

---

# Documents audiovisuels instrumentés

## Temporalités et détemporalisations dans les hypervidéos

Olivier Aubert — Yannick Prié

LIRIS - FRE 2672 CNRS

Université Claude Bernard Lyon 1

Bâtiment Nautibus, 43, Boulevard du 11 Novembre 1918

F-69622 Villeurbanne cedex

*prenom.nom@liris.univ-lyon1.fr*

---

*RÉSUMÉ.* Nous nous intéressons tout d'abord dans cet article aux documents audiovisuels et aux diverses temporalités qui leur sont liées. Après un état des lieux des possibilités actuelles des systèmes d'information audiovisuelle, nous définissons plus précisément ce que sont les hypervidéos, ou documents audiovisuels numériques instrumentés. Nous décrivons et illustrons alors un certain nombre des caractéristiques des objets hypervidéos en lien avec les temporalités définies, qui en font des objets complexes. Nous présentons alors un début de cadre d'analyse des synchronisations entre temporalités des hypervidéos générées et vécues.

*ABSTRACT.* First we analyse audiovisual documents and the temporalities they are concerned with. Then we define hypervideos as extended audiovisual documents that offer hypertextual capabilities in the framework of audiovisual information systems. We analyse some features of hypervideos regarding the temporalities we defined, and show that hypervideo are complex objects. So we finally present the beginning of a framework for analysing hypervideos, examining how two temporalities coincide : the one of the hypervideo played stream, and the one experienced by the reader.

*MOTS-CLÉS :* temps, temporalité, analyse temporelle, annotation, documents audiovisuels, hypervidéo, flux du jeu, flux vécu.

*KEYWORDS:* time, temporality, temporal analysis, annotation, audiovisual documents, hypervideo, played stream, experienced stream.

---

## 1. Introduction

Nous nous inscrivons dans cet article dans le contexte des *systèmes d'information audiovisuelle* (SIAV), terme générique par lequel nous désignerons l'ensemble des systèmes permettant, d'une manière ou d'une autre, d'exploiter des documents audiovisuels numériques (DAV). Les fonctionnalités des SIAV incluent le stockage et la diffusion de flux audiovisuels, l'indexation et la recherche de documents ou de fragments documentaires, la visualisation simple ou étendue de ces documents ou fragments documentaires. Par visualisation simple, nous entendons le simple jeu, muni des fonctions classiques d'un magnétoscope, d'un flux audiovisuel. Dans une visualisation étendue, par contre, se réalise l'intégration de documents ou de fragments de documents dans des documents hypermédias (par exemple un résultat de requête présentant des bandes-annonces de films), mais aussi la construction de documents hypermédias à partir de documents ou de fragments (par exemple un document permettant de naviguer à tout moment dans et entre le script d'un film et le film lui-même).

Un état des lieux rapide de la recherche en SIAV<sup>1</sup> nous permet de constater qu'un certain nombre de SIAV complets fonctionnent aujourd'hui. Les domaines visés sont essentiellement de l'actualité (portails d'information – TF1), de la diffusion de cours filmés (réseaux de diffusion internes à des campus [LEE 02]), ou de films (gratuits – Open-video, ou bien payants – FAI ADSL). Dans ces domaines, les documents audiovisuels sont considérés dans leur unité, et leur recherche est réalisée à partir des informations textuelles qui leur sont associées. Mais un grand nombre de systèmes « partiels »<sup>2</sup> prennent appui sur des structures d'information temporellement liées aux documents (structures d'annotations), construites dans une perspective d'analyse de ces derniers, au cours d'une lecture dite active [AUF 99] (voir [STO 03] pour des exemples de grilles de lecture systématique). On donnera comme exemple l'analyse des interactions dans des corpus de conversations ou de scènes de coopération enregistrées (références), le traitement de corpus anthropologiques, etc. Dans ces systèmes, outils d'annotation et de visualisation sont fondamentaux, au sens où les documents doivent être fragmentés, décrits, réorganisés avec des outils adaptés pour faire l'objet d'un travail intellectuel créatif. Ajoutons à ces systèmes d'autres SIAV plus limités, centrés sur des tâches plus précises, telles que le montage professionnel, la gestion de ses vidéos personnelles, le sous-titrage [MAG03], etc.

En parallèle au développement et à la maturation de ces SIAV, les formats audiovisuels font de même : les formats de codage prennent en compte la diffusion en streaming (QT, AVI, etc.) ; les formats de présentation de documents hypermédias intègrent de mieux en mieux les données audiovisuelles (MEPG4, SMIL) ; la norme DVD2 doit permettre d'intégrer facilement flux audiovisuels et HTML ou Flash pour des bonus DVD plus interactifs ; un standard de description de documents audiovisuels voit le

---

1. Voir également les actes de la journée consacrée aux *Modèles documentaires de l'audiovisuel*, dans le cadre de la « Semaine du Document Numérique » (La Rochelle, juin 2004) – <http://liris.cnrs.fr/yprie/Projets/SDN04/>

2. Au sens où ils ne prennent pas forcément à leur charge la diffusion des flux audiovisuels.

jour (MPEG7 [SAN 02]), qui doit permettre l'échange de ces descriptions ; un autre est à l'étude pour suivre à la trace les différents fragments documentaires dans leurs réutilisations (MPEG21).

En d'autres termes, les SIAV poursuivent leur développement, duquel on peut prévoir une intégration toujours plus forte des flux audiovisuels aux systèmes documentaires distribués s'appuyant pour une large part sur l'infrastructure du web. Les systèmes les plus industrialisés concernent la diffusion de documents considérés comme des unités, tandis que d'autres, nécessitant de fragmenter les documents pour leur attacher des informations en permettant une exploitation dépassant la simple visualisation, restent – pour l'instant – dans les marges de la recherche.

Dans tous les cas les particularités des documents audiovisuels demeurent, notamment leur(s) caractère(s) temporel(s), et doivent être prises en compte pour penser les systèmes d'information audiovisuelle, qui permettront de plus en plus une exploitation et une manipulation fragmentaires des documents et une visualisation étendue de ceux-ci. Nous nous efforcerons dans cet article – à vocation plus théorique et exploratoire que technique – de saisir une partie des temporalités liées aux documents audiovisuels, afin de mieux comprendre les problématiques qui se poseront avec ces étranges créatures documentaires hypermédiées qui émergent des différentes expériences de SIAV, et que nous regrouperons sous le terme générique d'hypervidéos, concept que nous définirons. Nous étudierons ensuite quelques caractéristiques essentielles des objets hypervidéos, considérés du point de vue des temporalités. Nous proposerons enfin un début de cadre d'analyse des hypervidéos suivant deux flux temporels, le flux du jeu et le flux vécu.

## 2. Les temps de l'audiovisuel

Cette partie préliminaire est consacrée aux différentes sortes de temps qu'il nous semble intéressant de considérer pour les documents audiovisuels, ainsi qu'à leurs manifestations informatiques dans le cadre de documents audiovisuels numériques.

Précisons tout d'abord ce que nous entendons par flux et document audiovisuels : un flux audiovisuel est un ensemble synchronisé d'informations audiovisuelles en train de se jouer, tandis qu'un document audiovisuel est un objet temporel, ayant une certaine durée, destiné à être joué. On pourra alors dire qu'un document AV se joue en un flux AV.

Considérons deux premières temporalités pour les documents audiovisuels : le temps du flux, et le temps du document.

### 2.1. Temps du flux

Les flux audiovisuels numériques (non générés automatiquement) sont construits à partir de la superposition de flux audio et vidéo synchronisés. Une succession suffi-

samment rapide des images du flux vidéo permet, de part l'« effet Phi »<sup>3</sup>, de donner à l'individu soumis à ces images l'illusion de mouvement des formes qu'il observe. Si le flux vidéo correspond à ce que l'individu est habitué à percevoir et qu'il est synchronisé avec le flux audio, alors le phénomène dit d'*illusion de réalité* se produit. Il est donc impératif pour la réception pleine et entière d'un document audiovisuel que le « flux de conscience » de l'individu soit synchronisé avec le flux audiovisuel, c'est-à-dire que l'individu se « soumette » à la temporalité propre de ce dernier<sup>4</sup>.

La temporalité du flux procède du caractère proprement physique de l'expérience audiovisuelle humaine, et résulte du jeu des flux d'images et sonores superposés, à cadence fixée. La description canonique en machine de ces flux consiste en la description plus ou moins compressée d'une suite d'images et du son correspondant. La génération d'une animation multimédia entraînant une expérience audiovisuelle (par exemple un cube qui se déplace, ou une animation Flash) ressort de la même temporalité en termes de rendu, tandis que le codage en machine diffère (il s'agit alors de décrire de façon plus ou moins explicite ce qu'il se passera dans le flux, et comment l'obtenir). Nous négligeons ici dans un souci de simplification toute considération sur les différences entre le flux tel qu'il est censé être joué, et le flux tel qu'il est effectivement joué par le dispositif de restitution (problèmes de réseaux, de puissance), et les questions de qualité de service associées.

Le temps du flux est un temps du *médium* audiovisuel, lié à la très courte durée (presque l'instant) de l'expérience audiovisuelle en tant que telle. Il concernera par exemple le rendu et le vécu d'un moment quelconque d'un flux issu d'un film.

## 2.2. Temps du document

Le *temps du document*<sup>5</sup>, concerne au contraire le document audiovisuel en tant qu'objet temporel séquentiel ayant une durée (par exemple, un film dure 1h34min).

