

TECHNOLOGIES D'ANNOTATION VIDÉO ET LEURS APPLICATIONS À LA PÉDAGOGIE DU PIANO

Bruno Emond

Institut de technologie de l'information, Conseil national de recherches Canada
Département des sciences de l'éducation, Université du Québec en Outaouais

Marion A. Barfurth

Faculté d'éducation, Université d'Ottawa

Gilles Comeau

Département de musique, Université d'Ottawa

Martin Brooks

Institut de technologie de l'information, Conseil national de recherches Canada

Résumé

La vidéo est utilisée pour soutenir l'apprentissage et l'enseignement dans plusieurs domaines. Cependant, la majorité des applications de la vidéo à des fins d'apprentissage et d'enseignement utilise peu les avantages de la vidéo numérique. Un élément fondamental associé à la vidéo numérique est la possibilité d'annoter de manière automatique ou manuelle le contenu visuel et sonore et d'amalgamer ces annotations à la vidéo pour créer de nouveaux documents multimédias. L'ajout d'annotations ouvre toutes sortes de possibilités sur le plan du développement de ressources médiatiques pour soutenir l'apprentissage, l'enseignement et la recherche sur la pédagogie du piano. Le présent article présente quelques exemples d'application d'annotations vidéo. Ces applications sont discutées dans le contexte des étapes clés de la création vidéo soit le tournage, le montage, la gestion des ressources, et la diffusion.

INTRODUCTION

Cet article présente une partie des recherches effectuées au Laboratoire de recherche en pédagogie du piano en collaboration avec le Conseil national de recherches Canada (CNRC), et plus particulièrement son programme de recherche sur la communication visuelle dans les réseaux à larges bandes passantes. L'objectif de cette collaboration est de mettre au point des technologies de communication vidéo qui faciliteront l'étude systématique d'une vaste gamme de modalités de téléapprentissage du piano, en mettant à profit le travail déjà réalisé en matière de téléconférences à large bande, de serveurs vidéo et outils pour l'apprentissage assisté par vidéo. C'est sur ce dernier élément que portera essentiellement le présent article en mettant en évidence certaines technologies d'annotation vidéo et d'indexation du contenu visuel et sonore de documents audiovisuels.

La vidéo est utilisée pour soutenir l'apprentissage et l'enseignement dans plusieurs domaines dont la physique (Constantinou et Papadouris, 2004), les mathématiques (Blisset et Atkins, 1993), les sports (Guadagnoli, Holcomb et Davis, 2002 ; Horn, Williams et Scott, 2002) les soins de santé (Hill, Hooper et Wahl, 2000), et la consultation thérapeutique (Urdang, 1999). Son utilisation en musique n'est pas non plus une pratique récente (Broyles, 1997 ; Fung, 1998 ; Wagner et Brick, 1993). Cependant, la majorité des applications de la vidéo à des fins d'apprentissage et d'enseignement utilise peu les avantages de la vidéo numérique (Shephard, 2003).

Comparée à la vidéo analogique, la vidéo numérique permet de stocker et de rendre disponible de manière plus efficace de grandes quantités de documents audiovisuels, soit localement ou à distance. La vidéo numérique permet aussi de parcourir le contenu de manière non linéaire, ce qui facilite grandement la sélection et le visionnement de segments vidéo, et de combiner rapidement des segments vidéo issus de sources différentes (Constantinou et Papadouris, 2004). L'accès en mode continu à distance est particulièrement intéressant parce qu'il permet de faire la lecture d'un flux audio ou vidéo à mesure qu'il est diffusé. Il s'oppose ainsi à la diffusion par téléchargement qui demande que le média soit transféré sur le disque local de l'utilisateur pour qu'il soit visionné.

