

# **Normes d'Accessibilité Canada**

## **Mieux saisir l'expérience des téléspectateurs du sous-titrage codé de supports médiatiques au rythme rapide, avec le compromis de vitesse/précision/délai de synchronisation**

Research Report #17324237

Proposé par

Deborah Fels, PhD., P.Eng.

## Résumé

Le sous-titrage codé (CC) au Canada a fait de grands progrès depuis les années 1990, mais il reste encore beaucoup à faire pour atteindre les critères d'équité et d'inclusion énoncés dans la Loi Canadienne sur l'Accessibilité. Ce projet a pour but d'examiner les trois plus importantes concessions faites par les personnes sourdes ou malentendantes dans le cadre des émissions en direct, en matière de vitesse, qualité et délai de synchronisation du sous-titrage.

Dans ce projet, nous nous concentrerons plus spécifiquement sur les sports à rythme rapide, domaine particulièrement problématique en raison du peu d'occasions de corriger les erreurs en temps voulu. En outre, la mise sur le marché d'un système de reconnaissance vocale plus fiable et plus précis (Speech-to-Text STT) et l'avènement de l'intelligence artificielle (IA) pourraient présenter des solutions pour certains de ces problèmes. Nous dressons un rapport sur trois études: 1) la comparaison entre le format de sous-titrage intégral (play-by-play) et le format transcrivant uniquement les commentaires (comment-only (CO) pour deux sports à rythme rapide; 2) la charge de travail du sous-titreur; et 3) une étude sur l'expérience des utilisateur·rice·s du logiciel PAVOCAT résultant de cette étude. Deux autres études, dans le cadre de mémoire de maîtrise, qui s'inscrivent dans la même continuité sont en voie de réalisation. Les principales constatations sont les suivantes : 1) la charge de travail des sous-titres pour les retransmissions des émissions au rythme rapide et en direct semble être élevée et devrait être réduite ; 2) le sous-titrage du format basé uniquement sur les commentaires des commentateurs (CO) d'un sport au rythme rapide diffusé en direct au lieu du format du commentaire intégral (PBP) pourrait présenter une solution aux problèmes de vitesse et de qualité pour les sous-titrages de ce type de retransmission; 3) bien que des travaux de recherche et de développement supplémentaires soient nécessaires pour assurer la fonctionnalité et la robustesse du logiciel PAVOCAT qui utilise l'intelligence artificielle pour générer des sous-titres en direct avec la supervision d'un sous-titreur, ce logiciel pourrait offrir une solution aux problèmes de synchronisation et de vitesse au sous-titrage conventionnel; et 4) les sous-titres ne font pas confiance au sous-titrage généré par l'intelligence artificielle. Les deux études restantes examineront la crédibilité des sous-titrages à long terme et le point de vue des téléspectateur·rice·s sourd·e·s et malentendant·e·s sur la fiabilité, la crédibilité, la qualité et la compréhension des sous-titrages générée par l'intelligence artificielle par rapport au sous-titrage conventionnelle dans le cadre des retransmissions en direct et à rythme rapide.

# Introduction

Depuis 1993, le sous-titrage codé (CC) pour tous les supports audiovisuels des multimédias, fait partie du règlement sur les licences du Conseil de la radiodiffusion et des télécommunications canadiennes (CRTC). Depuis, il y a eu une évolution de ces normes et règlements pour le nombre de programmes nécessitant des sous-titrages (CC), et plus récemment, pour leur qualité ; la dernière mise à jour de cette norme de qualité pour la programmation en direct est datée du mois d'octobre 2019. De plus, la norme a été élargie au contenu diffusé en ligne par les radiodiffuseurs, sur leur site. Toutefois, il manque le point de vue des téléspectateur·trice·s pour évaluer la qualité des sous-titrages, afin de compléter ces normes et règlements. Avec l'arrivée de nouvelles technologies, comme la production automatique de sous-titrage (CC), la reconnaissance vocale et l'intelligence artificielle (IA), les normes traditionnelles de production et de qualité sont en train d'être redéfinies et contestées, surtout dans le cadre des émissions retransmises en direct.

