

Sémiotique des Médias.

Le genre du documentaire audiovisuel

Cours VII :

Le vidéo numérique - une présentation générale

Peter Stockinger

**Séminaire de DESS à l'Institut National des Langues et
Civilisations Orientales (INaLCO)**

Paris, 2001 - 2002

Sommaire

1) INTRODUCTION.....	3
2) QUELQUES NOTIONS DE BASE	5
<u>2.1) SIGNAL ANALOGIQUE ET SIGNAL NUMERIQUE.....</u>	5
<u>2.2) LES COULEURS RVB.....</u>	6
<u>2.3) NOMBRE D'IMAGES PAR SECONDE ET RESOLUTION.....</u>	6
<u>2.4) FORMATS ET STANDARDS DE LA VIDEO ANALOGIQUE.....</u>	9
3) NUMERISATION ET COMPRESSION	11
<u>3.1) LA NUMERISATION.....</u>	11
<u>3.2) LA COMPRESSION</u>	12
<u>3.3) LES CODECS.....</u>	15
4) LA TECHNOLOGIE DV	17
5) VIDEO ET ENVIRONNEMENT INFORMATIQUE S.....	18
<u>5.1) LE CHOIX D'UNE CONFIGURATION.....</u>	18
<u>5.2) EQUIPEMENT:</u>	18
<u>5.3) LES CARTES D'ACQUISITION</u>	19
<u>5.4) LES LOGICIELS.....</u>	21
<u>5.5) UN EXEMPLE : LA TECHNOLOGIE MEDIA DE MICROSOFT</u>	22

1) Introduction

Enjeux autour du "contenu" audiovisuel :

Actuellement, nous sommes dans une phase de grande mutation dans l'industrie de l'audiovisuel - industrie qui couvre à la fois les producteurs de matériel et de technologie, les producteurs de "contenu" et les consommateurs de matériel, de technologie et de "contenu" audiovisuels :

- L'offre télévisuelle intègre une quantité de plus en plus importante de chaînes (thématiquement circonscrites) diffusant en mode numérique (cf., par exemple, en France, l'offre Noos ou Canalsatellite)
- La "Web-TV", la télévision sur Internet fait ses apparitions et commence à se populariser. Il y a deux grandes catégories de sites tv : ceux qui dépendent, qui représentent une chaîne (diffusant en mode analogique ou numérique) tel que les sites de TF1, de France 3, de Planète, de Voyage, etc. ceux qui sont purement "virtuels" tels que les sites Canal-U (émissions de cours, de colloques,); Alatele.com (site proposant des émissions de différentes catégories : vie quotidienne, culture, découverte, ...); citoyenne.net (site proposant des émissions sur la culture, la politique, l'économie, etc.)
- Côté "consommateur", le matériel nécessaire pour accéder aux "contenus" numériques se banalise (ordinateurs puissants, périphériques d'excellent niveau peu chers, outils (logiciels)

d'acquisition et de traitement du contenu audiovisuel "grand public";
meilleures connexions et bandes passantes plus grandes à des prix
relativement abordables,

- L'objectif du "tout numérique" dans la chaîne de la production, post-production, diffusion et conservation de données audiovisuelles commence à s'imposer dans un grand nombre d'institutions (universités, administrations, archives et bibliothèques, ...) et d'entreprises.

Afin de pouvoir aborder ces enjeux d'un point de vue technique, nous allons présenter, d'une manière générale, quelques aspects concernant la technologie de la vidéo numérique:

- Notions de base
- Questions relatives à l'acquisition et à la compression de l'image vidéo
- La technologie DV à proprement parler
- Le traitement et la diffusion de la vidéo dans un environnement informatique
- Les principales phases de la production - diffusion de vidéo en ligne
- La phase de la post-production (notamment celle du montage)

2) Quelques notions de base

2.1) Signal analogique et signal numérique

La caméra (le camescope) est l'instrument qui convertit une information optique en information électrique (= transduction de la lumière en courant électrique). L'**information optique** est constituée par deux aspects fondamentaux:

- les informations relatives à la couleur (la **chrominance**) et
- les informations relatives à la lumière (la **luminance**)

Mode analogique : signaux qui varient constamment entre un minimum et un maximum autorisé (une "courbe").

