

# Annotations pour l'accessibilité des vidéos dans le cas du handicap visuel

Benoît Encelle  
Université de Lyon, CNRS  
Université Lyon 1  
LIRIS, UMR5205, F-69622, France  
benoit.encelle@liris.cnrs.fr

Yannick Prié  
Université de Lyon, CNRS  
Université Lyon 1  
LIRIS, UMR5205, F-69622, France  
yannick.prie@liris.cnrs.fr

Olivier Aubert  
Université de Lyon, CNRS  
Université Lyon 1  
LIRIS, UMR5205, F-69622, France  
olivier.aubert@liris.cnrs.fr

**Résumé**—L'accessibilité des contenus vidéo aux personnes en situation de handicap représente un enjeu important pour l'égalité de l'accès aux ressources numériques. Cet article expose une approche à base de métadonnées visant à améliorer l'accessibilité des vidéos aux personnes mal et non-voyantes. On ajoute dans un premier temps à une vidéo des annotations qui décrivent textuellement des éléments visuels clés de son contenu. Dans un deuxième temps, on construit, à l'aide de modes de présentation de ces annotations et de la vidéo, des rendus enrichis de celle-ci, accessibles. Après le recueil et l'analyse des besoins des non-voyants, nous avons mené une étude préliminaire permettant d'évaluer notre approche et les mécanismes associés. Le développement de ces propositions et leur application à d'autres types de handicaps sont discutés et des perspectives, en cours d'étude au sein du projet ACAV (*Annotation Collaborative pour l'Accessibilité Vidéo*), sont exposées.

**Mots-clés**—Annotations, audiovisuel, braille, handicap visuel, vidéos enrichies

## I. INTRODUCTION

La part de l'audiovisuel numérique dans les contenus diffusés sur les réseaux augmente sans cesse. Les contenus audiovisuels ne sont cependant pas accessibles en pratique aux personnes en situation de handicap, telles que les non-voyants ou les sourds. Celles-ci ne peuvent que difficilement accéder de manière confortable aux vidéos, alors que le numérique permet maintenant d'offrir de nombreux modes d'accès à d'autres types de contenu (pages Web statiques par exemple).

Dans un contexte législatif favorisant l'égalité des personnes face à l'accès à l'information, améliorer l'accessibilité aux contenus audiovisuels apparaît comme un enjeu de société crucial, favorisant l'autonomie et, de ce fait, l'intégration des personnes en situation de handicap.

Pour tenter de répondre à cet enjeu de société, deux questions doivent être abordées. La première, qualitative, consiste à savoir comment améliorer l'accessibilité des vidéos. La seconde, quantitative, consiste à savoir comment augmenter à moindre coût le volume de vidéos accessibles.

Nous nous intéresserons plus particulièrement dans cet article à l'amélioration à l'aide d'annotations de l'accessibilité des contenus vidéo aux personnes non-voyantes.

Après avoir détaillé les solutions existantes et les projets, acteurs autour de l'accessibilité des contenus vidéo aux personnes non-voyantes (section II), nous posons le cadre général des questions de recherche associées à cette problématique

(section III) et présentons une étude ayant mené à des propositions pour l'accessibilité vidéo que nous avons évaluées (section IV). Une discussion de nos propositions et des perspectives ouvertes suit (sections V et VI).

## II. ACCESSIBILITÉ DES VIDÉOS AUX NON-VOYANTS

### A. Solutions classiques et limites

La technique de l'audiodescription permet aux non-voyants et malvoyants de mieux suivre et comprendre un film sans voir les images. C'est un procédé qui consiste à insérer dans « le blanc des dialogues » une voix-off décrivant les éléments purement visuels du film : actions, mouvements, expressions, décors, costumes, etc. L'association Valentin Haüy<sup>1</sup> (AVH) emploie régulièrement ce procédé, également appelé audiovision, pour décrire de nombreux films ou pièces de théâtre, descriptions effectuées par des acteurs (en général deux voix : un homme et une femme). L'audiodescription d'un film est un processus complexe en plusieurs étapes : traduction textuelle du contenu visuel, enregistrement en cabine insonorisée, puis mixage et pressage. Ce processus est de fait long et très coûteux : une audiodescription coûte entre 4000 et 6000 euros et se réalise en moyenne en un mois. Il n'est par conséquent pas raisonnable de vouloir décrire à l'aide de ce procédé une masse toujours grandissante de contenus vidéo (incluant notamment ceux situés sur le Web) au vu des coûts financiers et temporels que cela impliquerait. Il s'avère donc nécessaire de proposer des solutions alternatives pour améliorer l'accessibilité de ce type de contenu aux mal- et non-voyants.

