Andrea Bareggi*[1]

*ESME Sudria, France

[1]andrea.bareggi@esme.fr

Towards Objectivity in Automatic Segmentation of Music Score by XML/MIDI Language

ABSTRACT

Automatic segmentation of musical score is an area of ongoing research. The rapid growth of digitalised music collection raised the interest of the scientific community to comprehensive algorithm for structural analysis. An algorithm based on the research of Cambouropoulos was developed in a Matlab environment. The algorithm use a knowledge based strategy on a music score in MIDI/XML format. Results show good agreement with traditional structural analysis.

1. INTRODUCTION

Automatic music analysis is a general research area in which algorithms are developed to allow computer systems to understand the content of digital audio signals for further exploitation. Automatic music structural analysis is a specific subset of audio content analysis with its main task to discover the structure of music by analyzing musical data.

Music structure is a term that denotes the sounds organization of a composition by means of melody, harmony, rhythm and timbre. Repetitions, transformations and evolutions of music structure contribute to the specific identity of music itself. Recently music structural analysis has further extended its applications in the domain related to human cognition. Limitation of human memory makes us incapable to recall every single detail of all incidents that happen in our daily life. As human beings, we may only recall certain events, which have created a 'strong' impression in our mind. The same happens with music, we do not recall the music that we hear in its entirety but through a small number of distinctive excerpts that have left an impression on our mind.

For this reason some approaches are based on *Gestalt* psychology. The comprehension of music within the framework of Gestalt psychology is an exciting area of ongoing research. Early results lead to the representation of music as the object of perception, as shown by Dowling and Harwood (1986) and Butler (1992). Music analysis by the tools of Gestalt psychology is different from traditional musical analysis, and requires numerical computation. Music perception can be analyzed by two different approaches: the first is based on sensation (sensory driven or bottom-up process), the latter is on knowledge (knowledge based or top-down process).

Sensory driven processes are based on auditory nerve images, i.e. the representation of the sound filtered by the human auditory system, the cochlea in particular. The cochlea was extensively studied by Mammano (1998, 159) and modeled by Van Immerseel and Martens (1992, 3511). From a modeling point of view, and auditory nerve image (ANI) is a vector that represents the neural activity for each channel of the cochlea (40 channels, according to Van Immerseel).

Knowledge based processes are mostly based on numerical version of music scores. This kind of process uses a variation of musical elements, such as pitch, sound duration, rest duration, sound intensity, timbre, etc. Early attempts of this research approach were carried out on a statistical population with different level of musicianship by Sloboda (1985), Krumhansl (1995, 53), Dowling (1986), McAdams *et al.* (1993). These works are mainly based on western tonal music; however few attempts were made on contemporary music and extra European popular music (Castellano 1984, 394; Kessler 1984, 131).

The approach presented in this research paper is based on knowledge based processes. The method for temporal segmentation based on MIDI/XML technology within the framework of the general mathematics commercial software Matlab. The system aims to detect structural changes in music to provide a way of separating the different sections.

2. METHODS AND ALGORITHM

The segmentation of musical structures based on variation of musical elements was Eerola and Toiviainen 2004 in a MIDI based algorithm, extension of the work by Krumhansl and Kessler (1982, 334). The algorithm is called *Local Boundary Detection Model (*LBDM) is a variation of the model developed by Cambouropoulos (2003, 411) used for computing a segmentation of polyphonic sources starting by a MIDI source. Other attempts of segmentation based on the *Hidden Markov Model* (HMM) was applied to counterpoint music by Pardo 2005, however the LBDM shows a good adaptively to different musical styles.

The vector of the generic musical element (pitch, duration…) x can be defined for each beat (or time step, i.e. the shortest duration note in the score) n as Equation 1.

$$x = \begin{Bmatrix} \text{note pitch} \\ \text{note duration} \\ \text{rest duration} \\ \text{sound intensity} \end{Bmatrix}$$

Eq. 1. Vector of generic music elements.

For each time step n, the variation of vector x can be calculated as Equation 2.

$$\frac{\partial}{\partial n} x \Rightarrow \Delta x_n = \left| x_n - x_{n-1} \right|$$

Eq. 2. Temporal variation of the module of vector *x*.