Mode numérique : signaux sous forme de "points" sélectionnés par intervalles sur la courbe

Mode binaire : les points sont décrits sous forme de valeurs maximales (= "1") et minimales (= "0")

Avantage du numérique :

La très haute fidélité de l'information transmise (en principe, pas de bruit entre l'information transmise et l'information reçue, comme c'est le cas lors de la transmission, duplication, ... d'une information audiovisuelle en mode analogique)

2.2) Les couleurs RVB

L'information à la sortie de la caméra est constituée par trois signaux fondamentaux : le rouge, le vert et le bleu (RVB). Le RVB est, physiquement une juxtaposition de trois types de signaux que l'œil amalgame en un seul point coloré.

Le R, le V et le B constituent les valeurs d'un point coloré. On parle également en terme de *couches* d'une image informatisée : une image informatisée possède donc *trois* couches fondamentales.

Un point coloré (à trois couches) constitue une information binaire de 24 bits, pour l'ordinateur, 8 bits pour chaque valeur (i.e. $3 \times 8 = 24$). C'est dans ce sens que l'on parle de *couleur 24 bits*.

On peut représenter environ 16 millions de nuances de couleurs pour chaque pixel (chaque point coloré).

Note : pour les applications web, il n'en subsiste qu'un sous-ensemble infime - d'environ 210 couleurs - appelées "*web safe*".

2.3) Nombre d'images par seconde et résolution

Afin que l'œil humain puisse percevoir une animation, les images doivent défiler à une certaine vitesse (= *cadence* d'une animation en terme de nombre d'images par seconde)

- Environ 10 images par seconde : seuil inférieur pour une animation (en deçà : perception saccadée)
- 24 images par seconde : film (au cinéma)
- 29 ou 30 images par seconde : télévision (standards PAL ou NTSC)

Résolution

La quantité d'informations contenues dans une image (dans chaque image d'une animation) se calcule en **pixel** :

- un pixel : des petits espaces, des éléments individuels constituant une image (peuvent être plus ou moins grands);
- plus un pixel est "grand", moins élevé est son nombre, moins "lourde" est la quantité d'informations relatives à l'image mais aussi : moins bon est la qualité de l'image,
- plus un pixel est "petit", plus élevé est son nombre, plus importante est l'information relative à l'image (et plus important est aussi l'effort de calcul devant être fourni par l'ordinateur), mais aussi : plus grande est la qualité de l'image.

En pratique : il faut trouver un compromis entre les contraintes technologiques et physiques d'une part et la qualité optimale requise d'une information visuelle, d'autre part. Ainsi, pour la distribution - diffusion d'informations audiovisuelles, on tient compte de différents cas de figure :

- Diffusion sous forme de cassette VHS
- Diffusion sous forme de CD-ROM
- Diffusion sous forme de DVD
- Diffusion pour DTV "haute fidélité"

- Diffusion sur le web (et, ici, de nouveau, en tenant compte des contraintes de la bande passante).

Chacun de ces cas de figure exige une définition de la résolution d'images particulière.

2.4) Formats et standards de la vidéo analogique

Comme nous l'avons déjà vue, l'information optique :

- d'une part se compose des informations de chrominance (informations relatives aux couleurs) et des informations de luminance (informations relatives à la lumière, à la qualité de la lumière) et,
- d'autre part se construit à partir des couleurs primaires RVB

On sait également que l'œil humain est beaucoup plus sensible à la luminance qu'à la chrominance. Or, cette différence dans le traitement de l'information visuelle joue un rôle important :

- dans la définition de standards de la vidéo analogique (i.e. des standards réglant le transport des informations visuelles en mode analogique) et,
- dans celle de codecs (modules de compression-décompression) dans le monde de la vidéo numérique (cf. ci-après).

