

Vers une synchronisation du sous-titrage en direct français avec les mouvements de l'image

Claude Chapdelaine, Mario Beaulieu, Langis Gagnon

Département de R-D, CRIM, 550, rue Sherbrooke ouest, Suite 100, Montréal, Québec, H3A 1B9
claude.chapdelaine@crim.ca

RESUME

Nous présentons les premiers résultats d'une étude visant à explorer les facteurs pouvant permettre un couplage adéquat entre le sous-titrage en direct et différents degrés de mouvement dans l'image. Dans cet article, nous quantifions le débit du sous-titrage français et démontrons l'efficacité d'une tâche secondaire tactile pour l'analyse de l'attention visuelle dans ce contexte.

MOTS CLES: Sous-titrage en direct, surcharge attentionnelle, modalité tactile.

ABSTRACT

Our study analyses the factors involved in realizing a convivial embedding of online captioning with its images different levels of movement. Our primary results quantify the word rates for French captioning and ascertain the adequacy of a tactile double task in the analysis of visual attention in this context.

CATEGORIES AND SUBJECT DESCRIPTORS: H.5 [INFORMATION INTERFACES AND PRESENTATION], H.5.2 [User Interfaces].

GENERAL TERMS: Human factors

KEYWORDS: Online caption, attention overload, tactile modality.

INTRODUCTION

Le sous-titrage en direct pour les personnes sourdes et malentendantes nécessite une production dynamique qui laisse peu de place à une bonne coordination entre l'information de l'image et celle donnée par le sous-titrage. Cette coordination est d'autant plus critique lorsque l'image contient beaucoup de mouvement donc plus d'informations et que le sous-titrage doit refléter un haut débit de parole.

Préoccupé par un couplage adéquat entre ces deux

sources d'information, le CRIM a débuté une recherche portant sur l'attention visuelle requise pour effectuer la lecture de sous-titres par les téléspectateurs sourds et malentendants francophones. Cette recherche vise à augmenter la qualité des sous-titres produits en direct. Le direct ne permet pas la qualité d'édition des sous-titres préparés avant diffusion. C'est pourquoi, nous adressons le problème lié à la calibration entre le débit de sous-titres et le degré de mouvement de l'image. Nous tenterons de résoudre ce problème par l'étude de l'attention visuelle. Trois niveaux d'analyse de détection seront combinés soit : le mouvement des yeux, les expressions faciales et la surcharge attentionnelle (traduction de l'auteur pour le terme : *attention overload*).

Suite à la revue de littérature, deux éléments inconnus devaient être résolus pour nous permettre de compléter l'élaboration de nos conditions d'expérimentations. Il nous fallait identifier le débit (mots/minute) de sous-titres français pouvant être lu et nous assurer de l'efficacité d'utiliser une tâche secondaire tactile pour détecter la surcharge attentionnelle. Nous avons entrepris d'adresser ces questions dans une première expérimentation impliquant des personnes entendants, sourdes et malentendantes.

Cet article aborde les problématiques liées au débit du sous-titrage en direct et à la détection de la surcharge attentionnelle des personnes sourdes ou malentendantes. Et quoique l'expérimentation soit toujours en cours, nous présentons les résultats préliminaires obtenus à date.

DÉBIT DU SOUS-TITRAGE

L'ensemble des aspects ergonomiques du sous-titrage est un domaine d'étude restreint. À notre connaissance, aucune étude ne traite du sous-titrage pour le français [3,4,5,8]. Par conséquent, il y a peu de directives ergonomiques qui guident la production de sous-titrage. Les normes et standards en vigueur traitent principalement de l'affichage et de l'édition du texte pour le sous-titrage réalisé avant la diffusion. Dans ce cas, le contenu sonore est résumé afin d'être synchroniser avec le contenu visuel. Par contre lors de la production du sous-titrage en direct, le mode d'affichage et de synchronisation est très différent.

