The Use of Rhythmograms in the Analysis of Electro-acoustic Music, with Application to Normandeau's *Onomatopoeias* Cycle

David Hirst
School of Contemporary Music
University of Melbourne
dhirst@unimelb.edu.au

ABSTRACT

The rhythmogram is the visual output of an algorithm developed by Todd and Brown which is characterised as a "multi-scale auditory model" consisting of a number of stages that are meant to emulate the response of the lower levels of the auditory system. The aim of the current study is to continue the author's SIAM approach of employing a cognitive model, in combination with signal processing techniques, to analyse the "raw" audio signal of electroacoustic music works, and more specifically, to depict time-related phenomena in a visual manner. Such depictions should assist or enhance aural analysis of, what is essentially, an aural artform. After introducing the theoretical framework of the rhythmogram model, this paper applies it to a detailed analysis of a short segment of Normandeau's work called *Spleen*. The paper then briefly compares rhythmograms of the entirety of Normandeau's related works *Éclats de voix, Spleen* and *Le renard et la rose*. The paper concludes that rhythmograms are capable of showing both the details of short segments of electroacoustic works as well as the broader temporal feature of entire works. It also concludes that the rhythmogram has its limitations, but could be used in further analyses to enhance aural analysis.

1. INTRODUCTION

While undertaking a recent analysis of Jonty Harrison's electroacoustic musical work, *Unsound Objects* [1] the initial phase involved analysing the acoustic surface to identify sound objects. The next phase required an examination of relationships between sound objects, giving rise to the following question: What propels the work along from moment to moment, section to section, scene to scene? To help answer this question, I observed that an increase in sonic activity seems to elicit expectation in the listener that an important event is about to occur. There is a build up in tension that seems to require a release the longer the build up goes on. But how can we measure something I have called "sonic activity" and, even better, how can we display sonic activity in a way that is meaningful?

A follow-up paper [2] took the discussion further, in order to expand and refine the author's SIAM (Segrega-tion, Integration, Assimilation and Meaning) framework for the analysis of electro-acoustic music [3]. Clearly there was a need to expand the SIAM framework to consider the temporal dimension of an electroacoustic musical work in much more detail.

That follow-up paper outlined several methods for determining sound event activity. Beginning with the use of spectral irregularity [4] as a surrogate for activity, the paper then moved on to employ and compare various sound onset algorithms, which make use of a variety of permutations of inter-onset time (the time between the start of events). In terms of automating analysis, the raw inter-onset time plot is very effective in identifying sections in a long musical piece, while the inter-onset rate (events per second) does provide a measure of active versus inactive depiction for various passages in a long piece. The paper concluded that the next step in this work is to test the measurement of activity, and even more detailed rhythmic elements, in other works, especially more rhythmical pieces.

The aim of the current work, which this paper documents, is to continue the SIAM approach of employing a cognitive model, in combination with signal processing techniques, to analyse the "raw" audio signal, and more specifically, to depict time-related phenomena (beat, rhythm, accent, meter, phrase, section, motion, stasis, activity, tension, release, etc.). Such depictions should assist or enhance aural analysis of, what is essentially, an aural artform.

After an extensive literature search, the use of the "rhythmogram" in the analysis of speech rhythm, and the analysis of some tonal music, seemed to fulfill the requirement of a cognition-based method that uses an audio recording as its input signal to produce a plot of the strength of events at certain time points. While not being a "cure-all" for time-related organization within electroacoustic works, it seemed to show some promise within this realm.

2. THE RHYTHMOGRAM

The rhythmogram has been thoroughly described in Todd [5] and Todd & Brown [6]. Todd based his model on the visual edge detection work carried out by Marr [7], and Todd characterised his model as a "multi-scale auditory model". Consisting of a number of stages that are meant to emulate the response of the lower levels of the auditory and nervous systems, the first stage is a transfer function