Le monde de la vidéo analogique est régi par trois formats principaux :

- Le format composite
- Le format S-Vidéo
- Le format composante

Le *format composite* prévoit le transport de toutes les informations visuelles (luminance et chrominance) via un seul fil (cf. les VHS et les 8 mm). Le désavantage est la perte sensible en qualité d'image.

Le *format S-Video* distingue deux fil - un fil séparé pour le flux d'informations relatives à la luminance, à la qualité de la lumière d'une image, et un fil séparé pour le flux des informations relatives à la chrominance (= format YC). Ici : perte en qualité des informations relatives à la chrominance. Ce format est utilisé dans des équipements de type S-VHS et Hi-8.

Le *format composante* s'appuie sur trois fils - un fil réservé au flux des informations de luminance et deux fils réservés aux informations de chrominance (= format YUV). Ce format garantit la meilleure qualité et est utilisé par des équipements plutôt professionnels (Betacam).

La transformation d'une information codée en *RVB en une information codée en YUV* est très importante dans la mesure où elle permet un traitement approprié (à la sensibilité de l'œil humain) de la luminance (Y) et de la chrominance (U et V):

3) Numérisation et compression

3.1) La numérisation

Toute source audiovisuelle analogique doit être transformée en source numérique (binaire) afin qu'elle puisse être reconnue, traitée et diffusée via un ordinateur et un réseau informatique.

La numérisation recouvre :

- La décomposition et l'échantillonnage d'une image vidéo analogique en un ensemble de "points" (de pixels) selon une résolution donnée (ex. : 720 x 486 pixels pour une image vidéo normale);
- L'association d'une valeur numérique à chaque pixel (composé des trois couches RVB);
- L'utilisation d'une table de conversion des couleurs (= 24 bits par pixel pour environ 16 millions de couleurs possibles);
- La production d'un "fichier" informatique de l'image.

Ce processus doit se dérouler très vite étant donné que défile entre 25 et 30 images par seconde (pour la télé et la vidéo).

C'est pour cela qu'il faut des microprocesseurs très puissants et que la qualité de la conversion (numérisation) est limitée par la puissance du calcul du/des microprocesseurs !! (cf. l'acquisition en MPEG 2 sous Media Studio Pro 6).

L'image vidéo contient ainsi une quantité très impressionnante d'informations numériques (binaires) par exemple -

Environ 2 Go pour 9 minutes de flux DV

D'où donc des limitations techniques importantes :

- Limitation au niveau de la capacité du calcul des microprocesseurs (problème qui est en train de se résoudre grâce à des microprocesseurs de plus en plus puissants et peu chers);
- Limitation en terme de capacité de stockage (devenu quelque peu moins pressante avec la production de périphériques de stockage pouvant aller jusqu'aux Téra octets);
- Limitation la plus importante pour la diffusion sur le web : rapidité de transmission (cf. les limites drastiques pour les connexions modem et lignes téléphoniques)

3.2) La compression

Pour réduire la quantité des informations numériques relatives à une vidéo : on a recours à des procédures/stratégies de compression. "Compression" s'appelle aussi : *Bit Rate Reduction* (réduction du débit binaire).

a) Méthodes simples de réduction (mais souvent pas suffisantes):

- Réduction de la taille de l'image (par exemple de 640 x 480 pixels à 320 x 240 pixels)
- Réduction du nombre d'images par seconde de vidéo

b) Méthodes plus sophistiquées :

Constat : une image contient beaucoup d'informations redondantes (donc pouvant être éliminées sans que la qualité de l'image soit - pour l'œil humain (!!)) - affectée d'une manière significative)

L'œil humain est moins sensible à la couleur qu'à la luminance d'une image (donc : il est possible de réduire une information chromatique sans que l'œil humain s'en rende compte). D'où donc l'importance :

- de dissocier entre informations relatives à la luminance et informations relatives à la chrominance
- d'appliquer des *ratios luminance/chrominance* spécifiques et appropriés d'une part à des exigences de qualité et d'autre part à des contraintes techniques ou technologiques (capacité de traitement, espace de stockage, bande passante, ...)