The pseudo-derivative necessary to calculate the second order variation of the first order temporal variation of the vector x is calculated by using a back temporal step m (Equation 3).

$$\frac{d^*(\Delta x)}{dn} = \left|\frac{\Delta x_n - \Delta x_{n-1}}{\Delta x_n + \Delta x_{n-1}}\right|$$

$$\frac{\partial}{\partial n}\frac{d^*}{dn}\bar{x} = \Delta x_n \cdot \left(\left.\frac{d^*(\Delta x)}{dn}\right|_n + \left.\frac{d^*(\Delta x)}{dm}\right|_m \right) \qquad m = n-1$$

Eq. 3. Second order pseudo-derivative of vector *x*.

This quantity represents, for each time step *n*, the contribution of a single component of the vector x to the boundary strength for the segmentation algorithm. The overall boundary strength is the sum of each contribution. However, the weight associated to each component of the boundary strength depends on the musical style. A set of weighting factors *k* is defined in order take account of the musical style. The model needs to be calibrated with a traditional structural analysis before being used for automatic segmentation. A calibration carried out on western classical music style shows the following values of the weighting factors vector *k* Equation 4.

$$k = \begin{Bmatrix} 0.35 \\ 0.15 \\ 0.25 \\ 0.25 \end{Bmatrix}$$

Eq. 4. Weighting factors vector.

By combining the weighting factors vector *k* with the second order pseudo-derivative of the vector *x*, we obtain the boundary strength Equation 5.

$$boundary\ strength_n = \sum_{1,2,3,4} k_{1,2,3,4} \cdot \frac{\partial}{\partial n}\frac{d^*}{dn} x$$

Eq. 5. Boundary strength at temporal step *n*.

The *Local Boundary Detection Model* (LBDM) illustrated by the Equations 1 to 5 is an extension of the original LBDM developed by Cambouropoulos can be successfully applied to polyphonic MIDI and XML scores.

The model was implemented within the Matlab MIDI Toolbox environment, developed by Eerola and Toiviainen (2004). The input of the algorithm is a score in MIDI or XML format.

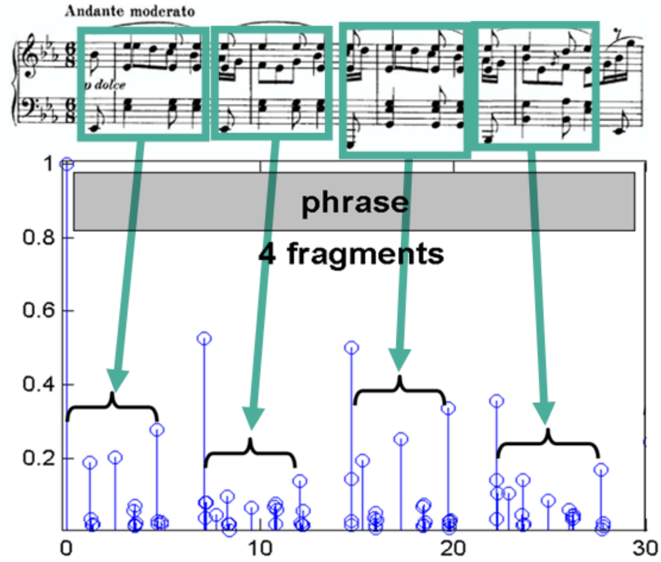
3. VALIDATION AND RESULTS

In order to validate the *LBDM* algorithm, three polyphonic scores of different length were analysed. The three example considered for the validation were :

1. J. Brahms, Intermezzo Op. 117 No. 1, mm. 1–4;
2. J. S. Bach, Italian Concerto in F, mm. 1–3;
3. L. V. Beethoven, Sonata 21 ‘Waldstein’, Rondo, mm. 1–109.

