COMPUTER MUSIC ANALYSIS VIA A MULTIDISCIPLINARY APPROACH

Francesca Nucibella
University of Padua
CSC - Department of
Information Engineering

Savino Porcelluzzi
University of Padua
CSC - Department of
Information Engineering

Laura Zattra
University of Padua
Department of Visual Arts and
Music

ABSTRACT

This multidisciplinary work aims to investigate the problem of the computer music analysis. It is based on the analysis of a computer music piece: *Winter leaves*, created in 1980 by Mauro Graziani at the CSC in Padova, using Music360 software. Listening, sonogram analysis and digital score analysis, represent the counterpart of the attempt to automatic analysing a fragment of computer music, a music which is characterized by "polyphony" of sound objects, any regular rhythm nor melody or timbre. Two researches (one with a Morphological Descriptor, the other with an algorithm which works via audio content and similarity computation) enlighten the practical problems analysis faces when it has to evaluate the difficult nature of this music.

1. INTRODUCTION

The musicological analysis of computer music is still a very complex issue. This highly depends on the identity of computer music itself: the timbre, the variety of software, the lack of a common musical notation for scores, the absence or undecipherable presence – for non specialists – of computer data. This justifies the development of two opposite analytical methods in musicology: one is the so-called aesthesic analysis, which approaches music from the point of view of perception, the other one (considering the famous semiologic tripartite by Jean-Jacques Nattiez [1]) is the poietic analysis which pays attention to the creative process. Regarding the first approach we can mention the major studies by Denis Smalley [2], Simon Emmerson [3], Michel Imberty [4], François Delalande [5], Francesco Giomi and Marco Ligabue [6]. These are all inspired by the pioneer researcher Pierre Schaeffer [7] and discussed in a recent book by Sthéphane Roy [8]. Other studies aim to graphically represent the electro-acoustic music flux in multimedia contexts, starting from the musicologist's personal listening. All these works aim to describe the listening in order to understand the musical structure and/or timbre. They sometimes use acoustic representation tools (time-amplitude representations, spectrograms, sonograms). The *poietic* analysis is a recent research trend which tries to contrast the inevitable individuality involved in the listening process. It studies the composition process [9] [10], or uses computer data as 'objective' material to

be analysed; one of the first and rare studies was Lorrain's analysis of *Inharmonique* by Jean-Claude Risset [11]. Nevertheless heterogeneity is the characteristics of these researches.

We firmly think that all this variety could in part or completely be clarified by a multidisciplinary approach which combines musicology with computer science and perception. Our proposal is to begin to operate towards this research paradigm. We want to explore how feature extraction and audio content description studies can be useful to the needs of musicology. Up to now the research done in this area is applied, from one hand, to describe single sound objects, on the other hand to automatically transcribe traditional western music or popular music. It is time to work also in the field of electro-acoustic music. We think that automatic analysis is a useful tool to study the classification of electro-acoustic sounds, their description, the structure derived from their polyphonic overlapping, the style of this music. Automatic analysis can help generating automatic segmentation and/or description of the sounds. All this study must be supported by the knowledge of the digital synthesis used by the author during the compositional process. This musicological competence aims to give further evidence of the musicologist's personal listening.

Our research focused on the analysis of a computer music piece: *Winter leaves* (EDIPAN-PRC-S20-16, 1984), for tape, created in 1980 by Mauro Graziani at the CSC (http://www.dei.unipd.it/ricerca/ csc/) in Padova, using Music360 software (the piece was created with an IBM S7 connected to an IBM 370/158, duration: 8'26"). This work derives from a precedent analysis by Laura Zattra based on listening (description and graphic score) and analysis of the digital score [12]. The automatic analysis, starting from a reflection on the sound objects, helps the listening and the identification of sound objects' flow, and above all, it is an important means to study the problems of the computer music analysis. *Winter leaves* is therefore a case study. The validity of its research's approach needs to be verified with other musical pieces in order to establish an analytical method for the analysis of electro-acoustic music.

We are going to show the results focused on a fragment of *Winter leaves*, so that we illustrate the method which was tested on the whole piece. The 2nd section shows the results obtained by Savino Porcelluzzi; the 3rd section follows and describes the

research made by Francesca Nucibella; the 4th section, by Laura Zattra, compares these results with the musicological analysis.

2. ANALYSIS OF COMPUTER MUSIC WITH MORPHOLOGICAL DESCRIPTOR TITLE

2.1. The Morphological Descriptor
The tool that we first experimented to automatically analyze *Winter leaves* is called "Morphological Descriptor" (from now MD), designed at Music Technology Group in the UPF of Barcelona by Julien Ricard and Perfecto Herrera [13][14]. It is based on the morphological theory of sounds objects by Pierre Schaeffer [7]. The choice of the MD is due to its performance in the analysis of singles sounds objects. The Schaeffer's typo-morphology and his computational implementation seems to describe quite well a sound object, so we decide to use an automatic approach to analyze a piece of electroacustic-music. Nevertheless, in order to adapt the MD to the analysis of an entire piece of music, to pass from the theory to the practice, we need to introduce some simplification and changes. In fact the MD could analyze only the following criteria:

• Dynamic profile: describes the shape of the temporal envelope.
• Pitchness profile: discriminates sounds with one predominant pitch (called Pitched), sounds with several pitches (called Complex) and sounds with no pitch (called Noisy).
• Pitchness profile: describes whether the pitchness is constant or varies in function of time (in that case, pitchness is the mean value).
• Pitch profile: describe the variation of the pitch, only specified for pitched sounds. For sounds with unvarying pitch, the pitch is given. Pitch-varying sounds are classified according to the type of variation (continuous or stepped) as well as the global envelope of the pitch (e.g. ascending, descending...).
• Harmonic timbre criteria, specified by a numerical value of brightness.
• Roughness, described by a numerical value.