Quant aux techniques de compression plus sophistiquées, on distingue entre :

- La compression "intratrame" ("spatiale") : on tient compte des similarités des informations dans des zones de l'image qui sont proches l'une de l'autre (dans une image, deux points voisins sont souvent similaires). Cette technique est utilisée, par exemple, par la technologie DV: chaque image est analysée, lors de sa capture par un

caméscope numérique, est comprimée. Par exemple, un flux d'environ 25 à 28 images (par seconde) représente, en tant que flux comprimé (à l'aide de la technique "intraframe", 25 Mb/s. D'où donc le problème de la *bande passante très importante* (i.e. entre 25 et 50 Mb/s) qu'exige un DV pour être visualisée, lue d'une manière satisfaisante par un spectateur/utilisateur

- La compression "interframe" ("temporelle") : méthode qui tire parti de la probable similitude entre une image donnée de la vidéo et les images qui l'entourent (qui la précèdent, qui lui succèdent); n'est stockée que la différence entre les images qui se suivent. A des intervalles choisis, est identifiée une *image clé* ("I-Frame")- une image de référence - par rapport à laquelle est calculée la différence. C'est le standard MPEG2 qui utilise cette technique et qui permet de maintenir une très grande qualité dans la transmission de la vidéo même sur des bandes passantes bien plus petites qu'exigent la DV ou le standard MPEG1 (i.e. entre 4,5 et 15 Mb/s). Dans le standard MPEG2, on connaît, à côté des *I-Frames*, deux autres types de frames: les P-Frames et les B-Frames. Les P-Frames ou encore les "predicted frames" sont des images dans lesquelles n'apparaissent que les différences visuelles prédites par rapport à un I-Frame précédent. Les B-Frames ou "bi-directional frames" sont des images dans lesquelles ne sont retenues que les différences entre un I-Frame précédent et un I-Frame suivant. I-Frame, B-Frame et P-Frame forment ensemble ce qu'on appelle des GoP ("group of pictures). Plus l'intervalle est grand, plus la compression d'information est importante, plus il y a aussi risque d'une déperdition sensible d'informations visuelles. Par exemple, une compression importante peut être effectuée sur des enregistrements de type conférences, cours, débats où l'information visuelle est hautement redondante. Par contre, pour des films de fiction ou encore pour la diffusion via la télévision numérique, seront choisis des intervalles bien plus courts, voire très courts.

3.3) Les codecs

Codec = **compression - décompression**.

Un module fourni soit avec le matériel (carte d'acquisition vidéo, caméscope numérique) soit avec les logiciels (Première d'Adobe, Vidéo Studio d'Ulead, ...)

Les trois rôles d'un codec sont (cf. Jean Charles Fouché, <http://www.repaire.net/savoir/essentiel/06.htm>):

- « 1. Afficher l'image DV sur l'ordinateur et/ou une sortie extérieure.
2. Calculer les effets liés au montage vidéo (fondus, volets...)
3. Capture et conversion Analogique>Numérique vice-versa ».

