A General Framework To Control Media That Only Exist In Time

Philippe JOLY and Philippe LEPAIN

Media Analysis and Interaction
IRIT- UPS, 118 rte de Narbonne 31062 Toulouse Cedex France
email : joly@irit.fr, lepain@irit.fr

Abstract

Perception needs comparison but comparing different parts of a temporal document is not an easy task though. We have build up a set of methods to split up temporal documents (such as digital videos or musical recordings) into discrete objects. Ergonomic interfaces are then developped to exploit these results. In these interfaces, browsing fonctionnalities and extracts selection are simple and fast. Those kind of representations help contents perception as well as documents analysis. In order to do so, we propose an interface that enables to program the rendering of the discrete objects. This "un-sequencing" of the document is brought out various comparing functionalities that can be used with extracts of documents of the same or different media.

Smart user interfaces based on the temporal structure of audiovisual and musical documents

Temporal media are hard to grasp and manipulate cognitively because they exist only in the time continuum. Human auditory and visual perception and cognition are limited in such a way as to make comparative perception and precise analysis difficult for autiovisual media. Even comparing two short extracts of different performances of the same musical work is difficult.

Developping ergonomic efficient tools to consult digital video or musical documents is a major issue for the spreading of these documents. So, we develop accessible visual representations of audiovisual documents. These representations fulfill a set of properties: they are objective, informative, multi-scale, multi-angle, affordable, consistent and discrete (Aigrain et al. 1995). Discrete objects segmentation methods we have build up are set on documents structure analysis. They take into account perception characteristics. They concern, for example, shot or sequences segmentation of a video or perceived pitch analysis of a musical piece. Resulting objects have many advantages. They enable a multi-scale indexation of the documents. They allow smart functionalities such as the fast browsing of the document, or the fast selection of an extract. As they are related to the way the document has been structured, they bring to the fore, for example, rythm or repetition effects or (un-)synchronization of different events. So, they make the general public aware of the artistic techniques of film-makers or editors or music composers or performers. They help as well specialists to build their own analysis on top of an objective and reliable basis.

Discrete temporal objects

Many research groups have aimed their works at discrete objects extraction in the time dimension of video or musical documents. Our works make it possible today to automatically extract, from any kind of document, a set of hierarchically structured discrete objects.

To analyse video documents we have developped several methods that automatically extract a three-level hierarchical structure from the document. The intermediate level corresponds to shots. Shots boundaries are identified with a method proposed in (Aigrain & Joly 1994). Cuts and wipes are detected by the analysis of high variations between images. This analysis takes into account lightning effects (flash of lightning, lamp turned on), editing artifacts (chromatic glides, cracked gelatine) and cases where images have the same mean light intensity (scenes in dull light). Fades are detected with the analysis of low pixel intensity variations. This method has been compared to some others and evaluated as one of the most efficient by (Dailianas, Allen & England 1995).

Considering that there are more than 1000 shots in a full-length feature film, we have develop ped a method that gathers shots into sequences or "macrosegments", based on parameters peculiar to the audiovisual production (Aigrain, Joly & Longueville 1995). For example, a single cross-dissolve among several cuts in the editing means (most of the time) that a new sequence is beginning. This method permits to reduce from three to ten times the size of the index produced from the shots segmentation as well as giving a synoptic view of the document.

Some shots (sequence-shots), may be longer than 10 minutes. It becomes then necessary to produce an index on

a sub-unit of a shot. To do so we have developped a method for the detection of camera movement. Then, we propose a micro-segmentation where each limit matches a change in the camera movement (Joly & Kim 1996).

Three years ago, our research group began to work on the segmentation of musical objects at different scale along the time axis.

The first level of segmentation automatically extracts segments - that we call strokes - corresponding to quasi-stationary phases of the signal (Lepain & Obrecht 1995). A vector of perceptive parameters is computed from each stroke (intensity, pitch, harmonicity, brightness, ...). A method to extract perceived pitch from a polyphonic context (Lepain 1994) let us draw out the pitch and harmonicity of each stroke.

