QUANTITATIVE CHARACTERISATION OF PERCEPTUALLY RELEVANT ARTIFACTS OF SYNTHETIC REVERBERATION USING THE EARWIG DISTRIBUTION

Dermot Furlong and Jonathan O'Donovan

Music and Media Technologies
Department of Electronic and Electrical Engineering
Trinity College Dublin
dermot.furlong@tcd.ie, jodonovn@tcd.ie

ABSTRACT

Synthetic reverberation is generally derived by using filter-based algorithms to mimic the types of impulse responses which are generated in natural room responses. Standard approaches would include Schroeder parallel comb filter, feedback delay networks and nested all-pass structures. One of the difficulties in specifying a useable reverb algorithm is that sonic quality is sensitive to the parameter choices used. For example, inappropriate choice of gain and delay in a Schroeder reverberator leads to what is referred to as 'graineyness' or 'fluttering'. These are subjective descriptions of perceived aural characteristics associated with reverb response. Because of the qualitative nature of these artifacts, parameter specification for any new reverb structure generally involves subjective assessment of sonic quality - a 'tweak-and-listen' approach. This paper attempts to take some initial steps toward objective assessment of synthetic reverberation by identifying possible quantitative measures which correlate with these perceptual features of reverberation response. These quantitative measures are derived from a novel joint time frequency distribution which incorporates accurate ear masking effects. This has been developed by generating ear-response based smoothing kernels for the Wigner Distribution leading to its designation as the EarWig Distribution (EWD). Thus, the EWD of an audio signal provides us with a representation which highlights perceptually relevant signal features only. From the EWD of reverb responses it is shown that it is possible to identify some objective signal characteristics which contribute to perceived graineyness and fluttering. This provides the basis for the establishment of associated quantitative measures which would be a significant contribution in synthetic reverberation design.

1. INTRODUCTION

Joint time-frequency (TF) analysis provides us with an intuitive 3D representation of the temporally varying spectral content which typifies non-stationary signals such as music and audio. One popular example of TF analysis is the spectrogram, developed in the 1940's for the analysis of speech. Unfortunately, the spectrogram suffers from a time versus frequency resolution trade-off - high frequency resolution necessitates poor temporal resolution and vice versa. Thus, the temporal and frequency resolutions of the spectrogram cannot be independently set and any choice of window length will of necessity introduce some degree of spectral smearing. An important family of TF distributions is Cohen's class [3], which has seen wide application in the fields of biomedical, sonar, radar and audio signal analysis. Cohen's class may be expressed in terms of the Wigner distribution and a smoothing kernel as follows:

$$C(t,\omega) = \int\int W(u,\xi)\Phi(t-u,\omega-\xi)\,du d\xi\ , \qquad (1)$$

where $W(t,\omega)$ is the Wigner Distribution and $\Phi(t,\omega)$ is the smoothing kernel. The Wigner distribution of a signal $s(t)$ may be defined as:

$$W(t,\omega) = \int_{-\infty}^{\infty} s^*(t-\frac{\tau}{2})\,s(t+\frac{\tau}{2})\,e^{-j\tau\omega}\,d\tau \qquad (2)$$

The main advantage of Cohen's class (CC) over linear TF analysis schemes such as the spectrogram or the wavelet transform is that the temporal and spectral resolution of CC member distributions may be independently specified. However, CC distributions typically suffer from high computational cost and memory requirements and they also feature cross-term interference. These disadvantages have been overcome by a newly developed CC member - the *EarWig* distribution (EWD) [5]. The EWD introduces a *frequency dependent* smoothing kernel that mimics the auditory masking behaviours of the ear. It therefore provides TF representations of audio signals which feature time and frequency resolutions matching those of our hearing.

2. AUDITORY MASKING AND THE EARWIG DISTRIBUTION

Masking [1] is a key feature in all auditory perception. It occurs when a relatively weak signal component is rendered inaudible (masked) by a relatively strong signal component (masker) if both are either temporally or spectrally close. Thus, in many cases much of a given audio signal's content is actually inaudible. This is, of course, the basis for most perceptual compression schemes.

Masking may be divided into two areas - *temporal* and *spectral* masking - deriving from the limited time and frequency resolution of the ear. Given the salience of masking in audio perception, much work in the psychoacoustics literature is concerned with achieving accurate models of auditory resolution capabilities. The frequency resolution of the ear is often likened to that of a bank of overlapping band pass filters, known as critical-band filters. The bandwidth of these filters increases with frequency in a manner designed to match the decreasing frequency selectivity of the auditory system with increasing frequency. The plot of the total output of such a filterbank with respect to frequency is known

as an *excitation pattern* and it reflects the internal representation of the ear's response to a stimulus.