of the outer and middle ears, approximated by a high pass filter. The basilar membrane is modelled by a bank of gammatone filters, and each cochlea channel is processed by the Meddis [8] inner hair cell model, which outputs the auditory nerve firing probability. The second stage pools the auditory nerve response across frequency and passes it to a multi-scale Gaussian low-pass filter system. In practice, the Guassian filters use a polynomial approximation. The last stage looks for peaks in the low pass response or zero-crossings of the first derivative of the response. Peaks are then summed and plotted on a time constant (corresponding to each frequency channel) versus time graph. This representation is referred to as a "rhythmogram". Figure 1 shows Silcock's schematic [9] for the Todd and Brown version of the model.
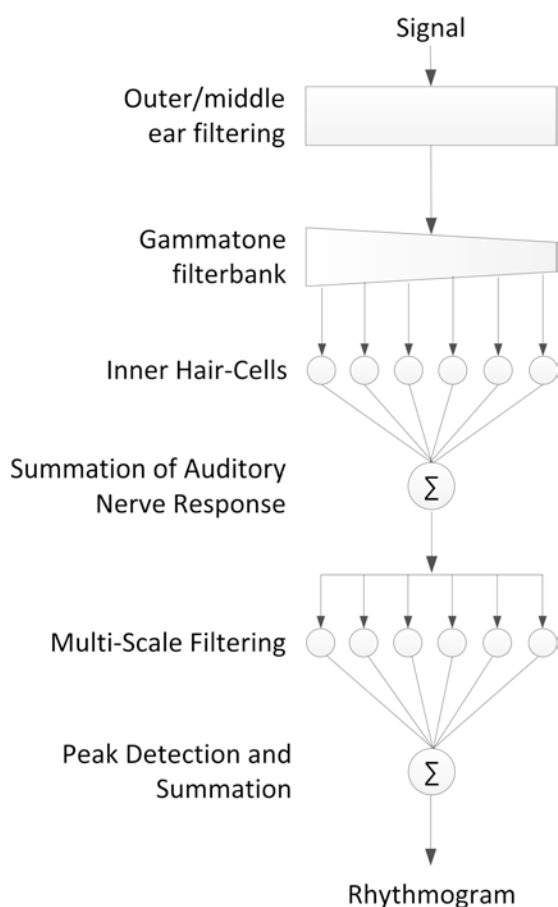

Figure 1. Rhythmogram algorithm[1].

An example of a rhythmogram is shown in figure 2. It is the output of one of the tests carried out in calibrating the software (see below) and shows a rhythmogram for a repeating pattern of three short 50ms tones, followed by a 550ms period of silence, lasting 7 seconds.
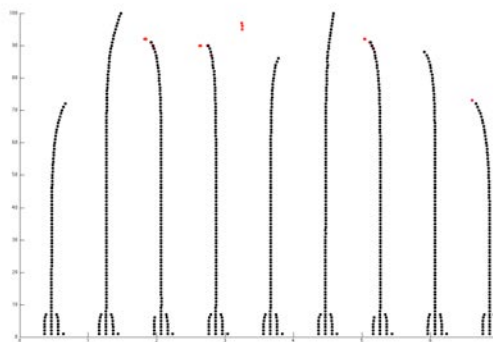

Figure 2. Rhythmogram for a repeating pattern of three short 50ms tones, followed by a 550ms period of silence.

Todd points out that the attraction of the rhythmogram is that it has some similarity to the familiar hierarchical tree diagrams of Lerdahl and Jackendoff [10]. Further, although there is not the space to go into detailed discussion here, Todd and Brown's model takes into account an auditory sensory memory consisting of a "short echoic store" lasting up to about 200 to 300 ms and a "long echoic store" lasting for several seconds or more[2]. Each "cell" (or filter channel) detects peaks in the response of the short term "memory" units. The sum of the peaks is accumulated in a simplified model of the "long echoic store". An "event" activation is associated with the number of memory units that have triggered the peak detector and the height of the memory unit responses. Thus, as Todd states: "Temporal integration relates to the growth of loudness with time. This is modelled as the increase in total neural activity associated with an event, which can be done by simply summing the peak responses of the memory units."[3]

The rhythmogram model not only detects onsets of events, but it can represent rhythmic grouping structures, influenced by a number of factors. The most fundamental of these is "temporal proximity", from which, rhythmic grouping of a sequence can be determined from relative interonset times. Changes in rhythm, or other phenomena, such as meter, can be inferred where there are contrasts, accents or varied articulations present in the signal: i.e. long-short; loud-soft; legato-staccato.