According to these criteria, a piano phrase of several low-frequency ascending notes, for instance, would be described as follows: dynamic profile = 'iterative', pitchness profile = 'pitched', pitchness profile = 'unvarying', pitch variation type = 'varyingstepped', pitch envelope = 'ascending', a low brightness value and a low roughness value.

2.2. Procedure of Analysis with MD
Our procedure of analysis needs 3 steps procedure:
1) Execution of the program (give the file.wav in input and receive a file.txt in output)
2) analysis of the results given by MD on the file.txt and confrontation with a listening analysis made with an audio editor.
3) if needed, re-compile the program varying threshold values and restart the procedure of analysis.

We list here the problems we encountered and the solutions we found:
• The Morphological Descriptors MD needs a heavy execution time. It needs more or less 3 hours for analyzing a fragment of 2 minutes, so we had to divide the piece in several tracks to reach a reasonable amount of execution time.
• The MD cannot analyze stereophonic tracks, so we had to mix-down every single track from stereo to mono with a balance of 50% between the right and left channel.
• It cannot point out sounds with a frequency over 5 Khz, but luckily in this piece there is any sounds over this frequency.
• The MD was designed to analyze singles sounds objects, but in this piece there is a lot of polyphony. So the only thing that we could do, was to put a higher threshold value. In this way the MD thought to observe less objects but with a good dynamic value. This allowed to take a description more efficient for objects which could not merge in the sound stream.

2.3. Results
We show here the results of the analysis of a fragment of *Winter Leaves* in table 1 (it begins at 3'02'' of the piece and ends at 4'02''). This is a list of the output description given by the MD.
The first 2 columns express the time of beginning and ending of the sound objects detected. The other columns indicate: pitchness, pitch profile, pitchness profile. In the last one a double slash indicates objects with a very short duration (< 1 sec.). This means that in this case we have not verified if the description was correct, because it is not possible for to ear to hear and manually describe a sound so short.
We point out some particular these detections:
• 0,0,9755(s) 8,7592(s) Pitched Varying_Other Varying_Stepped Varying Decreasing = this is a good detection, in fact it is the beginning of a fragment with a high dynamic and not so much polyphony.
• 37,4261 38,8384 Pitched Varying_Delta Unvarying Unvarying Undefined = the analysis detects a pitched sound but any percussive, iterative sound.