The second level of segmentation bring out elements we call patterns. It relies on the analysis of the evolution in time of the parameters vector of each stroke. Rules similar to those established by F. Lerdhal and R. Jackendoff (Lerdhal & Jackendoff 1983) can then be applied for melodic patterns discrimination. The analysis of strokes intensity variation made it possible to segment dynamic patterns by adaptative thresholding. We are now working on timbre classes discrimination. This will rely on intensity, harmonicity and brightness parameters.

There are other research works that concern musical documents automatic segmentation. For example, (Cambouropoulos 1996) is concerned with melodic contour extraction, (Todd & Brown 1995) deals with rhythmic groups multi-scale discrimination and (Mellinger 1991) is about auditory flow analysis.

To give an account of the documents structure is not sufficient for an analyst. The precise study of form and contents requires to have the opportunity to compare different extracts. It then makes necessary to be able to "un-sequence" the document in order to compare distant objects, whether they are in the same document or not. Considering these results, we propose new functionalities to organize and master time.

Programming the way the objects are played

The (dis)playing of time-linked media is modeled as a progression through a treelike structure. Each node is a temporal object or an operator. Both have attributes according to their type.

- Temporal objects $\boxed{O_i}$ are of musical or video type. They have at least three attributes: their begin ning, their duration and the reference to the document they belong to. They may be obtained with the fore-mentionned methods.
- Browsing operators are infinite loops (∞) or finite loops (n) (where n is the number of iterations).
- Processing operators ◇ are specific to the type of processed objects (scaling (video), pitch-shifting (music), filtering, slow / fast playing, mixing (both)).

The user is given the possibility to access a set of command for the real-time control of the nodes attributes values. He then can modifiy the progression through the tree as:

- he plays/pauses the objects. It could also be possible to give him a command to stop the displaying of the current object and play the next one at a proportionnal starting position. This is a really efficient command to compare two performances of the same works with different tempi.
- he acts on progression operators (immediate or differed exit of a loop).
- he modifies treatment operators (filters thresholds, acceleration factor, ...).

The a priori tree specification and objects initialization make it possible to estimate the calculus cost of the (dis)playing so that it can be operated in real-time.

Applications

Trees may be implemented through a simple language, or transparently proposed as functionalities of the user interface. For example, the traditionnal "play" function which is supposed to play a whole document is, in application of this formalism, associated to the tree 1, figure 1.

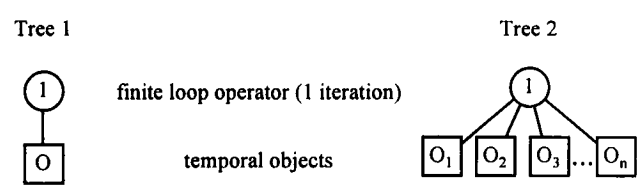

Figure 1 : Examples of simple trees

When reducing the duration attribute of consecutive temporal objects while conserving their beginning attribute, it becomes possible to produce a browsing functionnality that summarize the document as shown by the tree 2 in figure 1.

We are developping now an interface to give the user the choice between several browsing algorithms that can be interesting functionnalities to compare objects whether they are in the same or in different documents. The user

initializes objects which have to be played and modifies the processing operators parameters. Let us consider for example the following display tree:

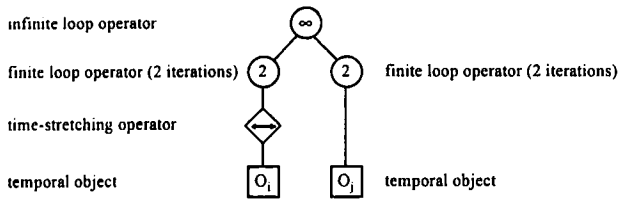

Figure 2 : Example of a tree to compare two temporal objects

Using drag and drop functionnalities, the user select the temporal objects he wants to compare among those which are proposed on a document representation as shown in figure 3. He can play the tree as soon as the time stretching factor is given. He can act on the display using control commands as described in paragraph 3.