By changing the analysis parameters, the algorithm can "zoom in" and focus on short-term rhythmic details, or "zoom out" and provide a representation of entire sections, or complete structural diagrams for entire works, with similarities to the generative grammar tree diagrams of Lerdahl and Jackendoff. Both of these levels of focus have been explored in the current study.

While this algorithm only attempts to model the auditory system, on its own, to make rhythmic inferences, Todd does attempt to make a link with the sensory motor system with regard to limb motion (foot tapping) and whole body motion (body sway) to speculate on how these may influence both meter and phrase perception.

---

[1] Diagram is reproduced from Silcock (2012) p. 11, and is a variation of the figure by Todd and Brown (1996) p. 257.

[2] Todd (1994) pp. 34-35.

[3] Todd(1994) p. 39.

3. EXPERIMENTAL METHODOLOGY

This study utilises the MATLAB code written by Guy Brown, and adapted by Vincent Aubanel for the LISTA project [11]. This code makes use of the fact that it is possible to increase the efficiency of the computation and still obtain a useful, meaningful rhythmogram plot by using a rectified version of the input signal directly, i.e. bypassing the Gammatone filterbank and inner hair cell stages[4].

3.1 Testing

Code testing used the same four tests as that were used by Silcock [9] in his real-time Pure Data (Pd) version. Each of the tests used the same analysis parameters. These parameters are critical in determining the level of focus desired in each application of the analysis procedure. Todd calls these "unit parameters", and, as short segments of sound were to be tested (seven seconds in this case), the following parameters were used: number of filters 100; minimum time constant 15ms (shortest window); maximum time constant 500ms (longest window); rhythmogram sampling frequency 1000 Hz (can be down-sampled from audio rate as we are only interested in rhythms). In this Brown and Aubanel implementation, the filters are spaced linearly, whereas both the Todd and Silcock versions use logarithmic spacing.

Four tests were carried out using repeating patters over seven seconds:

1. 250ms sine tones (440Hz) repeated at 1000ms intervals.
2. 500ms sine tones repeated at 1000ms intervals.
3. 150ms sine tones every 500ms.
4. A more complex pattern consisting of three 50ms tones, each separated by 50ms silence, programmed to repeat every 800ms.

The first three tests resulted in rhythmograms consisting of vertical spikes at the expected regular time intervals. The fourth test produced the pattern shown in figure 2. We can not only observe the pattern of three repeated spikes, but there are also accumulated, larger, spikes at the secondary 800ms period. Perhaps this could be interpreted as the basic beat.

These tests basically replicated the Silcock results, but with slight variations based upon the use of a more simplified algorithm.

3.2 Temporal Analysis of Electroacoustic works

The electroacoustic works chosen for analysis are collectively known as Robert Normandeau's *Onomatopoeias* Cycle, a cycle of four electroacoustic works dedicated to the voice. The *Onomatopoeias* Cycle consists of four works composed between 1991 and 2009, which share a similar structure of five sections and are of a similar duration of around 15 minutes. The works have been documented by Alexa Woloshyn [12], and Normandeau himself, in an interview with David Ogborn [13].

Two types of analysis were performed. The first is a detailed rhythmic analysis of a short segment of one of the works. The second analysis zooms out to examine the formal structure of three pieces in the cycle and make comparisons.

The work chosen for detailed rhythmic analysis was the second work in the cycle called *Spleen* [14]. This work was chosen as it has a very distinctive beat in various sections and it is slightly unusual for an electroacoustic work in that respect. The first section is called *musique et rythme* (Music and Rhythm) and, after an initial burst of accelerating activity, the piece settles into a rhythmical segment with a seemingly regular beat. This was the segment chosen for detailed examination and it consists of a segment of about 13.47 seconds duration, lasting from 9.25 seconds into the work until about 22.72 seconds[5]. Results and observations are detailed in the next section.