Un autre élément fondamental associé à la vidéo numérique est la possibilité d'annoter de manière automatique ou manuelle le contenu visuel ou sonore et d'amalgamer ces annotations à la vidéo pour créer de nouveaux documents multimédias. De grandes quantités d'information peuvent ainsi enrichir le contenu audiovisuel. L'ajout d'annotations à la vidéo ouvre toutes sortes de possibilités sur le plan du développement de ressources médiatiques pour soutenir l'apprentissage, l'enseignement et la recherche sur la pédagogie du piano. L'ajout d'annotations augmente de manière considérable l'interactivité de la vidéo (Geißler, 1995). Sans des moyens de naviguer dans un document audiovisuel, celui-ci demeure essentiellement un médium narratif qui supporte difficilement un apprentissage actif où l'apprenant a un contrôle sur le médium (Laurillard, 1995 ; Shephard, 2003). La présence d'annotations peut faciliter grandement cette interactivité et le contrôle par l'apprenant, en identifiant les segments vidéo pertinents qui correspondent à un but ou à une intention poursuivie par l'apprenant.

Les technologies avancées d'annotation permettent déjà de générer automatiquement par machine des métadonnées (Hunter et Newmarch, 1999), des résumés (Zhang et Nunamaker, 2004), des abstractions visuelles (Yeung et Yeo, 1997), ou d'offrir des représentations simples et concises du contenu qui permettront à l'utilisateur de sélectionner et de naviguer dans des segments vidéo sans avoir à visualiser des sections ayant moins d'intérêt. De telles applications sont essentielles à la consultation de banque de vidéos (Hunter et Newmarch, 1999). Ainsi, l'information textuelle ajoutée au document audiovisuel, comme l'analyse d'expert sur les techniques utilisées lors d'une leçon de piano, peut enrichir le contenu vidéo en offrant des légendes dynamiques qui serviront non

seulement d'outil d'apprentissage et d'enseignement mais aussi d'éléments de recherche sur la pédagogie du piano. L'annotation multimodale (Snoek et Worring, 2005) permet aussi d'associer le contenu MIDI au document audiovisuel, permettant ainsi de faire jouer le piano simultanément avec le visionnement, ou même d'utiliser le piano comme outil de recherche de document vidéo (comme rechercher tous les documents vidéo qui incluent une certaine séquence de notes). Les recherches en vidéo informatique permettent aussi de développer des applications où une ou plusieurs caméras peuvent indiquer si le professeur ou l'élève a ses mains sur le clavier et quel doigt est utilisé pour appuyer sur chacune des notes du piano (Gorodnichy et Yogeswaran, 2006).

Les prochaines sections du présent article visent à présenter une vue d'ensemble de la problématique de l'annotation vidéo et des différentes solutions qui peuvent y être apportées. La première section présente quelques notions de base touchant l'annotation vidéo. Par la suite, l'annotation sera discutée en relation avec la pédagogie du piano dans le contexte des étapes de saisie, de production et de diffusion de la vidéo.

QU'EST-CE QUE L'ANNOTATION VIDEO ?

L'annotation vidéo est le processus par lequel des informations textuelles ou autres sont associées à des segments déterminés de documents vidéo pour en enrichir le contenu. Ces informations ne modifient pas le document original, mais sont simplement mises en correspondance avec celui-ci. Définie ainsi, une annotation est un terme générique qui regroupe aussi bien l'ajout d'informations sans contraintes particulières, comme un échange de courriel à propos d'une vidéo, ou l'ajout d'informations qui doit respecter un format bien défini. Dans ce second cas, le terme métadonnées est plus approprié et spécifique. Ainsi plusieurs standards comme « Dublin Core » (Dublin Core Metadata Initiative, 2004) ou « IEEE Learning Object Model » (Learning Technology Standards Committee, 2002) permettent de décrire des contenus multimédias en utilisant une liste définie d'attributs comme la date de création, les auteurs, l'encodage utilisé par la vidéo, la résolution de l'image, ou la bande passante exigée pour sa transmission par réseaux. Dans le domaine de la vidéo, un standard de description du contenu d'un média tel que MPEG7 (JTC1/SC29/WG11, 2004) permet de décrire de manière fine les éléments sonores et visuels d'un document audiovisuel numérique.