Différents codecs

- *Le Motion Joint Photographic Experts Group (MJPEG)*

Standard de compression intratrame : une séquence vidéo est considérée comme une succession d'images fixes ; taux de compression relativement faible pour une résolution d'image de 720 x 486 pixels ; débit requis par seconde pour diffuser/recevoir une image vidéo : entre 0, 5 et 25 Mo/seconde

- *Le MPEG 1 (Moving Pictures Experts Group) :*

standard de compression intertrame pour des applications CD-ROM (et Internet haut débit) avec une résolution d'image de 352 x 240 pixels/image; débit requis par seconde pour diffuser/recevoir une image vidéo : entre 0,01 et 0,06 Mo/seconde

- *Le MPEG 2* :
standard de compression intertrame pour des applications DVD, TV numérique haute définition avec une résolution d'image de 720 x 480 pixels/image; débit requis par seconde pour diffuser/recevoir une image vidéo : entre 0,01 et 2 Mo/seconde
- *DV (25, ...)*:
standard de compression intratrame pour vidéo numérique; résolution de l'image : 720 x 480 pixels/image; débit requis par seconde pour diffuser/recevoir une image vidéo : autour de 3,5 Mo/seconde

4) La technologie DV

La technologie DV est la base du "tout numérique", i.e. du traitement, de la diffusion et de la conservation de données audiovisuelles. Elle se réalise, se manifeste à travers :

Un équipement

- Le caméscope DV (caméras) qui fonctionne avec une mini-cassette DV, compresse la vidéo au format DV 25 et dispose d'un port de connexion à un ordinateur
- La cassette DV

Des codecs DV

DV 25 qui s'appuie sur un débit de 25 méga-octets/seconde pour le stockage d'images vidéos à un taux fixe de compression de 5 à 1. Il utilise le modèle d'échantillonnage de couleurs 4 : 1 : 1

Digital8 : correspond au codec DV25 mais adapté aux caméscopes Digital8 enregistrant leurs données sur une cassette Hi-8. Une caméra Digital8 peut lire aussi des données analogiques.

DVCAM et **DVPRO** : codec pour un public professionnel

DV50 et **DV100** : compléments de DV25 rendant en fait obsolète la distinction entre technologie réservée à un grand public et technologie spécialisée.

Une norme

La norme IEEE 1394 - aussi connue sous le nom de FireWire (technologie Apple) et i.LINK (technologie Sony). Il s'agit de ports (sur l'ordinateur et la caméra DV) et un câble particulier permettant un transfert ultrarapide entre un caméscope DVC et l'ordinateur (vice-versa): autour de 400 millions de bits (i.e. 400 méga-octets) par seconde.

5) vidéo et environnement informatiques

5.1) le choix d'une configuration

Différentes configurations possibles selon les objectifs visés (i.e. vidéo grand public, vidéo pour la recherche et comme support à la recherche, vidéo comme support à la vie professionnelle, vidéo pour diffusion "broadcast", etc.)

Critères :

- usages de la vidéo
- types de la vidéo (uniquement DV ou aussi analogique; si analogique : composite, S-Vidéo ou composante ?)
- volumes de données à manipuler (9 minutes de DV = environ 2 giga octets...)
- types de travaux à effectuer sur ordinateur (post-production, indexation, ...)
- canal de distribution-diffusion (via le web, pour CD-ROM, pour DVD, ...)

5.2) Equipement:

a) station de travail : ordinateur relativement puissant (type Pentium III ou PowerMacG3 avec un microprocesseur de grande puissance (à partir de 500 Mhz), une mémoire vive d'au moins 128 Mo), un écran de 17 pouces au moins, etc.

b) un espace de stockage d'au moins 30 giga-octets (si possible : pile de disques durs)

c) cartes vidéo (carte DV temps réel ; carte temps réel pour entrée/sortie analogique ; carte E/S analogique/numérique)

d) caméscope DV

e) magnétoscope

f) moniteur télé (facultatif)

etc.

5.3) Les cartes d'acquisition

Six caractéristiques principales :

- Type d'entrée/sortie vidéo analogique pris en charge
- Type d'entrée/sortie vidéo numérique pris en charge
- Type de compression pris en charge
- Type de traitements spéciaux pris en charge
- Types de logiciels inclus
- Types d'audio pris en charge

Exemples de cartes d'acquisition

- **Pinnacle DV 500**

E/S analogique : S-Video et vidéo composite

E/S numérique : DV et IEEE 1394

Codec : DV 25

Traitements spéciaux : incrustation, effets de transition, ...