4. RESULTS AND OBSERVATIONS

4.1 Detailed analysis of a short segment of *Spleen*

Figure 3 shows a rhythmogram for the 13.5 second segment of *musique et rythme* from Normandeau's *Spleen*. The X-axis is time (in secs) and the Y-axis is filter number (from 1 to 100). The test parameters were:

- Rhythmogram sample frequency: 1000 Hz
- Minimum time constant 10 msec
- Maximum time constant 500 msec
- Number of smoothing filters: 100
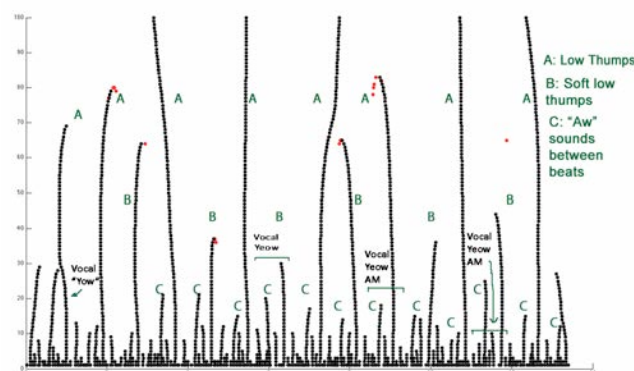- Spacing of filters: linear from .01" to .5"

Figure 3. Rhythmogram for 13.5" of *musique et rythme* from *Spleen*.

Some initial observations that we can make are that the vertical spikes occur at quite regular time intervals, and that there are four or five different height levels at regular intervals.

Labelled as 'A' in figure 3, the tallest spikes correspond with a "low thump", somewhat like a bass drum. Using these spikes we could even infer a tempo from their regularity. From around 2 secs to 12.7 secs, a time-span of 10.88 secs, at about filter #27, there are 12 spikes which are almost equally spaced at about 0.907 secs per spike, which could possibly equate to a tempo of around 66.17 beats per minute.

---

[4] See Todd (1994) "Appendix A.3.3 Input" p. 65.

[5] The first two mins of musique et rythme can be heard via the link on the electrocd site: http://www.electrocd.com/en/cat/imed_9920/

Labelled as 'B' and "soft low thumps" in figure 3, these softer peaks (B) are interspersed between the louder peaks (A) and are equidistant.

At about one second, a loud vocal "yow" shout enters just after a low thump, and from their two stems they seem to combine together into a higher order response, lending weight to this significant structural point. At about 3.7 secs, the vocal "yow" is repeated, but smeared out slightly in time. Another vocal "yeow" sound peaks at 6.3 secs, but it has been further smeared in time and is about 0.5 secs long. It precedes the "low thump" sound (at 6.3 secs), but is timed to reach its peak to coincide with the low thump (refer to the annotations on figure 3).

A further "yeow" vocal sound peaks at 9.1 secs, coinciding with a thump beat, but it begins at 8.5 secs, has a longer duration of 0.6 secs, and it now begins with an amplitude modulation as a decoration and variation. Yet another "yeow" begins at 11.1 secs and ends at 11.8 secs, coinciding with a thump beat. This instance of the vocal "yeow" exhibits even more amplitude modulation than the previous one. The amplitude modulation is represented on the rhythmogram as the small repeating peaks evident for its duration.

We could summarise our observations so far as : There is a rhythmic background of regular beats, consisting of low thumps, arranged in a hierarchy with softer low thumps interspersed. The "tempo" is around 66 bpm (or 132 bpm, depending how you want to count it). An implied duple meter results from the loud-soft thump beats alternating.

Against this regular background is a foreground of vocal "yow" shouts. Less regular in their placement, the shouts become elongated to "yeow", and then amplitude modulated to add colour and variety. Although less regular in their placement, the "shouts" always terminate on a "thump" beat and thereby reinforce the regular pulse.

There are finer embellishments too, labelled 'C' in figure 3. This third level of spikes in the rhythmogram depict events that are placed between thump beats and have a timbre that is somewhere between a saw and a squeaky gate. I'll describe these events as "aw" sounds, and they function as an upbeat to the main thump beat. This "one and two and three and four" pattern has a motoric effect on the passage. The presence of further, shorter, and regular spikes is an indication of more sound events which function to embellish the basic pattern.

Looking at the rhythmogram as a whole, for this passage, we can observe that it tells us there are regular time points in the sound, there is a hierarchy of emphasis in the time points (implying some meter), and a further hierarchy in the sense that there is a background of a regular part (the thumps) and a foreground of less regular vocal shouts. Both the background and the foreground have their own embellishments. Anticipation of the events in the case of the former, and an increase in length and use of amplitude modulation, in the case of the latter.