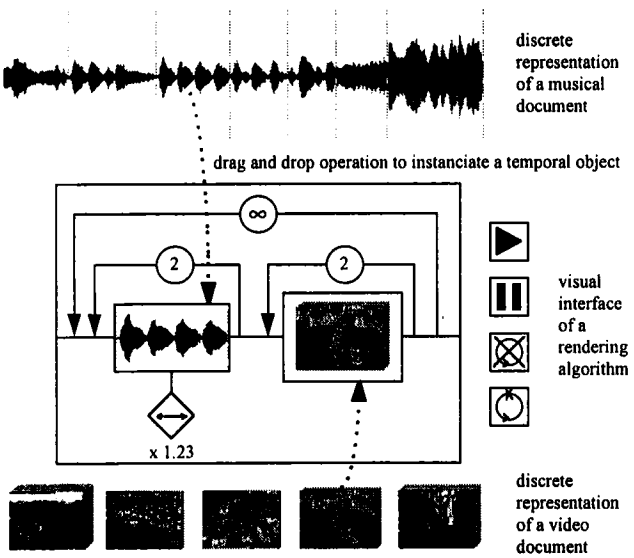

Figure 3 : Example of a user interface proposed to compare two temporal objects

It is now easily possible to compare:

- a theme and its repeat in a musical document, or the variation of the visual angle in a shot / reverse shot sequence in a video document (intra-document, mono-media comparison),
- two performances of the same musical piece, different editing versions of a video (inter-document, mono-media comparison),
- a soundtrack as it has been commercialized and as it has been mixed in the audiovisual document (multimedia, inter-document comparison).

To bring out this display specification and those functionnalities, it is necessary to consider a temporal document as made of discrete temporal objects. As we automate the segmentation on the basis of perception- or production-based criterion, we make it possible to realize a very powerful framework for media consulting and studying. The proposed discrete objects are basic units for the traditionnal analysis that is considerably simplified by the methods involved in our (dis)play strategy.

References

Aigrain, Ph.; Joly, Ph.; Lepain, Ph.; and Longueville, V. 1995. Representation-based user interfaces for the audiovisual library of year 2000. In Proceedings of IS&T / SPIE'95 Multimedia Computing and Networking, 2417:35-45. San Jose, Calif.

Aigrain, Ph; and Joly, Ph. 1994. The automatic real-time analysis of film editing and transition effects and its applications. *Computers & Graphics* 18(1):93-103.

Dailianas, A; Allen, R; and England, P. 1995. Comparison of automatic video segmentation algorithms. In Proceedings of SPIE Photonics West. Philadelphia.

Aigrain, Ph; Joly, Ph; and Longueville, V. 1995. Medium-Knowledge-Based Macro-Segmentation of Video into Sequences. In Proceedings of IJCAI Workshop on Intelligent Multimedia Information Retrieval - Ed. Mark Maybury - Montréal.

Joly, Ph; and Kim, H.K. 1996. Efficient automatic analysis of camera work and microsegmentation of video using spatiotemporal images - *Signal processing : Image communication* 8 (4)

Lepain, Ph; and André-Obrecht, R. 1995. Micro-segmentation d'enregistrements musicaux. In Proceedings of Journées d'Informatique Musicale. Laforia. 81-90. (In French).

Lepain, Ph. 1994. Extraction de fréquences fondamentales d'un document musical pour sa représentation dans un systême de consultation interactif. In Actes des Journées d'Informatique Musicale. Labri. 83-96. (In French).

Lerdhal, F; and Jackendoff, R. 1983. A generative Theory of Tonal Music. MIT Press.

Cambouropoulos, E. 1996. A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface. In Proceedings of Journées d'Informatique Musicale, Les cahiers du GREYC, n°4, p. 170-180, Caen, 1996.

Todd, N.P.McA; and Brown, G. 1995.Visualization of Rhythm, Time and Metre. *Artificial Intelligence Review*. Kluwer Academic Publishers, Netherlands. 00:1-21.

Mellinger, D.K. 1991. Event Formation and Separation in Musical Sound. Ph.D. diss., Dept. of Computer Science, Stanford Univ.