### B. Projets et acteurs

La majorité des travaux sur l'accessibilité vidéo s'est concentrée sur le handicap auditif, en proposant notamment des vidéos enrichies par des sous-titres ou du langage signé (voir par exemple le lecteur de lesite.tv<sup>2</sup> qui propose ces deux rendus). Des outils permettent de décrire des vidéos, tels que MAGPie<sup>3</sup>, Nico Nico Douga<sup>4</sup> ou YouTube subtitler<sup>5</sup>.

1. <http://www.avh.asso.fr/>

2. <http://pourtous.lesite.tv/>

3. [http://ncam.wgbh.org/invent\\_build/web\\_multimedia/tools-guidelines/magpie](http://ncam.wgbh.org/invent_build/web_multimedia/tools-guidelines/magpie)

4. <http://www.nicovideo.jp/>

5. <http://yt-subtitles.appspot.com/>

Les auteurs de [1] ont développé un navigateur dédié appelé aiBrowser proposant notamment une accessibilité minimale à base de métadonnées des vidéos pour les non-voyants : durée totale de la vidéo, timecode du moment en train de jouer, contrôles basiques du lecteur.

Cette approche de textualisation minimale des vidéos et de leur jeu est proposé notamment par le W3C qui s'intéresse à la fois aux problématiques de l'accessibilité via la Web Accessibility Initiative<sup>6</sup> (WAI) et à la standardisation de langages et protocoles liés à la vidéo en lien avec la future recommandation HTML5<sup>7</sup>. La WAI a notamment produit des recommandations concernant les applications riches<sup>8</sup> et les outils de production de contenu<sup>9</sup>. Depuis la conférence *Video on the Web*<sup>10</sup>, de nombreuses initiatives liées à HTML5 sont à signaler. La fondation Mozilla, pionnière dans la mise en œuvre des fonctionnalités vidéo de ce langage via son navigateur Firefox, a mené ses travaux liés à l'accessibilité vidéo en recensant les besoins et les outils existants et en développant des bases de solutions [2]. Les travaux de l'*open video alliance*<sup>11</sup> prennent également en compte l'accessibilité. La convergence de ces initiatives complémentaires devrait donner lieu à un futur groupe de travail dédié à l'accessibilité vidéo, dont les prémisses sont une réunion à Stanford en novembre 2009<sup>12</sup>.

Une approche plus ambitieuse d'enrichissement des vidéos est proposée par le projet canadien E-inclusion [3], qui regroupe un réseau d'utilisateurs, d'artistes, de producteurs et de chercheurs afin de développer des outils de traitement des contenus audiovisuels, pour permettre aux créateurs et producteurs de contenu d'améliorer la richesse de l'expérience des personnes ayant une déficience sensorielle, notamment en automatisant des aspects du processus de production et de post-production multimédia. Ce projet développe deux thématiques principales. La première vise au développement d'outils logiciels pour l'analyse et l'extraction automatique de contenu audiovisuel dans des documents vidéo (découpage en plan, visages, etc.) et la deuxième examine comment la reconnaissance de la parole et du contexte sonore peut aider à automatiser certains des processus de production et de post-production qui comptent parmi les plus coûteux en main-d'œuvre, comme la post-synchronisation, la correction des scripts, le sous-titrage en direct et en différé. Un outil de visualisation enrichie est proposé, qui utilise la synthèse vocale pour décrire aux mal/non-voyants le contenu visuel de la vidéo.

### III. UN CADRE GÉNÉRAL : L'ENRICHISSEMENT DES VIDÉOS À BASE D'ANNOTATIONS

Il apparaît de façon générale que pour être plus facilement accessibles aux personnes en situation de handicap sensoriel,

certaines informations des pistes vidéo ou audio doivent être transmises sur d'autres modalités sensorielles (i.e. transmodalité). La question qui se pose alors est de savoir quelles informations doivent être « transformées » et de quelle(s) manière(s).