It is important to note that the above interpretation was carried out using both a visual examination of the rhythmogram, plus aural analysis. This combination of approach was enhanced by the creation of a video which matched the rhythmogram image to the audio soundtrack using a vertical line to trace the time scale for the duration of the excerpt.

4.2 Comparison of whole works from the cycle

The second part of this study involves the use of the rhythmogram in the representation and analysis of whole works. It turns out that the works of Robert Normandeau are ideally suited to this application as well. The *Onomatopoeias* Cycle comprises four works (excluding the original Bédé) which consist of the same basic form. This originally came about because Normandeau used an Akai S-1000 sampler and a MIDI sequencing program (Master Tracks Pro) to create the 1991 piece *Éclats de Voix* using samples of children's voices. He then realised that he could use the same timeline, but different samples, to create a cycle of works [13]. In 1993 came *Spleen* using the voices of four teenage boys, and in 1995 *Le renard et la rose* used the same timeline with adult voices. The final piece in the cycle is *Palimpseste*, from 2005, and it is dedicated to old age. The first three works were analysed, and rhythmograms were created for them.

As these works are each about 15 minutes long, a different set of analysis parameters was required. After a lot of experimentation, the following parameters were found to produce a plot, within an acceptable computation time, that could be readily interpreted:

- Rhythmogram sample frequency: 100 Hz
- Minimum time constant 600 msec
- Maximum time constant 30,000 msec
- Number of smoothing filters: 100
- Threshold: 4500 ms

These parameters represent a "zoomed out" temporal view of the three pieces. The threshold value is a parameter that can be set in the Brown and Aubanel code for use in linking the peaks within their algorithm.

Figures 4-6 depict the rhythmograms[6] for *Éclat*, *Spleen* and *Le renard* for their full durations of around 15 minutes. The alternating grey and white areas mark out the five sections that each piece is divided into - as tabulated by Woloshyn in her paper [13]. With each section, Normandeau combined an emotion with a sonorous parameter. The first section of *Éclat*, for example, is called *Jeu et rythme* (Play and Rhythm). Alignment of these sections facilitates the comparison of the rhythmograms of the three works.

While there is not the space within the confines of this paper to go into a detailed analysis of the audio and visual representations as we did in the previous section, we can make some initial comparisons based upon a visual examination of the three rhythmograms.

Comparing *Spleen* (Fig 5) with *Le renard* (Fig 6) we can immediately see similarities between the rhythmic profiles of sections 1, 3, 4 and 5. To take a case in point, the section 5 of each of these two works seems to consist of three phrases.

