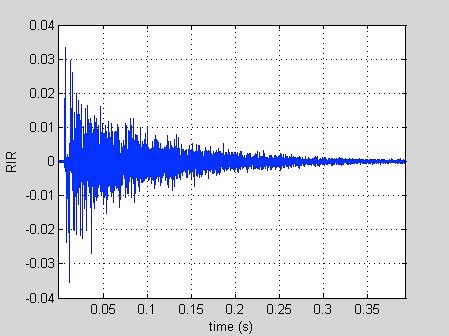
A room impulse response (RIR) describes the variation in the sound pressure level (SPL) over time at a single receiver position from the moment an impulse is input to an enclosed space. A hypothetical RIR profile is demonstrated in Figure 2, showing the SPL variation at one point in space; from the direct sound, to the early reflections, and finally the individually indistinguishable late reflections constituting the reverberation tail. The characteristics of any RIR are determined by the source and receiver positions, the absorption properties of surface materials, the geometry of the room, and the extent of acoustic diffusion due to uneven surfaces or objects in the room.

There are a number of methods available to procure a RIR of an acoustic space [Stan et al, 2002]. The operational principles of all these methods involve the input of a known and reproducible signal into an acoustic environment, and measurement of the system response at some receiver location.



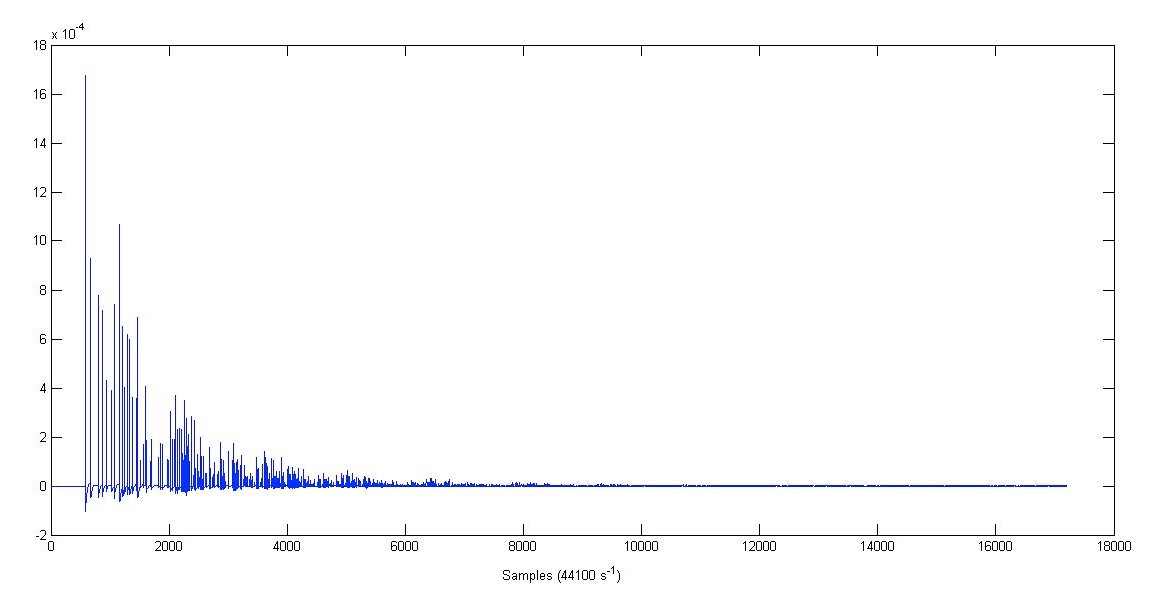
**Figure 2**: Idealized RIR showing distinct sections of direct sound, early, and late reflections.

A typical real RIR is shown in Figure 3.



**Figure 3**: A typical, real RIR

The RIR of a hypothetical room can be efficiently produced using the image-source method (ISM) [Allen & Berkley, 1979], which produces an idealised SPL profile, such as that shown in Figure 4.



