Le temps dans les documents audiovisuels

Jean Carrive

*Institut National de l'Audiovisuel
4, avenue de l'Europe
F-94360 Bry-sur-Marne
jcarrive@ina.fr*

RÉSUMÉ. Cet article présente différents aspects de la structure temporelle des documents audiovisuels, émissions télé ou radiodiffusées principalement. Il décrit tout d'abord les spécificités de ces documents quant à la notion de temps avant d'examiner la manière dont ces aspects peuvent être exploités automatiquement à des fins d'analyse et de documentation, et les problèmes que cela pose.

ABSTRACT. This article presents several aspects on the temporal structure of audiovisual documents, mainly radio and television programs. It first describes the specificities of these documents regarding the notion of time and then examines how these aspects may be exploited for analysis and documentation purposes, and the problems induced.

MOTS-CLÉS : documents audiovisuels, temps, représentation de connaissances, indexation automatique.

KEYWORDS: audiovisual documents, time, knowledge representation, automatic indexing.

1. Introduction

Nous nous intéressons ici aux aspects temporels des documents audiovisuels diffusés par voix hertzienne, câble ou satellite, c'est-à-dire essentiellement aux émissions télévisées et radiophoniques, et plus particulièrement aux émissions de télévision. Nous décrivons d'abord les principales caractéristiques temporelles de ces documents et nous examinerons ensuite la manière dont ces caractéristiques peuvent être exploitées par des processus automatiques à des fins d'analyse, de structuration et de documentation.

2. Structure temporelle des documents audiovisuels

Jusqu'aux dernières évolutions liées aux possibilités de la télévision interactive, le temps intrinsèque des documents audiovisuels est par nature continu, monodimensionnel et sans branchement, c'est-à-dire qu'il peut être vu comme un axe orienté. Le temps de *présentation* de ces documents, quant à lui, a considérablement évolué au fur et à mesure des avancées technologiques, depuis les émissions en direct des débuts de la télévision jusqu'aux possibilités de mise en pause d'émissions diffusées (*time shifting*) offertes par les magnétoscopes avec capacité de stockage.

Même si le temps intrinsèque de ces documents semble extrêmement simple, on peut toutefois différencier plusieurs niveaux de *structure*. Ainsi, pour une émission de télévision, on peut considérer :

- la succession temporelle des images (à 25 images par secondes pour la télévision française) ;
- la succession temporelle des plans. Un plan est ce qui est tourné en une seule prise de caméra : c'est ce qui est filmé entre l'instant où le réalisateur dit « moteur » et l'instant où il dit « coupez ». C'est dans la tradition du cinéma classique – tradition la plus souvent respectée à la télévision – l'unité syntaxique ;
- la succession temporelle des plans formant des scènes (ou séquences). La plupart du temps (sauf pour les *plans séquences*), une scène, unité sémantique, est composée de plusieurs plans ;
- la succession des séquences à l'intérieur d'un document : alternances des séquences de plateau et des séquences de reportage dans un magazine, par exemple ;
- la succession des émissions dans une journée de diffusion ou dans une grille de programmes.

Cohérence temporelle

Pour chacun de ces niveaux on peut définir une notion de cohérence. Ainsi les images consécutives d'un plan, unités discrètes, forment un flux continu pour le

spectateur. Le contenu des images ne varie en général qu'assez faiblement entre deux images, cette variation étant due à un mouvement d'objets à l'écran ou au mouvement de la caméra. Cette cohérence temporelle est exploitée par les outils de compression de la vidéo. Schématiquement, les outils de compression ne codent entièrement que certaines des images, seules les variations du reste des images par rapport à ces images de référence étant inscrites dans le flux. Cette cohérence temporelle à l'intérieur des plans est également exploitée par les algorithmes de segmentation qui cherchent à détecter les ruptures, que ce soit des ruptures brutales (*cuts*) ou des ruptures progressives (fondus, fondus enchaînés, volets, etc.).

Les plans constitutifs d'une scène respectent en général à la télévision un certain nombre de règles de montage qui aident à assurer la cohérence narrative de la scène. Ces règles constituent ce qu'on a pu appeler une « grammaire du film » (Arijon, 1976). Citons par exemple la règle des 180 degrés qui stipule que les différentes prises de vue d'une scène doivent toutes être faites à partir de points situés du même côté d'une ligne imaginaire coupant la scène en deux. Ainsi, sur un plateau de télévision, les différentes caméras sont la plupart du temps disposées en respectant cette règle. Les règles de raccord cherchent à rendre naturels les changements de plan :

- raccords regard : dans un dialogue filmé en champ contre-champ, les personnages doivent donner l'impression de se regarder l'un l'autre, même si les interventions sont tournées séparément ;
- raccords des entrées et sorties des personnages du champ de la caméra, raccords des dialogues, raccords des mouvements de caméra ;
- raccords de luminosité, d'ambiance sonore, de colorimétrie ;
- etc.

Aux plus hauts niveaux de structure, une cohérence est également assurée : cohérence du rythme de montage pour une émission, cohérence de l'habillage image sur une chaîne donnée, par exemple.

Notons qu'à la structure temporelle linéaire intrinsèque des documents audiovisuels vient dans certains cas se superposer une structure périodique. C'est le cas de beaucoup de magazines (magazines d'actualité, politiques ou de divertissement) qui sont constitués d'une succession de séquences de dialogues (séquences plateau, plans du présentateur, interviews) et de séquences de reportage ou d'interventions (musique, danse, sketches). De la même manière, la télévision peut être vue comme une alternance d'émissions et de publicités, et ce point de vue est loin d'être anecdotique dans la guerre commerciale que se livrent les diffuseurs et les concepteurs de magnétoscopes qui voudraient pouvoir détecter automatiquement les plages de publicité afin de pouvoir caler les débuts d'enregistrement. On peut également mentionner la structure périodique hebdomadaire, mensuelle pour certaines émissions, de la plupart des grilles de programme.