La temporalité du document prend appui sur cette durée et sur la linéarité qu'elle engendre. Elle est le plus souvent explicitée dans les formats de codage des documents audiovisuels (sinon, il n'y aura d'autre solution pour en faire usage que de la construire au moment du jeu du flux).

Considérer le temps du document permet toutes sortes d'opérations à caractère temporel : comparaison de durées de documents ; description de documents comme longs ou courts ; possibilité de considérer des moments et de les comparer (avant/après) ; possibilité de considérer des fragments et de les comparer (relations d'inter-

3. Effet psychologique mis en évidence par les théoriciens de la Gestalt : deux stimulus visuels suffisamment proches structurellement et temporellement sont confondus en un unique stimulus en vertu d'un principe de simplicité.

4. Acception dite phénoménologique de la nature temporelle des DAV d'après [BAC 99], qui la compare à l'acception empirique selon laquelle le rythme de lecture d'une séquence audiovisuelle est imposé par le document lui-même et non par le lecteur.

5. Qu'on pourrait appeler également *temps de la durée (du document)*.

valle de Allen [ALL 83]); possibilité de savoir où l'on se trouve (début, fin, milieu), etc.

Les temporalités du flux et du document sont directement liées à la nature même des documents audiovisuels, à leur représentation en machine (ou même sur bobine) et aux dispositifs de restitution, indépendamment des contenus véhiculés. Les temporalités analytiques et sociales auront au contraire à voir avec les contenus audiovisuels, et signent l'inscription des documents dans des usages.

### 2.3. Temps analytiques

Un document audiovisuel peut être décrit par une structure documentaire qui correspond à son genre, et qui explicite l'enchaînement de séquences identifiables. Ainsi, un journal télévisé d'une chaîne se déroule souvent de la même manière, une publicité se termine en général par le produit présenté, un film par un générique, etc. Nous pouvons appeler *temps documentaires* les temps liés aux structures de description des parties d'un document audiovisuel (par référence aux structures documentaires logiques des documents textuels).

Nous pouvons encore considérer d'autres temporalités, telle celle de la narration (ou temporalité diégétique) : quand une histoire est racontée se construit le temps de l'histoire, souvent cohérent avec la temporalité de la vie réelle (on parlera d'heures, de dates, d'événements, etc.).

Ces temporalités diffèrent radicalement de la temporalité du flux et de celle du document, car elle appartiennent à la catégorie des temporalités *analytiques*, au sens où elle réfèrent aux contenus et doivent être décrites (et le plus souvent écrites) pour pouvoir exister. Elles peuvent être considérées comme ayant trait à la nature documentaire des documents audiovisuels comme manifestations humaines, et étudiées dans des cadres équivalents à ceux utilisés pour les documents textuels à « textualité séquentielle » (structures logiques, sémantiques, etc.).

L'inscription informatique des temporalités analytiques se manifestera dans des structures analytiques de description, qui auront un rapport au flux à travers ses fragments temporels. La notion d'*annotation* est ici fondamentale, qui permet de faire le lien entre le flux et ses fragments (ou instants) et l'information ajoutée par l'analyse.

Ainsi, on pourra considérer une structure documentaire pour un épisode de série télévisée (le point de vue du réalisateur), ainsi que d'autres structures d'analyse par exemple diégétiques (description de l'histoire, de la progression narrative, temporalités des dialogues, etc.). Ces structures sont les manifestations de temporalités analytiques effectivement analysées et inscrites comme parties d'une structure d'annotation plus globale<sup>6</sup>.

6. Au sens où celle-ci pourra rendre compte d'autres analyses, non centrées directement sur la temporalité, comme par exemple celle du dispositif télévisuel : placement de caméra, décors, mouvements, etc.

#### **2.4. Temps sociaux**

La quatrième catégorie de temps qui nous intéresse concerne les temps liés au document audiovisuel en tant qu'il est document dans un monde, c'est-à-dire qu'il se rapporte au monde, appartient à un contexte de diffusion, est lié à d'autres documents au sein d'un corpus, etc. Nous qualifions ces temporalités de sociales, au sens où d'une part le document se retrouve dans la société d'autres documents (grille, corpus, collection) du fait d'une exploitation socialement située (diffusion : temps de la grille de programme, de la diffusion ; archivage : temps de l'intégration dans une archive, temps des accès ; réutilisation : temps de l'agrégation à d'autres documents) et où d'autre part le document réfère au monde réel et aux temporalités qu'on peut y considérer (« temps réel » : notion de direct/faux direct/différé, temps physique<sup>7</sup>, temps historiques, etc.).

Les temps sociaux peuvent se retrouver dans des modèles informatiques : dans TV-Anytime [TVA03] pour une grille de programme, dans un système d'information archivistique ou un modèle de génération de documents à partir de sources multiples.

#### **2.5. Discussion**

Le temps du flux est consubstantiel à l'expérience audiovisuelle. Le temps du document relève d'un objet ayant une durée. Les temps analytiques procèdent d'une activité de structuration et de description de contenu qui s'inscrit dans des formats d'annotations. Les temps sociaux considèrent le document en tant qu'il est inséré dans une société de documents et lié au temps réel et historique.

Tout d'abord, ces catégories de temps que nous proposons ont pour la plupart déjà été introduites, mais de façon disjointe. Le temps du flux est par exemple étudié par les chercheurs qui s'intéressent à la qualité de service dans la diffusion AV, afin d'éviter les ruptures de l'expérience audiovisuelle. Les chercheurs en SIAV, étudiant la possibilité de requêtes temporelles sur des fragments de documents, considèrent en premier lieu le temps des documents. Un grand nombre de modèles de description de documents audiovisuels vise à décrire ceux-ci suivant des structures temporelles analytiques (enchaînement de scènes pour une série télévisée, progression narrative, décomposition d'un JT, etc.). Les temps sociaux sont également l'objet de recherches actives au sens des études médiatiques (par exemple : temps du programme télévisuel en tant que tel, direct et temporalité, faux direct, différé, etc.<sup>8</sup>). D'autres auteurs enfin différencient plusieurs sortes de temps dans les documents hypermédias en général, tels [LUE 98] ou [HAR 99], nous y reviendrons dans la partie 5.3.

Remarquons ensuite que les catégories de temps audiovisuels présentées ne se veulent ni définitives, ni étanches. Il est possible de considérer par exemple comment

---

7. Temps des horloges, mesurable.

8. Voir par exemple le Forum *Les temps des médias* organisé par l'INA en juillet 2002 – <http://www.ina.fr/forum/index.fr.html>

le temps du direct réfère au monde directement dans le temps du flux : par exemple on peut *vivre* l'actualité<sup>9</sup>. On peut également étudier comment les temps analytiques audiovisuels peuvent être liés au temps social d'une grille de programmes (considérer par exemple un journal télévisé comme un document en soi, ou comme une agrégation documentaire) ; ou encore comment temps historique et diégétique peuvent mutuellement se façonner, etc.

Nous nous intéresserons désormais à ces temps audiovisuels uniquement si d'une part ils peuvent être saisis dans un système informatique à l'aide de structures d'informations disponibles dans et autour des documents audiovisuels numériques, d'autre part s'ils sont utilisés informatiquement en tant que tels (*i.e.* représentants informatiques de temporalités).

### 3. Hypervidéos : définitions

Dans le contexte des systèmes d'information audiovisuelle, notre objectif est de considérer les documents audiovisuels en compagnie des structures d'information qui viennent les compléter pour en permettre l'exploitation (recherche, visualisation, exploration, etc.). Plus particulièrement, il s'agit pour nous de se focaliser sur les documents hypermédias construits à partir des documents audiovisuels et des structures d'enrichissement, que nous appellerons hypervidéos.

Pour cela, nous définirons tout d'abord un *document audiovisuel annoté* (DAA) comme un document AV enrichi d'une structure d'annotation. Nous définirons ensuite une *vue* comme une façon de visualiser une partie du DAA. Ces deux définitions nous permettront d'introduire la notion d'*hypervidéo*, qui sera alors une vue particulière utilisant la structure d'annotation et donnant accès à la temporalité du flux issu du document audiovisuel.

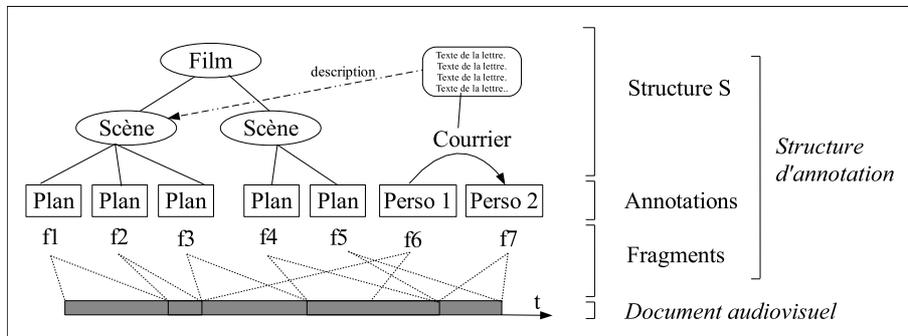
**Document audiovisuel annoté.** On appelle document audiovisuel annoté DAA un document audiovisuel  $D_{AV}$  enrichi d'une structure d'annotation  $S_A$  à portées spatio-temporelles dans le document. La structure d'annotation est la structure d'enrichissement du document audiovisuel, composée d'annotations  $a_i$  liées à des fragments spatio-temporels<sup>10</sup>  $f_i$  du document AV, ainsi que de la structure  $S$  mettant en relation les annotations.

$$DAA = \{D_{AV}, S_A\} = \{D_{AV}, \{\{a_i, f_i\}, S\}\} \text{ avec } f_i \in D_{AV}$$

9. Plus encore, on peut aussi vivre le flux *en même temps* que des millions de personnes, voir [STI 04] qui considère le danger d'intérioriser en masse des « objets temporels industriels » (des comportements de consommation), hypersynchronisation unifiant les consciences, et conduisant à la perte d'individualité.

