Projet Multimedia:
Détection des différentes phase de l'émission

Timothee, El Mostafa, Hamdi et Othman

October 2018

1 Introduction générale

Nous nous proposons d'analyser un support vidéo pour en caractériser le format et certains contenus. Notre objectif est de décrire le document en termes de séquences différentes qui le composent et d'en extraire l'équipe de réalisation au format text. Les différentes taches du projet sont listées ci-dessous:

- identifier la fin générique de début
- Situer le début et la fin du reportage de lancement
- Encadrer les instants du débat
- Identifier le déclenchement du générique de fin
- Récupérer en format Text l'équipe du film

Pour ce faire notre démarche on se propose de procéder à l'analyse du document vidéo pour en extraire différentes caractéristiques qui le composent. L'analyse et le traitement concerneront les features des images séquencées de même que celle de la bande son.
La démarche type pour les différents composants consisterait à:

- Détecter les différents éléments d'inérêt dans les différentes parties du signal vidéo et les extraire (image: segmentation, découpage en régions, audio: type de contenu).
- Caractériser ces éléments d'intérêt (dimensions, emplacement, ...) ;
- Identifier et classifier les caractéristiques du document.
- Analyser et Interpréter les scènes en fonction des résultats précédents. Les scènes pourront, entre autres, être décrites de façon quantitative et/ou qualitative.

2 Résumés du cours Audio et Image

2.1 Résumé de l'analyse audio

L'analyse audio peut être réalisée sur trois type de contenu : parole, music ou bruit environnemental. De plus, dans ce domaine, les analyses sont très différentes lorsqu'il s'agit de parole ou de musique/bruit environnemental. Il existe de nombreuses applications : classification de genre musicaux, analyse rythmique, audio encodage, reconnaissance vocale, linguistique... Dans le domaine de la parole, l'analyse audio représente entre autres une passerelle entre l'Homme et la Machine.
La plupart des applications repose sur des algorithmes de classification supervisé ou non supervisé.
Nous allons donc par la suite décrire les principales étapes d'une classification audio :

a. Digitization

b. Preprocession

c. Feature computation

d. Temporal integration

e. Classifier training

2.1.1 Digitization

Les analyses audios sont réalisées sur des signaux numériques qui été initialement des signaux acoustiques. Un signal acoustique est enregistré à partir d'un micro sou forme de signal analogique puis par une méthode d'échantillonnage nous obtenons un signal numérique. La quantité d'information est en générale codée sur 16 bits à une fréquence d'au moins deux fois 20 kHz pour ne pas perdre d'informations. En effet, la plage de fréquence audible pour un humain est de 20 Hz à 20 kHz. On remarquera que le débit obtenu est de 16 bit x 40 k Hz = 640 bit / s. D'où l'utilité d'algorithme de compression (mp3, mp4)

2.1.2 Preprocessing

L'objectif du preprocessing est à la fois de filtrer les bruits sur le signal, réduire le taux d'information (eg. subsampling) et la normalisation.

2.1.3 Feature computation

Pour définir un vecteur temporel de features, il est important de bien définir la fenêtre d'analyse. La segmentation temporelle peut être soit statique à intervalle de temps constant (eg. 20 ms) ou dynamique pour varier en fonction de la quantité d'information. Il existe trois catégories de features : 1.

a. Temporal feature : description de la forme de l'onde du signal (eg. Zero Crossing Rate),

b. Spectral features : représentation fréquentielle du signal (eg. DFT, STFT, Cepstrum ou MFC),

c. Preceptual features : représentation basée sur des considérations psychoacoustiques.