---

[6] Like Figures 2 & 3, in Figures 4-6, the X-axis is time (in secs) and the Y-axis is filter number (1 to 100)
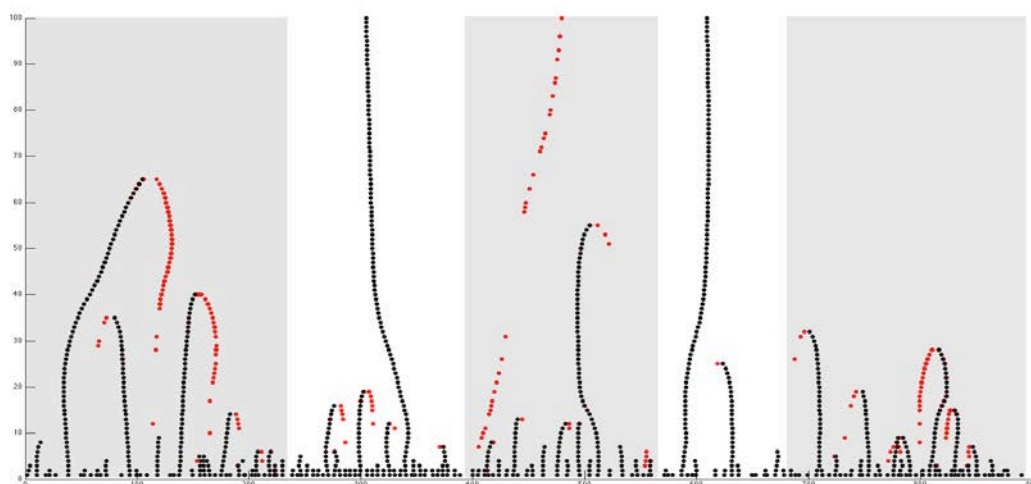
Figure 4. Rhythmogram of the whole of *Éclats de voix* from Normandeau's *Onomatopoeias* cycle.
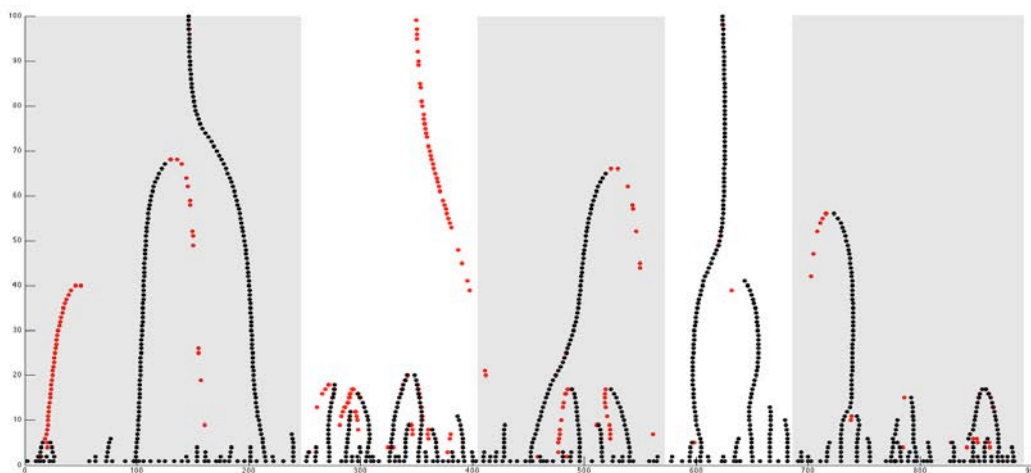

Figure 5. Rhythmogram of the whole of *Spleen* from Normandeau's *Onomatopoeias* cycle.
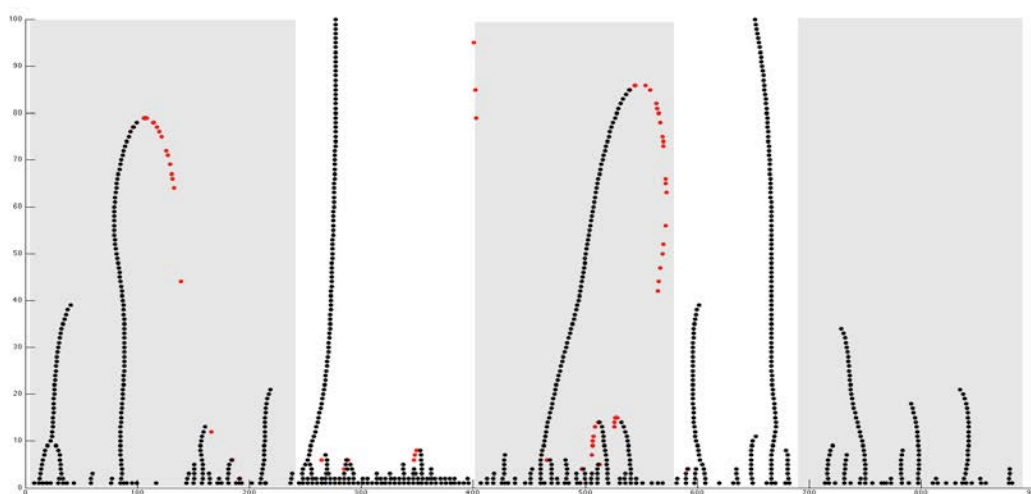

Figure 6. Rhythmogram of the whole of *Le Renard et la rose* from Normandeau's *Onomatopoeias* cycle.

With *Spleen* we have a minute of frenzied voices coming in three waves (one long spike with others clustered around it), followed by about 1'30" of more quiet vocal babbles (shorter, regular spikes), and finishing in the last minute with repetitive drips, punctuated by three soft pulsating gestures, which we can see in the final hierarchical traces of the *Spleen* rhythmogram. In section 5 of *Le renard,* this same scheme is much more exaggerated, so we can see the three distinct spikes, associated with these phrases more easily.