| T.B.	T.E.	PITCHNESS P=Pitched C=Complex N=Noisy	DYNAMIC PROFILE	PITCH PROFILE	PITCHNESS PROFILE U=Unvarying V=Varying	I=Increasing D=Decreasing O=Other Und=Undefined			
0	0,3714	P	Varyng_Decrescendo	Varying_Stepped	U	I	-1	0,10221	
0,3714	0,9755	P	Varying_Other	Varying_Stepped	U	O	-1	0,23769	
0,9755	8,7592	P	Varying_Other	Varying_Stepped	V	D	-1	0,37963	
8,7592	8,9510	C	Varying_Delta	Unvarying	U	Und	-1	0,31655	
8,9510	27,2735	P	Varying_Other	Varying_Stepped	V	I	-1	0,26064	
27,2735	27,4245	C	Varying_Delta	Unvarying	V	Und	-1	0,53926	
27,4245	29,7800	C	Varying_Delta	Unvarying	V	Und	-1	0,36321	
29,78	30,0616	N	Varying_Crescendo	Unvarying	V	Und	-1	0,69028	//
30,0616	30,2167	C	Varying_Delta	Unvarying	V	Und	-1	0,42656	//
30,2167	30,6167	P	Varying_Delta	Unvarying	U	Und	1225	0,23060	//
30,6167	32,0738	C	Varying_Decrescendo	Unvarying	V	Und	-1	0,36687	
32,0738	33,1024	P	Varying_Delta	Unvarying	U	Und	1188	0,39163	
33,1024	33,5228	C	Varying_Other	Unvarying	U	Und	-1	0,20126	//
33,5228	34,0208	P	Varying_Other	Unvarying	U	Und	393	0,27205	//
34,0208	35,0004	P	Varying_Decrescendo	Varying_Continuous	U	O	-1	0,17695	
35,0004	35,5024	P	Varying_Decrescendo	Unvarying	U	Und	1157	0,09849	//
35,5024	35,5881	P	Varying_Crescendo	Unvarying	U	Und	291	0,70269	//
35,5881	35,9881	C	Varying_Delta	Unvarying	U	Und	-1	0,13035	//
35,9881	36,1187	C	Varying_Delta	Unvarying	U	Und	-1	0,43168	//
36,1187	36,2412	C	Varying_Delta	Unvarying	V	Und	-1	0,49766	//
36,2412	36,3555	C	Varying_Delta	Unvarying	U	Und	-1	0,34555	//
36,3555	36,4861	C	Varying_Delta	Unvarying	U	Und	-1	0,44900	//
36,4861	36,6126	C	Varying_Delta	Unvarying	V	Und	-1	0,46434	//
36,6126	36,7391	C	Varying_Delta	Unvarying	U	Und	-1	0,49512	//
36,7391	36,8575	C	Unvarying	Unvarying	V	Und	-1	0,41123	//
36,8575	37,1065	C	Varying_Other	Unvarying	U	Und	-1	0,21969	//
37,1065	37,6004	C	Varying_Delta	Unvarying	U	Und	-1	0,13153	
37,6004	37,8371	C	Varying_Delta	Unvarying	V	Und	-1	0,19252	//
37,8371	38,0942	C	Varying_Delta	Unvarying	U	Und	-1	0,27311	//
38,0942	38,3391	C	Varying_Delta	Unvarying	U	Und	-1	0,16201	//
38,3391	38,7351	C	Varying_Decrescendo	Varying	U	Und	-1	0,15142	//
38,7351	40,4861	P	Varying_Other	Unvarying	U	Und	1188	0,15177	
40,4861	41,8984	P	Varying_Delta	Unvarying	U	Und	1181	0,20247	
41,8984	42,8780	P	Varying_Delta	Unvarying	V	Und	192	0,20836	
42,8780	42,9922	C	Varying_Crescendo	Unvarying	U	Und	-1	0,57216	//
42,9922	43,0861	C	Unvarying	Unvarying	U	Und	-1	0,46519	//
43,0861	43,2576	C	Varying_Delta	Unvarying	U	Und	-1	0,38774	//
43,2576	43,8290	C	Varying_Delta	Varying	V	Und	-1	0,43035	//
43,8290	44,8616	P	Varying_Other	Unvarying	U	Und	1208	0,23120	
44,8616	47,6657	P	Varying_Other	Varying_Stepped	U	I	-1	0,33272	
47,6657	48,4820	P	Varying_Other	Unvarying	U	Und	1181	0,48413	
48,4820	48,8208	P	Varying_Other	Unvarying	U	Und	1204	0,55355	//
48,8208	49,3024	P	Varying_Other	Unvarying	V	Und	1168	0,48615	//
49,3024	49,7106	P	Varying_Other	Unvarying	V	Und	1166	0,38821	//
49,7106	50,1229	C	Varying_Delta	Unvarying	U	Und	-1	0,43948	//
50,1229	50,5351	P	Varying_Decrescendo	Unvarying	U	Und	1177	0,41058	//
50,5351	51,7555	P	Varying_Other	Unvarying	U	Und	1168	0,37407	
51,7555	55,2127	P	Varying_Other	Unvarying	V	Und	1366	0,43397	
55,2127	56,7678	P	Varying_Other	Unvarying	U	Und	1158	0,76217	
56,7678	59,1229	N	Varying_Other	Unvarying	V	Und	-1	0,43566	
59,1229	59,5229	N	Varying_Crescendo	Unvarying	U	Und	-1	0,51220	//
59,5229	59,6371	N	Varying_Crescendo	Unvarying	U	Und	-1	0,58347	//
59,6371	59,7514	N	Varying_Other	Unvarying	U	Und	-1	0,53471	//
59,7514	59,8739	N	Varying_Delta	Unvarying	U	Und	-1	0,57636	//
59,8739	60	N	Varying_Impulsive	Unvarying	U	Und	-1	0,33112	//end

Table1. Description of sound objects detected

2.3. Discussion

When the sound stream shows single sounds objects (not polyphonic) or a dominant pitch the MD is almost exact. When there is polyphony, that is most of the time, the MD makes several mistakes and gives nonsense descriptions. It is easy to see, in Table 1, that many descriptions are too generic and useless.

To resolve the polyphony-problem we decided, as we said before, to rise the threshold level. However, with this choice the MD could not detect many objects like, for example, the glissando effect. On the other hand, setting the threshold lower, the MD would have find too many detections of less than 0.5 second, that – as we have seen – are useless for our purpose.

The most important result of this automatic-analysis is therefore not a good description but a good time-segmentation With the MD and the computational reliable criteria given by Schaeffer's Theory it is possible to focus on discontinuity and radical changing of the sound stream.
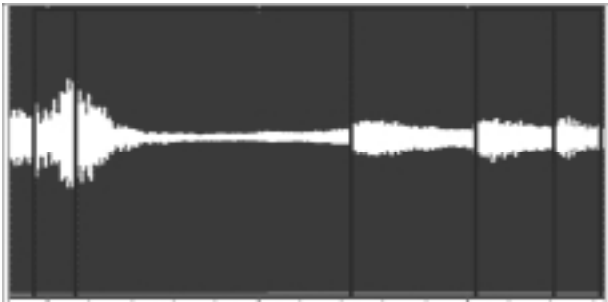
14,903 s	15,491 s	Noisy	Varying_Other	Varying	Unvarying	Undefined
15,491 s	22,181 s	Complex	Varying_Other	Varying	Varying	Undefined
22,181 s	25,144 s	Pitched	Varying_Other	Varying_Stepped	Varying	Decreasing
25,144 s	27,124 s	Pitched	Varying_Other	Unvarying	Varying	Undefined
27,124 s	29,291 s	Pitched	Varying_Decrescendo	Varying_Stepped	Unvarying	Increasing

Figure 1. Example of segmentation. The black line shows the beginning of a new object detected.