Nous allons détailler ici les spectral features les plus utilisés en analyse audio :

- Le Short-Time Fourier Transform (STFT) est un outil puissant pour le processing du signal audio. Il permet de définir une classe particulièrement utile de distributions fréquence-temps qui spécifie des amplitudes complexe par rapport au temps et à la fréquence pour n'importe quel signal :
  - But : Extraire des informations d'un signal analogique, ou numérique
  - Choix de la fenêtre : La largeur du lobe principal permet de contrôler la précision fréquentielle. Plus il est étroit, meilleur est la résolution fréquentielle. Les lobes secondaires doivent pouvoir être négligés par rapport au lobe principal.
  - Choix de la taille de la fenêtre : Obligation de prendre une résolution fréquentielle en fonction de la taille de la fenêtre qui permette de s'assurer les composantes du signal étudié soient séparées par une distance suffisamment importante.
  - Limitations de la représentation spectrale: l'analyse est globale et ne permet pas de capturer l'information temporelle comme le début et la fin du signal ou l'apparition d'une singularité, l'analyse STFT exige de connaitre l'intégralité du signal. La représentation engendrée est de grande dimensionnalité.
- Le modèle Source-Filter modélise la parole comme une combinaison de sources indépendantes :
  - une source de son par exemple le souffle pulmonaire qui fait entrer en vibrations les cordes vocales
  - un filtre acoustique linéaire comme par exemple le canal vocal qui joue le rôle d'amplificateur de certaines fréquences.

  Les deux font des contributions séparées aux features caractéristiques du son résultant. La source dans le son de la voix dans le discours exprimé est responsable de la hauteur du son alors que le filtre à cordes vocales est responsable de l'emplacement des formants et de la forme spectrale excessive. Dans un vrai spectre de discours, la forme globale du filtre et l'emplacement des formants sont souvent noyés par les effets du spectre source. S'il était possible de supprimer les effets de source, les deux

spectres pourraient être étudiés séparément pour donner une image plus précise des features du discours.
Enlever les effets de la source du spectre a pour conséquence d'enlever les sources de variabilité du signal. Il y a plusieurs techniques pour séparer la source et le filtre dans un signal audio comme par exemple l'analyse Cepstrale.

L'analyse Cepstrale repose sur l'observation qu'un spectre de discours logarithmique est constitué de la source et du spectre de filtres ajoutés ensemble. L'origine de cette idée est que le filtrage dans le domaine des fréquences est réalisé en multipliant les spectres ensemble.
La multiplication correspondant à l'ajout de logarithmes et un spectre filtré peut être dérivé en ajoutant des spectres logarithmiques (ou dB). La procédure pour l'analyse Cepstrale est de prendre la transformation inverse de Fourier du spectre dB, convertissant ainsi le signal en un domaine temporel. Ce signal de domaine temporel n'est pas un signal acoustique régulier, puisqu'il a été dérivé du spectre logarithmique; pour cette raison, il est appelé une cepstrum. Puisque le spectre des dB est la somme de deux spectres, La propriété importante du cepstrum est qu'il est la somme de deux composants correspondants à la source et au filtre. La convolution permet donc de caractériser la transformation entrée/sortie réalisée par un filtre linéaire invariant. Le tout est un système invariant dans le temps. La partie inférieure du cepstrum correspond au filtre tandis que la partie supérieure (ou plutôt le milieu du cepstrum reflétée) correspond à la source

- L'analyse Mel fréquence du discours est basée les expériences de perception humaine. Il a été observé que l'oreille humaine agissait comme un filtre en se concentrant que sur certaines composantes de la fréquence. Ces filtres sont non uniformément espacés dans l'axe de fréquence : Plus de filtres dans les régions basses fréquences et moins de nombre de filtres en régions de haute fréquence. Les MFCC sont utilisés dans deux domaines :
  - La synthèse de discours :
    - Utilisé pour joindre deux segments de discours S1 et S2
    - Représentation de S1 comme une séquence de MFCC
    - Représentation de S2 comme une séquence de MFCC
    - ointure au point où les MFCC de S1 et S2 ont la distance euclidienne minimale
  - La reconnaissance vocale:
    - Les MFCC sont surtout utilisées comme features dans les systèmes de pointe de reconnaissance de discours

Pour résumer, le discours est analysé sur des courtes fenêtres d'analyse dans lesquelles le spectre est obtenu en utilisant la FFT. Le spectre passe après par

un filtre Mel pour obtenir le spectre Mel (MelSpectrum), l'analyse Cepstrale est ainsi réalisée sur le spectre Mel pour obtenir les coefficients Mel-Frequence Cepstrales. En représentant le discours comme une séquence de vecteurs Cepstraux, on peut donner ces vecteurs au classifieurs pour réaliser la reconnaissance vocale.