10. Un fragment spatio-temporel d'un DAV est toute partie spatio-temporelle qu'on peut en extraire : fragment temporel entre deux instants, zone dans une image, zone fixe sur plusieurs images, zone en mouvement, etc.

Par exemple (cf. figure 1) : un film codé dans un fichier MPEG2, durant quatre minutes, pourra être accompagné d'une structure arborescente en décrivant la structure logique séquence/plan (un film est composé de séquences, qui chacune ont un titre, et contiennent des plans), en spécifiant pour ces derniers leurs bornes temporelles ; ainsi que d'une description des personnages et de leurs relations (ici, *Perso1*, présent à l'image tout le 3ème plan et une partie du 4ème, envoie à *Perso2* une lettre, dont le texte est connu et décrit ce qu'il s'est passé à la première scène du film). Alors les annotations  $a_i$  contiennent les informations concernant les plans et les personnages, les fragments associés  $f_i$  sont des fragments temporels,  $S$  décrit le reste de la structure arborescente (film, scène, arborescence, relation entre personnages, texte de la lettre, etc.).



**Figure 1.** Exemple de document audiovisuel annoté

Remarquons que nous restons très généraux dans la définition de la structure d'annotation : celle-ci regroupe toutes les informations liées à des fragments de flux de façon explicite, ainsi que les interrelations entre celles-ci. Notons qu'elle peut contenir des informations à caractère temporel, par exemple une annotation sonore.

**Vue d'un DAA.** On appelle vue d'un document audiovisuel annoté toute « manière de voir »  $Mv$  ou « façon de visualiser » ce DAA. Une vue sera construite à partir 1/ d'informations issues du document  $D_{AV}$ , 2/ de la structure d'annotation  $S_A$ , et 3/ de ses informations propres de visualisation. Ces dernières informations se distribuent entre deux pôles : elles sont toujours *a minima* inscrites dans les codes des outils/programmes de rendu des vues  $O_j$ , et peuvent également l'être dans des instructions déclaratives du type feuilles de style  $FS_j$ .

$$V_j = Mv_{(O_j, FS_j)}(D_{AV}, S_A)$$

Deux critères permettent d'analyser les vues : d'une part le fait de donner accès ou non au flux temporel issu du DAA ; d'autre part le fait d'utiliser soit les informations issues uniquement de la structure d'annotation  $S_A$ , soit du document audiovisuel  $D_{AV}$ , soit les deux. Le tableau 1 présente les cas de figure possibles et en donne des exemples dans le cadre du DAA décrit figure 1.

	Pas d'accès au flux temporel issu du DAA	Accès au flux temporel issu du DAA
$D_{AV}$	A- Visualisation d'un résumé statique du film : tableau d'images extraites du film toutes les 10 secondes, construit automatiquement. Visualisation de l'enveloppe du son du film.	B- Visualisation simple du film dans un lecteur vidéo standard, muni des commandes classiques de manipulation (pause, début, fin, avance rapide, etc.).
$S_A$	C- Visualisation de l'arborescence documentaire film/scènes/plans. Visualisation des noms de personnages impliqués dans le film et du texte de la lettre envoyée.	<i>Non pertinent</i>
$D_{AV}$ + $S_A$	D- Visualisation d'une table des matières du film, arborescente, avec pour chaque plan, une image-clé correspondant à la première image du plan. Visualisation du texte de la lettre avec 3 images d'illustration extraites des 3 plans de la première scène (cf. le lien <i>description</i> ).	E- Visualisation d'une table des matières hypertexte du film avec lancement d'un lecteur vidéo par activation d'un lien sur une image-clé. Visualisation du texte de la lettre, avec illustration animée de la première scène. Visualisation du film, avec surimpression des numéros des plans en cours, et possibilité de naviguer de plan en plan. Visualisation du film remonté, avec seulement les apparitions de personnages (plans 3, 4, 7).

**Tableau 1.** Cinq types de vues pour un document audiovisuel annoté

Les quelques exemples présentés dans le tableau 1 illustrent différentes possibilités de génération de vues utilisant des informations issues du DAA, mais aussi ce que nous avons appelé « informations propres de visualisation ». Par exemple, un lecteur vidéo standard contiendra toutes ces informations dans son propre code (vue type B-), une feuille de style pourra suffire à présenter dans un navigateur web une arborescence issue d'un fichier d'annotation XML (vue type C-). Par contre, il conviendra d'utiliser en plus un outil d'extraction des images-clés des fragments temporels AV si l'on veut compléter cette vue (vue type D-). D'autres vues encore plus complexes mêlent plus étroitement visualisation de flux audiovisuel et présentation d'informations issues de la structure d'annotation. On doit alors disposer d'outils nouveaux, par exemple de lecteurs vidéos capables d'afficher toutes sortes d'informations supplémentaires sur l'image, munis de possibilités de navigation originales, de montage pendant la lecture, etc. Ce sont ces vues que nous désignerons sous le nom d'*hypervidéos*.

**Hypervidéo.** On appelle hypervidéo toute vue d'un DAA qui d'une part utilise à la fois la structure d'annotation  $S_A$  et le document audiovisuel  $D_{AV}$ , d'autre part donne accès au flux temporel issu du DAA (fragments).

Quelques remarques s'imposent. La première tient à notre différenciation entre ce qui est une hypervidéo et ce qui n'en est pas une, laquelle tient d'une part (hyper) de la notion de rupture de la lecture linéaire standard de la vidéo, donc d'utilisation de la structure d'annotation<sup>11</sup>, d'autre part (vidéo), à la conservation du temps du flux, c'est-à-dire du caractère temporel de l'expérience du flux originel.

La seconde a trait à la composition de vues. Une hypervidéo peut être considérée comme un ensemble de vues d'un flux audiovisuel annoté, pourvu qu'on les articule, c'est-à-dire qu'on fournisse un moyen de passer de l'une à l'autre ou qu'on les synchronise. Ainsi, une vue récapitulative peut donner accès à d'autres vues plus élémentaires, fournissant alors un point d'articulation de l'hypervidéo. Dans notre exemple, une vue avec deux liens, donnant accès à une table des matières (vue type C-), et à un lecteur vidéo positionné au début du film (vue type B-) est une hypervidéo, tandis qu'aucune des deux vues élémentaires considérées individuellement ne l'est.

La troisième remarque concerne la généralisation de la notion d'hypervidéo à plusieurs documents audiovisuels sources. Ainsi, on peut considérer plusieurs documents audiovisuels annotés  $\{DAA_j\}$ , documents audiovisuels  $D_{AV_j}$  accompagnés d'une ou plusieurs structures d'annotation  $\{S_{A_j}\}$  (s'il n'y en a qu'une  $S_A$ , celle-ci doit avoir des portées spatio-temporelles – des fragments – dans tous les documents). Alors, une vue de ces DAA est définie comme une façon de les visualiser, tandis qu'une hypervidéo est une vue utilisant la ou les structures d'annotation *et* les documents AV, et donnant accès aux fragments de ceux-ci. Encore une fois, il s'agit de ne pas perdre ce caractère audiovisuel originel des flux audiovisuels source.

La quatrième concerne nos propres travaux autour de l'environnement Advène<sup>12</sup> et de ses modèles. Sans entrer dans les détails, nous y considérons des documents AV dont des fragments sont annotés par des annotations pouvant être mises en relation. Trois principaux types de vues y sont pour l'instant distingués : vues essentiellement statiques dans un navigateur web, construites à partir du flux et des annotations à l'aide d'un langage à base de modèles ; vues essentiellement dynamiques, construites autour d'un lecteur vidéo contrôlable et muni de fonctionnalités avancées comme afficher du texte ou des graphiques SVG, effectuer des captures d'écran ; vues ad-hoc enfin, codées en dur (par exemple, une vue en ligne de temps présentant sur un axe temporel l'ensemble des annotations de la structure  $S_A$ ). On se reportera à [AUB 04] pour plus d'informations sur les modèles et les expérimentations en cours.

Enfin, il nous faut nous situer par rapport aux diverses acceptions de la notion d'hypervidéo. Cette notion est apparue dans [SAW 96] pour faire allusion à une expérience de construction filmique interactive (Hypercafé) visant à combiner la vidéo classique avec les possibilités hypertextuelles. Il s'agissait là de briser la linéarité filmique standard en naviguant dans une trame narrative, et de permettre des changements de points

11. Nous ne considérons pas la possibilité de navigation suivant une ligne de temps, telle que fournie par tous les lecteurs vidéo numériques, comme un caractère d'hypervidéo, car celle-ci ne s'appuie pas sur le contenu de la vidéo, mais simplement sur sa durée.