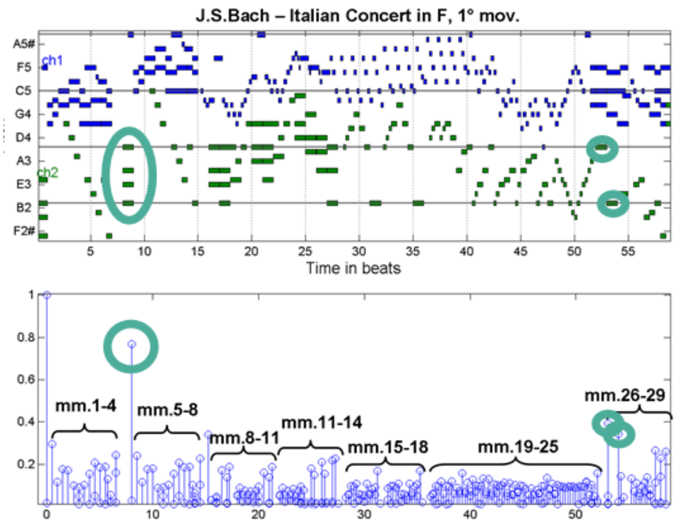
The length of the three examples was chosen in order to show the LBDM is a multiscale method. The output of the model is the vector of the boundary strength as a function of time.

Example 1 shows the output of the model for the Brahms example.



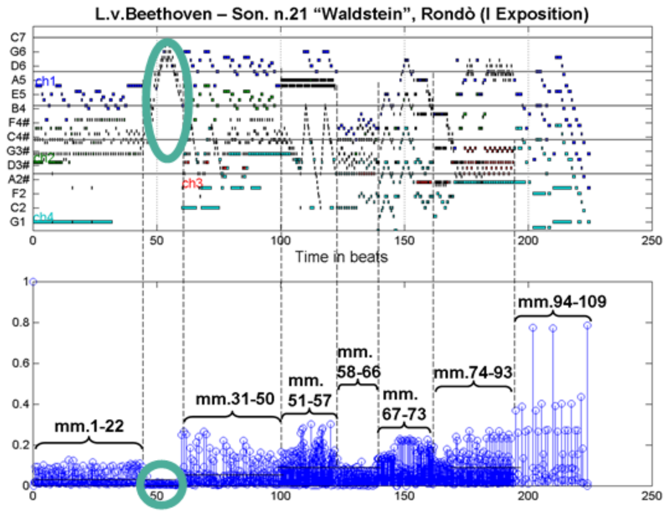
Ex. 1. The LBDM applied to the first four bars of Intermezzo Op. 117 No. 1 by J. Brahms.

The phrase is analysed by the LBDM, that recognise peaks of boundary strength in the same position the composer placed the slurs. We observe that the model take account only of the musical features in the vector *x*. At this scale, the model is able to detect small structures such as the fragments shown in the figure. In the second example, the LBDM shows the ability of identify sections at a bigger level, and even of detecting *cadenzas.* The first 30 measures of the Italian Concerto in F by J. S. Bach are analysed by the model (Example 2). The graphical output is composed by a piano roll (numerical representation of the score) and the boundary strength as a function of time.



Ex. 2. The LBDM applied to the first 40 bars of the Concerto Italiano in F by J. S. Bach.

At this scale the model can still be used for detecting semi-phrases as shown in the lower part of the figure. It can be observed that the dominant chord at the beginning of the second semi-phrase (m. 4), set a high boundary strength as a combination of voice patterns (consistent with the first chord on tonic (F major) of the composition. Also, the model detects the interrupted cadence at the beginning of the last section (m. 26). The third example shows the effectiveness of the automatic segmentation at bigger scale. An extract of the Rondo from Beethoven's piano sonata 'Waldstein' is represented in Example 3.



Ex. 3. The LBDM applied to the first 109 measures of the Rondo of Piano Sonata No. 21 'Waldstein' by L. V. Beethoven.

In this example, musical structures can be easily recognised by the LBDM output. The model calculate a small value of the boundary strength for section with poor harmonic and melodic character, as shown for the arpeggio at the bridge section in mm. 23 to 30.