Pour savoir quels features utiliser dans le cadre d'une analyse précise, nous pouvons nous baser sur un jugement d'expert, un algorithme de sélection automatique (Forward and Backward Stepwise selection) ou une méthode de réduction de la dimensionnalité (Wrapper, Filter or Embedded methods). Il existe aussi les réseaux de neurones qui, en se basant sur des raw data, permettent d'éviter ce type de questionnement. Temporal integration Une fois le vecteur temporel de features défini, nous effectuons une intégration temporelle (eg. La moyenne sur 30 s). L'intégration permet de :

- Supprimer les bruits et d'améliorer la robustesse du modèle,
- Synchroniser les features avec le bon choix de fenêtre,
- Capturer l'évolution temporelle des features.

Classifier training Nous définissons alors un échantillon d'apprentissage pour entrainer notre classifier et de test pour valider la justesse de la calibration du classifier en évitant l'overfitting

2.2 Traitement images

La video est une séquence d'images, superposées à un signal audio (bande son) et des annotations textuelles (sous-titres). les images correspondent à un signal 2D scalaire pour les images en niveau de gris ou vectoriel pour les images couleurs. Les images sont représentées par les valeurs communes suivantes:

Paramètre	**Symbole**	**Valeur**
Rows	N	256,512,525,625,1024,1035
Columns	M	256,512,768,1024,1320
Gray Levels	L	2,64,256,1024,4096,16384

Table 1: Images: valeurs communes

Les images et les vidéos sont des signaux complexes étant donnée leurs popriètés:

- Non stationnaire: le contenu en fréquences spatiales change avec les coordonnées spatiales.
- Non gaussien: les propriétés statistiques ne suivent pas une loi de probabilité gaussienne.
- Non isotrope: les propriétés du signal d'image ne sont pas les mêmes avec l'orientation.

Les images et le contenu visuel des videos possèdents des caractéristiques principales:

- les contours: changement abrupt dans une caractéristique importante
- la texture: variation spatiale (souvent la luminance) du signal 2-D pour laquelle la valeur moyenne ne change pas. Les bords sont localement des signaux 1-D, alors que la texture est toujours un signal 2-D
- les mouvements: pour la vidéo (variation temporelle du signal à une même coordonnée spatiale)

2.3 Différents niveaux de description

La gestion automatisée des contenus images et/ou vidéos nécessite la maitrise des contenus à decrire. Différents niveaux de description existent pour caractériser ce type d'objets.
La description bas-niveau traduit l'objet étudié en terme de variation de couleurs ou niveau de gris ou encore en distinguant les différentes formes et textures.
Un second niveau d'interprétation de plus haut niveau, rend compte de l'information véhiculée en fonction de connaissances préalables relatives aux données sémantiques ou descriptive.
La description des données images et vidéos sera une composante de la connaissance bas niveau et de la connaissance haut niveau.

2.4 Variété des éléments à identifier

Le traitement des données images/vidéos doit nous permettre d'identifier les différents types d'objets localisés ou répartis, des zones d'intérêt ou tout autres éléments, en tenant compte du contexte. Ces données doivent pouvoir être identifiées même pour des flux d'images dans des échelles et des niveaux de résolution différents.

Un bon niveau de déscription des documents vidéos permet de cataloguer leur données globales, les classifier et d'indexer leur contenus.