Afin de remédier à ces lacunes et à ces défis, nous présentons les résultats d'une série d'études et d'un projet de développement de logiciels financés en 2020 par Normes d'Accessibilité Canada et basés sur un sondage antérieur auprès des téléspectateur·trice·s sourd·e·s et malentendant·e·s sur leur point de vue de la qualité des sous-titrages (CC) pour les retransmissions en direct. La présente étude explore les habitudes de visionnage et l'expérience des téléspectateur·trice·s pour le sous-titrage en direct au sein de deux sports au rythme rapide (le hockey et le basketball), la charge de travail des sous-titreur·s et la compréhension des utilisateurs du sous-titrage généré par l'intelligence artificielle comparé à celui généré par l'homme avec les mêmes critères. Comme le compromis entre vitesse, précision et délai de synchronisation présente un défi considérable pour produire un sous-titrage (CC) de qualité supérieure et pertinente, nous nous sommes concentrés sur ces trois axes.

Les retards inévitables de la production du sous-titrage et le débit de parole (comme dans les commentaires de l'action en intégral (play-by-play - PBP), des commentateurs beaucoup plus rapides que la production de sous-titrage (CC) et la vitesse de lecture des téléspectateur·trice·s, font état d'erreurs et d'un sous-titrage (CC) de mauvaise qualité, interférant avec la compréhension, le plaisir et l'accès équitable au contenu verbal pour les téléspectateur·trice·s, sourd·e·s et malentendant·e·s. Par exemple, le débit maximal recommandé est de 180 mots par minute (WPM) (Jensema, 1998; Ofcom, 2005; Romero-Fresco, P., 2009; Szarkowska, A., Gerber-Moron, O., 2018), mais souvent le débit des commentaires de l'action en intégral (PBP) peut facilement dépasser 220 mots par minutes (WPM).

Les téléspectateur·trice·s ne sont pas en mesure de lire le sous-titrage (CC) à ce rythme et en même temps de regarder et comprendre le jeu. Le sous-titrage (CC) généré automatiquement à l'aide de la reconnaissance vocale (STT) et de l'intelligence artificielle pour le moment disponible pour les applications de visioconférence, telles que Zoom, pourrait présenter une solution pour les émissions avec un débit rapide, mais ces types de sous-titrage (CC) peuvent amener à

d'autres problèmes qui ne font qu'amplifier la lisibilité et la compréhensibilité de ce même sous-titrage (CC). Enfin, la reconnaissance vocale (STT) et l'intelligence artificielle pour le sous-titrage (CC) en direct nécessitent des logiciels qui intégreraient les compétences de l'humain à l'efficacité de l'intelligence artificielle, dans le but d'évaluer les préférences et les performances respectives des différentes méthodes de production de sous-titrage (CC).

Dans le présent rapport, nous présenterons les recherches qui ont été menées pour atteindre les objectifs énoncés dans la proposition de recherche financée par Normes d'Accessibilité Canada. Il convient de noter des retards considérables dans les recherches prévues en personne, en raison de la COVID-19. De plus, les figures et les tableaux demeurent en anglais, car les éléments d'information provenant de cette étude, comme le tableau thématique (tableau 2) était en anglais et ne se basait que sur un contenu en anglais.