**Figure 4**: Idealized RIR created using the image-source method.

ISM generated RIRs have been found to provide an adequate simulation of real room impulse responses, and have the advantage of allowing efficient computation. As such, they allow ready investigation of the mixing time for any room specification. Different definitions of mixing time have been adopted in the literature in order to accommodate different mixing time assessment techniques. It has been defined as the time required for the energy to become equidistributed in a room [Polack, 1993]. A second approach has been to adopt the time required for a sound wave to travel three times the mean free path [Blesser, 2001]. A third approach has been to adopt the time required for the reflection density to be such that10 reflections could be accommodated within the time resolution window of the auditory system [Polack, 1988]. Yet other algorithms for mixing time determination have relied on the matching pursuit numerical technique, or variants thereof [Defrance et al.,2009; Mallat & Zhang, 1993]. The effectiveness of model based RIR mixing time estimation, as compared to recorded RIRs has been examined in [Rubak & Johansen, 1999], with the results demonstrating accurate mixing time calculation comparisons. Given this, model based RIRs are used in this study to assess the effectiveness of a Fractal Dimension based estimation of acoustical response mixing time.

**3. FRACTAL ANALYSIS**

Fractal Geometry allows the accurate description of a broad range of natural processes. This mathematical vocabulary has facilitated technical advances and new levels of academic understanding in fields ranging from telecommunications to human-integrated technology. Benoit B. Mandelbrot first coined the term ‘fractal’ in 1982 to describe “*a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole*” [Mandelbrot, 1982]. The two defining aspects of a fractal are that it exhibits self-similarity at arbitrary levels of magnification, and that it can be described by a non-integer dimension whose value is greater than that of its topological dimension. Mathematically, a fractal can be generated by iterative calculation of an equation which by its nature will result in recursive self- similarity.

The dimension of a fractal is a measure of its complexity and denotes how completely the fractal fills a space. For real-world fractal systems, the fractal dimension (FD) must be estimated, as invariably the system will not be self-similar at all levels of magnification. The Higuchi Algorithm was used for estimating the FD in this investigation [Higuchi, 1988].

**4. PREDICTING THE MIXING TIME**

The methods for predicting the transition from early reflections to late reflections in a reverberant system fall into two categories. The first looks at the statistical behaviour of the reflections in the system by considering the behaviour of the room’s impulse response. These methods employ various mathematical algorithms to produce the mixing time (tmix). The second category of predictors considers how the human auditory system perceives information from reflections in an acoustic environment, and attempts to derive the perceptual mixing time (tmp) from a model that considers this factor and the characteristics of the environment. Alexander Lindau presented a comparative study of the perceptual mixing time (tmp) against the various predictors for tmix and tmp. [Lindau et al, 2010]. The results of this study are used for comparison in this paper.

**5. METHOD**

It was intended that the fractal dimension (FD) of room impulse responses (RIRs) be investigated for correlation with known values for the perceptual mixing time (tmp) for a set of rooms. Lindau’s study was chosen for investigation because not only did his study provide values for tmp for a variety of acoustic environments of known dimensions and average absorption coefficients (*αave*) from which the rooms could be modelled, but it also presented results for the gamut of mixing time predictors against which the effectiveness of the fractal dimension method could be measured.