12. Annotation de DVDs Échangées sur le Net – <http://liris.cnrs.fr/advene>

de vue. Si ce terme est toujours utilisé pour désigner des films interactifs (terme qui est d'ailleurs préféré en français, cf. [REN 03]), d'autres auteurs [CHA 02] préfèrent l'utiliser en opposition au terme hypermédia, trop général. En effet, il ne s'agit pas seulement d'intégrer des morceaux-blocs de documents audiovisuels dans des constructions multimédias, que le lecteur peut jouer et arrêter au cours de sa navigation, mais bien de désigner une « vraie » intégration de la vidéo et de l'image animée en tant que telle dans les hypermédias, en étudiant de nouvelles possibilités de navigation et d'interaction avec le flux. La navigation dans des hypervidéos sur le web est un des sujets importants du moment, et pose des questions aussi bien ergonomiques en termes de navigation que techniques liées au système réparti et à la nécessaire qualité de diffusion [GRI 02].

Notre définition nous semble cohérente avec le faisceau de sens décrit par ces différents travaux : nous mettons l'accent sur le caractère audiovisuel de l'hypervidéo, tout en insistant sur la rupture de la linéarité rendue possible par la structure d'information supplémentaire servant de support à la manipulation et à l'intégration dans un document hypermédia, ainsi que par les vues elles-mêmes, et les connaissances plus ou moins déclaratives qu'elles incluent. Ceci nous semble pertinent pour au moins trois raisons. La première est qu'avec notre notion d'hypervidéo, nous sommes à même de récupérer et d'étudier dans un cadre unificateur beaucoup des tentatives existantes de mêler audiovisuel et exploitation interactive (par exemple des « bricolages » ad-hoc utilisant Flash dans des bonus de films DVD, des façons partagées de présenter vidéo et informations textuelles à la télévision, etc.). La seconde est que cette approche nous semble largement transdomaniale, au sens où nous ne nous intéressons pas spécifiquement à la fiction interactive ou à l'enseignement hypermédia, mais aussi à tout ce qui relève de l'indexation ou de la lecture active des documents, base d'archives, analyse historique d'émissions, commentaires filmiques, analyse de corpus d'interactions sociales, etc. Notre troisième raison est qu'en situant respectivement le plus clairement possible les documents audiovisuels, les structures d'annotation et les vues, il nous semble qu'il devient possible de considérer une ou des implantations génériques de ces concepts, permettant d'une part d'encourager le développement d'idées nouvelles, d'autre part d'ouvrir la voie à une méthodologie voire une ingénierie de création des hypervidéos.

#### **4. Temporalités et détemporalisation dans les (hyper-)vidéos**

Après avoir défini ce que sont les vues de documents audiovisuels annotés (notamment ces vues particulières que nous nommons hypervidéos), nous nous proposons dans cette partie d'en présenter certaines caractéristiques mettant en jeu les différentes temporalités audiovisuelles que nous avons définies précédemment. Ceci nous permettra de présenter quelques éléments d'une « grammaire hypervidéo » encore à définir, et d'illustrer la très grande complexité potentielle des documents qu'il est possible de construire.

Nous discuterons de ces caractéristiques relativement indépendamment de l'usage qui est fait des hypervidéos (que ce soit par exemple voir une œuvre à base de vidéo, discuter un film, faire une recherche de document AV dans une archive et visualiser des résumés, naviguer dans une base de documents, etc). L'objectif est de se situer du côté du document numérique dans l'objectif d'une ingénierie, point de vue adopté dans notre définition d'hypervidéo comme vue d'un ou plusieurs flux audiovisuels enrichis.

#### **4.1. Jouer avec le temps du flux**

Toute vue d'un DAA donnant accès au flux temporel issu du document audiovisuel annoté donne la possibilité de *jouer* avec le temps du flux (vues de type B et E).

Ainsi, on peut accélérer ou ralentir celui-ci, ce qui procure une perception audiovisuelle différente de la perception canonique, pouvant tendre à l'incompréhensible ou faire naître de nouveaux modes de compréhension. Le dispositif technique utilisé n'est évidemment pas neutre dans le rendu audiovisuel, notamment le format numérique de représentation des documents AV (par exemple, le codage bidirectionnel de MPEG a pour conséquence qu'il est aisé d'accéder à certaines images, mais plus difficile pour d'autres). Dans le cas du ralenti, on utilisera moins d'images par seconde qu'un flux normal, et le rendu pourra tendre à la suite d'images, perdant une partie de sa continuité, tandis que le son (parole, musique, bruit) dépassera la limite du compréhensible. Dans le cas de l'avance rapide, une image sélectionnée périodiquement pourra appartenir à un flux présenté à un certain rythme<sup>13</sup>, tandis que le son montera en ton, rendant vite la parole inaudible<sup>14</sup>.

Parmi les autres possibilités de jeu avec le temps du flux et avec l'expérience physique de l'audiovisuel figurent l'arrêt, qui suspend le flux sur une image, tandis que disparaît le son ; l'inversion de la temporalité due au jeu en sens inverse ; mais aussi le bouclage, dans la mesure où il permet de monter en boucle un court fragment audiovisuel (par exemple moins de quelques secondes), lequel tend à acquérir une qualité temporelle propre du fait de la répétition.

#### **4.2. Spatialiser une temporalité audiovisuelle**

La vue la plus naturelle d'un DAA consiste à présenter le flux audiovisuel correspondant dans une fenêtre. De nombreux outils proposent un enrichissement de cette vue, sous la forme d'une spatialisation des temporalités qui s'y manifestent, qu'elles

---

13. Dans le domaine de la visualisation de résumés, [WIL 03] observe ainsi que les utilisateurs trouvent pertinent le fait de présenter à un rythme standard de 24 images/seconde un image extraite sur 64.

14. Mais pas forcément inutilisable, il reste possible d'identifier un changement de locuteur, de différencier musique et dialogue, etc.

soient liées au document AV (temporalité du flux, du document) ou externes (temporalités analytiques, sociales).

Toute temporalité peut être spatialisée, pourvu que sa description numérique soit munie d'une échelle de temps, et d'instantanés ou d'intervalles situés sur celle-ci. On devra alors se demander d'une part comment représenter l'échelle de temps et la forme des instantanés ou intervalles qui s'y rapportent ; d'autre part comment représenter l'information liée aux instantanés ou intervalles audiovisuels eux-mêmes.

Les lignes de temps sont la réponse standard à la première question : de direction horizontale (de gauche à droite) ou verticale (de haut en bas), éventuellement contraintes par les limites du support de représentation. Elles peuvent représenter un temps continu (se reporter à une échelle), ou bien se contenter d'aligner des moments, ne représentant alors qu'une succession.

#### 4.2.1. Représenter un fragment temporel audiovisuel

La question de la représentation des contenus audiovisuels correspondant aux instantanés et intervalles considérés est plus difficile. Cette représentation est la plus souvent *statique*, et si elle ne peut prendre appui sur les informations de la structure d'annotation (qui documentent les fragments) elle devra obligatoirement utiliser les informations audiovisuelles, au premier rang desquelles les images.

L'utilisation d'une image comme représentant (voire index si elle permet d'y accéder) d'un fragment temporel duquel elle est extraite est en effet répandue. Comme le remarque [BAC 99], cela résulte du dispositif technique qui discrétise le flux, « ce sont les unités signifiantes proposées par le système technique, à l'instar des linguistiques qui ne considèrent que les mots, unités prescrites par le découpage technologique de l'écriture. »

Très souvent, l'image choisie pour représenter un fragment est alors la première (à laquelle on pourra adjoindre la dernière). Le cas plus particulier du plan filmique<sup>15</sup> a cependant été le plus étudié. Des travaux s'attachent ainsi à essayer de détecter automatiquement « le bon représentant » d'un plan, en se basant par exemple sur des caractéristiques visuelles (choisir l'image où il y a le plus de mouvement) ou de filmage (choisir la dernière image d'un zoom arrière). D'autres essayent de saisir une part de temporalité des plans, par exemple en construisant par interpolation des images nouvelles, prenant en considération les mouvements de la caméra ou des objets [MAS 96].

Malgré la richesse de ces travaux, la question de la possibilité de représenter un fragment temporel à l'aide d'une image n'est cependant pas vraiment résolue en général. Si elle reste de grande importance dans l'objectif de construire des « résumés vidéos » permettant de décider rapidement si un document est d'intérêt ou non lors

---

15. Unité de montage audiovisuel, correspondant à un enregistrement continu de la caméra. En tant que telle, souvent utilisée comme unité *a priori* d'analyse et de représentation dans les systèmes à base de documents audiovisuels numériques, d'autant plus que les changements de plans sont assez facilement détectables automatiquement.

d'une recherche, le type de document et l'utilisation qui en est prévue joueront également un rôle non négligeable. Dans le cas d'images extraites de documents déjà visualisés par un utilisateur, il est certain qu'elles pourront avoir un rôle de représentant, du fait du rappel induit chez celui-ci.

En ce qui concerne les représentations *dynamiques* de fragments, visant à garder tout ou partie de la temporalité du flux, on pourra considérer comme représentants des extraits de fragments audiovisuels joués en boucle, ou bien des extraits sonores de fragments associés à des images fixes, qui se jouent lors du survol de l'image. Ces représentants pourront à ce titre être alignés sur une ligne de temps, et on en trouvera des exemples dans certaines pages d'accès aux chapitres de DVD.