## **Les objectifs de la recherche**

### **Objectif 1. Etude comparative**

**Cette étude vise à comparer un ensemble de facteurs de précision des sous-titres modifiés et à évaluer l'expérience des téléspectateur·trice·s, sourd·e·s et malentendant·e·s pour les retransmissions en direct à rythme rapide pour trois sports et/ou événements olympiques avec des équipes masculines et féminines, et deux talk-shows (1 d'intérêt général et 1 sportif en anglais). Les facteurs ajustés concernent l'annonce et les commentaires de l'action intégrale (play-by-play - PBP) par opposition à l'analyse des commentaires sous-titrés (CC) uniquement pour les sports, la suppression des mots à haute fréquence par opposition aux mots à basse fréquence et la quantité de paraphrases.**

Une étude a été conçue sur la base de différentes expérimentations, combinant des données précises de suivi oculaire pour utilisateur·trice·s de sous-titre participant à l'étude avec des données qualitatives de préférence et de compréhension. La conception de l'étude a été approuvée par le comité d'éthique de l'Université Ryerson (aujourd'hui l'Université Métropolitaine de Toronto). Nous avons recueilli des retours subjectifs sur la qualité des sous-titres, l'engagement et l'intérêt des utilisateur·trice·s après le visionnement des séquences de sport, à l'aide d'une série de questionnaires. Nous avons également mené des entrevues de type conversationnel avec les participant·e·s afin d'évaluer leur compréhension en rapport ou non avec le match; nous avons discuté des événements clés qui se sont produits pendant le jeu (par exemple, les pénalités et les buts) et les commentaires des commentateurs sur des informations non liées au match. L'étude comprenait deux extraits de match sous-titré (CC) (hockey ou basketball), chaque extrait présentant un type de sous-titrage différent aux participant·e·s. L'ensemble de l'étude a pris environ 1 heure à chaque participant·e et a été menée sur le campus de l'Université Métropolitaine de Toronto dans l'un des laboratoires d'utilisabilité. Les participantes ont reçu pour leur participation, une rémunération de 60 \$ ainsi que leurs frais de transport.

Cette étude a introduit une nouvelle forme de style de sous-titrage (CC) pour tenter de réduire le nombre de mots, sans réduire l'accessibilité à l'information. Pour illustrer la façon dont les différents modes de sous-titrage (CC) apparaissent à l'écran, la figure 1 montre que lors d'une retransmission de type intégrale (PBP), il n'y a pas de sous-titrage de type transcription des commentaires (CO). Lorsque le jeu s'arrête, alors seulement les légendes de type transcription des commentaires (CO) s'affichent, conformément à ceux pendant le match. Les sous-titres de type intégrale (PBP) sont toujours affichés chaque fois que l'un des commentateurs parle. Notez que dans les images ci-dessous, les sous-titres de type transcription des commentaires (CO) sont centrées sur la page, tandis que les sous-titres de type intégrale (PBP) à droite sont réparties sur toute la partie supérieure de l'écran.

**Figure 1:** Affichage à l'écran des sous-titres de type transcription des commentaires et des sous-titres de type intégral pendant le déroulement du match.