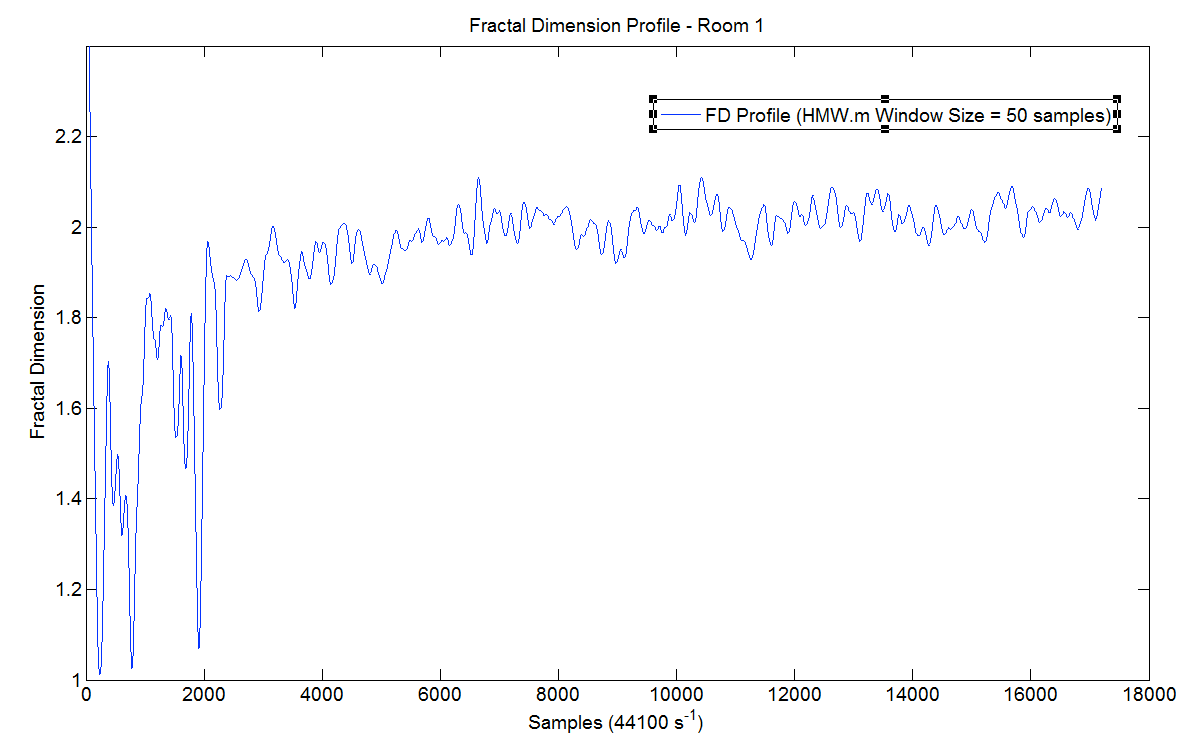
To investigate how the FD of RIRs correlates with tmp, MATLAB models for the first eight rooms from the Lindau study were created using Allen & Berkley’s image-source method (ISM).

How the FD varied over the lifetime of a RIR was of particular interest {~~for~~} to this study. To obtain this FD ‘profile’ of a RIR, a MATLAB function, *HMW.m,* was created that generated a moving window of a pre-defined sample length, which moves incrementally along the length of the RIR. As the moving window advances along the RIR, it implements the Higuchi algorithm to calculate the FD over the sample length, resulting in a profile showing how the FD of the RIR changes over its length. *HMW.m* was applied to the eight modelled RIRs from the Lindau study to produce eight FD profiles.