L'affichage du sous-titrage en direct se distingue par son défilement dans une zone fixe de l'écran. La plupart du

sous-titrage est affiché sur trois lignes au bas de l'écran. Sauf pour les émissions de sports où la norme est l'affichage sur deux lignes en haut de l'écran. De son côté, la synchronisation est nécessairement décalée et le rythme est lié aux capacités de l'intermédiaire humain utilisant la sténotypie ou la reconnaissance automatique de la parole. En direct, on tient rarement compte du contenu de l'image pour établir le débit du sous-titrage et c'est précisément cette problématique que nous désirons adresser. Mais d'abord, il faut se préoccuper du débit de sous-titrage qu'un humain peut adéquatement traiter.

Les études de Jensema [2] sur le sous-titrage américain ont démontré que la plupart des émissions anglophones diffusent à un débit moyen de 141 mots/minutes avec des débits variant de 74 à 231 mots/minutes. Dans une étude subséquente [3], Jensema a démontré que le débit confortable pour les sourds-malentendants se situe entre 140 à 156 mots/minutes avec une moyenne 145 mots/minutes. Il a aussi observé que ce rythme était jugé trop rapide par les personnes entendant. Mentionnons également que ces tests ont été faits sur des extraits de courtes durées. Hors, il peut être facile de lire 145 mots/minutes pendant 30 secondes mais beaucoup plus difficile de soutenir ce rythme pendant plusieurs minutes.

Est-ce que cette moyenne de 145 mots/minutes est applicable au débit possible pour le français? Pas nécessairement, puisque la lecture du français est plus facile que celle de l'anglais. En effet, les linguistes ont démontré par une approche de médiation phonologique (transformation des phonèmes en graphèmes) que non seulement l'apprentissage de lecture du français est plus facile et rapide que l'anglais mais qu'également cette approche est un « très bon prédicteur des performances ultérieures en lecture et en écriture » [1]. Puisque la performance acquise à l'apprentissage demeure, nous en déduisons qu'elle influence le débit de lecture. Selon nous, le débit moyen en français pour la lecture de sous-titrage devrait être plus élevé que pour l'anglais. Nous allons donc tester des débits allant de 120 à 220 mots/minutes.

SURCHAGE ATTENTIONNELLE

La théorie la plus répandue qui décrit les phénomènes de partage des ressources de l'attention humaine est la *Multiple-Resource Theory* proposée par Wickens [7]. Elle stipule que la répartition des ressources d'attention lors de l'exécution d'une tâche se définit selon trois dimensions: étapes de traitement (*stages*), modalités et processus (*process codes*). Une modalité pouvant être de nature visuelle, auditive ou tactile.

Une façon de détecter la fluctuation d'allocation d'attention est de doubler une tâche primaire d'une tâche secondaire. Ainsi, lorsqu'une personne tarde ou néglige d'exécuter la tâche secondaire ceci indique que les ressources d'attention sont monopolisées par la tâche

primaire. L'exemple classique est celui du pilote, qui tout en naviguant, doit réciter une liste de chiffres. Si les manœuvres deviennent difficiles automatiquement le pilote abandonne la récitation pour allouer le maximum d'attention à la tâche de pilotage.

Pour notre expérimentation, nous avons décidé d'avoir recours à cette approche. Cependant, le choix la tâche secondaire devait se faire judicieusement. En effet, celle-ci ne pouvait pas impliquer une modalité visuelle puisque ce mécanisme aurait impliqué la direction du regard et aurait ainsi influencé à l'analyse du mouvement des yeux. D'un autre côté, la modalité auditive était aussi à proscrire puisque nos participants étant sourds ou malentendants, ils leur étaient donc impossibles d'effectuer une tâche secondaire en répondant à un stimulus sonore. Par conséquent, nous avons opté pour une tâche secondaire tactile. Mais comme le souligne Sklar [6], il y a peu d'études qui ont exploré l'efficacité d'un mécanisme tactile comme mode de présentation d'information. Nous avons donc intégré une tâche secondaire tactile dans notre expérimentation afin de vérifier son potentiel à détecter adéquatement une surcharge attentionnelle.