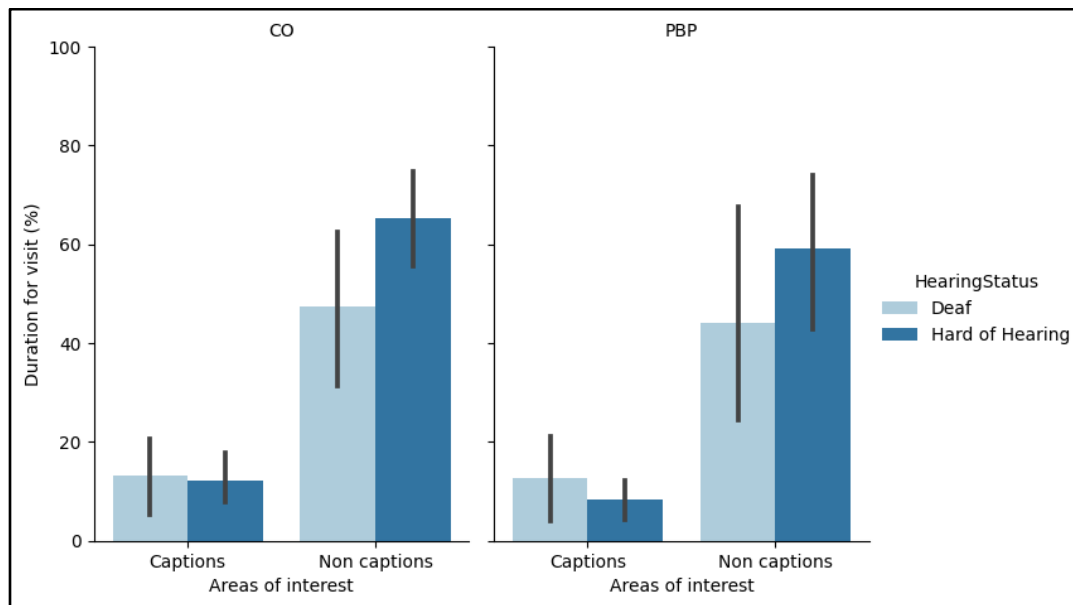


**Figure 2:** Apparition à l'écran des sous-titres de type transcription des commentaires (CO) et des sous-titres de type intégral (PBP) pendant les commentaires en direct.