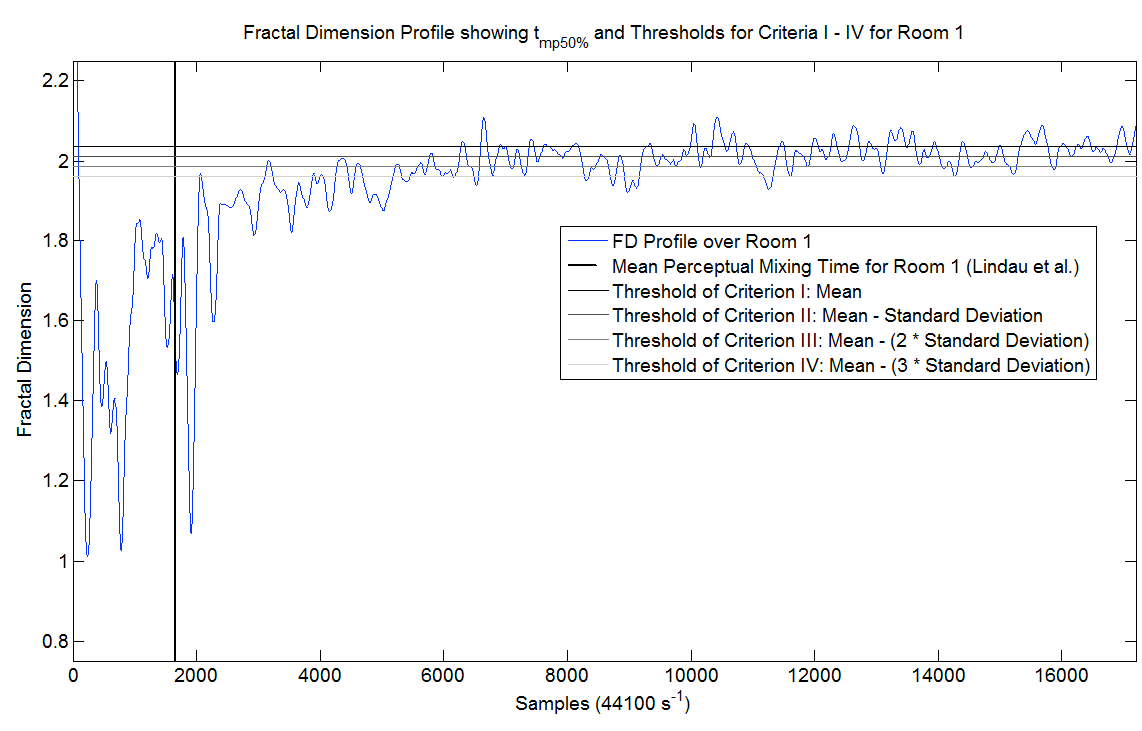
Following the production of the eight FD profiles, a mathematical analysis was carried out on each profile relating the behaviour of the profiles to the respective mean perceptual mixing times (tmp50%) from Lindau’s study. The results from this analysis were then compared against Lindau’s results for predictors of tmp50% to determine the benefit, if any, from using the fractal dimension of RIRs as a basis for predicting the perceptual mixing time.

**5.1 Fractal Dimension Profile of Room Impulse Responses**

The FD profiles for each RIR were created using the *HMW.m* function with a window length of 50 samples (~1ms). This value was chosen as it was large enough for the Higuchi algorithm to work effectively to provide a highly detailed FD profile (see Figure 5.2). Before the FD profile is plotted it is smoothed with a moving window of 200 samples to reduce noise. In Figure 5.1 the FD profile for the ISM modelled RIR for room 1 from Lindau’s study can be seen.



**Figure 5.1**: Fractal Dimension Profile from the RIR of Room 1 created using *HMW.m,* with a window size of 50 samples.



**Figure 5.2**: FD profile for Room 1 including tmp50%

and thresholds for Criteria I – IV.

**6. Analysis**

The eight FD profiles were investigated to see if there was a correlation between the shape of the FD profile and the tmp50% {~~figures~~} values from the Lindau study. It was presumed from theory that the RIR would be most diffuse towards its end. Based on this assumption, the mean and standard deviation for the last 10% of each FD profile were calculated. Once these values had been established, four criteria relating to the FD profiles were identified in order to investigate the relationship between the shape of the curves and tmp50%. Criterion I specified the sample number at which the magnitude of the FD profile first reaches the mean value calculated from the last 10% of its length. Criteria II, III & IV specified the sample number at which the magnitude of the FD profile first reaches the mean value calculated from the last 10% of its length minus 1, 2, and 3 times the standard deviation of the last 10% of its length, respectively. Figure 5.2 shows tmp50% and the thresholds for criteria I – IV on the FD profile for Room 1.

Once the various criteria thresholds had been established for each room, the corresponding sample numbers required to reach the mixing time thresholds were derived for each room using a MATLAB function. With the sample numbers obtained for each criterion, these values were plotted against tmp50% from the Lindau study, and a linear regression analysis conducted to test for correlation.

**7. Results**

The values for the mean and standard deviation for the last 10% of each room’s fractal dimension (FD) profile and the FD thresholds for the four criteria from Section 6 are presented in Table 7.1a.