EXPÉRIMENTATION

Participants

Les participants seront au nombre de 12 soient six personnes entendant et six autres sourdes ou malentendant. Les tests sont en cours et six personnes entendant ont déjà participé. Ce groupe est composé de quatre hommes et deux femmes. Tous regardent la télévision avec peu ou sans sous-titrage.

Procédure

La tâche primaire effectuée par les participants consiste à visionner des vidéos stimuli sous-titrés et sans audio. La directive donnée au participant est qu'il doit tenter de comprendre le contenu à partir de l'image et des sous-titres comme le ferait une personne sourde ou malentendante. Pendant la projection, la manette émet des vibrations auxquelles le participant doit répondre en appuyant sur une touche de la manette. Après chaque projection, le participant doit évaluer la présentation du sous-titrage et répondre à des questions spécifiques sur le contenu de l'image et des sous-titres.

Vidéo stimulus

Puisque notre intérêt porte sur le problème de balancer le débit de sous-titres avec le degré de mouvement de l'image, nous avons choisi des extraits d'un match de hockey (Figure 1). Ce type de contenu représente un défi particulier. En effet, ces images contiennent beaucoup de changements de plans et les séquences sont très mouvementées. Elles contiennent aussi un copieux débit de parole composé de la description du jeu et des commentaires des animateurs; débit pouvant parfois atteindre 300 à 350 mots/minutes. Nous avons donc composé un jeu de six vidéos stimuli; dont quatre sont

des séquences de jeu et deux proviennent des entractes contenant surtout des interviews mais aussi de courtes séquences de jeu.

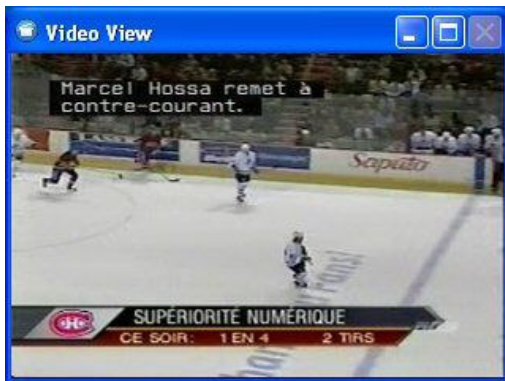


Figure 1 : Échantillon de vidéo sous-titrée

Ensuite, le sous-titrage a été édité à partir d'une transcription intégrale. Le sous-titrage respectait un décalage de trois à quatre secondes pour imiter la diffusion en direct. Le texte a été divisé pour former cinq paliers de débits différents décrits dans le Tableau 1.

Palier	Durée	Mots/minute
1	60 secondes	120
2	30 secondes	140
3	30 secondes	160
4	60 secondes	220
5	60 secondes	130

Tableau 1 : Description des paliers de sous-titrage

Équipement

Une manette de jeu Logitech permettait d'émettre l'information tactile sous forme de vibration. Un logiciel avec interface a été développé avec les bibliothèques DirectX wxWidget, pour synchroniser la vidéo avec les vibrations envoyées (Figure 2).

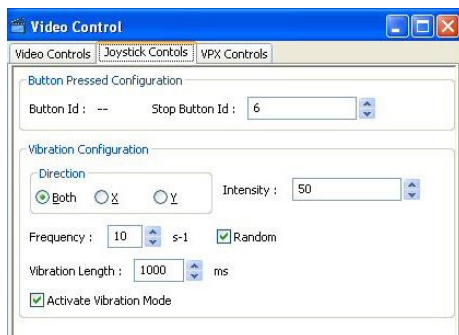


Figure 2 : Interface du logiciel de contrôle

Les vibrations étaient émises à une intensité de 50 Hz pendant 250 ms, et ce, aléatoirement selon une fréquence variant entre cinq à quinze secondes. La manette enregistrait également la réponse tactile du participant. Les vidéos stimuli étaient présentées sur le deuxième

écran d'un ordinateur. Le premier écran permettait de démarrer la projection qui synchronisait les vibrations de la manette et les enregistrements des réponses tactiles des participants.