- Catalogue: exploitation de informations globales sur le document visuel, indépendemment du contenu.
- Classification: exploitation d'une caractéristique globale du document visuel.
- Indexation: exploitation d'une analyse fine du contenu du document visuel.

2.4.1 Indexation

Indexer un objet audiovisuel consiste à extraire une information synthétique des images qui le composent afin de faciliter l'accès à leur contenus. Conetenus qui peuvent concerner des domaines variés tels que la fouille de données (data mining), la classification, ingénierie des connaissances, vision artificielle, SGBD, ...
L'information ainsi processée peut-être encoder pour être conservée, ou traitée dans le cadre des cas d'utilisation variés ou échangées pour les besoins globaux. L'indexation génère une clé d'accès à l'information contenu dans l'objet multimédia représenté par des identifiants représentatifs du contenu de l'image/vidéo.

2.4.2 Comment faire

- L'indexation textuelle: consiste à annoter les contenus de l'objet image ou vidéo manuellement pour enrichir sa description et faciliter la recherche ou la classification.
  Cette approche, couteuse en temps de labalisation, ne peut pas s'appliquer à des volumes trop important d'images ou de support video (des 10 aine de milliers d'images: BDD google, INA, ...). De plus, il faut gérer les problématiques sémantiques entre ce que peut représenter le docuent multimédia et ce que pourrais évoquer la description qui lui est jointe.

- Indexation par le contenu: elle s'opère via l'utilisation de modèles mathématiques pour extraire les caractéristiques des contenus multimédia et gérer leur indexation automatique.
  L'indexation par le contenu est adaptée à la manipulation des données en masse et permet l'exploitation par les outils d'apprentissage.
  La modélisation doit tenir compte des expertises métiers pour rendre compte de la compléxité d'extraire les caractérisiques esse,ntiels des objets étudiés. Certains aspects des caractéristiques des contenus restent complexes à modéliser (émotions par exemple).

2.5 Indexation par le contenu

Le principe de l'indexation par le contenu se décompose en deux phase distinctes. La première correspond à la génération des indexes et leur persistence dans un support de données. On parle alors d'indexation Off-line.
La seconde phase de l'indexation prend en charge les requêtes des utilisateurs, par différents biais (textuel, contenus exemplaires), pour fournir les documents indexés répondant aux critères de recherches.

- Off-line : production d'index issus de l'analyse du contenu des images
  - Extraction de caractéristiques pertinentes ("signatures")
  - Réduction de la dimension
  - Organisation par classification

- On-line : gestion des requêtes d'un utilisateur
  - Requêtes exprimées
    - soit par du texte (si des étiquettes sont disponibles)
    - soit par des images exemplaires (voire contre-exemplaires)

2.5.1 Extraction des caractéristiques

Le terme d'extraction de caractéristiques recouvre en fait deux probl'emes distincts : la quantification de texture et l'extraction des caractéristiques des structures ou régions préesentes sur une image. Ces problèmes sont distincts, non pas tant dans les méthodes de résolution auxquelles ils font appel, mais dans la manière même qu'on a de les aborder. L'extraction des caractéristiques d'un signal procède géenéeralement par une transformation de ce signal: sa réponse à une transformation donnée est utilisée pour en déduire une caractérisation.
L'information est pertinente à partir du moment où la transformation est discriminante : deux signaux distincts (dans un contexte donné) ont des réeponses distinctes à la transformation. Dans le cas de l'analyse de textures, cette condition est généralement suffisante. Pour la satisfaire, on peut être amené à considérer non pas une transformation mais plusieurs transformations et l'ensemble des réponses à ces transformations. Lorsque le problème se pose en termes d'extraction de caractéristiques des structures ou régions présentes dans l'image, cette condition ne suffit généralement pas. En effet, il faut également être en mesure d'interprêter les caractéristiques déduites.