In this time/amplitude graph we show the results of the segmentation. We can see that a visual variation is separated from the other like the times of MD advise.

3. A TWO-PHASED APPROACH BASED ON AUDIO CONTENT AND SIMILARITY COMPUTATION

3.1. Basic approach

This chapter follows Savino Pocelluzzi's results and problems and tries another, we hope useful, approach. It describes the application of a previous research in music segmentation of an electro-acoustic music piece. This experiment has been carried out with the method proposed by the study partially funded by the project SIMAC (Semantic Interaction with Music Audio Contents), developed by the researchers Bee Suan Ong and Perfecto Herrera at the Music Technology Group in the UPF of Barcelona [15]. It's a novel, two-phased approach to detect structural changes in music audio signals based on audio content analysis and similarity computation.

The method is based on the assumption that, although music structure creates the uniqueness identity for each music piece, it is possible to detect non-trivial/significant structural changes in music audio signals. In order to obtain appropriate musical content descriptions to detect structural changes, a combination set of low-level descriptors are proposed to be extracted from music audio signal [17][18]. In this way, it is addressed the problem of finding acceptable structural boundaries, without prior knowledge about musical structure.

The application of this segmentation tool through repeated adjustments of analytical parameters, due to the problematic timbre of such a music, leads encouraging results. Manual segmentation is take as a reference for the estimation. Until now this approach has been evaluated positively on a database of "60's pop music" audio files, providing a way to separate the different .sections. of a piece, such as .intro., .verse., .chorus., etc. The attempt of testing the performance of a different music genre, such as electro-acoustic one, represents a novelty. For the application of the algorithm we chose an audio segment of *Winter leaves* by Mauro Graziani which corresponds to the section analysed by S.Porcelluzzi.

3.2. Basic idea of the algorithm

This section describes the basic idea of the proposed algorithm as well as its strong points to better understand obtained results in the subsequent section. It gives the whole picture without going to much detail. To go deeper in details about its implementations procedure , please refer to [16].

In order to detect boundaries candidates of segment changes of music audio this novel approach proceeds computing significant audio descriptors for fixed-length audio frames (i.e. MFCC).

The algorithm works selecting the most significant segment boundaries from the similarity representations computed from each one of the used features. Similarity matrix [19] is a non-parametric technique for studying the global structure of time-ordered streams. It is done by measuring the distance measure between feature vectors using Euclidean distance or the cosine angle between the parameter vectors. It's a two-dimensional representation that contains all the distance measures for all the possibilities of frame combinations. As every frame will be maximally similar to itself, the similarity matrix will have a maximum value along its diagonal. The segment process considers the time instant of audio novelty, which is useful for identifying the immediate changes of audio structure.

Segment boundaries are extracted by detecting peaks where the novelty score exceeds a local or global threshold. In order to obtain the final segment boundaries,it further refines the obtained boudaries by using some dynamics features. It computes the similarity measures between each segment and its neighbouring segment. The adjoint segments with high similarities measure will be merged while those with low similaries measure will be treated as significant segment boundaries.

3.3. Procedure of analysis

The algorithm is applied to a fragment sampled at 44.1 kHz, 16-bit mono. The section lasts 1'00" sec (extracted from 2'55" to 3'55" within the piece) and it is quite meaningful for our purpose: first, it considers events which come slowly one after the other, followed by a change of timbre at 29,76" with a set of percussive sounds and small metal objects. It is therefore an excerpt which concerns different sound typologies.

Once the algorithm has processed the section, the output text file reports the boundaries detected in seconds. L.Zattra supervised results in order to evaluate the algorithm's accuracy. The resulting manual segmentation is taken as a reference for the evaluation.

Table 2 exhibits the segments automatically found, as well as the resulting improvement following the comparison with a naïf listening.

Number of segments	Output text file (sec)	Corrections (sec)
1	0.0 1.63	none
2	1.63 10.22	Detection at 8,48
3	10.22 23.8	Glissando at 18.19
4	23.8 38.89	Detection at 29,76
5	38.89 46.09	New segmentation 38,40 - 41,54 41,54 - 43,49
6	46.09 56.31	New segmentation 45.59 - 51.43 51.43 - 53..38 53.38 - 55.00 55.00 - 56.31
7	56.31 60.00	none

Table 2. Output text file and manual segmentation.

At first glance, it is clearly visible that the algorithm better detects evident audio variations when they are enough separated: this confirms the original purpose for
which the algorithm was created, that is separate different "sections" of a western tonal piece.

It also seems to work appropriately with objects that come slowly one after the other. Hence, the first three segmentations show satisfying results. On the other hand, the dynamics change at 29,76 sec but the algorithm seems unable to separate events since they are too close.

To further improve the results of the detection algorithm it has not been done any optimization, rather it looked convenient working over on the method of

selection. In fact, we chose to bring down the threshold of selection in order to increase the detail of choice and finding more boundaries.

This evaluation line is based on the computed distance measure between segment and it is useful for selecting significant segment boundaries finding the peaks in the novelty score.