Logiciels inclus : Première 5.1 d'Adobe pour montage ; Hollywood FX (pour création de titres, génériques et effets en 3D), Minerva Impression (pour production de DVD)

E/S audio : IEEE 1394

Coût : environ 7000 frs.

Pinnacle DV 500 peut travailler avec une résolution image : 720 x 576 (PAL), compatible Betacam (qualité "studio"); elle peut être utilisée pour la production de DVD et CD-ROM (exportation en fichier MPEG 2)

- **DV EZ**

E/S analogique : S-Video et vidéo composite

E/S numérique : DV et IEEE 1394

Codec : DV 25

Traitements spéciaux : incrustation, effets de transition, ... (plate-forme Mac et Windows 95, 98, 2000)

Logiciels inclus : Premiere 5.1

E/S audio : IEEE 1394

Coût : environ 10 000 frs.

Pinnacle DV 500 peut travailler avec une résolution image

- **Amber**

E/S analogique : Composite, S-Video, composante; standards : PAL et NTSC

E/S numérique : non

Codec : MPEG 1 et 2

Traitements spéciaux : non (destiné surtout pour archivage et.ou export pour applications DVD, TVHD)

Logiciels inclus : non

E/S audio : MPEG 1 et 2

Coût : environ 20 000 frs.

5.4) Les logiciels

Il faut distinguer trois grands groupes/composants de logiciels/services

a) le client

logiciel ou plug-in permettant de visualiser une donnée audiovisuelle, un clip, une animation, une séquence, ... :

Exemples : Windows Media Player, Quick Time, Real Player

b) logiciels pour la production de contenu audiovisuel

Pour pouvoir classer et comprendre la diversité des logiciels spécialisés, il faut garder en tête qu'ils répondent à des besoins précis des principales activités du traitement de la vidéo - principales activités qui sont les suivantes :

- *Acquisition* (i.e. numérisation-compression à partir d'une source analogique et/ou numérique)
- *Edition* (montage, production d'effets spéciaux, développement de produits hypermédias interactifs, indexation-annotation, traduction, ...)
- *Exportation* (pour le web, CD-ROM, DVD, TVHD, ...)

Différents types de logiciels:

- logiciels d'acquisition de vidéo (analogique et/ou numérique);
- logiciels de montage "virtuel" (montage de scènes et séquences, production de titres, d'intertitres, production d'effets spéciaux, retouche d'image, création de scènes 3D, utilisation d'effets VR, ...);

- logiciels de description (indexation automatique ou manuelle, production et gestion des espaces d'annotations, outils de traduction et de sous-titrage...);
- logiciels de développement hypermédia, de produits interactifs, de produits destinés pour le web, pour CD-ROM ou DVD,
- logiciels spécialisés en l'exportation et le traitement de la vidéo numérique vers des fichiers ASF (advanced streaming format) pour le web.

c) les services serveur multimédia

- base de données (de type médiathèque en ligne)
- vidéo "on demand"
- protocole "streaming" (flux)
- monocast/multicast
- protection et sécurité
- etc.

5.5) Un exemple : la technologie média de Microsoft

= *Windows Media Technologies* : plate-forme composée d'un ensemble de composants, outils et services pour

- la création,
- la distribution et
- le visionnement et l'audition

de produits multimédias (clips, animations, bandes son, conférences, ...) pour le web (en format "stream avancé": asf, wmv, wmx, ...) et/ou d'événements à temps réel ("live stream")

Windows Media Technologies - quatre composants principaux :

- **Client** (pour le "playback" de clips, animations, ...) Windows Media Player
- **Outils pour le développement** de contenu multimédia : Windows Media Tools
- **Server** : Windows Media Services
- **Outils pour le développement d'applications** : Software development kits (SDKs)