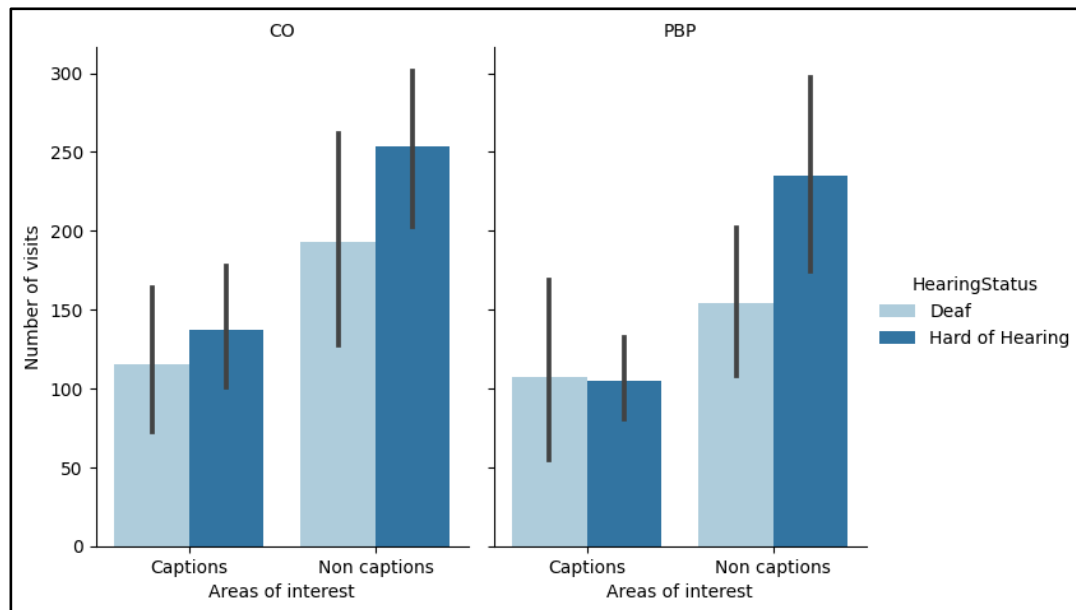


À l'aide du système de suivi oculaire Tobii, nous avons enregistré la durée et la fréquence des regards portés sur les zones d'intérêt (AOI); Chaque regard correspondant à la totalité des temps de fixations, des saccades et des mouvements de l'œil en millisecondes. Les zones d'Intérêt, dans le contexte de cette étude, font référence à la zone de l'écran où sont placés les sous-titres, ainsi que les zones ne comprenant pas de sous-titres (le match en cours, les gradins, le tableau des scores, etc.). Grâce aux enregistrements de suivi oculaire, il a été possible d'effectuer une analyse statistique qui a révélé le manque de différences statistiquement significatives dans la durée et la fréquence des regards entre les deux formats de sous-titres (intégral (PBP) et transcription des commentaires (CO), quel que soit le statut auditif des participant.e.s. L'étude comparative a été réalisée avec succès auprès de 16 participantes sourd.e.s et 11 participantes malentendant.e.s. La durée totale des regards portés sur les sous-titres de type transcription des commentaires (CO) était de 11 %, contre 10 % pour le sous-titrage intégral (PBP). Malgré l'absence de signification statistique, la figure 3 présente les valeurs moyennes du visionnement des participantes sourdes et malentendantes pour chaque type de sous-titres. Les participantes semblent consulter davantage les sous-titres de type transcription des commentaires (CO) que les sous-titres de type intégral (PBP), quel que soit le sport ou leur capacité auditive. Ces résultats insinuent que les sous-titres de type intégral (PBP) semblent perturber le visionnement du match, poussant les participantes à ne pas s'y attarder longtemps à consulter moins fréquemment cette partie de l'écran, indépendamment du format de sous-titres. Les sous-titres en rapport avec l'action du match n'étaient pas affichés dans le format transcription des commentaires (CO) ce qui a peut-être incité les participant.e.s à penser que ces derniers ne valaient pas la peine d'être lus. Le nombre moyen de regards pour les sous-titres de type transcription des commentaires (CO) était de 124, tandis que le nombre moyen de regards pour les sous-titres de type intégral (PBP) était de 106 (voir la figure 4).

**Figure 3:** Moyenne et l'écart-type (lignes) de la durée des visites sur les deux zones d'intérêt pour les deux types de sous-titrage.



**Figure 4:** Moyenne et l'écart-type de la fréquence oculaire portée sur les deux zones d'intérêt pour les deux types de sous-titrage.



Comme les participants étaient habitués à regarder les commentaires de type intégral (PBP), le manque de différence statistique entre les deux types de sous-titrage a révélé une potentielle tendance vers la lecture du sous-titrage de l'analyse des commentaires (CO), l'effet de nouveauté influençant probablement de manière négative l'expérience des téléspectatrices (p. ex., "c'était différent de ce à quoi j'étais habitué"). Il semblerait que ce potentiel soit plus évident en comparant les commentaires positifs des participants du sous-titrage des commentaires (CO) et les commentaires négatifs du sous-titrage de type intégral (PBP), après avoir visionné le match avec les deux types de sous-titrage.

*"Je préfère le sous-titrage du deuxième clip [faisant référence au sous-titrage des commentaires [CO]. C'était bien parce que je comprenais la progression du jeu. Je n'ai pas besoin du format intégral. Quand il y a du direct, les commentateurs n'en parlent même pas, donc je me laisse facilement distraire. J'aime le format transcription des commentaires [CO] parce qu'il n'y avait pas de sous-titres pendant le match en direct. Comme je ne peux pas entendre, je ne pouvais pas savoir s'ils parlaient. J'ai pu rattraper ce qui s'était passé seulement quand le match était interrompu (reprises et temps morts)."*

Un autre participant a décrit comment le sous-titrage de type transcription des commentaires (CO) lui permettait de suivre le match sans avoir à se soucier de rater les moments forts.

