Université François Rabelais de Tours

Aide à la transsciption audio de vidéos

TRAVAUX PRATIQUES

Enseignant **Jean-Yves ANTOINE**

(Jean-Yves.Antoine AT univ-tours.fr)

Transcription audio de videos : VirtualDub et Transcriber

1. Présentation

Si l'on se limite au logiciel libre, il n'existe pas d'outils réellement satisfaisants d'un point de vue ergonomique pour la transcription orthographique d'interactions filmés en vidéo. Certains outils (ELAN, CLAN) permettent effectivement une transcription synchronisée des échanges tout en visualisant la vidéo, mais leur manque d'utilisabilité pose encore problème : le transcripteur voit sa charge cognitive avant tout mobilisée par la manipulation du logiciel, ce qui détourne son attention d'une transcription de qualité. Par ailleurs, pour des raisons purement computationnelles, les outils existants ne peuvent le plus souvent pas travailler sur des vidéos de plus de 10 minutes de long. C'est la raison pour laquelle le transcripteur de vidéo se voit contraint de réaliser sa tâche en se contentant d'écouter la bande sonore liée à la vidéo. Quelque soit l'outil utilisé, la procédure reste la même :

1. Extraire la bande sonore sur l'ensemble ou partie de la vidéo
2. Travailler ensuite sur la bande sonore, sachant que la transcription sera synchronisée avec la bande sonore, donc également la vidéo.

Certains outils de transcription, tels Praat, peuvent même produire une sortie exportable vers des logiciels de transcription vidéo tels que CLAN.

Cette possibilité n'est cependant intéressante que si vous voulez annoter ensuite des évènements purement vidéos synchronisés avec la transcription audio, cas de figure qui ne sera pas étudié au cours de ce TP. Nous utiliserons donc les outils suivants :

1. *VirtualDub* pour l'extraction de la bande sonore
2. *Transcriber*, pour la transcription orthographique de l'audio.

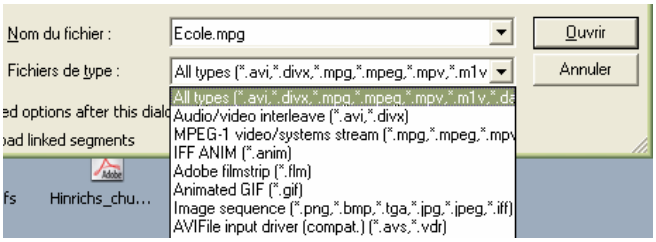
2. Extraction d'une bande sonore : VirtualDub

Il existe une multitude d'outils gratuits permettant d'extraire la bande sonore liée à une vidéo, et ce quel que soit le format vidéo utilisé (Mpeg, DivX etc…). Au cours de ce TP, nous utiliserons *VirtualDub*, qui est un outil très répandu qui est intéressant par sa richesse en termes de formats gérés et de fonctionnalités. Ce TP n'a pour seul objectif qu'une illustration du problème d'extraction, de nombreux outils équivalents à *VirtualDub* pouvant être utilisés suivant vos envies. *VirtualDub* peut être récupéré à l'adresse URL suivante :

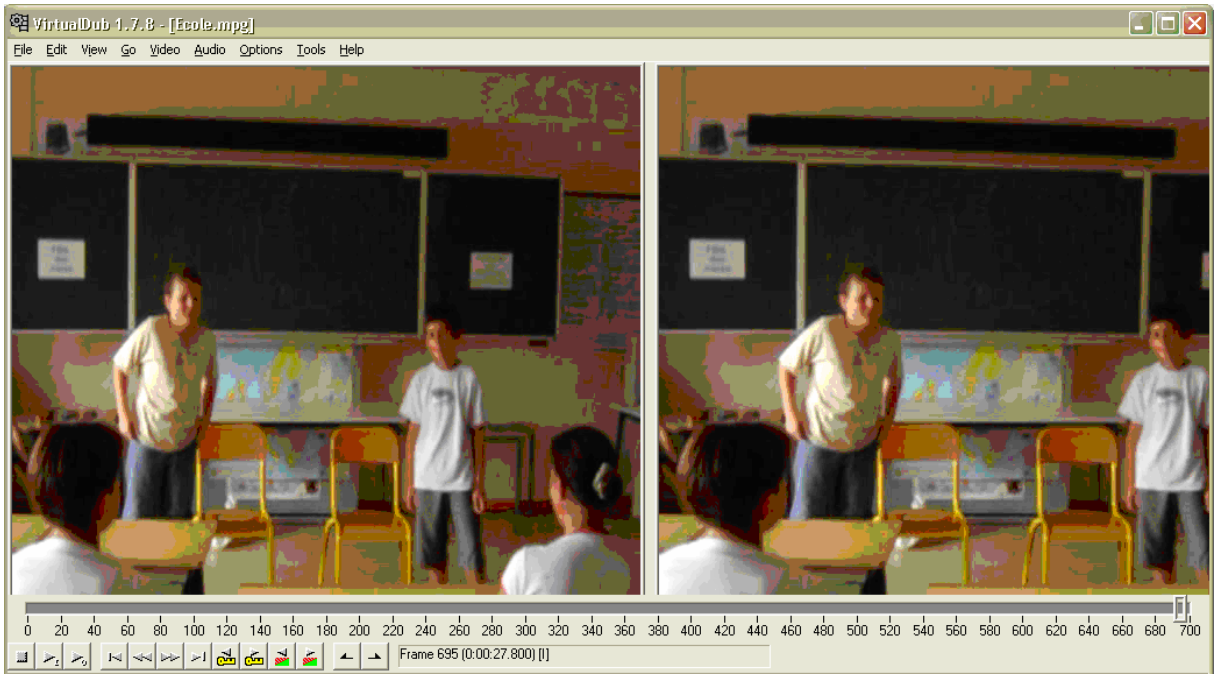
http://virtualdub.sourceforge.net/.