For initialization, our adaptive threshold holds a default value of 0.5. It means that we would only consider similarity measures with values more or equal to 0.5. This is the normal value for popular music for which the algorithm provides a way to separate the different "sections" of a piece. We decide to lower the threshold down to 0,02.

Figure 2 shows the (dis)similarity representation computed for the fragment, whereas the graph of Figure 3 represents the novelty measures computed from the (dis)similarity representations between segments.
It is based on the similarity matrix showed above. In these figures the number of segments is 30 which means that the highest local maxima found from novelty measure plot are less than 40, so they are all selected, (please refer to [19]).

In Figure 3 large peaks are detected in the segment number-indexed correlation and labelled as segment boundaries.
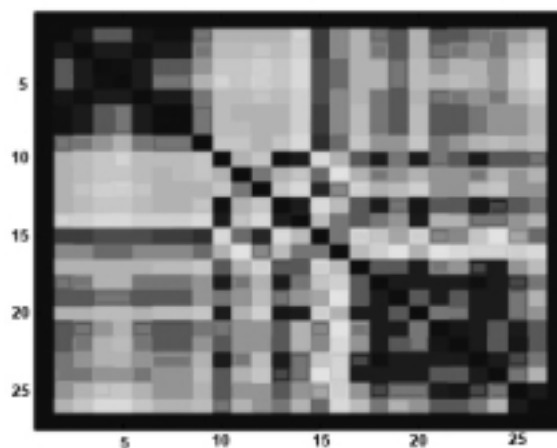
Figure 2. The (dis)similarity representations between detected segments. Axes comprise the number of segments.
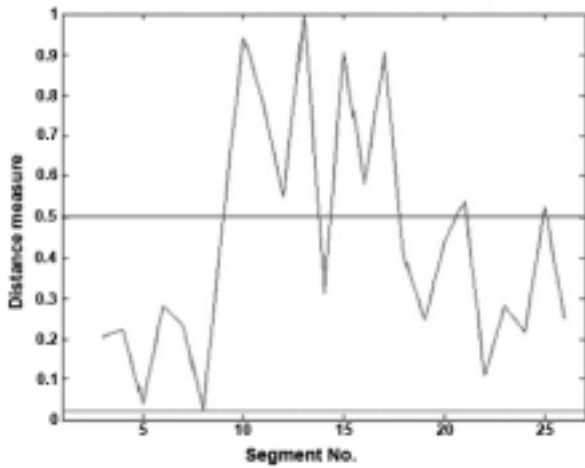
Figure 3. Distance measure vs. Segment Number. Segmentation is performed by detecting peaks over the threshold. Bringing down the threshold to 0.02 means to select basically everything.

Therefore the new results are quoted in Table 3, as well as in Figure 4.

Segments	New results (sec)	Detections (sec)
1	0.0 2.09	1,63 is detected
2	2.09 3.02	3.02 is detected
3	3.02 3.37	
4	3.37 8.48	8,48 is perfectly detected
5	8.48 18.69	18.19 is detected
6	18.69 30.19	29,76 is detected
7	30.19 35.99	
8	35.99 39.71	38,39 is detected
9	39.71 47.02	
10	47.02 50.74	
11	50.74 53.06	53.38 is detected
12	53.06 55.5	55.00 is detected
13	55.5 60.00	

Table 3. New results and annotation of detections. While in the first part of section boundaries are all revealed, the second one give imprecise results.
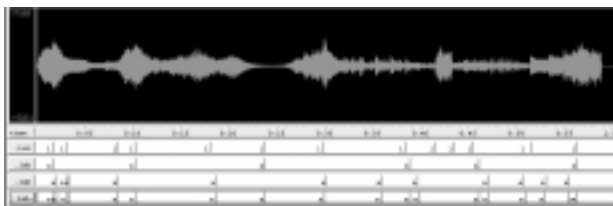
Figure 4.Manually labelled segment boundaries (top line), detected segment boundaries by the algorithm (second line), , post-processed results (third line).and total number of boundaries (final line).

As in this first interaction of the algorith when slow transformations of the flux of sound objects (layers, beats, glissando) are well marked, the detection process is more accurate. This is the case of the beginning of the fragments then timbre differences make analysis inevitably imprecise.

The correct position of a segment boundary is not exactly defined. Any segment boundary within the silence period should be regarded as correct. Therefore, a tolerance Δt is defined. In evaluating the identified segments, the algorithm normally works with a tolerance deviation of ± 3 seconds from the manually labelled boundaries, tested with popular music pieces of about 3 minutes. If a segment boundary is hypothesized within the time interval $t_0 - \Delta t < t < t_0 + \Delta t$ of the reference boundary, t_0 it is judged correct. For our experiment, a reasonable choice seems to be a setting $\Delta t = 500$ms. In fact, even if this setting is too high for the analysed piece, it could not find encouraging results with a lower deviation.

The final step consists in combining both results of the first run and of the re-run of the algorithm together, considering the significant segment boundaries founded in both steps.