Temporal masking occurs when a masker influences the audibility of a preceding or following signal. Backwards masking occurs when a masker raises the threshold of audibility for a signal which precedes the masker in time. Forward masking occurs when the masker alters the audibility threshold for signals which follow the masker. Forward masking is more marked that backward masking. One current model of temporal resolution is the *temporal window model* [2]. This is used as a running averager of stimulus energy and the temporal window shape is designed to narrow with increasing frequency.

Although auditory masking is well understood, it has rarely been incorporated into TF distributions intended for audio analysis. If we wish to arrive at an accurate representation of auditory perception using TF analysis, incorporating the resolutions of the ear is necessary. The EarWig distribution (EWD) achieves this by using a frequency dependent smoothing kernel with time and frequency resolutions exactly matching two well estabished models of auditory temporal and spectral resolution - namely the temporal window model and the *gammatone filterbank* [1]. This means that the EWD simultaneously incorporates both spectral and temporal masking and registers only unmasked (i.e. perceptually relevant) signal detail.

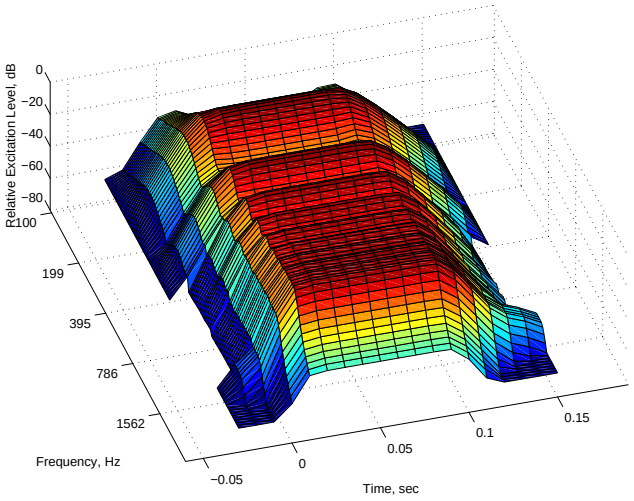

Figure 1: *EWD of a complex tone consisting of the first 10 harmonics of a 200 Hz fundamental. Both spectral and temporal masking are simultaneously and accurately modelled.*

In order to illustrate the use of the EWD in modelling both spectral and temporal masking, Figure 1 presents the EWD of the first 10 equal-level harmonics of a 200 Hz fundamental. The data is plotted on a logarithmic frequency axis. The signal duration is 100ms, but it is clear from the plot that the response to the signal extends outside of this duration both forward and backwards in time, indicating temporal masking. The decreasing inter-ripple height of the signal harmonics indicates the decreasing frequency resolution of the ear with frequency. Overall, what is obvious from this EWD is that both spectral and temporal masking can be effectively presented on a single 3D representation. This is significant for any studies relating to auditory perception as we can now view audible features of any audio signal. Also of note is the fact that the smoothing function kernel used in the EWD has the effect of eliminating the cross-term interference [3] which features in the Wigner Distribution (WD), and which has been often previously been identified as a major problem obviating against use of the WD for audio or acoustical signal representation. It is also noteworthy that it is possible to extract the temporal and spectral marginals from the more general EWD, giving temporal and spectral excitation patterns, respectively [5] [2]. So, we can appreciate that the EWD is a generalised, perceptually significant representation from which various temporal and spectral characteristics may be derived.

3. REVERBERATION ARTIFACTS

Almost all music recorded nowadays is processed using artificial reverberation, typically implemented on DSPs using digital filtering techniques. Some popular approaches are the Schroeder parallel comb filter, nested all-pass structures and feedback delay networks. Unfortunately, the sonic quality of artificial reverb algorithms is typically quite sensitive to parameter choice. For example, inappropriate values of gain and delay in a Schroeder reverberator leads to audible degradations of the reverberant decay known as 'graineyness' and 'fluttering'. In order to avoid these artifacts it is desirable that the delays of the parallel comb filter section of a Schroeder reverberator be incommensurate (mutually prime) so that excessive echo superposition does not occur. Also, the Schroeder reverberator does not generally perform well for long reverb times [4]. Parameter specification for any new reverb setting typically involves time consuming subjective assessment of sonic quality - 'a tweak-and-tune' approach - as there are no agreed objective measures which correlate with the noted artifacts of graineyness and flutter. Given that reverberation is inherently a joint TF process it would seem appropriate to begin any analysis of reverberation with a suitable TF representation. If we are interested in perceptual artifacts, we are further put upon to use a TF representation which highlights perceptually relevant features only.