RÉSULTATS ET DISCUSSION

Pendant les 144 minutes de diffusion (total des six vidéos de quatre minutes des six participants), 1,022 vibrations ont été émises. Dans 37 cas (Figure 3), la réponse du participant a été omise. Le fait qu'il y ait eu omission indique qu'à certains moments le visionnement de l'image et la lecture des sous-titres ont nécessité plus d'attention et que la tâche secondaire a été abandonnée. Dans certains cas, nous avons observé des séquences de trois à cinq vibrations consécutives qui ont été ignorées.

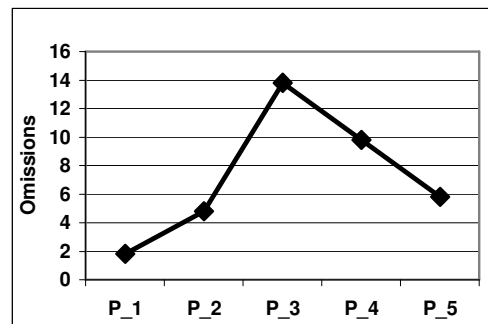


Figure 3 : Nombre d'omissions par palier de sous-titrage

Nous avons aussi effectué une étude des délais. Un délai est déterminé par la durée entre le moment où la vibration est émise et le moment où le participant répond en appuyant sur une touche de la manette. Le délai minimum observé est de 0,297 seconde et la maximum de 6,032. La Figure 4 présente le délai moyen par palier de sous-titrage. On constate que les paliers 4 et 5 qui contiennent les débits élevés (160 à 220 mots/minutes) correspondent aussi aux délais moyens les plus longs.

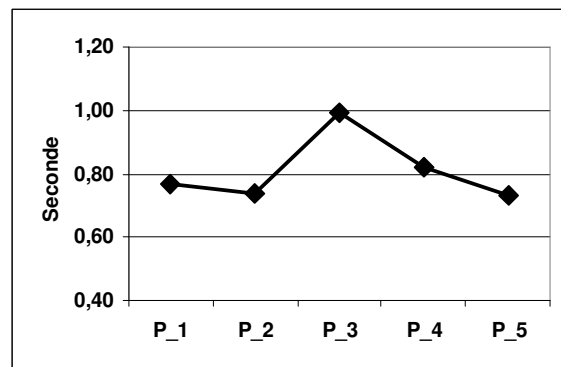


Figure 4 : Délai moyen par palier de sous-titrage

Ce résultat indique que le mécanisme tactile nous a permis de détecter adéquatement une fluctuation dans l'allocation de l'attention. Bien que l'on pouvait s'attendre à retrouver le plus grand nombre d'omissions

et les plus longs délais au palier 4 (220 mots/minute), ceux-ci ont été observés au palier 3 (180 mots/minutes). Ce résultat suggère une limite concernant le débit qui peut être lu. En effet, on observe que le débit de 140 mots/minute du palier 2 ne cause pas de trop longs délais. Ensuite, on observe une augmentation significative au palier 3. On peut supposer qu'à ce niveau, le fort débit demandait beaucoup plus d'attention peut-être même jusqu'à saturation. Ceci expliquerait également la baisse du palier 4 (220 mots/minute) où l'attention portée aux sous-titres semble relâcher, indiquant que ce seuil serait au-delà de la capacité des participants. Cette hypothèse a été corroborée par les commentaires des participants.

Notre dernière analyse a porté sur les délais associés aux vidéos stimuli. Comme le montre la Figure 5, certains vidéo ont causé des délais beaucoup plus longs que d'autres. Notamment, les vidéos 2 et 5 surtout mais aussi le vidéo 6. Hors, seul la vidéo 2 contient une séquence de jeu, les vidéos 5 et 6 sont des extraits d'entracte.