Le processus d'annotation est automatique ou manuel. L'annotation automatique procède par traitement informatique du signal vidéo ou sonore pour en extraire des éléments servant à l'indexation. L'intervention humaine dans ce contexte est minimale et consiste uniquement à lancer et interrompre les applications informatiques. L'annotation automatique est effectuée par machine et elle peut être plus ou moins complexe selon les techniques et la sophistication des outils d'intelligence artificielle utilisés. L'annotation la plus simple à ce niveau consiste à extraire des informations sur les caractéristiques du

média comme la durée de la vidéo. Ce type d'extraction automatique n'implique aucune technique d'analyse du signal vidéo. Par contre, la détection des mouvements des mains et des doigts au-dessus du clavier et l'appariement de ces mouvements aux messages MIDI transmis par le piano permettent de produire des annotations en temps réel des images vidéo. Ainsi, lors de la saisie, un système de capture des images des mains jouant sur les notes peut spécifier si les mains de l'élève ou du professeur de piano se trouvent au-dessus du clavier et quels doigts sont utilisés pour enfoncer les notes (Gorodnichy et Yogeswaran, 2006).

L'annotation manuelle quant à elle, exige une intervention humaine. Ce type d'intervention est nécessaire lorsque le traitement automatique ne peut suffire pour catégoriser des segments vidéo. L'annotation manuelle n'a pas de contrainte particulière sauf qu'elle peut demander un certain niveau d'expertise de la part de la personne annotant un document audiovisuel. Ainsi, l'identification d'une technique de doigté ou l'approche pédagogique utilisée demandent l'intervention d'un expert en pédagogie du piano. Cette information peut être saisie à l'aide de logiciels spécialisés d'édition vidéo (comme *Final Cut Pro*) ou par l'intermédiaire d'une interface Web, interface qui permet aux usagers de commenter le contenu de vidéos livrées par des serveurs (Emond et Brooks, 2003 ; Emond, Brooks et Smith, 2001 ; Schroeter, Hunter et Kosovic, 2004).

ANNOTATIONS ET PRODUCTION VIDEO

Production vidéo

Il y a plusieurs points dans le cycle de vie qui va de la saisie à la diffusion d'images vidéo où le contenu peut être annoté. Les trois étapes principales de ce cycle de vie sont : le tournage, le montage, et la diffusion. Lors du tournage, les images sont enregistrées par l'entremise d'une caméra soit sur ruban magnétique, soit directement sur le disque d'un ordinateur. L'étape de montage consiste à numériser, à sélectionner et à organiser des segments vidéo, et finalement à compresser le document vidéo dans un format adéquat pour diffusion tel MPEG4 (JTC1/SC29/WG11, 2002). L'étape de diffusion consiste à rendre disponible le document vidéo par l'intermédiaire de serveurs Web et de serveurs vidéo.

Même si en principe il est possible d'effectuer des annotations automatiques et manuelles à chacune des étapes du cycle de vie des documents vidéo, l'annotation automatique se retrouve surtout aux étapes de tournage, et de montage, alors que l'annotation manuelle se retrouve surtout aux étapes de montage (par des experts) et de diffusion (par des utilisateurs). En fait, l'annotation automatique s'effectue sur des documents vidéo en cours de saisie ou lorsqu'ils ont été finalisés à l'étape de montage puisque le contenu vidéo est déterminé suite à cette dernière étape. L'annotation manuelle par contre est difficile lors du

tournage puisqu'elle demande souvent une analyse du contenu en fonction d'un domaine d'expertise. Par contre, cette expertise peut s'exercer beaucoup plus facilement à l'étape du montage et de la diffusion puisque l'opérateur humain peut naviguer librement dans la ligne temporelle de la vidéo et a donc le temps nécessaire pour faire l'analyse de contenu.

La figure 1 représente certains processus et produits du cycle de vie de la production vidéo et d'annotations qui peut y être ajoutées. La figure est divisée en trois étapes principales, la gestion des médias étant associée à la phase de montage. La gauche de la figure présente les moments importants qui mènent de l'enregistrement vidéo jusqu'au visionnement de documents multimédias en passant par les étapes spécifiques au montage. Les sections du centre et de droite présentent différents processus d'annotations et les résultats de ces processus. Les sections suivantes décrivent ces méthodes d'annotation vidéo.