We have noted earlier that perceptually meaningful information such as a signal's spectral or temporal excitation patterns [2] may be extracted from the EWD. For example, the temporal excitation pattern (TEP) of the individual partials or harmonic components of a given signal reflects the ability of the auditory system to register changes in the amplitude of such components. In [5] a study of the TEPs of signal partials for reverberated signals exhibiting the percepts of graineyness or fluttering demonstrated a strong correlation between the artifacts of graineyness and fluttering and signal partial modulation. It further demonstrated that graineyness can be attributed to a noise-like modulation of signal partials while fluttering is produced by a more regular, less rapid modulation.

As just one particular example, Figure 2 presents the TEPs of the 4th partial (1480 Hz) of a vibraphone signal of fundamental 370 Hz. The lowest curve is the TEP for the unreverberated signal. The 2nd, 3rd and 4th curves from the bottom are the TEPs produced by a real concert hall, a Schroeder reverb with 'non-prime' delays and a Schroeder reverb with 'prime' delays, respectively, where the reverb time is 0.5 seconds. Curves 5-13 repeat the Curve 2-4 pattern with reverb times of 1, 2 and 3 seconds. It is clear from the plots that there is little partial modulation caused by the 0.5 and 1 second reverbs (Curves 2-4 and 5-7). This is not the case for the 2 and 3 second Schroeder reverbs, where the 'non-prime'

reverberations (Curves 8 and 11) exhibit a noise-like modulation. These examples were judged to sound grainey. Curves 9 and 12 are for the prime reverbs and they present a slower pseudo-periodic modulation which corresponds with a percept of a fluttering decay.

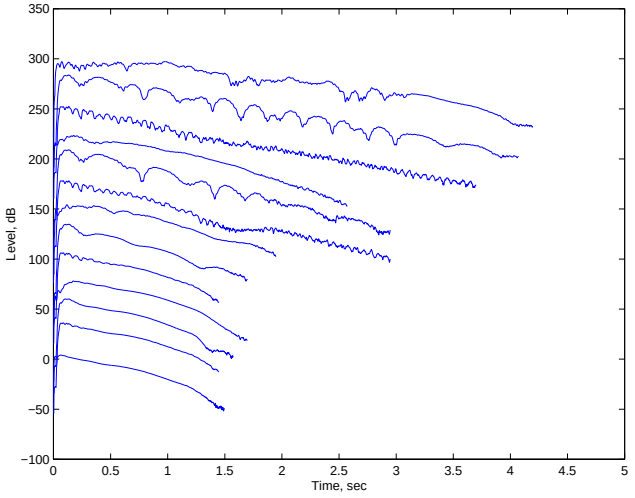

Figure 2: *The temporal marginals for vibraphone partial, $4f_0$ (1480 Hz).*

4. ARTEFACT MEASURES

The above analysis is for just one particular partial of one particular note of one particular instrument, but it is presented as a typical finding. It is suggested that the spectra of the TEPs of reverberated signals may be used as a basis for an objective measure of the percepts of graineyness and fluttering. Figure 3 presents such spectra, for the 2 and 3 second reverbs, respectively, as described earlier. It is clear that the spectral content of non-prime, prime and real reverbs has a higher bandwidth than that of the unreverberated signal. The unreverberated signal's TEP spectral content is concentrated in a 0-5Hz bandwidth, whereas the TEP spectra for the prime and non-prime reverberations demonstrate more significant spectral content above 5Hz. Notable is the fact that the non-prime reverberations for the 2 and 3 second cases demonstrate significant peaks in the 15Hz region. Both these examples were judged to demonstrate graineyness. Likewise, both the prime reverberations for the 2 and 3 second cases demonstrate obvious spectral peaks in the 3-6Hz region and both were judged to be 'fluttery' in their aural quality. It is again emphasised that while a very small number of examples are being presented here by way of explanation, other tests have demonstrated corresponding findings.

While a more thorough study awaits completion, it is here suggested that the perceptual artifacts of graineyness and fluttering can be correlated with modulations of the TEPs which can be measured in terms of spectral peaks in TEP spectra. Furthermore, these preliminary studies suggest that the difference between graineyness and fluttering is in fact one of TEP modulation rate, with lower frequency modulations (approximately 2-8Hz) being heard as flutter, while higher frequency modulations (10-15Hz) giving rise to judgements of graineyness.