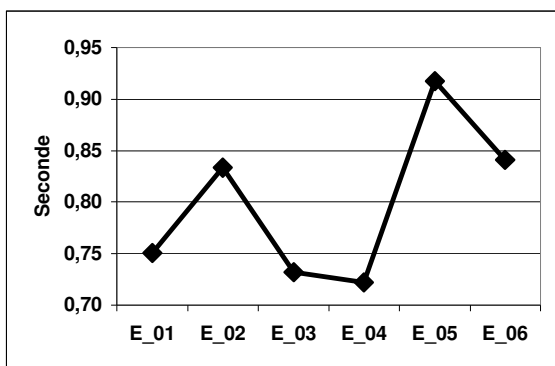


Figure 5 : Délais moyen par vidéo stimulus.

A priori, nous avons supposé que les séquences de jeu solliciteraient beaucoup plus l'attention que l'entracte. Ce n'est pas ce qui a été observé. Ceci peut-être relatif au fait que nos premiers résultats proviennent d'un groupe de personnes entendant qui n'ont pas une lecture de sous-titre expérimentée. Donc, dans les cas où l'information est surtout dans le sous-titrage, ces personnes sont désavantagées et elles ont besoin de plus d'attention pour comprendre. La suite de l'expérimentation avec des personnes sourdes et malentendantes devrait nous permettre de confirmer cette hypothèse.

CONCLUSION

Ces résultats préliminaires nous indiquent que le recours à l'information tactile comme tâche secondaire semble être efficace. On a constaté que lorsque la tâche primaire se complexifie le mécanisme a permis de détecter des ralentissements et des omissions. Du côté du débit du sous-titrage, les premiers résultats indiquent que la limite pour les entendants inexpérimentés à la lecture de sous-

titre serait proche de 160 mots/minute. Ainsi ce débit pourrait être plus élevé pour le français que pour l'anglais. Mais ce sont les futurs tests avec les personnes sourdes et malentendantes qui seront déterminantes.

En plus de poursuivre les tests avec les participants, d'autres analyses devront être effectuées. Par exemple, la détection automatique des changements de plan et des degrés de mouvement dans les séquences nous permettraient de mieux comprendre les phénomènes concurrents dans ce partage des ressources d'attention.

REMERCIEMENTS

Ce travail est supporté en partie par (1) le Ministère du patrimoine canadien via le programme Culture canadienne en ligne, (2) le Conseil de recherche en science naturelle et Génie (CRSNG) du Canada et (2) le Ministère du Développement Économique de l'Innovation et de l'Exportation (MDEIE) du Gouvernement du Québec. Le laborieux travail de production des vidéos stimuli n'aurait pu être complété sans l'assistance précieuse de Sophie Leclerc et Simon Dupuis. Merci pour leur collaboration.

BIBLIOGRAPHIE

1. Fayol, M., Gombert, J.E. *L'apprentissage de la lecture et de l'écriture*. Dans J.A. Rondal et E. Espéret (Eds.), *Manuel de Psychologie de l'enfant*. Bruxelles: Mardaga. 1999, pp. 565-594
2. Harkins, J. *Ergonomic Considerations for Communication Technologies for Deaf and Hard of Hearing People*. HCI 1993: pp. 504-509
3. Jensema, C., McCann R., Ramsey S. *Closed-captioned television presentation speed and vocabulary*. *American Annals of the Deaf*, 141(4), 1996, pp. 284-292.
4. Jensema, C. *Viewer Reaction to Different Television Captioning Speeds*. *American Annals of the Deaf*, 143 (4), 1998, pp. 318-324
5. Maruyama, I., Abe, Y., Sawamura, E., Mitsuhashi, T., Ehara, T., Shirai, K. *Cognitive experiments on timing lag for superimposing closed captions*. Dans EUROPEECH'99, 1999, pp. 575-578.
6. Sklar, A. *Good Vibrations: Tactile feedback in support attention allocation and human-automation coordination in event-driven domains*. *Human factors*, 41(4), 1999, pp. 543-552.
7. Wickens, C. , *Engineering psychology and human performance*, Third edition, Harpers Collins, 1999
8. Woodcock, K. *Ergonomics and automatic speech recognition applications for deaf and hard-of-hearing users*. *Technology and Disability*, 7: , 1997, pp. 147-164