2.6 Outils de traitement

Différents modèles mathématiques sont employés pour traiter les images dans le but d'extraire les caractéristiques qu'elle contient.
Aux différents modèles correspondent un certain nombre d'outils fondamentaux, qui se sont révélés au cours du temps plus ou moins incontournables, que ce soit d'un point de vue pratique ou théorique. Citons : la convolution, la transformée de Fourier, l'histogramme, les pyramides, la corrélation, la transformée en Tout-Ou-Rien, les ondelettes...

2.7 Types d'opérations

Différentes catégories d'opérations peuvent êtres réalisées sur une image numérisée pour réaliser les opérations d'indexation ou de classification. Ces opérations portent sur des périmètres différents de l'image.

- Point: la valeur de sortie pour un point spécific dépend uniquement du point d'entrée à la même position. On trouve dans cette cat egorie, les fonctions de recadrage ou d'egalisation de dynamique, de binarisation.
- Local: La sortie dépend non seulement des coordonnées du point en entrée mais également de ces points voisins.

- Global: La sortie pour un point en entrée dépend de la globalité de des points de l'image.

2.8 Approches standars

2.8.1 Détection de contours

La détection de contours permet de repérer les différents objets qui constituent la scène de l'image. Il existe de nombreuses méthodes pour trouver les contours des objets, la plupart sont basées sur les dérivées premières et secondes de l'image.
La détection de contours permet de repérer dans les images les objets qui s'y trouvent avant d'appliquer le traitement uniquement sur ces objets.

2.8.2 Histogramme - seuillage

L'histogramme d'une image donne la répartition de ses niveaux de gris. Ainsi pour une image qui possède 256 niveaux de gris, l'histogramme représente le niveau de gris en fonction du nombre de pixels à ce niveau de gris dans l'image. On sait que les niveaux de gris à zéro correspondent au noir et que les niveaux de gris à 1 indiquent le blanc. L'histogramme donne donc une excellente idée de la séparation entre quelque chose qui est clair et quelque chose qui est foncé dans l'image. Typiquement, une utilisation de ce fait est le seuillage d'une image, ce terme désigne la définition d'un seuil au-dessus ou en-dessous duquel on va garder certaines valeurs de niveaux de gris.

2.8.3 Espace de couleurs

La couleur est une donnée importante pour une image, elle modifie la perception que l'on a de l'image. L'espace de représentation standard décompose une image en trois plans de couleur: le rouge, le vert et le bleu - Red/Green/Blue RGB en anglais. Les couleurs finales sont obtenues par synthèse additive de ces trois couleurs primaires.
Il existe cependant des problèmes qui peuvent nécessiter de changer d'espace de couleur pour percevoir différemment l'image. Il y a des images où la couleur importe peu, par exemple des photographies de cellules vivantes (pseudo-transparentes), des images radar, des images satellites...
Dans ce cas, l'espace RGB n'est plus utilisé. On lui préfère d'autres espaces comme HSV Hue/Saturation/Value ou YCbCr Luminance/Chrominance bleue/Chrominance rouge.

2.8.4 Transformée de Fourier

La transformée de Fourier est un outil mathématique de traitement du signal qui permet de passer d'une représentation temporelle à une représentation fréquentielle du signal.
Cette théorie est basée sur le fait que toute fonction périodique est décomposable sur une base de sinus et de cosinus. Ainsi, on peut passer d'une représentation temporelle du signal (dans le repère temporel classique) à une représentation en fréquence sur une base de sinus et de cosinus (dans le repère fréquentiel).
La puissance de cet outil réside dans le fait que cette transformée est réversible et qu'elle peut être étendue aux signaux non périodiques (qu'on considère alors comme de période infinie).

3 Utilisation des résultats de caractérisation

3.1 Comparaison d'images

Une fois la caractérisation de l'image est réalisée est les features représetnatifs extrait, il devient possible d'utiliser ce résultat d'indexation pour réaliser une comparaison entre une image en entrée et le référentiel d'images indéxées dans une base de données notamment.
Les modèles comparatifs d'images mettent en place des algorithmes de mesures de similarités entre les objets étudiés en tenant comptent d'un certains nombre de caractéristiques pour définir des distances les séparant. Le matching peut être défini ensuite en fonctuion des critères utilisés en cherchant les images les plus proches du point de vue métier souhaité. Différentes méthodes de clacul de distances ont été définies pour tenir compte des différents domaines et modèles méties étudiés. Ci-dessous une liste non exhaustive des méthodes de calcul de distance.