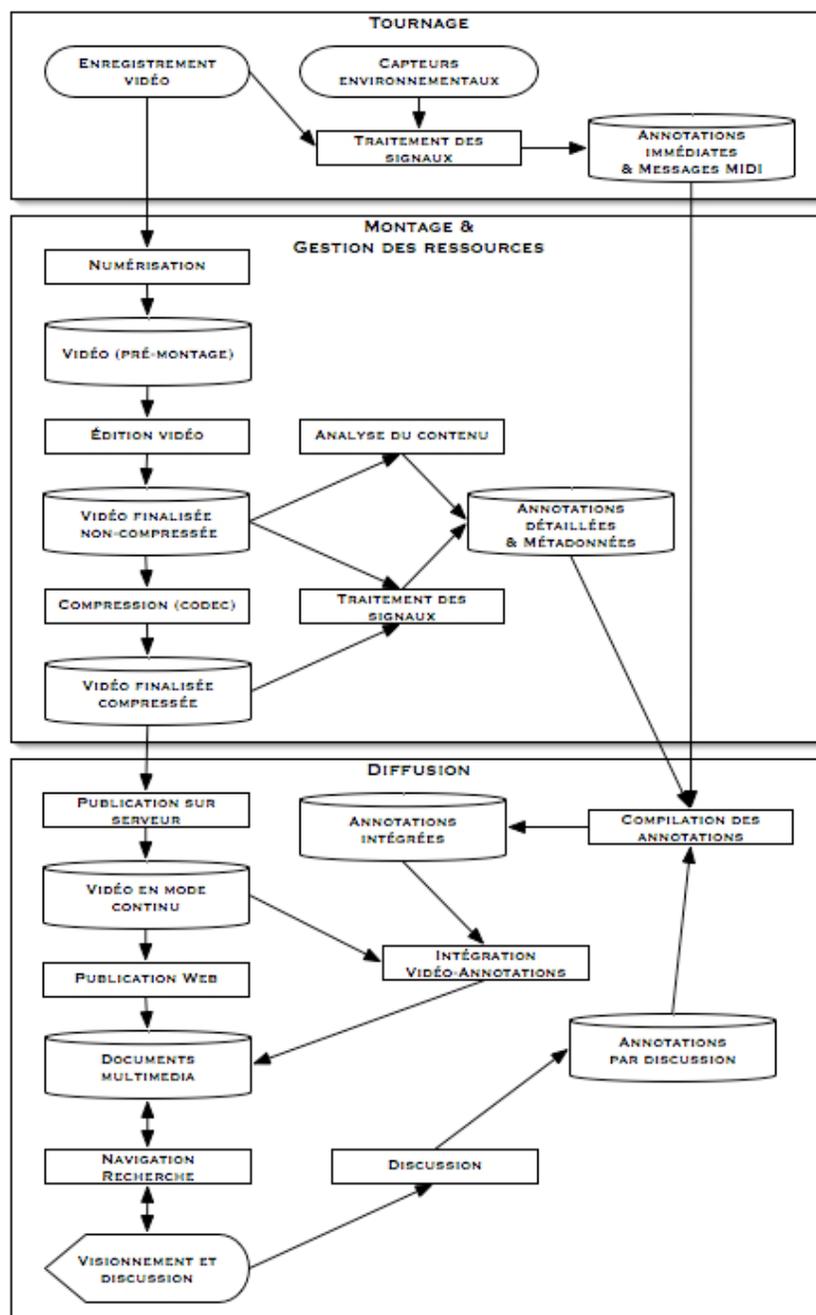


FIGURE 1. Processus, documents audiovisuels et annotations générés lors du tournage, du montage, de la gestion des médias, et de la diffusion de documents multimédias.