#### 4.2.1.1. Temporalités audiovisuelles et spatialisation

Il semble relativement peu pertinent de spatialiser la *temporalité du flux*, du fait de son caractère quasi instantané. La ligne de temps ne peut en effet y être que locale et centrée sur l'instant présent, tandis que défilent les informations de spatialisation du flux AV. On peut cependant en trouver une illustration dans le domaine sonore, en considérant par exemple la visualisation d'une onde sonore, qu'un spécialiste sera à même d'exploiter. Dans le domaine de l'image, on pourrait imaginer de présenter à un utilisateur des éléments de métrique AV autour de l'instant présent s'appuyant par exemple sur une « quantité de mouvement instantanée ».

La *temporalité du document* audiovisuel, qui tient à sa durée, est souvent spatialisée dans des vues dynamiques, par exemple dans les lecteurs audiovisuels (vues type B) où une ligne représente les limites du document, tandis qu'une marque situe l'instant présent du jeu, permettant de le contextualiser temporellement par rapport au document dans son ensemble. Remarquons que le fait que la marque se déplace peut alors être considéré comme spatialisation du temps du flux.

La temporalité du document est également souvent spatialisée dans des vues plus statiques. Elle peut prendre la forme d'une ligne de temps indiquant des timecodes de début et de fin du document, sur laquelle sont placés des éléments visuels représentant l'instant ou le fragment du flux audiovisuel. Elle peut également s'affranchir de l'échelle temporelle en effectuant simplement une disposition séquentielle des éléments. Comme vu précédemment, ces éléments seront souvent des images extraites automatiquement du document, qui permettront de construire des résumés imagés du document AV. Certains pourront être très évolués, par exemple une vue telle que décrite dans [GIR 01] affiche en permanence un tableau composé d'un nombre fixe d'images extraites (à intervalles réguliers) du document entier, ou bien du fragment sur lequel on choisit de zoomer. On pourra également extraire des représentants visuels du contenu sonore du document pour proposer une représentation de l'enveloppe d'un son, qui pourra être utile par exemple pour le calage temporel de transcriptions, pour lesquelles il convient d'être précis au centième de seconde. Si on bénéficie d'une structure d'annotations, on pourra visualiser les contenus des annotations le long de la ligne de temps.

Dans le cas des *temporalités analytiques*, la ligne de temps pourra prendre appui sur l'échelle du temps analytique, ou bien utiliser celle du temps du document. Dans le premier cas, on pourra par exemple avoir une représentation du temps diégétique d'un film indépendante du temps du document filmique, en considérant les événements de l'histoire situés par rapport à leur propre échelle de temps<sup>16</sup>. Dans le deuxième cas, on pourra souhaiter représenter le temps diégétique en relation avec la construction du film, et l'ordre d'apparition des éléments (ce qui permet d'analyser le style du film). Un autre exemple concerne le cas du travail sur une interaction interpersonnelle filmée, où le temps de l'interaction analysée (tours de paroles, gestes) sera lié au temps du document audiovisuel, du fait de son statut d'enregistrement de la réalité.

Terminons par un exemple de spatialisation de *temps social*, disponible sur une base de reportages télévisés indexés dans le cadre du projet Informedia [CHR 02]. En fonction de mots-clés, des imageries représentant les résultats de recherche sont alignés sur une ligne de temps social chronologique divisée en années et mois.

### 4.3. *Hypertextualisation et ruptures temporelles*

Les documents hypertextes « classiques » permettent l'intégration de documents audiovisuels, mais ceux-ci y sont en général considérés dans leur totalité, comme des objets audiovisuels qu'il convient d'intégrer comme des tous. Ainsi, au détour d'une page hypertexte, on lancera une vidéo, qu'on pourra manipuler comme telle, avant de repartir dans une navigation standard.

La question de l'hypertexte dans les hypervidéos va beaucoup plus loin. A partir du moment où le noyau d'une hypervidéo comprend un ou plusieurs documents audiovisuels, l'enjeu est plus d'hypertextualiser les documents audiovisuels, de les accompagner d'un matériau hypertexte, que de les intégrer dans des documents hypertextuels. Le flux audiovisuel doit alors être étroitement lié aux possibilités de navigation, de documents statiques vers des fragments audiovisuels et vice-versa, mais aussi entre fragments audiovisuels. C'est en ce sens qu'il est instrumenté.

Du point de vue des temporalités audiovisuelles, les premières possibilités de navigation dans *un* document audiovisuel concernent la possibilité d'utiliser la ligne de temps du lecteur vidéo pour atteindre un instant repéré sur celle-ci. Il s'agit d'une navigation dans le temps du document qui se caractérise par un contrôle direct de l'utilisateur, qui doit gérer lui-même les conséquences des ruptures de flux qu'il induit.

Lorsqu'une structure d'annotation est présente – qui donc permet le repérage de fragments temporels – les représentants des fragments (images, contenus des annotations) sont autant d'index qui peuvent être utilisés comme liens hypertextuels, de na-

16. Ce qui permet par exemple de comparer les trois films de la trilogie de Luc Belvaux <http://www.uncouplepatant-cavale-apreslavie.com/index2.html> et leurs points de synchronisation par rapport à la chronologie générale.

vigation d'un document statique (ligne de temps, texte, etc.) vers les fragments considérés (en général le début). Les ruptures de temporalités principales introduites ici tiennent au passage d'une proposition de navigation (lien statique) à la visualisation d'un flux (dynamique).

Si la structure d'annotation offre de plus la possibilité de mettre en place des liens entre fragments, il devient possible de considérer des fragments comme les ancres de liens hypertextuels, tandis que d'autres seront des destinations. La question de la navigation dans un document, ou dans un ensemble de document se pose alors, qui commence à être étudiée en tant que telle [CHA 02]. On imagine en effet les problèmes de visualisation et de manipulation que cela entraîne : comment présenter un ou plusieurs liens avec une ancre temporelle (*hotspot*), comment naviguer, comment revenir en arrière, etc. La rupture de temporalité se fait ici dans la continuité du flux audiovisuel visionné. Cette rupture peut cependant être médiée par la manière de présenter le lien, ainsi que par l'action de navigation. Une navigation suivant une temporalité analytique peut ainsi fournir une continuité dans la rupture.

#### **4.4. *Soumission et insoumission temporelle***

Ce que nous appelons soumission temporelle à un document audiovisuel correspond à la synchronisation du flux de conscience du spectateur avec le temps du flux audiovisuel. Ce phénomène tient à plusieurs facteurs. Le premier est celui de l'expérience physique audiovisuelle, liée au temps du flux, qui est imposée par l'enchaînement des flux visuel et sonore, et qui d'une certaine manière tient de la fascination réflexe pour ce qui bouge et fait du bruit. Le second tient au document lui-même, en tant qu'artefact, qui a une certaine durée, et a en général été réalisé, monté avec art en vue de maintenir l'intérêt du spectateur du début à la fin.

La rupture de l'expérience de temps du flux et du document que constitue l'arrêt du document dans le cadre d'une hypervidéo met en valeur cette soumission. Par exemple quand en classe un film s'arrête à un instant annoté par le professeur ménageant une pause pour que celui-ci puisse poser une question [AUB 03] ; ou dans le cas d'un film interactif lorsque le « spect-acteur n'est plus dans un espace symbolique libéré des contraintes de la réalité [mais] face à une image proposant des éléments présents aux fonctions tactiles » [REN 03], expérience limite du cinéma dans laquelle le spectateur quitte un court instant le flux et l'histoire auxquels il se soumet.

La lecture d'une hypervidéo peut profiter de la structure d'annotation qui est associée au(x) flux pour proposer un résumé d'un document (on s'abstrait du temps du document), un remontage temporel automatique d'un film (le lecteur jouant par exemple des fragments dans un ordre défini par une description analytique du temps diégétique), voire un montage nouveau de fragments de document différents. Alors la soumission de l'utilisateur à la temporalité tient du temps du flux, mais aussi du temps du document et du temps analytique utilisés pour la construction.

L'insoumission temporelle permise par les hypervidéos tient à la possibilité de contrôler le lecteur audiovisuel, de naviguer, de négliger une fenêtre jouant un document audiovisuel pour s'intéresser à un document statique non loin (alors qu'il est plus difficile d'en négliger la bande sonore), de s'intéresser à un flux plutôt qu'un autre quand plusieurs sont présentés en même temps, voire de manipuler une temporalité analytique pour créer son propre remontage temporel d'un document audiovisuel (attitude par laquelle l'utilisateur se rapproche du créateur de l'hypervidéo).

Alors que la soumission temporelle tient à une réception passive d'hypervidéo qui impose ses temps (temps originel et temps de la vue du flux annoté), l'insoumission tient aux possibilités d'action et de navigation offertes par l'extériorisation (par les liens et les vues statiques) des structures d'enrichissement des flux audiovisuels.

#### **4.5. Temporalités audiovisuelles et hypervidéos**

Il nous apparaît au terme de cette courte étude qu'instrumenter les documents audiovisuels aux sein de systèmes d'information audiovisuelle pour les manipuler comme des documents multimédias (que nous avons appelés hypervidéos) met au jour de nombreuses caractéristiques temporellement intéressantes : jeu avec le temps du flux, spatialisation des documents suivant des temporalités internes (temps du document) ou externes (temps analytiques ou sociaux), hypertextualisation des documents audiovisuels entraînant des ruptures temporelles, multiplication des possibilités d'insoumission ou de nouvelles soumissions aux temporalités variées des hypervidéos.