- Distance euclidienne
- Distance euclidienne généralisée
- Malahanobis
- Chi2
- Similarité en cosinus
- Combinaisons linéaires de similarités (ou distances)
- ...

4 Etudes de cas

4.1 Analyse audio

L'objectif de l'analyse audio est d'identifier la séquence musicale de l'introduction et du générique de fin d'une émission télévisée. En effet, l'émission est divisée en quatre sections :

- Introduction : musique
- Reportage : voix
- Débat : voix
- Générique de fin : musique

Nous avons choisi de définir un classifieur entre deux catégories : voix et musique. Pour cela nous avons récupéré des morceaux de musique sur internet et un jeu de données de voix basé sur l'émission puis avons construit des observations de 5 secondes en Python en utilisant la librairie MoviePy. Par la suite, nous avons extrait des features en utilisant la libraire Librosa. En affichant, le chronogramme des 20 premières secondes de l'émission (10 secondes de musique puis 10 secondes de voix), nous observons des différences entre ces 2 types de classes audio. De même avec le spectrogramme.

La librairie open source Librosa fournit plusieurs méthodes d'extraction de features de piste audio : melspectrogram, mfcc et chroma-stft. Ces méthodes ont d'ailleurs été présentées dans le résumé de cours sur l'analyse audio. La librairie open source Librosa fournit plusieurs méthodes d'extraction de features de piste audio : melspectrogram, mfcc et chroma-stft. Ces méthodes ont d'ailleurs été présentées dans le résumé de cours sur l'analyse audio.

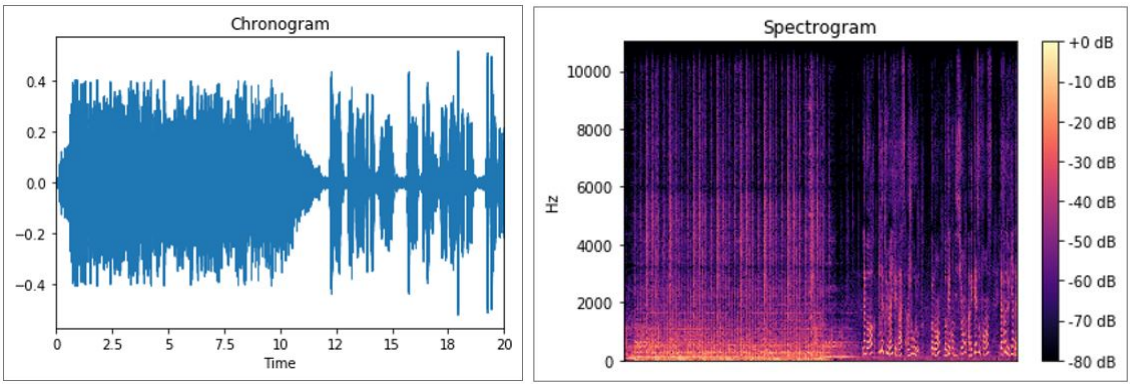

Figure 1: Chronogramme et Spectrogramme

La librairie open source Librosa fournit plusieurs méthodes d'extraction de features de piste audio : melspectrogram, mfcc et chroma-stft. Ces méthodes ont d'ailleurs été présentées dans le résumé de cours sur l'analyse audio.

```
def extract_feature(file_name):
    y, sr = librosa.load(file_name)
    stft = np.abs(librosa.stft(y))
    mfccs = np.mean(librosa.feature.mfcc(y=y, sr=sr, n_mfcc=40).T,axis=0)
    chroma = np.mean(librosa.feature.chroma_stft(S=stft, sr=sr).T,axis=0)
    mel = np.mean(librosa.feature.melspectrogram(y, sr=sr).T,axis=0)
    contrast = np.mean(librosa.feature.spectral_contrast(S=stft, sr=sr).T,axis=0)
    tonnetz = np.mean(librosa.feature.tonnetz(y=librosa.effects.harmonic(y),
    sr=sr).T,axis=0)
    return mfccs,chroma,mel,contrast,tonnetz
```