Tournage

La complexité inhérente aux situations d'apprentissage exige souvent la présence de plusieurs caméras (Comeau, Brooks et Spence, 2004 ; Stigler, Gonzales, Kawanaka, Knoll et Serano, 1999), et la richesse en contenu des situations d'apprentissage limite de manière importante la possibilité pour un ou plusieurs observateurs humains d'annoter les événements qui se déroulent en temps réel. Cependant, le traitement automatique des signaux vidéo et sonore permet d'appliquer des catégories en parallèle et simultanément au processus d'enregistrement des données audiovisuelles sur bandes ou sur disques magnétiques. De plus, lorsque les instruments de musique ont des capacités MIDI (International MIDI Association, 1983, 1988), comme le piano, il est possible d'associer au contenu audiovisuel la performance musicale saisie sous la forme de messages MIDI. Une application de ce type a été développée au Laboratoire de recherche en pédagogie du piano, en collaboration avec le CNRC, pour identifier les doigts utilisés pour appuyer sur les notes du piano. Une caméra, située au-dessus du clavier, traite le signal vidéo en temps réel, repère les doigts et apparie cette classification aux messages MIDI qu'émet le piano. L'image vidéo qui surplombe le clavier peut ainsi être annotée en établissant une correspondance entre le message MIDI et l'image de la caméra située au-dessus du piano (Gorodnichy et Yogeswaran, 2006). Les annotations immédiates ainsi générées peuvent être compilées par la suite afin de produire des documents multimédias et guider l'utilisateur dans la découverte et la recherche de segments vidéo.

Montage et gestion des médias

La production d'annotations et de métadonnées est une activité importante lors du montage et de la gestion des médias puisque les métadonnées sont un élément essentiel de la gestion des médias numériques (Hunter et Newmarch, 1999). Cependant, les schémas de métadonnées sont souvent limités quant à leur capacité de décrire spécifiquement le contenu de domaines particuliers tels que l'enseignement du jeu pianistique. L'intention principale des schémas de métadonnées comme Dublin Core (Dublin Core Metadata Initiative, 2004) ou IEEE LOM (Learning Technology Standards Committee, 2002) vise l'interopérabilité, soit le partage de ressources entre différentes plateformes et domaines d'application. D'où le besoin de développer des outils et des vocabulaires spécialisés d'annotation vidéo, en complément aux applications génériques de métadonnées.

Le CNRC a mis au point une application d'acquisition de connaissances d'experts qui s'intègre à un logiciel spécialisé d'édition vidéo. Lors de l'étape de montage, un expert peut ajouter des étiquettes à des segments vidéo dans le même environnement où se fait l'édition vidéo. Le résultat de l'ajout de ces annotations à une vidéo en cours d'édition est exporté dans des fichiers de liste d'édition qui sont par la suite traités pour générer des documents multimédias utilisant le format SMIL (World Wide Web Consortium, 2005). Le Laboratoire de recherche en pédagogie du piano est en train de développer, en

collaboration avec le CNRC, des catégories pour annoter des leçons de piano afin de faciliter l'apprentissage et l'enseignement à distance (Brook, 2006). Cette technologie a déjà été appliquée avec succès pour l'indexation de leçons de violon. Cette dernière application comporte une interface en quatre langues, l'anglais, le chinois, le japonais et le français.

Diffusion

Une fois que les annotations ont été produites automatiquement ou manuellement aux étapes du tournage et du montage, elles doivent être compilées et intégrées pour produire des documents multimédias. À l'étape de diffusion, les documents vidéo ont été finalisés et ils sont prêts à être livrés par des serveurs vidéo. Cependant, il reste que les usagers qui visionnent les documents peuvent aussi contribuer à enrichir ces documents en les annotant. Une application Web, servant d'interface avec les serveurs vidéo, peut permettre à l'élève et au professeur de piano, ainsi qu'au chercheur en pédagogie du piano d'apporter leur contribution et de discuter du contenu des documents audiovisuels. Le serveur d'annotations vidéo (Emond et Brooks, 2003 ; Emond et al., 2001) permet aux usagers qui consultent un répertoire de ressources vidéo d'ajouter des annotations textuelles. Cette application a été utilisée dans le cadre de deux projets impliquant plusieurs commissions scolaires à travers le Canada. Le premier projet, LearnCanada (www.learncanada.ca), portait sur le développement professionnel d'enseignants, alors que le second, MusicGrid (www.musicgrid.ca), cherchait à soutenir les programmes de musique à l'aide de la vidéoconférence et des réseaux à larges bandes passantes (Masum, Brooks et Spence, 2005).