En ce qui concerne les non-voyants, des informations visuelles spécifiques doivent être transcrites puis transmises par exemple sur des modalités audio ou tactile. Si certaines informations visuelles peuvent être de manière automatique repérées relativement facilement (les changements de plan, les visages notamment [4]), les résultats ne sont pas obtenus en temps réel et leur mauvaise qualité nécessite le plus souvent une correction manuelle. D'autres informations visuelles sont quant à elles quasi-impossibles à obtenir automatiquement (e.g. décrire en audio l'habillement d'un acteur). La transformation à la volée des informations visuelles dans d'autres modalités semble donc pour l'instant hors de portée.

Il convient alors de prendre une approche plus modeste, qui consiste à décrire le contenu visuel des vidéos dans des métadonnées ayant un format facilement transformable pour pouvoir être transmises sur telle et/ou telle modalité(s).

Ces métadonnées sont appelées *annotations*, que l'on définit comme des informations temporellement situées dans une vidéo. Par exemple, on pourra avoir une annotation de type « Vêtements » qui a pour contenu « Pantalon gris et chemise blanche, chapeau » associée au fragment d'un film défini entre les *timecodes* 1min52 et 2min34. Les annotations permettent de mettre en place des *descriptions* des vidéos.

On appelle alors *rendu enrichi* d'une vidéo tout jeu de la vidéo auquel on associe une ou plusieurs pistes de jeu des annotations. Par exemple, on jouera le film et on oralisera les annotations de type vêtement au moment où débutent les annotations considérées. Un spectateur non-voyant pourra alors écouter la bande-son du film enrichie par une voix lui décrivant à 1min52 les vêtements de l'acteur en scène. Les rendus enrichis constituent des *modes de présentation* des descriptions réalisées.

Il y a deux grands types d'approche à base d'annotations pour obtenir des rendus enrichis de vidéos.

En ce qui concerne l'annotation de fragments de la vidéo, Annodex [5] définit un format basé sur des standards ouverts, permettant d'insérer des annotations HTML dans un flux vidéo, pour ensuite les afficher dans un lecteur dédié tout en jouant la vidéo. Dans cette approche, comme dans d'autres telles que SMIL ou Youtube annotation, les annotations sont directement destinées à un mode de visualisation particulier, et les interfaces qui permettent de les créer sont également très liées à ce mode de visualisation. Ces annotations ne sont donc pas indépendantes de leur rendu.

D'autres approches, comme Anvil [6], Advène [7] ou Ligne de Temps [8], proposent des structures d'annotations génériques et extensibles. Dans ces modèles, la structure d'annotation est relativement orthogonale aux modes de visualisation, ce qui permet de réutiliser les mêmes annotations quelques soient leurs modalités de présentation. Cette dernière approche est plus porteuse dans le cadre de l'accessibilité,

6. <http://www.w3.org/WAI/>

7. <http://dev.w3.org/html5/spec/Overview.html>

8. <http://www.w3.org/TR/wai-aria/>

9. <http://www.w3.org/TR/ATAG20/>

10. <http://www.w3.org/2007/08/video/>

11. <http://openvideoalliance.org/open-video-conference/>

12. <http://www.w3.org/2009/11/01-media-minutes>.

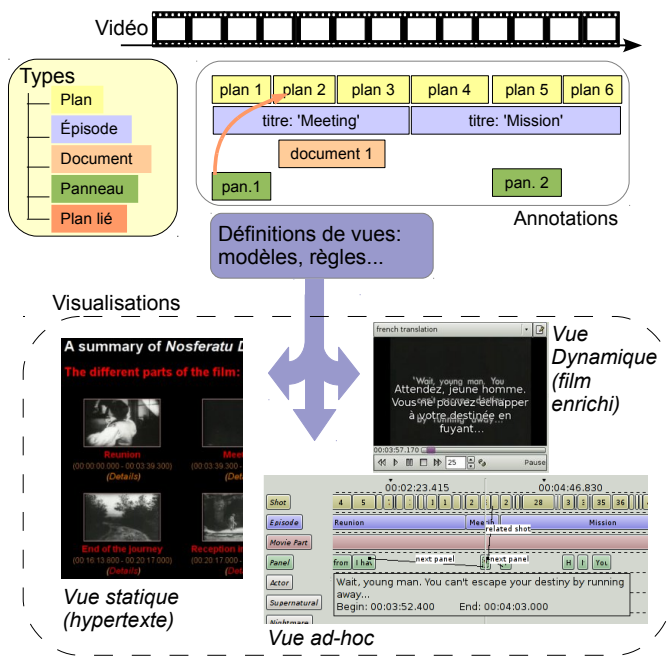


FIGURE 1. Principe de fonctionnement d'Advene : des annotations catégorisées sont associées à la vidéo, puis remobilisées pour générer des visualisations

où les modalités de présentation doivent pouvoir varier d'un utilisateur à l'autre, sans nécessiter une ré-édition complète des annotations.