In order to investigate these issues further, a Schroeder reverberator algorithm which allowed control of the 'non-primeness' of

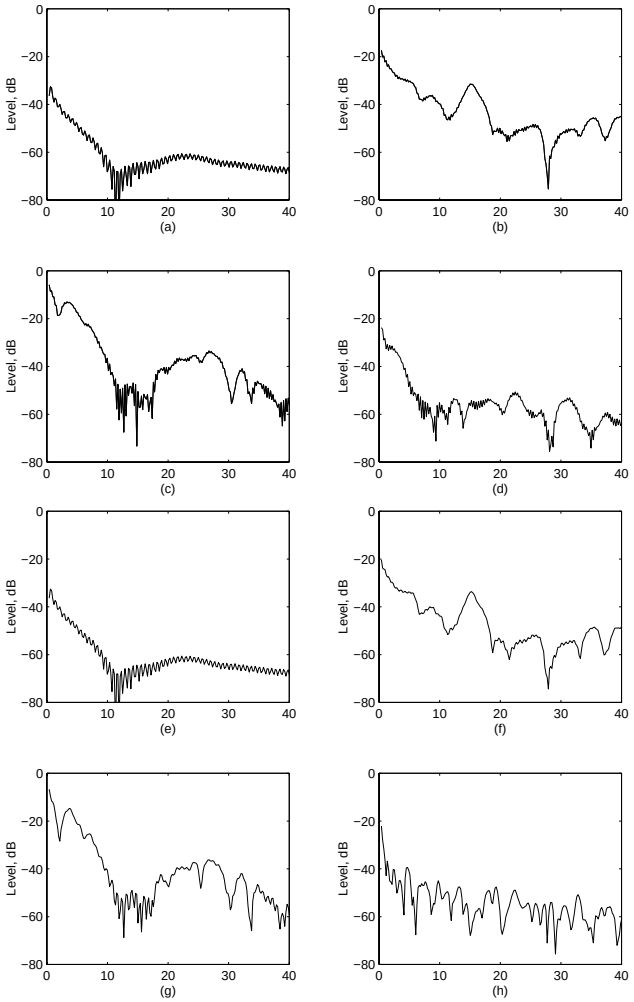

Figure 3: *The TEP spectra for the partial at $4f_0$ for the vibraphone signal; curves (a)-(d) are the unreverberated, 'non-prime' reverberated, 'prime' reverberated and real concert hall reverberated vibraphone signals for a reverb time of 2 seconds. Curves (e)-(h) present the corresponding TEPs for a reverb time of 3 seconds.*

the comb filter delays was developed. This reverberator is illustrated in Figure 4 and is in fact the parallel combination of 2 reverberators - a reverberator consisting of prime combfilter delays and a second reverberator consisting of non-prime delays. The gain, G, of the non-prime reverberator was adjustable and was set to 5 possible values: 0, 0.25, 0.5, 0.75 and 1. By varying the gain G associated with the non-prime sections it was possible to introduce an increasing influence of non-prime comb filter sections to an otherwise prime comb filter reverberator. Reverb times of 0.7,1,2 and 3 seconds were investigated. The effect on the reverberation responses for a reverb time of 3 seconds are catalogued here in terms of the TEP marginals, as shown in Figures 5 (a) and (b).

Although difficult to appreciate without colour, Figure 5 (b) highlights a gradual increase in the spectral energy above 8 Hz, and gradual decrease in spectral energy below 8 Hz as G is increased from 0 to 1. This supports the perception of an initial flutter response gradually becoming increasingly grainey as G approaches

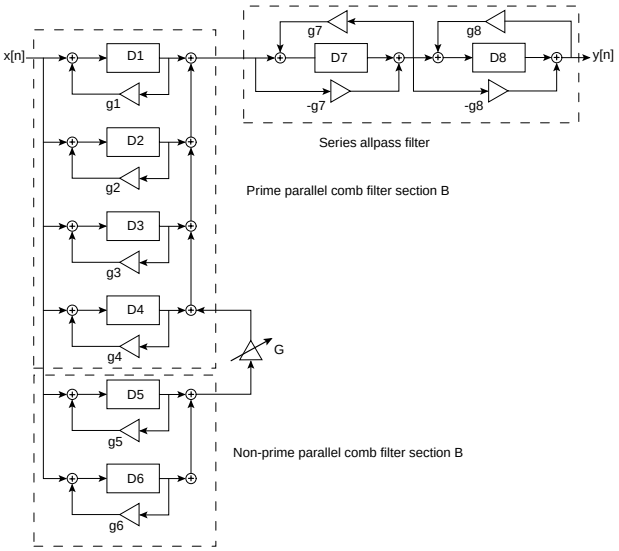

Figure 4: *Schroeder's reverberator with additional variable-gain, 'non-prime' comb filter section.*