CONCLUSION

Cet article a présenté une partie des recherches effectuées au Laboratoire de recherche en pédagogie du piano, en collaboration avec le Conseil national de recherches Canada (CNRC). L'objectif de cette collaboration est de mettre au point des technologies de communication vidéo qui faciliteront l'étude systématique d'une vaste gamme de modalités de téléapprentissage du piano, en mettant à profit le travail déjà réalisé en matière de téléconférences à large bande, de serveurs vidéo et outils pour l'apprentissage assisté par vidéo. L'utilisation de la vidéo comme soutien à l'apprentissage peut exploiter davantage les possibilités de la vidéo numérique. Plus particulièrement, la possibilité d'enrichir le contenu audiovisuel par l'ajout d'annotations et de métadonnées joue un rôle essentiel pour la gestion des médias, l'accès et la recherche de documents pertinents. Ainsi, l'indexation de leçons de piano, en y ajoutant des annotations, permet d'accéder à des segments particuliers de vidéos à partir de divers critères de recherche comme le type d'intervention pédagogique, une séquence de notes jouée au piano, le geste technique, ou même la qualité expressive du jeu musical.

Un autre aspect important de l'utilisation généralisée d'annotations de la vidéo est qu'elle ouvre la possibilité d'accéder aux mêmes documents multimédias dans plusieurs contextes. Les annotations peuvent s'appliquer à un contexte d'apprentissage, où l'élève peut accéder à des séquences vidéo contenant des instructions spécifiques, en compléments des leçons reçues en face à face ou à distance. Au stade initial de sa formation, l'élève peut aussi utiliser la vidéo comme modèle pour acquérir certains gestes techniques, ou réviser le doigté d'un segment d'une pièce musicale. Les annotations peuvent aussi soutenir le contexte d'enseignement, où l'enseignant peut accéder à des séquences vidéo lors de leçons à titre d'exemple, où à des fins de développement professionnel pour comparer ses interventions pédagogiques à celles répertoriées dans la base de données multimédias, ou ajouter des commentaires à une vidéo. Finalement, les annotations peuvent être utiles dans un contexte de recherche sur la pédagogie du piano, où le chercheur peut comparer les interventions pédagogiques et ajouter des annotations caractérisant les situations d'enseignement et d'apprentissage, et ce, afin d'exploiter le potentiel de la vidéo pour analyser et comprendre l'apprentissage non seulement en fonction de ses productions mais du processus (Constantinou et Papadouris, 2004).

Références

- Blisset, G. et Atkins, M. (1993). « Are they thinking? Are they learning? A study of the use of interactive video ». *Computers and Education*, 21 (1) : 31-39.
- Brook, J. (2006). *The development of an asynchronous video reference database to support Suzuki piano instruction*. Communication présentée lors de la Graduate Music Students' Association (GMSA) Conference, Université d'Ottawa.
- Broyles, J. W. (1997). « Effects of videotape analysis on role development of student teachers in music ». Thèse de doctorat (Ph.D.) non publiée, Université d'Oklahoma.
- Comeau, G., Brooks, M. et Spence, J. (2004). *Video and broadband video conference in professional development*. Communication présentée lors de la Biennial Conference of the International Consortium for Educational Development (ICED), Université d'Ottawa.
- Constantinou, C. P. et Papadouris, N. (2004). « Potential contribution of digital video to the analysis of the learning process in physics: a case study in the context of electric circuits ». *Educational research and Evaluation*, 10 (1) : 21-39.
- Dublin Core Metadata Initiative. (2004). *Dublin Core metadata element set, version 1.1: reference description* [en ligne]. <http://dublincore.org/documents/dces/>
- Emond, B. et Brooks, M. (2003). « The Private Video Sharing and Annotation Server: A broadband application for teacher training and music education ». Communication présentée lors de la International Lisp Conference, New York, NY.
- Emond, B., Brooks, M. et Smith, A. (2001). « A broadband web-based application for video sharing and annotation ». *Proceedings of Ninth International ACM Conference on Multimedia*, p. 603-604.
- Fung, V. (1998). « Effect of video presentation on asian music perceptual dimensions ». *Psychology of Music*, 26 (1) : 61-77.
- Geißler, J. (1995). « Surfing the movie space: advanced navigation in movie-only hypermedia ». Communication présentée aux Proceedings of the 3rd ACM International Multimedia Conference (Multimedia 95), San Francisco, CA.
- Gorodnichy, D. O. et Yogeswaran, A. (2006). « Detection and tracking of pianist hands and fingers ». Communication présentée aux Proceedings of the Canadian Conference Computer et Robot Vision (CRV'06), Québec, Canada.