La plateforme libre Advene<sup>13</sup> développée au LIRIS permet d'une part d'associer des annotations (contenant des descriptions) à des contenus vidéo et, d'autre part, de spécifier des modes de présentation de ces annotations. Une fois saisies, les annotations, liées à un fragment d'une vidéo et catégorisées par la notion de *type d'annotation*, sont remobilisées afin de permettre dans un premier temps une navigation dans la vidéo, et dans un deuxième temps la définition de multiples visualisations conjointes de la vidéo et des annotations. Comme le présente la figure 1, ces visualisations peuvent prendre la forme d'interfaces dédiées, de document hypertextes permettant une navigation dans la vidéo, ou d'un enrichissement du rendu de la vidéo elle-même, comme par exemple l'affichage du contenu des annotations en tant que sous-titres ou un résumé (remontage) de la vidéo généré à partir des annotations. Les visualisations sont spécifiables par l'utilisateur, afin de permettre d'expérimenter rapidement de nouvelles formes d'interaction avec les vidéos. De plus, la plate-forme elle-même est extensible via un système de plugins, ce qui a permis de l'étendre facilement avec des fonctionnalités de pilotage d'une plage Braille ou d'un moteur de synthèse vocale comme nous le verrons par la suite.

13. <http://advene.org>

#### IV. ÉTUDE SUR L'ACCESSIBILITÉ DES VIDÉOS AUX NON-VOYANTS, PROPOSITIONS

Nous avons mené une étude pour mieux comprendre les avantages et inconvénients des modes d'accès aux vidéos employés actuellement par les non-voyants (section IV-A). Cette étude nous a permis d'identifier plusieurs besoins en matière d'accessibilité (section IV-B) et nous a conduit à développer plusieurs propositions pour y répondre (section IV-C). Ces propositions ont ensuite été évaluées (section IV-D).

##### A. Recueil des besoins

Des entretiens semi-guidés ont été effectués avec deux sujets non-voyants contactés avec l'aide de la Mission Handicap de l'Université Claude Bernard Lyon 1<sup>14</sup>.

Ces entretiens ont été enregistrés pour ensuite être analysés, et les questions suivantes ont été abordées. Premièrement, un ensemble de questions préliminaires a été posé afin de mieux caractériser les deux sujets. Outre les questions classiques (âge, profession, etc.), des questions concernant leur capacité à lire le Braille et leur connaissance en Braille ont été posées : lisez-vous le Braille ? Quels codes Braille connaissez-vous ? (e.g. code 6 points, code abrégé, code informatique 8 points, code musical).

Ensuite, la première question (*Q1*) portait sur la fréquence à laquelle ils regardent des vidéos (émissions télévisuelles, films) que celles-ci soient audio-décrites ou non et dans quelles conditions ils les regardent.

La question suivante (*Q2*) les invitait à parler de l'audio-vision en précisant les avantages et les inconvénients de ce système.

Enfin, la dernière question (*Q3*) visait à recueillir leurs avis concernant l'intérêt, la pertinence, de présenter les descriptions sur différentes modalités (par exemple toutes les descriptions en audio ou en Braille, ou certaines en audio et d'autres en Braille).

##### B. Analyses

Par rapport à la première question (*Q1*), le premier sujet regarde fréquemment des vidéos non audio-décrites (films à la télévision) et mentionne qu'il n'a suivi que deux ou trois fois des vidéos/spectacles audio-décrits (films en audiovision, pièces de théâtre). Le deuxième sujet est amateur de cinéma et assiste régulièrement à des séances pour voyants. Il regarde également beaucoup de films en audiovision.

Les sujets, lorsqu'ils regardent des vidéos non audio-décrites (à la maison, au cinéma), demandent à des personnes proches (conjoint, amis) des descriptions orales supplémentaires. Ils précisent toutefois qu'ils ne peuvent avoir ces descriptions que dans certains contextes (i.e. lorsque cela ne dérange pas les autres spectateurs).