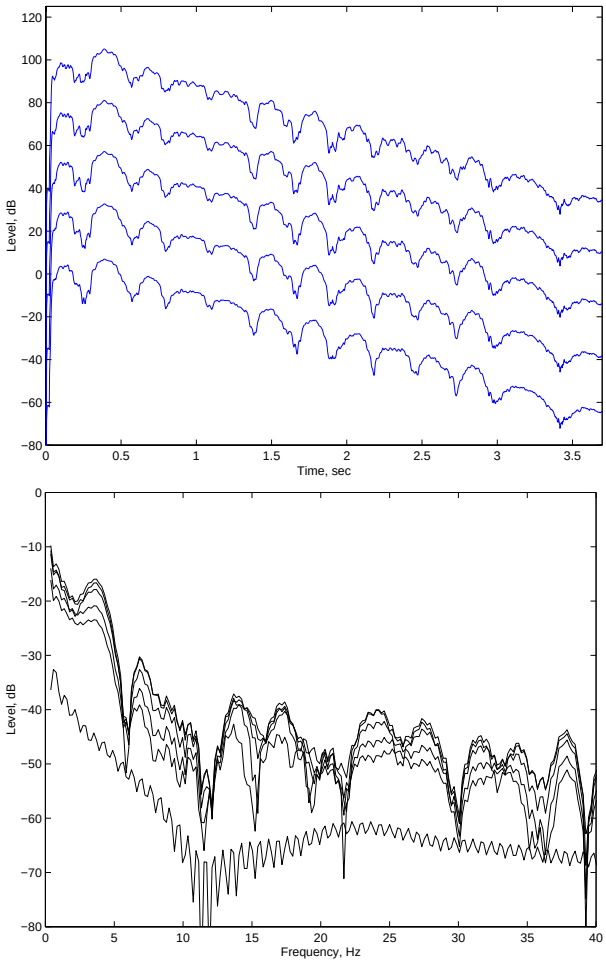

Figure 5: *(a) The temporal marginals for the vibraphone partial, 4f0 (1480 Hz) reverberated with the 3 second Schroeder reverb. The value of gain, G, from bottom to top is 0, 0.25, 0.5, 0.75 and 1. Each curve is offset by 20 dB for clarity. (b) The spectra of the TEPs plotted in (a). Also included is the unreverberated TEP (lowest curve).*

1. This example can be interpreted using the above insights regarding fluttering and graineyness as follows. As the non-prime section gain factor, G, is increased we see an initial 'flutter' response for G=0 gradually take a form which suggests an increasing graineyness as G approaches 1. On listening, this effect can in fact can be identified.

5. CONCLUSIONS

The work outlined here is but a preliminary investigation of the quantification of the perceptual artifacts that manifest in synthetic reverberation. It has been suggested that the EarWig distribution (EWD) offers some definite advantages as a joint time-frequency distribution for audio and music signal analysis. Having a TF representation which presents perceptually salient features only is a good place to start in the quest for objective measures of reverberation characteristics. Clearly, much remains to be done in this area but the results achieved so far are encouraging. The approach adopted has focussed on the extraction of temporal excitation pattern (TEP) spectra from EWD representations of synthetic reverberation. The identification of spectral peaks in TEP spectra which correlate with the presence of graineyness and fluttering suggests the possibility of establishing indicative measures of both from reverberation EWDs. Future work will focus on refining the suggested approach and on undertaking more stringent listening tests under controlled conditions. For now, it is hoped that the presentation of these findings, even at this very preliminary stage, might stimulate interest in the area.

6. REFERENCES

[1] Moore, B. C. J., An Introduction to the Psychology of Hearing, Academic Press, 1997.

[2] Plack, C. J., Moore, B. C. J., "Temporal window shape as a function of frequency and level", J. Acoust. Soc. Amer., Vol. 87, 1990, pp. 2178-2187.

[3] Cohen, L., Time-Frequency Analysis : Theory and Applications, Prentice Hall, 1995.

[4] Moorer, J. A., "About This Reverberation Business", Computer Music Journal, pp. 13-28, 1979.

[5] O'Donovan, J. J., Perceptually Motivated Audio Time-Frequency Analysis, PhD Thesis, Trinity College Dublin, 2000.