From the values in Table 7.1a following, the sample numbers from the eight FD profiles for Criteria I – IV were obtained. These sample numbers, and the corresponding time (in ms) are presented in Table 7.1b.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Room**  **No.** | **Mean**  **(FD)** | **Standard Deviation (FD)** | **Threshold (Criterion I) (FD)** | **Threshold (Criterion II) (FD)** | **Threshold (Criterion**  **III) (FD)** | **Threshold (Criterion IV) (FD)** |
| **1** | 2.0364 | 0.0252 | 2.0364 | 2.0112 | 1.986 | 1.9608 |
| **2** | 2.0450 | 0.0265 | 2.0450 | 2.0185 | 1.9921 | 1.9656 |
| **3** | 2.0250 | 0.0321 | 2.0250 | 1.9927 | 1.9605 | 1.9284 |
| **4** | 1.9833 | 0.0394 | 1.9833 | 1.9439 | 1.9046 | 1.8652 |
| **5** | 1.9782 | 0.0404 | 1.9782 | 1.9377 | 1.8973 | 1.8569 |
| **6** | 1.9866 | 0.0432 | 1.9866 | 1.9435 | 1.9003 | 1.8571 |
| **7** | 1.9682 | 0.0394 | 1.9682 | 1.9288 | 1.8894 | 1.8500 |
| **8** | 1.9644 | 0.0506 | 1.9644 | 1.9138 | 1.8631 | 1.8125 |

**Table 7.1a**: Values of mean and standard deviation for the last 10% of the FD profiles of Rooms 1 – 8.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Room** | **Sample**  **Number (Criterion I)** | **Time**  **(ms) (Criterion I)** | **Sample**  **Number (Criterion II)** | **Time**  **(ms) (Criterion II)** | **Sample**  **Number (Criterion III)** | **Time**  **(ms) (Criterion III)** | **Sample**  **Number (Criterion IV)** | **Time**  **(ms) (Criterion IV)** |
| **1** | 6277 | 142.34 | 5772 | 130.88 | 3136 | 71.11 | 2053 | 46.55 |
| **2** | 4573 | 103.70 | 3197 | 72.49 | 3170 | 71.88 | 3049 | 69.14 |
| **3** | 6056 | 137.53 | 4335 | 98.30 | 2494 | 56.55 | 1567 | 35.53 |
| **4** | 13724 | 311.20 | 7027 | 160.36 | 7013 | 159.02 | 7004 | 158.82 |
| **5** | 7996 | 181.32 | 7971 | 180.75 | 7906 | 179.27 | 7661 | 173.72 |
| **6** | 12366 | 280.41 | 9486 | 215.10 | 8060 | 182.77 | 2052 | 46.53 |
| **7** | 14850 | 336.73 | 12308 | 279.09 | 11708 | 265.49 | 2192 | 49.71 |
| **8** | 16194 | 367.21 | 12571 | 285.06 | 8465 | 191.95 | 5209 | 118.12 |

**Table 7.1b**: Sample number (44100 samples s-1) from the eight FD profiles for the four threshold criteria.

**8. Correlation with Mean Perceptual Mixing Time**

Using the values in Table 7.1b, linear regression analysis was performed using MATLAB’s curve fitting toolbox for each of the four criteria, {~~and tmp50% from the Lindau study~~}. Figure 8.1 shows plots of the sample values obtained from the four criteria against tmp50%. The linear regression equations and explained variance (R2) for the four criteria against tmp50% are presented in Table 8.2. From Figure 8.1 and Table 8.1 it can be seen that Criterion I produced a correlation with tmp50%, with an explained variance (R2) of 72.29%. Criterion II produced a correlation with tmp50%, with an R2 of 78.82%. Criterion III produced the best correlation with tmp50%, with an R2 of 93.49% and a linear regression equation of the form,

tmp50% = 0.3197· III + 325 (eq. 8.1)