De ce fait, il apparaît d'une part nécessaire de développer des solutions donnant aux non-voyants des descriptions supplémentaires (besoin **1a**). D'autre part, ces solutions

14. Service d'accueil et d'aide aux étudiants de l'Université en situation de handicap quel que soit celui-ci <http://handy.univ-lyon1.fr/>.

doivent, dans la mesure du possible, donner aux non-voyants ces descriptions de manière à ne pas déranger les autres spectateurs (besoin **1b**).

Concernant l’audiovision (*Q2*), le premier sujet trouve que les descriptions données par l’audiovision sont parfois trop verbeuses et contiennent trop de détails (e.g. sur les décors par exemple) non nécessaires à la compréhension de l’histoire, pouvant de ce fait réduire son plaisir d’écoute. Le deuxième sujet souligne la pertinence des descriptions apportées par l’audiovision et apprécie leur neutralité. Par contre, il indique que des descriptions chevauchent parfois les dialogues ou des bruitages ce qui amène une certaine gêne.

De ce fait, il apparaît important de trouver un juste équilibre entre compréhension de la vidéo (en donnant suffisamment de descriptions) et plaisir d’écoute (en ne donnant que les descriptions nécessaires, en faisant attention à leur longueur). De plus, il semble être pertinent de personnaliser les enrichissements : par exemple, le premier sujet peut ne vouloir que des descriptions courtes pour les décors alors que le deuxième, au contraire, en désirera des plus détaillées (besoin **2**).

Concernant la possibilité d’employer différentes modalités pour présenter les descriptions (*Q3*), les sujets ont présenté un certain enthousiasme même s’ils ont naturellement demandé à expérimenter cette possibilité avant de se prononcer sur sa pertinence.

### C. Propositions

Pour répondre aux besoins précédemment identifiés, deux propositions ont été faites.

**Qualification des descriptions** : cette proposition vise à répondre aux besoins **1a** et **2**. Selon Turner et Mathieu [9], cinq types d’information doivent être privilégiés dans une description pour les personnes non-voyantes, par ordre d’importance : les personnages et leurs relations, les actions, les lieux, le temps et les décors. Suite à notre entretien avec un des sujets, amateur de cinéma, nous avons ajouté un sixième type d’information à décrire : les mouvements de caméra et de plan. Une description pourra alors être qualifiée, caractérisée par un des types énoncés. En plus de cette qualification des descriptions, nous proposons de définir trois niveaux de verbosité (minimal, normal et complet) pour les caractériser. De cette manière, une description sera qualifiée en fonction de son type et en fonction de son degré de verbosité (i.e. niveau de détail). Ce système de qualification permettra de ne présenter, en fonction des préférences du non-voyant, que les descriptions souhaitées.

**Modes de présentation des descriptions** : pour répondre au besoin **1b** et pour étudier l’éventuelle pertinence d’utiliser différentes modalités lors de la présentation des descriptions, plusieurs modes de présentation ont été développés. Un mode de présentation est défini par un ensemble de modalités et, pour chaque modalité, par les types et degrés de verbosité des descriptions qu’elle véhicule.

La liste suivante récapitule différents modes de présentation développés, certains n’employant qu’une modalité, d’autres plusieurs.

- présentation des descriptions à l’aide d’une synthèse vocale (1)
- descriptions complètes (tous types confondus)
- descriptions mixtes (complètes et minimales pour les décors)
- présentation des descriptions en Braille intégral (2)
- descriptions complètes (tous types confondus)
- descriptions mixtes (complètes et minimales pour les décors)
- présentation des informations de caméra, de plan en Braille et à l’aide d’icônes sonores (3)
- présentation des descriptions en Braille et à l’aide d’une synthèse vocale (4)

### D. Evaluation

Nous avons mené une étude informelle dans le but d’évaluer les précédentes propositions auprès de deux sujets non-voyants.

Nous avons employé comme contenu vidéo le film *Les enfants du paradis*<sup>15</sup> qui a été annoté en s’inspirant de son audiodescription par l’AVH. Les annotations ont ensuite été qualifiées selon leur type et leur niveau de verbosité.

Nous avons enfin présenté différents rendus à des non-voyants suivant les différents modes proposés.

**Extension des rendus dans Advène** : la plateforme Advène a été employée pour mettre en œuvre nos propositions. Nous avons tout d’abord utilisé la plate-forme comme outil d’annotation (i.e. description) des vidéos. Nous avons également étendu ses possibilités d’enrichissement de rendu des vidéos avec :

- des rendus oralisés (faire prononcer par une voix synthétique le contenu d’une annotation au moment de son temps de début) ;
- des rendus sonores (jouer un son par exemple au début d’un plan) ;
- des rendus Braille (afficher sur une plage Braille un contenu d’annotation).