Figure 2: Code: Extraction features

Pour chaque piste audio, nous définissons ainsi les features et le label associé (voix ou musique).

```
def parse_audio_files(parent_dir, mylabel, file_ext="*.wav"):
    features, labels = np.empty((0,193)), np.empty(0)
    for fn in glob.glob(os.path.join(parent_dir, file_ext)):
        try:
          mfccs, chroma, mel, contrast,tonnetz = extract_feature(fn)
        except Exception as e:
          print ("Error encountered while parsing file: ", fn)
          continue
        ext_features = np.hstack([mfccs,chroma,mel,contrast,tonnetz])
        features = np.vstack([features,ext_features])
        labels = np.append(labels,mylabel)
    return np.array(features), np.array(labels)
```

Figure 3: Code: Parsing des fichiers audio

On obtient alors un dataset de 200 observations ayant chacune 193 features. Dans le domaine de la classification supervisée, il existe de nombreuses méthodes : KNN, Naive Bayes, LDA/QDA, Logistic, SVM, Neural Network...
Pour notre étude, nous aurions pu tester l'implémentation d'un neural network en Tensorflow. Cette méthode est en effet devenue très populaires pour le traitement de signal et nous avons trouvé beaucoup d'articles sur ce sujet.
Le problème de classification étant assez simple, nous avons opté pour un modèle de régression logistique.

```
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.5)
clf = LR()
clf.fit(x_train, y_train)
y_predict = clf.predict(x_test)
print("Score model:", accuracy_score(y_test,y_predict))
```

Figure 4: Code: Application du classifieur

Avec ce modèle, nous obtenons un score de précision de 100Il ne nous reste maintenant plus qu'à appliquer notre classifieur sur la piste audio de l'émission pour identifier les séquences musicales d'introduction et de générique de fin.

Par soucis d'affichage, nous n'avons représenté que les 20 premières et 20 dernières secondes de l'émission. Le classifieur donne l'introduction musicale sur les 10 premières secondes et le générique de fin à partir de la 2342 ème seconde.

```
clf = classification_music_voice(features,labels)

emission_features, emission_timing = parse_audio_files_timing('audio_emission')
emission_labels = clf.predict(emission_features)
```

Figure 5: Code: Application du classifieur

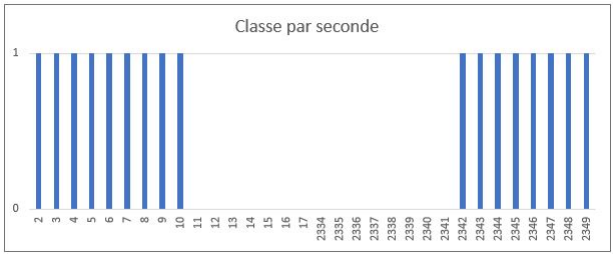

Figure 6: Code: Application du classifieur

4.2 Analyse image et vidéo

Objectif: Notre objectif dans la partie image du projet est d'identifer les plans de la video qui caractérise le début du générique de fin.
Onsuppose dans ce cas, que l'analyse audio de la piste son a pu repré une zone d'interêt qui caractérise la fin de la partie Débat et le debut de la séquence du Générique. Le traitement réalisé ensuite consiste à extraire les images contenus dans la séquence réduite proposée par l'analyse audio et den extraire les planches conetenant le texte détaillant l'équipe de l'émission.