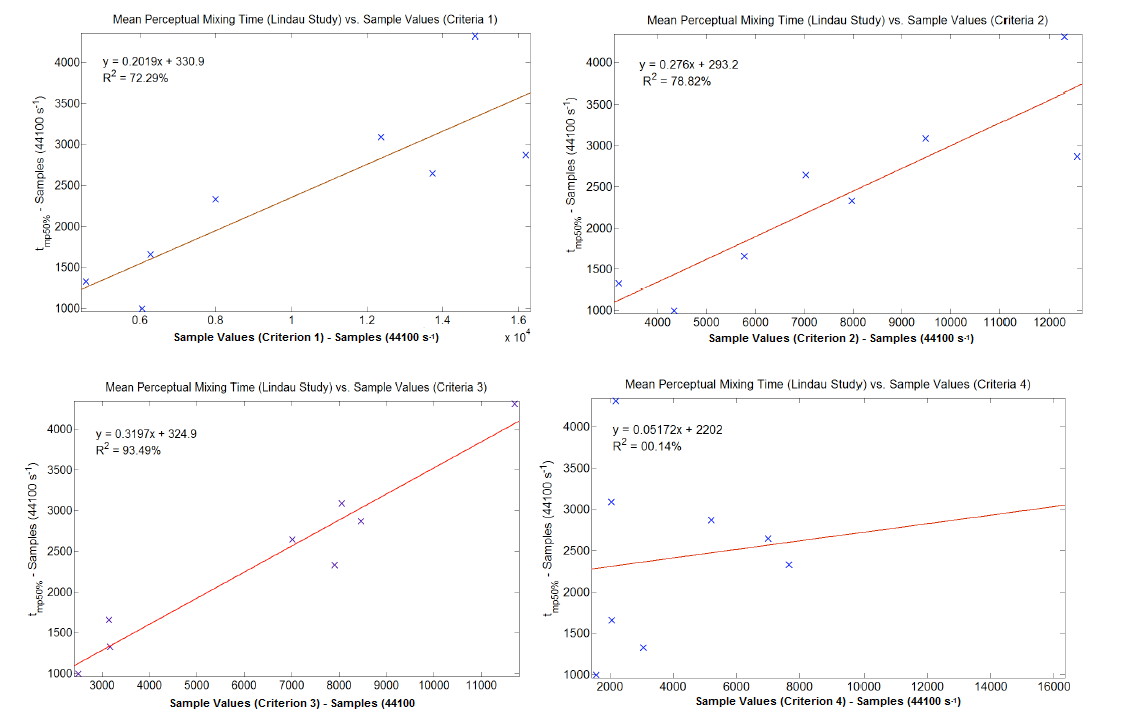
where III indicates the sample number at the point the FD profile first crosses the FD threshold dictated by Criterion III. Criterion IV produced no correlation with tmp50%, with an R2 of 0.14%.

It would be expected for a strong correlation, as observed for Criterion III, that there would be little or no constant in {~~its~~} the linear regression equation, because {~~if~~} when the criterion value is zero {~~so should~~} tmp50% {~~be~~} should be too. However, the constant observed in Equation 8.1 {~~with the constant of~~} is 325 samples, introducing a small uncertainty of less than 1ms (1.3% to 0.35% of tmp50%).

Table 8.2 shows the values for from the Lindau study against the adjusted values for Criterion III sample values after the linear regression equation (Eq. 8.1) had been applied.

Figure 8.2 shows comparative results between values for tmp50% and the results from applying equation 8.1 to the results from Criterion III.

Table 8.3 shows how the results for the explained variance for Criterion I - III (FDI - FDIII) measure up against other predictors of mixing time as presented in the Lindau study.



**Figure 8.1**: Plots of tmp50% against sample values from Criteria I – IV with Linear Regression

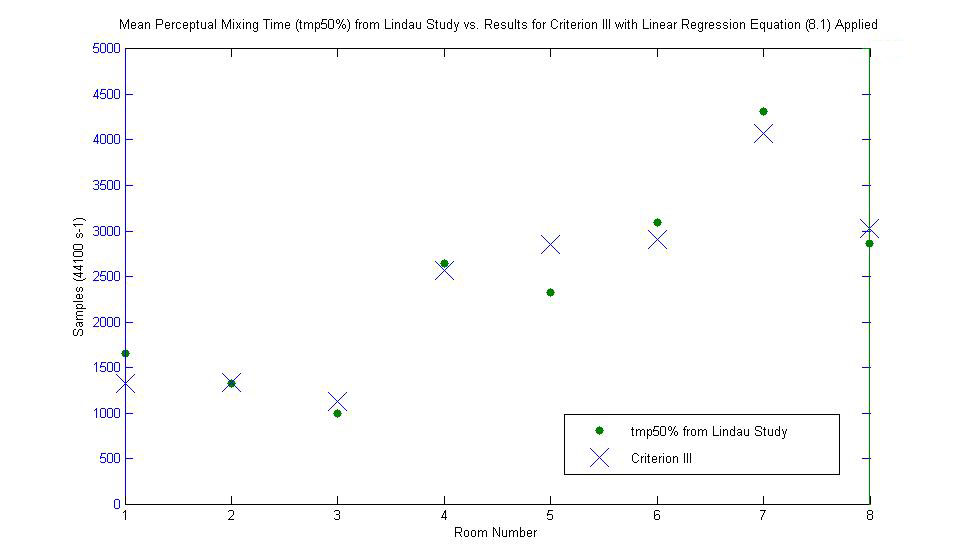
Equation and R2 values.