3.3. Recall and Precision Measure

The result of a segmentation can contain two possible types of errors [20]. Type-I-errors occur if a true segment boundary has not been spotted by the segmenter (deletion). Type-II-errors occur if a found segment boundary does not correspond to a segment boundary in the reference (false alarm, or segment insertion). The information retrieval community uses two closely related numbers, precision (PRC) and recall (RCL). Precision and recall can be expressed by Type-I-error rate and Type-II-error rate, and vice versa. They are defined as:

- RCL = number of correctly found boundaries/ total number of boundaries
- PRC = number of correctly found boundaries/ number of hypothesized boundaries

Sometimes it is desirable to have one single number for the performance of an algorithm instead of two. In such cases, the Fmeasure is frequently used. This measures overall effectiveness of detection by combining recall and precision with an equal weight.
F is defined as:

$$F = \frac{2*PRC*RCL}{PRC+RCL} \quad .$$

Like RCL and PRC, it is bounded between 0 and 1. In order to evaluate quantitatively the detected segments from the proposed algorithm against the ground truth, we calculate the precision, recall and F-measure measures of the test.
After a tool.s further interaction the percentage of the boundaries the algorithm returns that are correct has fallen to 0.66% , due to the increase of the proportion of   false alarms, namely   segments insertions; meanwhile the rate of correct segments' identification is considerably grow up   to 0.88%. The overall F-

measure reached almost 73% after the re-run of the algorithm, while in the first step of this approach it was at least 60%. In another words, post processing results revealed that with 10 detected segment boundaries about 6 of them are correctly detected compared to a ground truth data. Whilst 2 out of 10 manually labelled boundaries are missed by our automatic boundaries detector. Reprocess of data lead accurate results: neighbouring sound events are almost revealed.

3.4. Discussions

Evaluation results have shown an important validity of the performance of this application. The results obtained are encouraging; they show that this approach has achieved almost 70% in accuracy and reliability in identifying structural boundaries. Although the algorithm has been previously tested on a database of popular music audio files, it seems working well for shorter fragment which does not show a repetitiveness structure, as it is in *Winter leaves*. The algorithm reaches an appropriate evaluation especially in the first part of the section we analysed, where the sound stream is characterized by slowness and division of singles sounds objects, whereas the fast and brief transitions in the second part of the fragment gives nonsense results.

For this reason this approach makes assumptions regarding the content or structure of the source: analysis results depend also on the choice of the analysed section as well as the nature of the changes in the fragment. Self-similarity analysis approach has the advantage of providing a clear and intelligibility view of audio structure,but it is not efficient for spotting repetition with certain degree of tempo change. Another problem with this approach is its threshold dependency in reducing noise for line segment detection. Threshold setting may vary from one song to another,thus, a general setting threshold may not be valid for a wide range of audio.

Perhaps the integration of some previously disregarded lower-level feature attributes could further improve the detection algorithm, as well as the improvement of the selection's detail in the analysis of fragments shorter than 4 seconds.

However evaluation results of this approach can give an overlook on the possible applications of this work, such as automatic labelly and sound classification.

4. A MUSICOLOGICAL LOOK: THE LISTENING AND THE SYNTESIS PROGRAM USED IN *WINTER LEAVES*

4.1. Aesthesic analysis

We present here the musicological analysis of the musical fragment taken from *Winter leaves*, served as a basis to the automatic analysis' work [12]. Through the course of our research, we have compared automatic analysis's results with this musicological view.

The first type of material which can be useful for setting an algorithm of automatic analysis, is the listening process and its effects. The Aesthesic Analysis of the electro-acoustic music affirms that this music does not have, and probably never will have, a neutral level, a musical text with a strict connection between its graphical representation and its sound text. So, the listener becomes the only element of the musical electro-acoustic phenomenon that it is possible to study. We will see that this cannot be the only approach.

However, when listening to *Winter leaves*, we have the impression of hearing a flux of sound objects (layers, beats, glissando) which change slowly and transform themselves into one another. The whole piece can be divided into three parts: a first part in which "chords" are slowly transforming themselves, a second more complex part (from which we took the fragment that we have automatically analysed), which is made up of short sound objects, and a last part in which the chords seem to reappear, enriched by short objects.
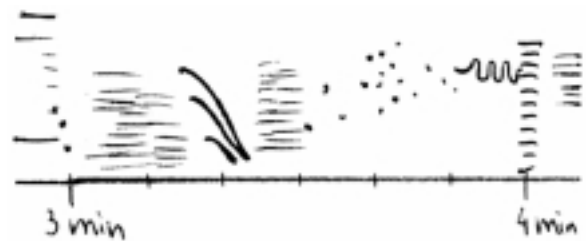
Figure 5. Graphical Score of the fragment 2'55"-3'58"

In this fragment, our listening finds long percussive sounds (dots), inharmonic "chords" (horizontal lines), glissandi. Sonogram (Figure 6) confirms our listening and helps in tracking down accurate tempos (most evident objects: glissandi at 18", sudden change in timbre at 30"; chord at 41'57" and 43'56", oscillating glissandi at 49'50"). Evidently, original sonogram is stereo, whereas automatic analysis needs a mono file.
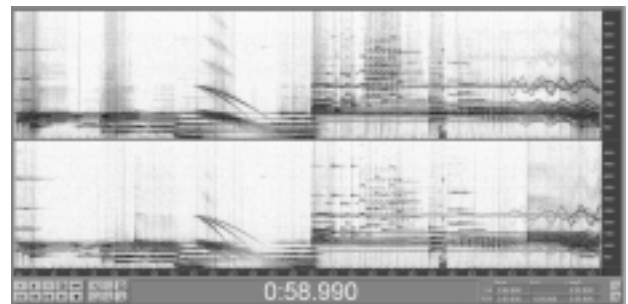
Figure 6. Sonogram of the 1-minute fragment (Hamming window, a *FFT size* of 8,192 tapes and a dynamic *range* of visualisation of 120 dB).