Remarquons que dans le cadre d'hypervidéos complexes, il est imaginable de faire jouer de façons multiples les principes évoqués. Imaginons par exemple qu'on ralentisse le temps de jeu d'un fragment filmique de deux secondes à quinze secondes, pour proposer par dessus l'image plusieurs ancres hypertextes (par exemple des boutons), dont les fragments de destination sont liés à l'analyse diégétique du document original, en rapport avec le moment diégétique courant. Supposons de plus que chacune des ancres clignote à son tour, tout en proposant un aperçu sonore de trois secondes du contenu de son fragment de destination : les temporalités évoquées sont multiples et entremêlées, et difficiles à analyser. Sans aller jusqu'à ces extrêmes qui restent à étudier, retenons que dans le domaines des SIAV et des hypervidéos, au moins deux directions de recherche sont actuellement actives, une première concernant les interfaces de visualisation rapide de documents audiovisuels, l'autre proposant les linéaments d'une analyse de la navigation hypertextuelle au sein de documents audiovisuels.

### **5. Un pas vers un cadre d'analyse des hypervidéos**

Potentiellement, une hypervidéo peut devenir un document multimédia extrêmement complexe. Qu'on imagine par exemple de généraliser l'affichage de classiques sous-titres textuels à une présentation de multiples annotations, aussi bien textuelles que graphiques, spatialement et temporellement situées sur un flux audiovisuel, éven-

tuellement munies de leur propre temporalité (son, mouvement). Qu'on imagine encore des documents mêlant à la fois présentation de plusieurs flux audiovisuels de façon simultanée, éventuellement incrustés les uns dans les autres, et présentation d'informations (hyper) textuelles relevant de ces flux, proposant des possibilités multiples de navigation. La complexité potentielle extrême des hypervidéos entraîne que dans le champ des possibles, seules quelques formes se révéleront pertinentes.

S'il est certain que c'est des seuls usages que pourront *in fine* émerger ces formes, voire des genres d'hypervidéos stables, en parallèle avec le développement éventuel de nouvelles compétences de lecture (on doit bien apprendre à lire des sous-titres tout en regardant un film en version originale), il nous semble qu'on peut d'ores et déjà essayer de les appréhender de plusieurs façons.

Une première consiste à réaliser des hypervidéos, et à en constater à l'usage les bonnes et les mauvaises propriétés<sup>17</sup>. Une seconde consiste à reconnaître et à étudier les hypervidéos telles qu'elles se trouvent dans la nature (bonus de DVD, sites web, télévision, films expérimentaux, installations artistiques interactives, etc.) afin d'en étudier bonnes et mauvaises caractéristiques, qu'on sera à même de reproduire. Une troisième, aura pour objectif de comprendre et de modéliser, au moins en partie, la lecture des hypervidéos, permettant d'en dégager une certaine prédictibilité de la validité.

Il va de soi que toute pratique procédera plus ou moins de ces trois démarches en parallèle. Cependant, dans le champ de la troisième, nous voudrions présenter les premières idées d'une approche qui nous semble d'intérêt pour la compréhension de l'articulation des temporalités audiovisuelles, et des différentes dimensions que nous avons mises en évidence dans la section précédente.

### 5.1. Flux du jeu et flux vécu

L'enjeu est donc d'être à même de comprendre la lecture des hypervidéos par l'utilisateur dans les cas complexes, pour lesquels les possibilités perceptives et intégratives humaines risquent d'être sollicitées de façon soutenue. Alors il devrait être possible de prévoir ce qu'il est raisonnable ou non de proposer dans une hypervidéo et dans quelle mesure.

Notre proposition de démarche d'analyse nécessite de considérer à la fois le « point de vue de la machine » (ce qu'il s'y passe) sur les hypervidéos générées, et celui de l'utilisateur dans la lecture et la réception de celles-ci. Dans un premier temps, nous ne nous intéressons pas en effet aux documents tels qu'ils sont décrits dans les modèles documentaires et les vues et outils de lecture qui leur sont associés, ni au niveau de

---

17. Dans le cadre de la prise de connaissance rapide (*browsing*) d'une bibliothèque numérique vidéo, et considérant que ces nouvelles interfaces restent arbitraires et non prédictibles, [LEE 02] applique une démarche analytique systématique pour construire des interfaces, testées systématiquement.

leur conception. L'analyse portera plus directement sur les *documents tels qu'ils sont effectivement joués et tels qu'ils sont vécus*, les résultats pouvant dans un deuxième temps avoir une influence sur les modèles et les techniques de conception.

Un point nodal de l'analyse nous semble résulter du caractère audiovisuel du matériau principal des hypervidéos, qui relève du temps du flux tel que nous l'avons décrit, de l'expérience audiovisuelle, de la soumission au flux du lecteur, en bref de la synchronisation entre le flux de conscience du lecteur et le flux du jeu audiovisuel. Notre hypothèse sera que *sous certaines conditions, cette synchronisation peut être attestée ou prévue, et contribuer à la compréhension – peut-être à la prévision – des rapports entre un document numérique hypertextuel (l'hypervidéo) et un humain (le lecteur)*.

Nous considérons pour cela deux flux temporels, que nous nommons flux du jeu et flux vécu.

Le *flux du jeu* correspond au document tel qu'il est effectivement joué par le lecteur, à l'*instance d'hypervidéo*, ou *hypervidéo générée*. Il procède du temps propre du document multimédia généré. De ce point de vue, on observe l'enchaînement des événements automatiques ou déclenchés par l'utilisateur. Le début et la fin de l'affichage d'éléments (fragment de vidéo, texte, image...) composant le document hypervidéo constituent par exemple des événements. Certains éléments du flux du jeu seront des éléments audiovisuels (morceaux du document AV annoté, annotations à caractère audiovisuel, éléments dynamiques générés). On pourra, si le besoin s'en fait sentir pour l'analyse, considérer des sous-éléments et événements (par exemple des éléments images d'une vidéo, et les événements d'affichage-enchaînement de nouvelles images). Cependant, le caractère particulier de l'expérience audiovisuelle nous incitera à les considérer globalement le plus souvent. Certains éléments du flux du jeu seront des éléments plus « conceptuels », car ils auront été définis à partir de structures d'annotation et/ou des définitions des vues et outils d'affichage.

Le *flux vécu* correspond au temps vécu du document multimédia généré et est en relation avec le flux de conscience de l'utilisateur et l'*hypervidéo vécue*. Il s'agit d'observer le document vécu par l'utilisateur de manière passive (en termes de soumission à ce qui se déroule automatiquement au niveau des flux audiovisuels et des changements automatiques ; et de reconnaissance des possibilités d'actions<sup>18</sup>) et de manière active (actions effectives de l'utilisateur).

L'objectif de l'analyse en flux du jeu et flux vécu est d'étudier (observer, comprendre, prévoir) les moments de synchronisation, de désynchronisation, de rupture, de surprise, de prise en contrôle par l'hypervidéo ou par le lecteur humain de la manière dont se poursuit l'expérience multimédia.

Les éléments qu'il s'agira de prendre en compte dans l'analyse seront d'abord une *temporalité commune* aux deux flux, qui sera celle de l'observation, qu'on représentera avec deux lignes de temps horizontales parallèles ; ensuite des *événements*, tant

18. Ceci pourrait être fait par exemple par analyse oculométrique : certaines études existent déjà, pour analyser les mouvements de regard sur un flux audiovisuel.

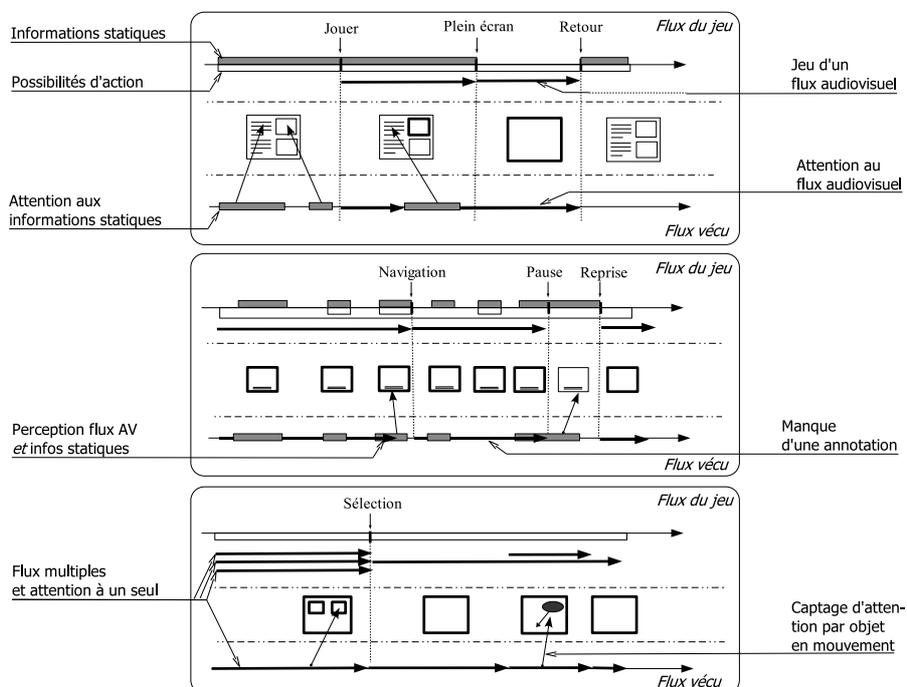
du côté flux du jeu que flux vécu. Les événements déterminant des états, on représentera si besoin ceux-ci pour clarifier la compréhension. Par exemple du côté du flux du jeu, l'état d'un écran statique, ou bien le fait qu'un flux vidéo se joue. Du côté du flux vécu, de l'utilisateur, un état pourra être un état d'attention à un flux vidéo, de visualisation d'une image, etc. Enfin, des *synchronisations* entre les deux flux, qui attesteront leurs influences réciproques. La principale synchronisation découlera de la soumission audiovisuelle au flux du jeu.