4. CONCLUSION AND FURTHER RESEARCH

A Local Boundary Detection Model based on works by Cambouropoulos was developed in a Matlab environment. The model can be used in multiscale analysis. The three examples analysed by using the LBDM shows a good agreement with traditional structural analysis at different scales semi-phrase for the first example (extract of the Intermezzo No. 1 by J. Brahms), period for the second example (first Tutti of Italian Concerto in F by J. S. Bach), and larger structures for the third example (exposition in the Rondo of Piano Sonata No. 21 'Waldstein' by L. V. Beethoven). Thanks to the choice of the weighting factors *k*, the model can be applied to a wide range of musical styles.

Further research include the extension of the model to different musical style, with different choices of the coefficient vector *k*; Graphical output as a function of measure, not of beats (measure length automatic detection); coupling the graphical output analysis with the LBDM (pattern recognition); creation of a tone map as a function of musical structures; eventually creation of stand-alone application (smartphone app or executable, no Matlab required).

KEYWORDS

Computational Musicology, Music Segmentation, Music Cognition.

REFERENCES

Butler, David, 1992. *The Musician's Guide to Perception and Cognition*. New York (NY): Schirmer.

Cambouropoulos, Emilios, 1996. 'A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface', Proceedings of *Journées d'Informatique Musicale 1996 (JIM 1996)*, <http://jim.afim-asso.org/jim96/actes.html>, accessed 24/04/2023.

———, 2003. 'Pitch Spelling: A Computational Model', *Music Perception* 20/4: 411–30.

Castellano, Mary A., Bharucha, Jamshed J., and Krumhansl, Carol L., 1984. 'Tonal Hierarchies in the Music of North India', *Journal of Experimental Psychology, General* 113: 394–412.

Dowling, Jay, and Harwood, Dane L., 1986. *Music Cognition*. London: Academic Press.

Eerola, Tuomas, and Toiviainen, Petri, 2004. *MIDI Toolbox: MATLAB Tools for Music*. Jyväskylä: Department of Music, University of Jyväskylä.

Greenwood, Donald D., and Joris, Philip X., 1996. 'Mechanical and 'Temporal'Filtering as Code Terminants of the Response by Cat Primary Fibers to Amplitude-modulated Signals', *Journal of the Acoustical Society of America* 99/2: 1029–39.

Javel, E., McGee, J., Horst, J., and Farley, G., 1988. 'Temporal Mechanisms in Auditory Stimulus Coding', in Gerald M. Edelman, Einar W. Gall, Maxwell W. Cowan (eds.), *Auditory Function: Neurobiological Bases of Hearing*. New York (NY): John Wiley and Sons.

Kessler, Edward J., Hansen, Christa, and Shepard, Roger N., 1984. 'Tonal Schemata in the Perception of Music in Bali and the West', *Music Perception* 2: 131–65.

Krumhansl, Carol L., 1990. *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

———, 1991. 'Music Psychology: Tonal Structures in Perception and Memory', *Annual Review of Psychology* 42: 277–303.

———, 1995. 'Music Psychology and Music Theory: Problems and Prospects', *Music Theory Spectrum* 17/1: 53–80.

Krumhansl, Carol L., and Kessler, Edward J., 1982. 'Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys', *Psychological Review* 89: 334–68.

Mammano, Fabio, Nobili, Renato, and Ashore, Jonathan F., 1998. 'How Well do we Understand the Cochlea?', *Trends in Neurosciences* 21/4: 159–67.

McAdams, Stephen, and Bigand, Emmanuel, 1993. *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford/New York: Oxford University Press.

Pardo, Bryan, and Birmingham, William, 2005. 'Modeling Form for On-line Following of Musical Performances', in Anthony Cohn (ed.), *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*. Pittsburgh (PA). (<https://fileadmin.cs.lth.se/ai/Proceedings/aaai05/PAPERS/AAAI05161.pdf>, accessed 24/04/2023.)

Sloboda, John, 1985. *The Musical Mind: The Cognitive Psychology of Music*. Oxford/New York: Oxford University Press.

Van Immerseel, Luc M., and Martens, Jean-Pierre, 1992. 'Pitch and Voiced/Unvoiced Determination with an Auditory Model', *The Journal of the Acoustical Society of America* 91: 3511–26.