Les annotations ont été catégorisées suivant différents types (actions, personnages, décors, échelle de plan, etc.) et qualifiées par leur niveau de verbosité (minimal, normal, complet).

La capture d’écran figure 2 présente une partie de l’interface d’Advène. Le lecteur vidéo (1) dispose des contrôles standards d’un lecteur, et le titre du mode de présentation en cours est affiché au dessus (ici, un affichage en Braille littéraire de la description de la scène). Différents modes de présentation sont disponibles (2). Une *timeline* (3) offre une vue conventionnelle des différentes annotations suivant leur position dans la vidéo : chaque ligne représente un type d’annotation (e.g. personnages, actions, décors...), identifiable par un nom et par une couleur. Enfin, nous avons utilisé un émulateur de plage Braille (4), placé ici au dessus de l’interface d’Advène, pour présenter en noir les informations envoyées à la plage Braille. En plus de l’affichage d’informations, la plage Braille peut également être utilisée pour naviguer dans la vidéo, en utilisant

15. Marcel Carné, 1945

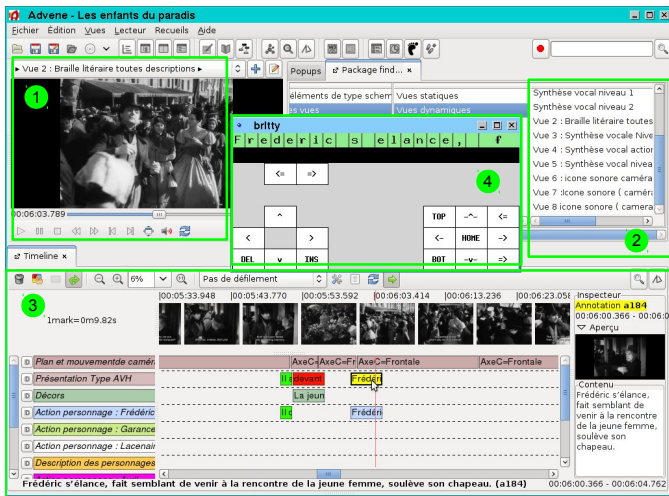


FIGURE 2. Interface d'Advène : lecteur vidéo (1), liste de rendus enrichis (2), timeline présentant les annotations catégorisées (3), émulateur de page Braille BrTTY (4)

les annotations ou bien en posant des signets pour revenir à un moment donné.

**Déroulement** : une séance d'évaluation de nos propositions avec un sujet non-voyant était composée de trois phases. La première était une phase de discussion préliminaire pour mieux connaître le sujet (profession, intérêts pour le média vidéo, connaissances du Braille), de prise en main du logiciel Advène et d'explication du déroulement de la séance.

Ensuite, la seconde phase était une phase d'observation du sujet non-voyant employant le logiciel. Au cours de cette phase, des séquences du film ont été présentées au sujet, chaque séquence étant associée à un mode donné de présentation des annotations. Après le test d'un mode de présentation, un court entretien semi-guidé a été effectué avec le sujet afin de vérifier sa compréhension de la séquence et de recueillir ses impressions « à chaud ».

Étant donné le peu de disponibilités des sujets non-voyants, seuls les modes de présentation des familles (1), (2) et (3) ont pu être testés (cf. section IV-C).

La dernière phase d'une séance d'évaluation consistait en une discussion générale sur la séance, les tests effectués et les perspectives envisagées.

**Résultats** : concernant la qualification des annotations et son usage en cours de lecture pour sélectionner les annotations à présenter, les sujets ont grandement apprécié les différents niveaux de verbosité leur permettant d'avoir une description minimaliste des éléments visuels importants de la vidéo ou, au contraire, une description très « riche ».

Concernant la description des mouvements de caméra et des changements de plan les avis sont mitigés. Un sujet trouve que cette information est importante pour transmettre le « rythme » du film, l'autre n'en voit pas l'intérêt. La modalité audio employée (*earcons*) pour transmettre ce genre d'information ne semble pas être adéquate pour tout le monde.

Concernant les modalités de présentation, la synthèse vocale utilisée, pourtant de qualité médiocre, a été rapidement et

facilement appréhendée.