- Guadagnoli, M., Holcomb, W. et Davis, M. (2002). « The efficacy of video feedback for learning the golf swing ». *Journal of Sports Sciences*, 20 (8) : 615-622.
- Hill, R., Hooper, C. et Wahl, S. (2000). « Look, learn, and be satisfied: video playback as a learning strategy to improve clinical skills performance ». *Journal for Nurses in Staff Development*, 16 (5) : 232-239.
- Horn, R. R., Williams, M. A. et Scott, M. A. (2002). « Learning from demonstrations: the role of visual search during observational learning from video and point-light models ». *Journal of Sports Sciences*, 20 (3) : 253-269.
- Hunter, J. et Newmarch, J. (1999). « An indexing, browsing, search and retrieval system for audiovisual libraries ». *Lecture Notes in Computer Science*, 1696 : 76-91.
- International MIDI Association (1983). *MIDI Musical Instrument Digital Interface Specification 1.0*. Los Angeles, CA : International MIDI Association.
- International MIDI Association. (1988). *Standard MIDI Files 1.0*. Los Angeles, CA : International MIDI Association.
- JTC1/SC29/WG11, I. I. (2002). *MPEG-4 Overview - (V.21 – Jeju Version)* [en ligne]. <http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm>
- JTC1/SC29/WG11, I. I. (2004). *MPEG-7 Overview (version 10)* [en ligne]. <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- Laurillard, D. (1995). *Rethinking university teaching*. London : Routledge.
- Learning Technology Standards Committee. (2002). *IEEE Standard for Learning Object Metadata (1484.12.1-2002)* [en ligne]. <http://ieeeltsc.org/wg12LOM/>
- Masum, H., Brooks, M. et Spence, J. (2005). « MusicGrid: A case study in broadband video collaboration ». *First Monday*, 10 (5).
- Schroeter, R., Hunter, J. et Kosovic, D. (2004). « The FilmEd Project - Collaborative annotation and discussion tools for film/video over broadband networks ». Communication présentée lors de la 10th International Multimedia Modelling Conference MMM2004, Brisbane, Australia.
- Shephard, K. (2003). « Questioning, promoting and evaluating the use of streaming video to support student learning ». *British Journal of Educational Technology*, 34 (3) : 295-308.

- Snoek, C. G. M. et Worrying, M. (2005). « Multimodal video indexing: a review of the state-of-the-art ». *Multimedia Tools and Applications*, 25 : 5-35.
- Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S. et Serano, A. (1999). *The TIMSS videotape classroom study: methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. Washington : National Center for Education Statistics Report, US Department of Education.
- Urdang, E. (1999). « The video lab: mirroring reflections of self and the other ». *Clinical Supervisor*, 18 (2) : 143-165.
- Wagner, M. J. et Brick, J. S. (1993). « Using karakoe in the classroom ». *Music Educators Journal*, 79 (7) : 44-46.
- World Wide Web Consortium. (2005). *Synchronized Multimedia Integration Language (SMIL 2.0)* [en ligne]. <http://www.w3.org/TR/2005/REC-SMIL2-20050107/>
- Yeung, M.M. et Yeo, B. L. (1997). « Video visualization for compact presentation and fast browsing of pictorial content ». *IEEE Transactions on Circuits Systems and Video Technology* 7 : 771-785.
- Zhang, D. et Nunamaker, J. F. (2004). « A natural language approach to content-based indexing and retrieval for interactive e-learning ». *IEEE Transactions on Multimedia*, 6 (3) : 450-458.