3. Exploitation de la structure temporelle

La structure temporelle des documents audiovisuels peut être exploitée dans un processus d'analyse visant à la structuration ou à la documentation de ces documents. Les technologies du numérique permettent dans certains cas d'automatiser ce processus. Après avoir présenté les usages dans lesquels un tel processus d'analyse peut intervenir, nous décrivons les difficultés que rencontre l'automatisation de ce processus.

3.1. Usages visés

La connaissance de la structure temporelle peut parfois faciliter le travail documentaire. Citons par exemple :

- la recherche dans des bases documentaires audiovisuelles. Ainsi, à l'INA, les notices documentaires sont à présent liées aux contenus audiovisuels au sein de la même application. Il est donc désormais possible pour un(e) documentaliste de visionner directement les documents résultant d'une requête ;
- la thématisation d'un fonds audiovisuel. Les séquences, les extraits, les émissions, une fois repérés et documentés dans leur contexte, peuvent être regroupés selon différentes grilles de lecture ;
- la navigation dans les documents. La structure temporelle peut être utilisée pour délinéariser les documents et offrir une navigation par séquence, par table des matières, etc. ;
- le résumé de documents. Il est ainsi possible de condenser un match de tennis en ne montrant que les points importants, de faire un bout à bout des airs d'un opéra ou des reportages d'un magazine, des interventions d'un invité particulier dans un talk-show, etc.

Dans tous ces cas, la structure temporelle doit être explicitée afin de pouvoir être mise en œuvre dans le processus documentaire. La première difficulté à surmonter est donc la question de la représentation de cette structure.

3.2. Représentation

Représenter la structure temporelle des documents audiovisuels est plus difficile qu'il ne peut paraître. En particulier, la norme MPEG-7 (ISO, 2003) de description de contenus multimédias ne permet pas de représenter finement des structures temporelles pourtant relativement simples. Nous proposons dans (Troncy *et al.*, 2004) un formalisme de description permettant de représenter les aspects ontologiques non temporels des documents audiovisuels et les moyens d'étendre ce formalisme pour représenter les aspects structurels temporels. Les caractéristiques non temporelles (nombre de personnes à l'écran, présence de texte dans l'image,

informations de catalogage, par exemple) peuvent être aisément représentées à l'aide de formalismes classiques de représentation de connaissances. En revanche, les aspects temporels nécessitent des langages de représentation plus puissants qui permettent d'exprimer différents types de relations temporelles, telles que la succession, le recouvrement ou la répétition (Carrive *et al.*, 2000).

3.3. Structuration automatique

Afin de pouvoir assister au mieux les processus documentaires décrits ci-dessus, il est envisageable de détecter automatiquement la structure temporelle de certains types d'émissions. On parle alors de *structuration automatique*. L'indexation automatique de documents audiovisuels est une discipline scientifique maintenant établie (Brunelli *et al.*, 1999 ; Snoek *et al.*, 2002) et la structuration automatique en est un des champs. Il peut s'agir soit de détecter automatiquement la structure temporelle d'une émission soit de retrouver une structure connue dans un document. On peut ainsi chercher les séquences d'une fiction sans information *a priori*, activité appelée parfois macrosegmentation (Hammoud *et al.*, 2000), à retrouver les plateaux dans un journal télévisé ou à aligner un flux diffusé sur une grille de programmes annoncée. Ces algorithmes font en général appel à une analyse multimodale du flux où l'on cherche à extraire des descripteurs des différents canaux et à les combiner entre eux : détection de logos, de textes incrustés, détection de visages, reconnaissance de la parole, segmentation en plans, segmentation en locuteurs, segmentation de la bande son en parole/bruit/musique, par exemple.

4. Perspectives

Les avancées technologiques récentes, et en particulier l'apparition des magnétoscopes numériques avec capacité de stockage, offrent des possibilités d'utilisation remettant en cause le modèle purement linéaire des émissions de télévision traditionnelles. Il est ainsi techniquement possible à certains moments d'une émission diffusée d'accéder à des compléments de programme diffusés sur des canaux complémentaires et, après consultation de ces compléments, de reprendre l'émission initiale à l'endroit où elle avait été interrompue. Le modèle temporel sous-jacent reste profondément linéaire, avec quelques branchements possibles, et n'est pas, en cela, réductible aux modèles temporels utilisés pour le multimédia.

5. Bibliographie

Arijon D., *Grammar of film language*, Focal Press, London & Boston, 1976.

Brunelli R., Mich O., Modena C. M., « A Survey on the Automatic Indexing of Video Data », *Journal of Visual and Image Representation*, 10, 1999, p. 78-112.

Carrive J., Roy P., Pachet F., Ronfard R., « A language for Audiovisual Template Recognition », in Springer Verlag LNCS, editor, *Proceedings of the Sixth International Conference on Principles and Practice of Constraint Programming*, vol. 1894, Singapore, September 18-22, 2000, p. 128-142.

ISO, Information technology – *Multimedia content description interface*, ISO, Final Draft International Standard ISO/IEC 15938, 2003.

Snoek C.G.M., Worring M., « A Review on Multimodal Video Indexing », in *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME)*, Lausanne, Switzerland, 2002.

Troncy R., Carrive J., « A Reduced Yet Extensible Audio-Visual Description Language: How to Escape From the MPEG-7 Bottleneck » (to appear), in *ACM Symposium on Document Engineering*, Milwaukee, Wisconsin, USA, October, 28-30 2004.