### 5.2. Trois exemples d'analyse en flux du jeu / flux vécu

La figure 2 présente quelques exemples d'analyse en flux du jeu et flux vécu, suivant un premier formalisme graphique d'analyse. La ligne de temps du flux du jeu (en haut) est complétée par l'indication de la présence à l'écran d'informations statiques (fragments sombres) ; l'indication de la présence à l'écran d'un ou plusieurs flux vidéo en train de se jouer (flèches grasses) ; l'indication de la présence dans l'hypervidéo générée de possibilités d'action pour l'utilisateur (fragments clairs) ; l'indication des actions utilisateur (flèches verticales). La ligne de temps du flux vécu est complétée par un indicateur d'attention aux informations statiques (fragments sombres) ; un indicateur d'attention aux flux vidéo (flèches grasses) ; des explications éventuelles sur ce sur quoi porte l'attention. Entre les deux lignes de temps se trouve toute vue explicative résumée de l'écran permettant de mieux comprendre et expliquer l'analyse (on pourra envisager d'y mettre des copies d'écran).

Dans le premier exemple, un document statique comprend le script d'un film : pour chaque plan, une description textuelle est accolée à une image-clé de ce plan, sur laquelle l'utilisateur peut agir pour jouer le plan correspondant à la place de l'image. Il peut également passer le flux AV en plein écran, et revenir au document statique initial. C'est ce qu'il se passe dans le scénario décrit. Une première action « jouer », démarre le film à la place de l'image fixe. Ce flux audiovisuel joué est représenté par un rectangle gras dans la vue explicative, ainsi que par la flèche grasse, qui représente le temps du flux en train de se jouer. La deuxième action passe en plein écran, et l'action de retour ramène à l'état statique initial. Les indicateurs d'éléments statiques concernent les descriptions textuelles de plan, les indicateurs de possibilité d'action concernent la possibilité de jouer un extrait, de l'arrêter, de passer en plein écran, etc. La ligne de temps du flux vécu comprend des fragments sombres, qui montrent l'attention du lecteur à telle ou telle zone du document, les flèches grasses signalent la lecture du flux, c'est-à-dire la soumission à celui-ci. On remarque que le lecteur regarde d'abord le texte d'un plan, puis l'image, puis le flux en vignette, s'en détache pour revenir au texte, reporte son attention sur le flux, qui passe en plein écran sans rupture d'attention notable.

Le second exemple montre un outil de lecture audiovisuelle permettant de présenter des annotations (fragments sombres) superposées à un flux audiovisuel. Certaines annotations offrent la possibilité de liens (par exemple la deuxième et la troisième, fragments clairs), tandis que le lecteur a tout loisir d'arrêter le lecteur quand il le veut



**Figure 2.** Trois exemples d'analyse d'hypervidéos selon flux du jeu et flux vécu

(fragment clair global). Dans le scénario, une action de navigation a lieu, ainsi que la mise en pause du jeu, puis sa reprise. Au niveau du flux vécu, l'utilisateur y est soumis (flèche grasse) et lit les annotations en même temps, sauf quand il s'intéresse plus précisément à l'un d'entre eux, perdant le contact fort avec le flux avant son action de navigation. Après celle-ci, il manque une annotation (ne la voit pas). Plus tard, il met en pause afin de lire plus précisément l'annotation, s'en détache, puis reprend le jeu.

Le troisième exemple montre un flux audiovisuel en plein écran, sur lequel deux vignettes présentent deux autres flux. La sélection d'un de ces flux passe celui-ci en plein écran, sur lequel se superpose automatiquement un objet en mouvement (objet audiovisuel). Le flux vécu considère qu'avant la sélection, un seul flux de vignette a été vu (sur les trois), puis que l'attention de l'utilisateur a été capté par le flux en plein écran, puis par le mouvement de la forme, puis à nouveau par le flux en plein écran.

Il va de soi que ce que nous ne prétendons pas ici présenter un langage graphique complet, encore moins un formalisme de description des flux du jeu et flux vécu et de leurs interrelations<sup>19</sup>. Nous entendons simplement montrer l'intérêt potentiel d'une analyse des hypervidéos prenant en compte leur caractère audiovisuel et ses liens avec

19. D'une part il n'est pas certain qu'une formalisation trop rigide soit nécessaire quand il s'agit d'abord de comprendre ce qu'il se passe lors de la lecture, d'autre part il convient maintenant

la manière dont un flux est reçu et dans une certaine mesure *capte l'attention* de l'utilisateur. Il s'agit alors d'être à même de mieux apprécier comment une hypervidéo générée (description documentaire + vue + programme) est ou pourra être vécue (utilisateur). Dans les cas les plus simples, les flux se recouperont (documents statiques, simple lecture audiovisuelle) sans grand intérêt. Dans d'autres cas, l'analyse permettra d'envisager la compréhension de la réception du document, et donc d'influer sur sa conception, voire sur la conception des modèles documentaires sur lesquels celle-ci s'appuiera.

### 5.3. Discussion

La proposition que nous faisons trouve des échos dans un certain nombre de domaines de recherche.

Parmi les nombreux travaux existant sur la question de la temporalité et des hypermédias, certains auteurs ont cherché à catégoriser différents temps, aussi bien liés aux documents (temps narratif, diégétique, etc.) qu'à la lecture elle-même. Par exemple, [LUE 98] différencie le « temps interface », lié à la lecture, du « temps cognitif », lié au récit. Le temps interface peut s'analyser suivant le « temps mécanique » (attente d'une réponse de la machine ou du réseau) ; le « temps de lecture » (passé à lire l'hypertexte) ; et le « temps interactif » (passé à interagir avec le document). Cette analyse se base sur des soumissions plus ou moins grandes de l'utilisateur à l'hyperdocument généré, mais ne considère pas le cas particulier de la vidéo, et de sa temporalité physique particulière. On pourrait sans doute considérer un « temps de visionnage » dont la soumission serait intermédiaire entre les temps mécanique et de lecture. Si la notion de flux vécu nous semble peu utilisée dans la communauté hypermédia, les études sur les temps des documents s'intéressent plus à ce que nous avons appelé flux du jeu. [HAR 99] différencie par exemple le « temps d'exécution » (*runtime*) du « temps du document » (ligne de temps du concepteur). Le flux vécu n'est ainsi pris en compte qu'implicitement à travers l'influence qu'il a sur le flux du jeu. Des langages comme SMIL [W3C 01] représentent l'aboutissement de travaux actifs de la communauté hypermédia pour la mise au point de langages de présentation de documents multimédias et d'outils visuels de programmation multimédia prenant en compte de façon intégrée les objets audiovisuels. Le contrôle de la temporalité finale du document revient cependant à l'auteur, et la temporalité vécue n'est pas étudiée en tant que telle.

Dans le domaine des hypertextes au contraire, un certain nombre de travaux se concentrent sur la réception, et la temporalité de l'utilisateur construisant son texte à partir des possibilités offertes. On parlera alors de lecture élargie, de sous-détermination *a priori* de la séquentialité, d'énonciation, d'appropriation, d'interactivité, avec une perspective dans laquelle l'utilisateur contrôle sa navigation, en se concentrant sur le lien hypertexte et ses promesses. Un modèle tel que présenté dans [BOO 03] pro-

---

d'appliquer l'analyse à des exemples d'hypervidéos de complexités variées afin d'éprouver ce premier langage graphique.

pose les notions de « texte-à-voir », « texte-écrit », « texte-lu » qui se rapprochent de nos flux du jeu ou flux vécu. L'interaction entre logique du concepteur et temporalité du vécu du lecteur est étudiée. Cependant, dans les hypervidéos, la composante dynamique du document, son aspect de temporalité *programmée* (nativement pour la vidéo, programmatiquement pour un déroulement multimédia) font que la machine dépasse la proposition hypertexte, pour également *disposer* du lecteur, fait qui doit être pris en compte.

Nous signalons ci-après plusieurs pistes intéressantes pour l'analyse des instances d'hypervidéos en lien avec les hypervidéos vécues.

Les théories cognitives et ergonomiques tout d'abord, théories de l'attention ou de la double tâche par exemple qui permettent d'analyser justement les liens entre l'attraction pour l'audiovisuel, le captage d'attention induit et la difficulté par exemple à lire des informations proposées en même temps (l'exemple des pages web et des publicités en Flash à côté de textes est à cet égard instructif). Un certain nombre de modèles cognitifs d'attention à vocation prédictive pour la conception d'interfaces ont été développés, mais ne conviennent que pour l'analyse d'éléments simples [SUT 00]. Des travaux dans le domaine de l'attention (*change blindness*, *inattention blindness*), la surveillance d'événements (*monitoring*), par exemple pour le contrôle aérien [NIK 01], pourraient également s'avérer utiles.