|  |  |  |
| --- | --- | --- |
| Criterion | Linear Regression Equation  (Samples: 44100s-1) | Explained Variance (R2) |
| I | tmp50% = 0.2090 · I + 331 | 72.29% |
| II | tmp50% = 0.2760 · II + 293 | 78.82% |
| III | tmp50% = 0.3197· III + 325 | 93.49% |
| IV | tmp50% = 0.0517 · IV +2202 | 00.14% |

**Table 8.1**: Linear Regression Equations and Explained Variance (R2) for Criteria I – IV.

|  |  |  |
| --- | --- | --- |
| Room Number | tmp50% from Lindau Study  (Samples: 44100s-1) | Sample Number from Criterion III with Linear Regression Equation (eq. 8.1) Applied  (Samples: 44100s-1) |
| 1 | 1655 | 1328 |
| 2 | 1323 | 1338 |
| 3 | 993 | 1122 |
| 4 | 2646 | 2567 |
| 5 | 2327 | 2853 |
| 6 | 3088 | 2901 |
| 7 | 4311 | 4068 |
| 8 | 2867 | 3031 |

**Table 8.2**: Mean Perceptual Mixing Time from Lindau Study against values for sample number obtained from applying Linear Regression Equation (eq. 8.1) to the sample numbers obtained from Criterion III



**Figure 8.2:** Mean Perceptual Mixing Time (tmp50%) vs. Results for Criterion III with Linear Regression Equation (Eqn. 8.1) Applied

|  |  |  |
| --- | --- | --- |
| **Rank** | **Model Predictors** | **Empirical Predictors** |
| **#** | tmp50% | tmp50% |
| **1** | V/S  R2 = 81.5% | FDC III  R2 = 93.49% |
| **2** | √V  R2 = 78.6% | FDC II  R2 = 78.82% |
| **3** | V  R2 = 77.4% | Abel & Huang I  R2 = 74.7% |
| **4** | S  R2 = 73.3% | FDC I  R2 = 72.29% |
| **5** | RT  R2 = 53.4% | Hidaka 1k  R2 = 57.3% |
| **6** | αmean  R2 = 4.3% | Hidaka 500    R2 = 56.5% |
| **7** |  | Abel & Huang II  R2 = 50.7% |
| **8** |  | Stewart & Sandler  R2 = 37.6% |

**Table 8.3**: Summary of correlation results for tmp50% against predictors, in ranked order.

**9. Discussion and Conclusions**

In this investigation an analysis of the fractal dimension (FD) profile of room impulse responses (RIRs) was investigated as a potential tool for predicting the perceptual mixing time (tmp) of room impulse responses for use in hybrid artificial reverberation or other applications. RIRs were created from models of eight similarly shaped acoustic environments, which were previously used in a study evaluating the performance of various model- and RIR- based predictors for tmp [Lindau *et al*., 2010]. These RIRs {~~were here~~} in this investigation were created using the image source method [Allen & Berkley, 1979]. Fractal dimension profiles of each of these RIRs were generated using the Higuchi algorithm [Higuchi, 1988] for estimating FD. The eight FD profiles were compared, using four criteria, to mean perceptual mixing time (tmp50%) values from the evaluation study which had been obtained from listening tests. Using the third criterion, a correlation between the FD profile and tmp50% was found, with an explained variance (R2) of 93.4%, and a linear regression equation of the form:

tmp50% = 0.3197· III + 325, (6.2)

where III indicates the sample number at the point the FD profile first crosses the FD threshold imposed by Criterion III.