De plus, la modalité tactile (i.e. Braille) a été unanimement appréciée : les sujets ont facilement compris et mis en relation les informations Braille et audio (i.e. informations de la bande son de la vidéo). Certaines remarques indiquent qu'il faut faire attention à la longueur des annotations présentées en Braille, celle-ci devant être proportionnelle à la vitesse de lecture du spectateur. De ceci découle naturellement la proposition d'emploi du Braille abrégé, lu par la plupart des jeunes non-voyants.

## V. DISCUSSION

Les premiers résultats de l'étude menée sont plutôt encourageants : les observations et remarques des testeurs amènent à penser que les propositions exposées doivent être développées. Cependant, cette étude est à considérer comme une étude préliminaire : une évaluation à plus grande échelle doit être accomplie pour valider les précédentes propositions. Il convient également d'évaluer certains modes de présentation que nous n'avons pas encore pu tester (e.g. présentation des annotations en Braille et à l'aide d'une synthèse vocale) et d'en proposer d'autres. Une étude en cours a notamment pour objet d'évaluer l'intérêt d'icônes sonores (i.e. *auditory icons*) pour la présentation des informations de caméra et de plan car celles-ci sont a priori plus intuitives et appréhendables que les *earcons* [10]. Une autre étude vise, en fonction du profil du spectateur (e.g. ses préférences, ses connaissances à un instant donné), à adapter en cours de lecture l'enrichissement d'une vidéo (quelles annotations présenter et de quelles manières ?) afin de minimiser l'effort cognitif et de ce fait maximiser le confort (i.e. plaisir) du spectateur.

L'approche à base d'annotations peut évidemment être employée pour rendre accessibles des vidéos aux malentendants et sourds : des annotations peuvent représenter le contenu audio de la vidéo, tels que les dialogues et les bruitages, et être présentées sous la forme de sous-titres ou de télétextes. Si ceci n'est pas nouveau (cf. section II-B) l'enjeu ici est que nous souhaitons étudier la possibilité de générer des rendus, à partir d'une *base d'annotations unique* et en utilisant différentes modalités, pour différents types de handicaps.

## VI. CONCLUSION ET PERSPECTIVES

Cet article expose une approche à base d'annotations et des propositions associées pour améliorer l'accessibilité des contenus vidéo aux non et malvoyants. Après une revue de quelques solutions existantes, projets et acteurs du domaine, nous avons présenté notre approche d'enrichissement de vidéos à base d'annotations : des annotations, contenant des descriptions textuelles, sont attachées à des fragments de la vidéo pour être ensuite présentées en cours de lecture selon différentes modalités (e.g. audio — synthèse vocale, sons — et/ou tactile). Une étude préliminaire nous a permis d'évaluer certains modes de présentation des annotations visant à améliorer l'accessibilité des vidéos (e.g. en Braille, à l'aide d'une synthèse vocale). Ces modes de présentation ont été maquetés à l'aide du

logiciel Advène. Si les résultats confirment le potentiel de l'approche retenue et des propositions, mécanismes développés, il convient de continuer leur développement et leur évaluation.

Nous nous sommes dans cet article concentrés sur la question de l'amélioration de l'accessibilité des vidéos, mais n'avons pas abordé celle de l'augmentation du nombre de vidéos accessibles. L'approche adéquate dans ce dernier cas nous semble passer par la notion d'« accessibilité sociale ». Certains travaux comme [11] ou [12] (du projet d'IBM appelé précisément *Social Accessibility*<sup>16</sup>), visent en effet à utiliser les réseaux sociaux pour maintenir collaborativement un annuaire des ressources web accessibles ou rendre des ressources plus accessibles.

Tout en continuant à travailler sur les rendus enrichis de vidéos à base d'annotations, nous allons désormais explorer la voie de l'accessibilité sociale au sein du projet ACAV<sup>17</sup> (Annotation Collaborative pour l'Accessibilité Vidéo), qui regroupe le LIRIS, EURECOM et DailyMotion, ainsi que la mission Handicap de Lyon, l'INSHEA et la Cité Scolaire pour jeunes déficients visuels René Pellet. Ce projet vise à mettre en place des outils directement accessibles dans les navigateurs permettant d'une part l'annotation collaborative de vidéos sur le web en vue de leur accessibilité et, d'autre part, l'obtention de rendus enrichis de ces vidéos. Ce projet vise plus spécifiquement trois classes de scénarios liés à l'utilisation de vidéos en situation de handicap sensoriel dans des contextes différents : scolaire, grand-public et en lien avec les ayants-droit de DailyMotion. Ce projet sera l'occasion de pousser les tests de nos modes de description et de présentation sur des populations plus nombreuses et variées, de mener des études plus fines sur la perception multimodale, de tester les possibilités des traitements automatiques pour une pré-description des paroles, etc. C'est également dans le contexte de ce projet que nous participerons au groupe de travail W3C sur l'accessibilité vidéo si celui-ci est lancé.