Le domaine des études sur le cinéma ou la télévision nous paraissent également d'intérêt, car de nombreuses expérimentations y ont été menées qui mériteraient d'être analysées à l'aune des hypervidéos comme évolution de ces médiums [GRO 02]. La question du cas limite de l'*animation multimedia* pourrait également être au cœur de la problématique, du fait d'un versant technologique plus transparent qu'un flux d'images (il est plus facile d'imaginer les objets tels qu'ils ont été conçus que les images qui s'enchaînent), mais aussi de la possibilité d'être lue/perçue comme de la vidéo, la dimension mouvement/son y apparaissant comme primordiale et dégagée de celle de « représentation de la réalité » propre au flux vidéo filmé. Enfin, le statut des jeux, des propositions qu'ils contiennent et des multiples temporalités qu'ils intègrent, de la notion de jouabilité même pourraient amener d'intéressants apports à l'étude et à la conception des hypervidéos.

## 6. Conclusion

Au terme de cet article, il nous faut retracer le chemin suivi, les propositions développées et les pistes ouvertes. Nous plaçant dans le contexte des systèmes d'information audiovisuelle (SI AV), nous commençons par dégager un certain nombre de temporalités liées aux documents audiovisuels (section 2) : temps du flux et du document, déjà inscrits dans les formats audiovisuels ; temps analytiques ensuite, qui doivent être inscrits comme annotations en accompagnement des documents ; temps sociaux enfin, qui résultent de l'inscription des documents dans des collections, et de leurs liens avec le temps historique. Nous constatons que l'entour des documents audiovisuels

dans les SIAV en permet une *instrumentation*, c'est-à-dire une intégration toujours plus fine dans des systèmes d'information autorisant une recherche et une utilisation dépassant la simple visualisation, notamment du fait de la rupture introduite par la possibilité de liens hypertextuels. Nous proposons ensuite (section 3) de désigner de façon générale ces documents audiovisuels instrumentés sous le terme d'hypervidéos, documents multimédias offrant des vues statiques et dynamiques sur un ou plusieurs flux audiovisuels. Nous étudions alors (section 4) les hypervidéos sous l'angle des temporalités précédemment définies, en proposant quelques réflexions sur les caractéristiques possibles de celles-ci. Dans notre dernière section (5), sur la constatation du caractère fondamental de l'expérience audiovisuelle et de la possibilité de synchronisation avec le lecteur qu'elle induit, nous proposons de considérer dans l'analyse des hypervidéos deux temporalités, qui sont le temps du jeu (lié à l'instance d'hypervidéo) et le temps du flux (lié à l'hypervidéo telle qu'elle est vécue), qui permettraient d'étudier plus avant les rapports entre les modèles documentaires d'hypervidéos et leur réception.

Les propositions faites dans cet article sont de plusieurs ordres, relativement indépendantes, et – étant donné la complexité du problème – ne doivent être considérés que comme un cadre de réflexion, qu'il convient maintenant de développer et d'éprouver. Nous y contribuons pour notre part au sein du projet Advène [AUB 04] dans lequel un modèle d'hypervidéo est proposé, à base d'annotation de documents audiovisuels, dont le prototype permet d'expérimenter l'instrumentation pratique de documents audiovisuels filmiques, mais aussi de corpus d'interactions communicatives et leurs analyses (en lien avec le laboratoire ICAR - ENS-LSH/Lyon2), et autour duquel se greffent divers études et projets.

Le début de cadre d'analyse proposé, s'il ne prétend pas permettre de spécifier comment une hypervidéo sera reçue dès le niveau de la conception, pourrait au moins permettre de définir les limites de ce que l'on peut faire et ne pas faire dans les hypervidéos standard. Ceci dans le cas général (on ne peut *a priori* expérimenter deux flux audiovisuels en même temps) et dans des domaines particuliers ou avec des outils particuliers, pour lesquels il est probable qu'une construction socio-technique de nouveaux rapports aux hypervidéos naisse, résultant en modèles, genres et contextes sociaux (tâches) adéquats.

## 7. Bibliographie

- [ALL 83] ALLEN J. F., « Maintaining knowledge about temporal intervals », *Communications of the ACM*, vol. 26, n° 11, 1983, p. 832-843.
- [AUB 03] AUBERT O., CHAMPIN P.-A., PRIÉ Y., « Instrumentation de documents audiovisuels : temporalisations et détemporalisation dans le projet Advène », *Actes Workshop "Temps et documents numériques"*, Grenoble (FR), 2003.
- [AUB 04] AUBERT O., CHAMPIN P.-A., PRIÉ Y., « Advène : une plate-forme pour l'édition et la visualisation d'hypervidéos », *Journée "Modèles documentaires de l'audiovisuel"*, La Rochelle (FR), Jun 2004.

- [AUF 99] AUFFRET G., PRIÉ Y., « Managing Full-indexed Audiovisual Documents : a New Perspective for the Humanities », *Computer and the Humanities, special issue on Digital Images*, vol. 33, n° 4, 1999, p. 319-344.
- [BAC 99] BACHIMONT B., « Bibliothèques numériques audiovisuelles : des enjeux scientifiques et techniques », *Document Numérique, numéro spécial Les Bibliothèques Numériques*, vol. 2, n° 3-4, 1999, p. 219-242.
- [BOO 03] BOOTZ P., « Construction du sens par les processus dans le modèle procédural », *Hypertextes hypermédias. Créer du sens à l'ère numérique. H2PTM'03*, Paris, 2003, p. 79-88.
- [CHA 02] CHAMBEL T., GUIMARAES N., « Context perception in video-based hypermedia spaces », *Proceedings of the thirteenth conference on Hypertext and hypermedia*, ACM Press, 2002, p. 85-94.
- [CHR 02] CHRISTEL M. G., HAUPTMANN A. G., WACTLAR H. D., NG. T. D., « Collages as dynamic summaries for news video », *Proceedings of the tenth ACM international conference on Multimedia*, Juan-les-Pins, France, Jun 2002, p. 561-569.
- [GIR 01] GIRGENSOHN A., BORECZKY J., WILCOX L., « Keyframe-Based User Interfaces for Digital Video », *IEEE Computer*, vol. 34, n° 9, 2001, p. 61-67.
- [GRI 02] GRIGORAS R., CHARVILLAT V., DOUZE M., « Optimizing hypervideo navigation using a Markov decision process approach », *ACM Multimedia*, Juan-Les-Pins, France, Dec 2002, p. 39-48.
- [GRO 02] GRODAL T., *Moving Pictures : A New Theory of Film, Genres, Feeling, and Cognition*, Oxford University Press, New York, 2002.
- [HAR 99] HARDMAN L., VAN OSSENBRUGGEN J., MULLENDER K. S., RUTLEDGE L., BULTERMAN D. C. A., « Do you have the time ? Composition and linking in time-based hypermedia », *Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots*, Darmstadt, Germany, 1999, p. 189-196.
- [LEE 02] LEE H., SMEATON A. F., « Designing the User Interface for the Físchlár Digital Video Library », *Journal of Digital Information*, vol. 2, n° 4, 2002.
- [LUE 98] LUESEBRINK M. C., « The moment in hypertext : a brief lexicon of time », *Proceedings of the ninth ACM conference on Hypertext and hypermedia : links, objects, time and space*, Pittsburgh, Pennsylvania, 1998, p. 106-112.
- [MAG03] « Media Access Generator (MAGpie) », 2003, <http://ncam.wgbh.org/webaccess/magpie/>.
- [MAS 96] MASSEY M., BENDER W., « Salient stills : process and practice », *IBM Systems Journal*, vol. 35, n° 3-4, 1996.
- [NIK 01] NIKOLIC M., SARTER N. B., « Peripheral visual feedback : A powerful means of supporting effective attention allocation in event-driven, data-rich environments », *Human Factors*, vol. 43, n° 1, 2001, p. 30-38.
- [REN 03] RENUCCI F., GASTE D., « Sens d'un film interactif et position de l'auteur face au spectateur », *Hypertextes hypermédias. Créer du sens à l'ère numérique. H2PTM'03*, Paris, 2003, p. 311-323.
- [SAN 02] SANCHEZ J. M. M., KOENEN R., PEREIRA F., « MPEG-7 : The Generic Multimedia Content Description Standard, Part 1. », *IEEE Multimedia*, vol. 9, n° 2, 2002, p. 78-87.

- [SAW 96] SAWHNEY N. N., BALCOM D., SMITH I. E., « HyperCafe : Narrative and Aesthetic Properties of Hypervideo », *UK Conference on Hypertext*, 1996, p. 1-10.
- [STI 04] STIEGLER B., « Le désir asphyxié, ou comment l'industrie culturelle détruit l'individu », *Le Monde Diplomatique*, vol. 603, 2004, <http://www.monde-diplomatique.fr/2004/06/STIEGLER/11261>.
- [STO 03] STOCKINGER P., *Le Document audiovisuel : procédures de description et exploitation*, Hermès Science, Paris, 2003.
- [SUT 00] SUTCLIFFE A., « On the effective use and reuse of HCI knowledge », *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, n° 2, 2000, p. 197-221.
- [TVA03] TV Anytime, « TV-Anytime Metadata Specification Version 1.3 », 2003, <ftp://tva:tva@ftp.bbc.co.uk/pub/Specifications/SP003v13.zip>.
- [W3C 01] W3C, « Synchronized Multimedia Integration Language (SMIL 2.0) », W3C, 2001, <http://www.w3.org/TR/smil20/>.
- [WIL 03] WILDEMUTH C. M., MARCHIONINI G., YANG M., GEISLER G., WILKENS T., HUGHES A., GRUSS R., « How fast is too fast ? : evaluating fast forward surrogates for digital video », *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*, Houston, Texas, 2003, p. 221-230.