Frequencies are seen to range from 0 to 5,000 or 6,000 Hz. So, it can be understood that the piece is made up of synthesized sounds going from 0 to 5,000 or

6,000 Hz, which also gives us information about the sampling frequency of the work.

4.2. Music360 score

Only via the analysis of the computer score can the composition totally be explained and the listening precisely analysed. This is our new contribution to the musicological analysis of the electro-acoustic music. The musicological analysis cannot avoid to admit that computer music pieces *have* scores, even if they are not evidently understandable. In the case of *Winter leaves*, we own the complete output of the Music360 score, printed on 24 January 1980. It is divided into three parts: in the first (29 pages) we find the score of the Music360 orchestra; in the second, we find two sub-routines (3 pp.) written in the FORTRAN language; in the third we can see the event lists of the score (68 pp.). We cannot find instructions for the final mix of the work, or the sub-routines (Macro) created by Graziani, which in *Winter leaves* are indicated by ZGx, even if we can understand them from the expansions that appear in the orchestra score (instructions beginning with the symbol +).

We can see from the Music360 score that the piece is dived into eight sections each having one minute duration (the 8th section is in its turn divided in 4 shorter sections). The composer says that the mixing simply makes one section follow its precedent. As the score does not present any other value, the sampling frequency can be seen to be 10,000 Hz, i.e. the lowest possible sampling frequency for good audio quality at that time. According to Shannon theorem, the piece's sharpest frequency is 5,000 Hz.

Eleven instruments, often controlled by macros (indicated as ZDx), create spectra born of 3 ratios (2, 2.24, 2.8856).

- A, B, C and F instruments realize additive synthesis via the subroutine ZGSEL (this modifies a basic frequency, according to different parameters indicated in the event list);
- D (ZGADD performs the additive synthesis) and H instruments are similar, but H have the amplitude of each components dependent to the input frequency;
- E creates additive synthesis (25 components) modified by a resonator filter;
- G produces pulses;
- I generates glissandi;
- L and M transform other instruments (delay and reverb).

Music360 score shows an instruction "I" for each of the synthesised sounds. So it is clear that, beyond the complexity of each instrument (e.g. additive synthesis with 4 or more components), the program can synthesise more instructions I at the same time (as Musis360 could not produce lots of 'notes' I from the

same instrument at the very same time, each instrument is used with several numbers: 1, 2, 3, 4, 5, 6, etc.).

A plan has been conceived for each section showing the temporal development (init time) of the instruments. This representation can only be read from a temporal point of view, one-dimensionally and not as a time/frequency relation. Figure 7 shows the plan of the fragment we analysed and correspond to the 4th section generated by Music360 program (p. 17 of the score).
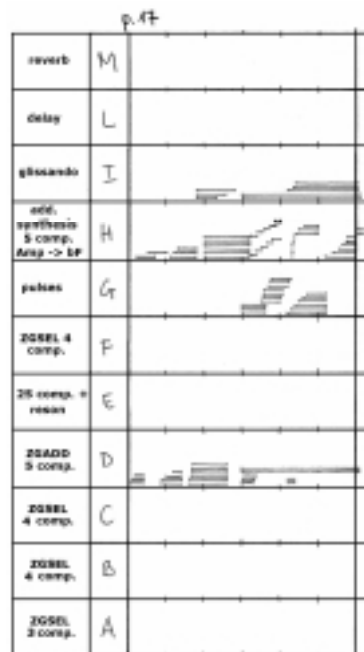

Figure 7. Plan of init-time of I instructions in section 4.

This section contains instructions only for the following instruments: I (glissandi), H (additive synthesis, 5 comp. depending on the basic frequency), G (pulses) and D (additive synthesis, 5 comp.).

4.3. Comparison between the data

Listening is useful to translate the digital data and understand the timbre. In fact, the pure reading of the score sometimes is hard to be related to the real effect. Thanks to the collection and analysis of all the data collected, we can trace the different instrumental effect in the following figure.
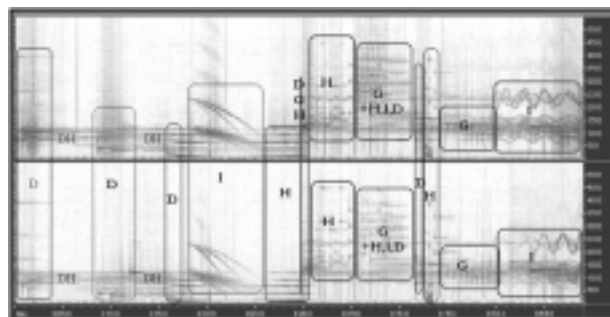

Figure 8. Comparison between the data collected in Figure 5, 6 and 7.

This comparison permits to go very deep inside the listening of the fragment, to understand the initialization time of each timbre and to relate each effect to its digital instrument.

We can also see that the very detailed, even if sometimes imprecise, analysis obtained in the table 1 (through MD algorithm) tended to track down each single input of sounds. Nevertheless, what we can easily find through the listening and the score, does not always correspond to the automatic analysis of pitchness and dynamics.