*"Je pouvais me concentrer sur le jeu et ne pas m'inquiéter de passer à côté de quelque chose en essayant de lire les deux sous-titres (...) Comme c'était à hauteur*

*des yeux même lorsque le match a commencé, je pouvais en terminer la lecture et en même temps voir ce qui se passait dans ma [vision] périphérique.”*

Pour le sous-titrage de type descriptif (PBP), les participantes ont mentionné combien il était difficile de suivre en même temps le jeu et les sous-titres (CC) affichés à l'écran. Voici plusieurs commentaires des participants sur le sous-titrage de type intégral (PBP):

*“Trop distrayant” / “trop stressant”*

*“Vraiment fatigant” / “Beaucoup d'effort pour les yeux” / “Sans pauses”*

*“Beaucoup de va-et-vient visuel [entre les sous-titres et le jeu].”*

*“Je me sens un peu stressé parce que je n'veux pas passer à côté d'une information importante avant qu'elle ne disparaisse.”*

*“C'est comme si je lisais un livre et que quand je levais les yeux, j'avais déjà manqué les sous-titres.”*

Certains participants ont également ajouté que le sous-titrage habituel (PBP) les avait amenés à éviter tout simplement les sous-titres (CC) et à “seulement regarder le match.” Les commentaires des participantes nous amènent à comprendre la raison de la durée des visites oculaires dans la zone sans sous-titres, les participantes voulant se concentrer sur le visionnement des images et du match, avec le moins de distractions possible.

Cependant, toutes les opinions sur les sous-titres de type transcription des commentaires (CO) n'étaient pas positives et toutes les opinions sur les sous-titres de type intégral (PBP) n'étaient pas négatives. Une personne préférait le sous-titrage de type intégral (PBP) parce qu'elle en avait l'habitude, et surtout parce qu'elle “savait que tout y était”. D'autres commentaires ont fait référence au fait que le sous-titrage du type transcription des commentaires (CO) pouvait omettre de mentionner le nom et le numéro des joueuses, alors que les participantes utilisent aussi le sous-titrage (CC) pour se rappeler les noms.

D'après les réponses du questionnaire, la plupart des participantes (65 %) n'ont indiqué aucune préférence pour les deux types de sous-titres. Les autres participantes préféraient le sous-titrage de la transcription des commentaires (CO) (27 %) au sous-titrage de type intégral (PBP) (8 %). Soixante pour cent des participantes sourdes et 73 % des participantes malentendantes n'ont exprimé aucune préférence. Dans le groupe qui a déclaré une préférence, environ 33 % des participantes sourdes et 18 % des participantes malentendantes préféraient le sous-titrage de la transcription des commentaires (CO), alors que 7 % des participantes sourdes et 9 % des participantes malentendantes préféraient les sous-titres de type intégral (PBP). Après que les participantes aient visionné les extraits, nous avons conclu que la qualité, l'expérience de visionnement et d'utilité du sous-titrage (CC) étaient meilleures pour les sous-titres de type transcription des commentaires (CO) que pour le format intégral (PBP), même si ces différences ne sont pas statistiquement significatives (voir le tableau 3).



**Tableau 1 :** Statistiques descriptives de la satisfaction globale et de l'expérience de visionnage après avoir regardé les extraits pour les deux types de sous-titres.

Sports	Les conditions de visionnement	Status d'audition	Moyenne de satisfaction des sous-titres (SD)	Moyenne de l'expérience visuelle (SD)
Basketball	PBP	Deaf	3.75 (0.96)	4.50 (0.58)
		Hard of Hearing	2.86 (1.35)	4.13 (0.64)
	CO	Deaf	3.75 (0.96)	4.25 (0.96)
		Hard of Hearing	3.63 (1.19)	4.38 (1.06)
Hockey	PBP	Deaf	2.91 (1.38)	3.45 (1.13)
		Hard of Hearing	3.33 (1.53)	3.00 (1.00)
	CO	Deaf	3.36 (1.75)	3.73 (1.27)
		Hard of Hearing	2.33 (1.53)	3.33 (1.53)