#### REMERCIEMENTS

Les auteurs remercient Olivier Helleu, qui a participé aux travaux présentés dans la section IV lors de son stage de M2 Professionnel de Sciences Cognitives de février à juin 2009.

#### RÉFÉRENCES

- [1] H. Miyashita, D. Sato, H. Takagi, and C. Asakawa, "aiBrowser for Multimedia: Introducing Multimedia Content Accessibility for Visually Impaired Users," in *9<sup>th</sup> International Conference on Computers and Accessibility (Assets'07)*, 2007, pp. 91–98.
- [2] S. Pfeiffer and C. Parker, "Accessibility for the HTML5 <video> element," in *6<sup>th</sup> International cross-disciplinary conference on Web accessibility (W4A'09)*, Madrid, Spain, 2009, pp. 98–100.
- [3] L. Gagnon, S. Foucher, M. Heritier, M. Lalonde, D. Byrns, C. Chapdelaine, J. Turner, S. Mathieu, D. Laurendeau, N. T. Nguyenand, and D. Ouellet, "Towards computer-vision software tools to increase production and accessibility of video description for people with vision loss," *Universal Access in the Information Society*, vol. 8, no. 3, pp. 199–218, 2009.

16. <http://sa.watson.ibm.com/>

17. <http://blog.dailymotion.com/acav/>, projet financé par la DGCIS dans le cadre de l'appel à projet *Web innovant*.

- [4] S. Foucher and L. Gagnon, "Automatic detection and clustering of actor faces based on spectral clustering techniques," in *CRV '07: Proceedings of the Fourth Canadian Conference on Computer and Robot Vision*. Washington, DC, USA : IEEE Computer Society, 2007, pp. 113–122.
- [5] S. Pfeiffer, C. Parker, and C. Schremmer, "Annodex : a simple architecture to enable hyperlinking, search and retrieval of time-continuous data on the web," in *5th ACM SIGMM International workshop on Multimedia information retrieval*, 2003, pp. 87–93.
- [6] M. Kipp, "ANVIL - A Generic Annotation Tool for Multimodal Dialogue," in *Proceedings of Eurospeech 2001*, Aalborg, Sep 2001, pp. 1367–1370.
- [7] O. Aubert and Y. Prié, "Advène : Active Reading Through Hypervideo," in *16<sup>th</sup> ACM Conference on Hypertext and Hypermedia*, Salzburg, Austria, 2005, pp. 235–244.
- [8] V. Puig, "Lignes de temps, une plateforme collaborative pour l'annotation de films et d'objets temporels," in *Actes de l'université d'automne, Collection "À propos de ...", Les représentations du réel à l'écran, Ministère de l'Education Nationale/Cinémathèque française*, oct 2007.
- [9] J. Turner and S. Mathieu, "Audiovision ou comment faire voir l'information par les personnes aveugles et malvoyantes : lignes directrices pour la description d'images en mouvement," in *Congrès annuel de l'ACSI*. Montreal, Canada : Canadian Association for Information Science, 2007.
- [10] S. Garzonis, S. Jones, T. Jay, and E. O'Neill, "Auditory icon and earcon mobile service notifications : intuitiveness, learnability, memorability and preference," in *CHI '09 : Proceedings of the 27th international conference on Human factors in computing systems*. New York, NY, USA : ACM, 2009, pp. 1513–1522.
- [11] S. Ferretti, S. Mirri, M. Rocchetti, and P. Salomoni, "Notes for a Collaboration : On the Design of a Wiki-type Educational Video Lecture Annotation System," in *International Conference on Semantic Computing (ICSC'07)*, Irvine, CA, USA, 2007, pp. 651–656.
- [12] H. Takagi, S. Kawanaka, M. Kobayashi, D. Sato, and C. Asakawa, "Collaborative Web Accessibility Improvement : Challenges and Possibilities," in *11<sup>th</sup> International Conference on Computers and Accessibility (Assets'09)*, 2009, pp. 195–202.