1.1 Extraction de l'ensemble de la piste sonore d'une vidéo

VirtualDub est un outil dont les fonctionnalités vont bien au-delà de la seule extraction audio. Son utilisation va donc vous paraître très simple. Dans un premier temps, nous allons extraire l'ensemble de la bande sonore d'un vidéo enregistrée lors des portes ouvertes d'une école primaire. Cette vidéo, qui a été enregistrée au format MPEG est disponible sur la page WWW consacrée à cet enseignement : fichier `ecole.mpg`. Sauvegardez le fichier sur votre ordinateur, puis lancer *VirtualDub*. L'interface de l'application s'ouvre.



En allant dans le menu `File`, sous-menu `Open Video File`, vous allez pouvoir charger le fichier vidéo sur lequel vous désirez faire l'extraction audio. Au passage, vous pouvez constater que VirtualDub comprend tous les formats vidéos les plus courants : AVI, MPEG, DIVX, MPV etc… Sélectionnez le fichier sur lequel vous voulez travailler et chargez-le en cliquant sur le bouton `Ouvrir`.



Après importation du fichier, l'interface *VirtualDub* affiche la première image de la vidéo (cf. ci-dessus où l'image est volontairement floutée et ne correspond donc pas au rendu réel). Deux scènes sont en fait affichées en parallèle. Cette fonctionnalité sert à contrôler la vidéo produite (sur la droite) lorsqu'on lance par exemple une exportation de vidéo dans un autre format. Pour nos besoins, on ne considèrera ici que les images à gauche de l'interface. Au bas de l'interface on trouve :

- **Une barre de défilement temporel** – Son échelle est en nombre de trames vidéos (frames) et non pas en millisecondes (ici, le fichier comprend 1773 trames). En cliquant sur une position de la barre, vous accédez directement à la scène débutant à l'instant sélectionné.
- **Une barre d'outils** qui permet entre autres l'écoute, la pause, l'avance ou le recul rapide dans le fichier. Vous pouvez utiliser cette barre pour lire la vidéo sur laquelle nous travaillerons.

Compte-tenu des objectifs de ce TP, nous n'allons pas étudier ici toutes les fonctionnalités de VirtualDub. Nous nous contenterons d'extraire la bande audio de la vidéo. Sous VirtualDub, la façon la plus simple est de sauver l'audio sous le format WAV, qui est utilisé par les CD audios et est compris de *Transcriber*. Pour cela, allez dans le menu `File`, sous-menu `Save Wav`. VirtualDub vous demande de préciser le répertoire de sauvegarde du fichier audio extrait, et de préciser son nom. Choisissez par exemple le nom de fihier `ecole.wav`. Ecoutez le fichier ainsi créé (par exemple avec le lecteur *Windows Media*) : il correspond bien à la bande sonore de la vidéo. Vous utiliserez ce fichier sonore lors du TP sur la transcription de parole.

1.2 Extraction d'une zone temporelle de la piste sonore d'une vidéo

L'opération précédente a consisté à extraire l'intégralité de la bande sonore de la vidéo étudiée. Il se peut cependant que seule un extrait de la vidéo soit intéressante. *VirtualDub* permet une extraction limitée à une zone définie dans la barre de défilement temporel. Supposons par exemple que seule nous intéresse la vidéo entre les trames 300 et 1000. L'extraction va consister à définir un début et une fin de zone de sélection :

1. En cliquant à la position correspondante, placez le curseur de défilement sur la trame 300 et définissez le début sélection dans le menu général l'option `Edit > Set selection Start`.
2. Faites de même pour la fin de sélection, en sélectionnant `Edit > Set selection End`.

Vous remarquez que la zone temporelle qui vient d'être définie est grisée dans la barre de défilement. Procédez maintenant à l'extraction du signal audio comme précédemment (`File > Save Wav`). Ecoutez le signal obtenu : il correspond bien à la bande sonore de la zone temporelle définie.

Cette extraction partielle aurait pû être obtenue d'une autre manière. *VirtualDub* dispose en effet de fonctionnalités basiques de montage vidéo. Il est donc possible de couper la vidéo en enlevant les zones correspondantes aux trames 0 à 300, puis celles correspondant aux trames 1000 à 1773. Pour

cela, on sélectionne comme précédemment une zone temporelle (par exemple, trames 0 à 300), puis on coupe les séquences correspondantes dans le film en choisissant dans le menu `Edit > Cut`. Une fois la vidéo réduite à la zone intéressante, il ne reste plus qu'à extraire la bande sonore sur toute la vidéo restante. Il est par ailleurs possible de générer un nouveau fichier vidéo correspondant aux scènes conservées. Pour cela, il suffit que sélectionner dans le menu `File` l'option `Save as Avi` : la vidéo « réduite » est ainsi créée au format AVI. Cette démarche est très utile lorsqu'on souhaite synchroniser la bande son extraite avec les séquences vidéos correspondantes.

Une fois cette opération réalisée, vous disposée donc d'un fichier audio sur lequel va pouvoir être réalisée la transcription orthographique de l'interaction. Celle-ci se fera à l'aide du logiciel libre *Transcriber* (TP suivant).