Cette étude, la première du genre au Canada, a révélé des informations sur l'expérience de l'utilisateur trice avec le sous-titrage codé (CC) lorsqu'il s'agissait de regarder des matchs sportifs en direct et à rythme rapide pour le hockey et le basketball. Bien que l'industrie et les téléspectateur.trice.s comprennent la nécessité de résoudre les problèmes généraux liés au sous-titrage (CC) en direct, tels que le délai de synchronisation, la précision, la vitesse, l'identification des intervenant.e.s, l'emplacement des sous-titres sur l'écran et la lisibilité, un aspect spécifique au sous-titre des émissions en direct des sports à rythme rapide génère des défis liés à la production et à la lecture d'un nombre plus élevé de mots par minute (WPM). Dans cette étude, nous avons introduit un nouveau style de sous-titrage (CC) pour les retransmissions sportives en direct, n'affichant que les sous-titres basés sur les commentaires en direct, avec moins de mots que dans le contexte de type intégral (PBP) habituel.

Les résultats de la partie de l'étude portant sur les mouvements des yeux ont montré qu'il n'y avait pas de différences statistiquement significatives entre les deux styles de sous-titrage (CC), mais l'analyse descriptive a révélé une différence dans la durée et la fréquence des visites

oculaires, où les participantes ont passé plus de temps à regarder et ont visionné plus fréquemment le sous-titrage de la transcription des commentaires (CO).

De plus, les réponses des participants sur leur satisfaction à l'égard de la qualité du sous-titrage (CC), l'expérience de visionnement et l'aide du sous-titrage (CC) à la compréhension, montrent une préférence pour les sous-titres des commentaires des commentateurs (CO). De plus, bien que beaucoup n'aient pas indiqué de préférence pour le style de sous-titrage (CC), les commentaires des participantes ont permis d'approfondir le raisonnement qui sous-tend les préférences subjectives. Dans le cadre de cette étude, il a été surprenant de constater que les participantes ont très bien accueilli le sous-titrage basé seulement sur les commentaires avec des commentaires positifs, malgré le fait que l'étude était biaisée en faveur du sous-titre de type intégral (PBP) qui jusqu'à présent, constituait le seul type de sous-titrage disponible pour les téléspectateur.trice.s.

Enfin, les résultats de cette étude fournissent des preuves suffisantes pour poursuivre la conception de «commentaire en couleur», l'unique style de sous-titrage dans une étude longitudinale visant à déterminer si le sous-titrage de la transcription des commentaires (CO) est viable et si les téléspectateur.trices étaient satisfaites après avoir surmonté la nouveauté de ce style. De plus, une étude longitudinale pourrait indiquer si les sous-titres de l'analyse des commentaires (CO) amenuisent suffisamment certains des compromis entre la vitesse, la précision et le délai de synchronisation dans les sous-titrages (CC) des émissions sportives en direct et au rythme rapide dans le but d'améliorer l'expérience des téléspectateur.trice.s.

Une deuxième étude dans le cadre d'un projet de mémoire de maîtrise en cours, qui s'étendra au-delà de l'achèvement de ce projet, se penche sur l'impact du sous-titrage (CC) automatique, concept moins familier pour les téléspectateurs, pour les sports rapides, comme ceux conçus pour les Jeux olympiques. Dans cette étude, le sous-titrage (CC) sera produit à l'aide d'une technologie de reconnaissance vocale indépendante du locuteur, telle que l'IA Speech-to-Text de Google (Google Cloud Speech-to-Text, 2024). Les participants regarderont quatre extraits pour quatre sports olympiques d'hiver à rythme rapide, tels que le ski freestyle et le snowboard féminin, avec un sous-titrage (CC) attribué au hasard (de type transcription de commentaires ou intégrale), soit généré par l'homme soit généré automatiquement. Cinq facteurs seront mesurés à l'aide de critères normalisés et d'une méthode de compréhension, développée dans la première étude comparative: la justesse (Bowers & Phillips, 1967, Simpson & Kahler, R. (1981), la satisfaction (Battacherjee et al. 2001), la fiabilité (Bowers & Phillips, 1967), la qualité et la compréhension.