Methods for predicting the mixing time (tmix) that look at RIRs have possibly failed in the past [Abel & Huang, 2006; Hidaka, 2007; Stewart & Sandler, 2007; Defrance *et al*., 2009] because by basing an analysis on the statistical behaviour of the reflections in the room, the behaviour (and limitations) of the human cognitive and auditory systems were not being fully taken into account. Model-based methods for predicting tmp could have lacked accuracy due to the fact that they may not have considered all the factors contributing to the perception of reflections mixing in a room. The results from this investigation suggest that the complexity of reflections in an acoustic environment, i.e. {~~the amount of reflection similarity over time~~} the similarity of reflections over time, may contribute to the perception of mixing in an acoustic environment.

The proposed method for determining tmp from the fractal dimension of a RIR offers a way to measure the t­mp from the complexity of an RIR, and, by establishing a criterion from experiment to relate human perception to the complexity of a RIR, it provides a marked improvement over previous prediction methods as presented in [Lindau *et al.*, 2010]. This method has the added advantage that the FD profile of any room can easily and quickly be derived from a measured or modelled RIR, and the efficacy of the method is independent of a room’s volume or absorption characteristics.

**References:**

Allen, J.B., Berkley, D.A., 1979, *Image method for efficiently simulating small-room acoustics*, Acoustical Society of America, 65 (4), 943-950.

Blesser, B., 2001, *An Interdisciplinary Synthesis*

*of Reverberation Viewpoints.* J. Audio Eng.

Soc., Vol. 49, No. 10, 867- 903

Defrance, G., Daudet, L., Polack, J. D., 2009, *Using Matched Pursuit for Estimating Mixing Time Within Room Impulse Responses*, Acta Acustica united with Acustica, 95 (6), 1071 – 1081.

Hidaka, T., Yamada, Y., Nakagawa, T., 2007, *A new definition of boundary point between early reflections and late reverberation in room impulse responses*, Journal of the Acoustical Society of America, 22 (1), 326-332.

Higuchi, T., 1988, *Approach to an irregular time series on the basis of the fractal theory*, Physica D, 31, 277–283.

Lindau, A., Kosanke, L., Weinzierl, S., 2010, *Perceptual Evaluation of Physical* *Predictors of the Mixing Time in Binaural Room Impulse Responses*, Proceedings of the128th AES Convention, 8089.

Mallat, Stephane and Zhang, Zhifeng, 1993 , *Matching pursuit with time-frequency dictionaries,* IEEE Transactions on Signal Processing, vol. 41, pp. 3397-3415.

Mandelbrot, B. B., 1982, *The Fractal Geometry of Nature*, W.H. Freeman and Company.

Murphy, D., Stewart, R., 2007, *A Hybrid Artificial Reverberation Algorithm*, Proceedings of the 122nd Audio Engineering Society Convention, 7021.

Polack, J.D., 1992, *Modifying Chambers to play Billiards; the Foundations of* *Reverberation Theory*, Acustica, 76, 257-272.

Polack, J.D., 1993, *Playing billiards in the concert hall*: *The mathematical foundations of geometrical room acoustics.* Applied Acoustics, Vol. 38, pp. 235-244

Primavera, A. Palestini, L., Cecchi, S., Piazza, F., Moschetti, M., 2010, *A Hybrid Approach for Real-Time Room Acoustic Response Simulation,* Proceedings of the 128th Audio Engineering Society Convention, 8055.

Rubak, P., Johansen, L.G., 1999, *Artificial Reverberation based on a Pseudo-random* *Impulse Response II*, Proceedings of the 106th Audio Engineering Society Convention, 4900.

Stan, G. B.; Embrechts, J.J; Archambeau, D., 2002, *Comparison of different impulse response measurement techniques*, Journal of the Audio Engineering Society, 50 (4), 249, 262.

Stewart, R., Sandler, M., 2007, *Statistical Measures of Early Reflections of Room Impulse* *Responses*, Proceedings of the 10th International Conference on Digital Audio Effects, 59.