Ayant travaillé séparemment sur les parties audio et vidéo, notre démarche dans le cadre de traitement image, pour isoler la séqeunce du générique de fin et d'enextraite les informations sur l'équipe de l'émission suit le plan ci-dessous:

- Découpage de la vidéo en un jeu d'images de manière uniforme
- Isoler un certain nombre d'images qui contiennent une partie du débat jusqu'à la fin de l'émission
- Calcul des Histogrammes de l'ensemble des images.
- Comparer les images ainsi caractériser avec une image en entrée représentant une plance avec les données de l'équipe de l'émission.
  - différentes méthodes de calcul de diatnces seront utilisées dans le but d'identifier celles qui sont les mieux à mêmes de nous permettre les résultats les puls performants.

,

4.3 Résultats obtenus

Ci-dessous les résultats obletnus pour les différentes méthodes de calcul de similarité. Trois types de retour seront présentés; ceux sans match de la partie Texte mais match sur le arrière pla de l'image, les matchs partiels et enfin les matchs totals.

4.3.1 Graphes


```
In [77]: print ("Equipe de réalisation de l'image de test \n", texteTest

Equipe de réalisation de l'image de test
 Jouraliste: Isabelle Gay
 Réalisateur : Raphael Theler
 Son : Yannick Bacher
 Caméra : Amauld Roux
 Goretty Albasini

 I Canal 9 Novembre 2006
```

Figure 7: Images de test et son text extrait

Le texte extrait reprend les différents éléments textuels de l'image avec quelques erreurs mineures sur les composantes graphiques autre que texte, des symboles globalement. Il en ressort qu'une fois la zone d'intéêrt spécifiée par l'analyse audio, la caractérisation des images générées par découpage nous permettera d'extraire celle contenant du texte, à supposer que seules les planches avec l'intitulé de l'équipe de tournage en contienne.

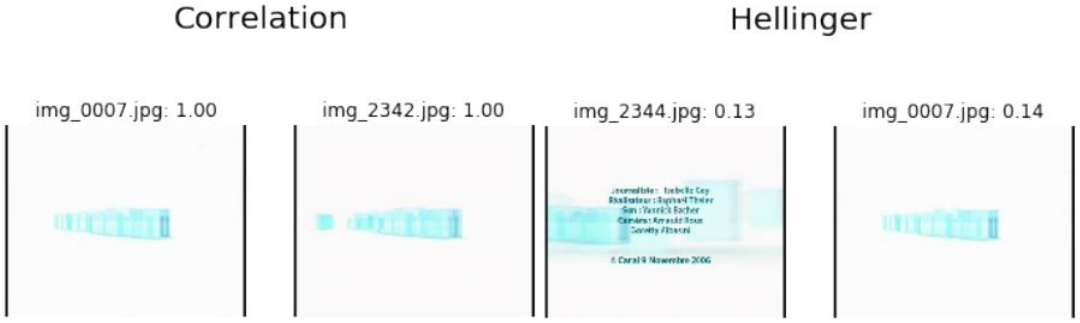

Figure 8: Retour des méthodes Correlation et Hellinger

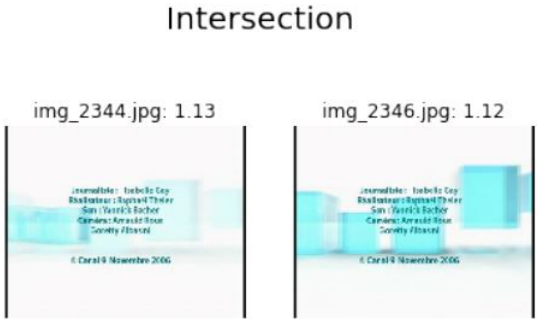

Figure 9: Retour de la méthode Intersection

5 Conclusion Génarale

L'analyse composée des données audios et images d'une video avec un format constant permet d'identifier les différentes phases de l'émission ainsi que de récupérer certaines données d'inté^{r} $et tellequecellesdel'équipederéalisation$.