The automatic analysis is more accurate with the algorithm which functions via audio content and similarity computation. We can compare Table 3 with the following Table 4.

Sections	Detections (sec) (Table 3)	Music 360 score (Figure 7)	Listening (Figure 5)
1	0.0 - 2.09	D	Percussive sounds
2	2.09 - 3.02	DH	Chords
3	3.02 - 3.37		Chords
4	3.37 - 8.48		Chords
5	8.48 - 18.69	D	Chords
6	18.69 - 29.76	I	Glissandi (followed by chords)
7	29.76 - 35.99	DGH	Percussive sounds
8	35.99 - 39.71	G+HID	Percussive sounds
9	39.71 - 47.02		Glissandi
10	47.02 - 50.74	G	Silence
11	50.74 - 53.06	I	
12	53.06 - 55.5	H	
13	55.5 - 60.89	Silence in the audio file	Silence in the audio file

Table 4. Comparison between Table 3, figure 5, 7 and 8.

Some inaccuracies still happen: the automatic analysis is sometimes more precise than the listening or vice-versa. Nevertheless, the high number of coincidences let us know that the results of this step of research is positive.

5. CONCLUSIONS

This was meant to be a proposal for a new multidisciplinary paradigm in the electro-acoustic music analysis domain. Automatic analysis can be an excellent tool for the musicological analysis.

The positioning of the analysis of computer scores as a counterpart to listening and sonogram analysis grows from the observation that it is possible to check the results of our own perception when listening to a computer piece by reading the calculation data. Automatic analysis can help and justify the listening. Using features extraction it would be possible to classify sound objects which characterize electro-acoustic music and make therefore possible different classifications and style studies.

REFERENCES

[1] Nattiez, J.-J. *Musicologie génerale et sémiologie*, 1987 (trad. ital. *Musicologia generale e semiologia*, Torino, EDT, 1989).

[2] Smalley D. "La spettromorfologia: una spiegazione delle forme del suono (I)" "(II)", M/R Musica/Realtà n. 50 1996/3, n. 51 1996/3, pp. 121-137 / pp. 87-110, LIM - Quaderni di Musica/Realtà, 1996.

[3] Emmerson S. "The relation of language to materials", *The language of electroacoustic* music, London, MacMillan Press, pp. 17-39, 1986.

[4] Imberty M. "Continuità e discontinuità", *Enciclopedia della musica. Il Novecento*, Torino, Einaudi, pp. 526-547, 2001.

[5] Delalande F. "Music analysis and reception behaviours: *Sommeil* de Pierre Henry", *Journal of new music research*, vol. 27, no. 1-2, pp. 13-66, 1988.

[6] Giomi, F. & Ligabue, M. "Un approccio estesico-cognitivo alla descrizione dell'objet sonore", R. Dalmonte - M. Baroni (a cura di) - *Secondo convegno europeo di analisi musicale*, Trento - Università degli studi di Trento, pp. 435-448, 1991.

[7] Schaeffer P. *Traité des objets musicaux. Essais interdisciplinaires*, Paris, Seuil, 1966.

[8] Roy S. *L'analyse des musiques électroacoustiques: Modèles et propositions*, Paris, L'Harmattan, Univers Musical, 2003.

[9] Delalande *Le condotte musicali*, Bologna, Clueb, 1993.

[10] Analyses musicales (1996-2005), http://mediatheque. ircam.fr/

[11] Lorrain, D. *Analyse de la bande magnetique de l'œuvre de Jean-Claude Risset*

Inharmonique – Rapports IRCAM 26/80, 1980.

[12] Zattra, L. "Science et technologie comme sources d'inspiration au CSC de Padoue et à l'IRCAM de Paris", Ph.D. Thesis, Paris IV-Sorbonne – Trento University, 2003.

[13] Ricard J., Towards computational morphological description of sound, DEA pre-thesis research work, Universitat Pompeu Fabra, Barcelona, September 2004.

[14] Ricard, J. - Herrera, P. "Using morphological description for generic sound retrieval". Proceedings of Fourth International Conference on Music Information Retrieval; Baltimore, Maryland, USA, 2003.

[15] Suan Ong B. & Herrera P., "Semantic segmentation of music audio contents", Proceedings of the ICMC 2005, Barcelona.

[16] Bee Suan Ong "Towards Automatic Music Structural Analysis:Identifying Characteristic Within-Song Excerpts in Popular Music". *Master Thesis Doctoral Pre-Thesis Work.* UPF. Barcelona, 2005.

[17] Foote, J. "Visualizing Music and Audio using Self-Similarity", *Proceedings of ACM Multimedia Conference,* Orlando, Finland, 1999, pages 77-80.

[18] Tzanetakis, G., and Cook, P. "Multifeature Audio Segmentation for Browsing and Annotation", *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* New Paltz, New York, Oct 1999, pages 103-106.

[19] Foote, J. "Automatic Audio Segmentation using a Measure of Audio Novelty", *Proceedings of IEEE International Conference on Multimedia and Expo,* New York, USA, 2000, pages 452-455.

[20] Kemp T., Schmidt M., Westphal M, Waibel A., "Strategies for automatic segmentation od audio data", in press.