Comparing the rhythmograms from *Éclats de voix* (Fig 4) and *Spleen* (Fig 5), there are some similarities of shape, especially in sections 3, 4 and 5. In glancing down the three rhtyhmograms we can see that *Éclats* is more busy than *Spleen,* which is busier than *Le renard et la rose.* One might conclude that *Éclats* contains more subtleties and then there is a progression to starker contrasts with *Spleen.* Then with *Le renard,* the contrasts are even more exaggerated. This is born out by Normandeau's own statement: "One of the characteristics of this cycle is the use of pulses and rhythms. The use of rhythm is not so obvious in *Éclats de voix*, but in *Spleen*, because the boys were so much more energetic and rhythmic in the studio, I decided to push the boundaries a little bit: the sound is raw, the rhythms are more evident, more "in the face". In *Le renard et la rose*, the boundaries are pushed further again, with minimal sound treatments." [13]

5. CONCLUDING REMARKS

This initial use of the rhythmogram in the analysis of electroacoustic music has demonstrated that the algorithm is capable of displaying the temporal organization of a short segment in with details that may enhance analysis through listening. The algorithm is also flexible, given the careful selection of analysis parameters, in the sense that it can also be used on entire pieces to help elicit information regarding more formal temporal organisational aspects, and to make comparisons with other works.

Some of its short-comings are that it can't solve the separation problems of polyphonic music, rhythmograms can be awkward to interpret, and they also rely on aural analysis. Careful selection of analysis parameters is crucial in obtaining meaningful plots.

A logical next step for this work is to make a more detailed comparative analysis of the Normandeau pieces, and then move on to other electroacoustic works.

Acknowledgments

6. REFERENCES

[1] Hirst, D. 2013. Connecting the Objects in Jonty Harrison's Unsound Objects. *eOREMA Journal, Vol 1, April, 2013*. Available open access, online: http://www.orema.dmu.ac.uk/?q=eorema_journal.

[2] Hirst, D. 2014. *Determining Sonic Activity in Electroacoutic Music* Paper submitted to the Australasian Computer Music Association Conference 2014 for consideration by the review panel.

[3] Hirst, D. 2008. *A Cognitive Framework for the Analysis of Acousmatic Music: Analysing Wind Chimes by Denis Smalley* VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG. Saarbrücken.

[4] Jensen, K. 1999. *Timbre Models of Musical Sounds*, Ph.D. dissertation, University of Copenhagen, Rapport Nr. 99/7.

[5] Todd, N. 1994. The auditory "Primal Sketch": A multiscale model of rhythmic grouping, *Journal of New Music Research*, 23: 1, 25-70.

[6] Todd, N. & Brown, G. 1996. Visualization of Rhythm, Time and Metre. *Artificial Intelligence Review* 10: 253-273.

[7] Marr, D. 1982. *Vision.* Freeman. New York.

[8] Meddis, R. 1988. Simulation of Auditory-Neural Transduction: Further Studies. *J. Acoust. Soc. Am*. 83(3): 1056-1063.

[9] Silcock, A. 2012. *Real-Time 'Rhythmogram' Display*. Report submitted in partial fulfillment of the requirement for the degree of Master of Computing with Honours in Computer Science, Dept. of Computer Science, University of Sheffield.

[10] Lerdahl, F. & Jackendoff, R. 1983. *A Generative Theory of Tonal Music*. Cambridge, Mass. MIT Press.

[11] Brown, G. & Aubanel, V. Rhythmogram MATLAB code written and adapted for the Listening Talker, LISTA, Project (See http://listening-talker.org/)

[12] Woloshyn, A. 2010. Wallace Berry's Structural Processes and Electroacoustic Music: A Case study analysis of Robert Normandeau's "*Onomatopoeias*" cycle. *eContact!* 13(3) http://cec.sonus.ca/econtact/13_3/woloshyn_onomatopoeias.html

[13] Ogborn, D. 2009. Interview with Robert Normandeau. *eContact!* 11(2) http://cec.sonus.ca/econtact/11_2/normandeauro_ogborn.html

[14] Normandeau, R. 1993. Spleen. On music CD *Tangram*. Empreintes DIGITALes, Montréal (Québec), 1994, IMED-9419/20-CD.