## **Objectif 2. Etude longitudinale**

**Réaliser une étude longitudinale (étude 2) afin de comprendre les effets de l'expérience du téléspectateur.trice et de l'évaluation de la qualité à long terme, auprès de 30 téléspectateur.trice.s sourdes/malentendantes dont les préférences sont déduites de l'objectif 1. Il pourrait s'agir d'une étude d'un à trois mois par participant.e.**

Comme les résultats de la première étude comparative ont révélé une préférence dans l'utilisation des sous-titres des commentaires (CO) au lieu du sous-titrage de type intégral (PBP) conventionnel pour les personnes sourdes/malentendantes (D/HoH), une étude longitudinale portant sur l'exposition à long terme du sous-titrage des commentaires (CO) a été conçue pour un deuxième mémoire de maîtrise, en cours de réalisation. Contrairement à la première étude, cette étude longitudinale vise à fournir aux participants une période d'exposition plus longue (20 minutes par jour pendant trois jours) au sous-titrage des commentaires (CO) pour les sports à rythme rapide. Nous prévoyons que les individus auront suffisamment de temps pour s'habituer à regarder avec moins de mots. La demande d'évaluation éthique a été préparée et soumise au Comité d'éthique de la recherche de l'Université de Toronto.

### **Objectif 3. Amélioration de l'outil de sous -titrage**

**Élaborer et évaluer un logiciel de production de sous -titrage (CC) permet tant de générer des sous -titres (CC) en fonction des résultats des objectifs 1 et 2.**

#### **Part 1. Comprendre la charge de travail des sous -titreurs**

En plus d'analyser les points de vue des téléspectateur·trices sur les sous-titres (CC), nous avons conçu une étude portant sur le sous-titrage en général. Le but de cette étude était d'évaluer la charge de travail mentale subjective (SMW) du sous -titrage (CC) des émissions en direct. Les questions de recherche de cette étude sont les suivantes :

1. Quelle est la charge de travail mentale subjective (SMW) des sous-titreur·s des émissions en direct qui sous-titrent différents genres de contenu en direct ?
2. Quel est l'impact de la paraphrase sur la charge de travail mentale subjective (SMW)?

#### *La méthodologie*

Avec l'approbation par l'université des questions d'éthique, les participant·es ont été invité·es à répondre à un sondage en ligne de 55 questions, comprenant 21 questions de pondération appariées et six questions de classement pour l'outil standard NASA -TLX Task Load Index afin d'évaluer la charge de travail mentale subjective (SMW) (Hart, 1986), quatre questions démographiques, telles que le sexe et l'éducation, 12 questions sur les habitudes et l'expérience générale avec le sous -titrage (CC) des émissions en direct, et 10 questions sur les comportements et les préférences pour la paraphrase des sous -titrages (CC) des émissions en direct. Le NASA -TLX considère six critères (effort mental, demande physique, restrictions temporelles, performance, concentration et frustration) pour évaluer la charge mentale d'un individu effectuant une tâche. Chaque critère est évalué sur une échelle allant de 0 à 20. De plus, il existe une composante de pondération demandant aux participant·es d'évaluer l'importance des paires pour les six aspects.

Nous avons mené une série d'entrevues afin de recueillir d'autres données qualitatives auprès des personnes qui avaient consenti à participer à cette étude. Les entrevues de type conversationnel comportaient des questions sur la configuration du poste de travail, les défis et leurs opinions sur la paraphrase.

### *Les participants*

Trente personnes (27 femmes et 3 hommes) ont répondu à l'ensemble du sondage. Les participantes représentaient quatre pays différents: le Canada (16 sur 30), les États-Unis (11 sur 30), l'Australie (2 sur 30) et l'Union européenne (1 sur 30). Les personnes étaient âgées de 18 à plus de 60 ans. Le niveau d'éducation allait de l'école secondaire aux diplômes d'études supérieures. La majorité des participantes (17 sur 30) avaient plus de huit ans d'expérience avec la production de sous-titrages (CC) des émissions en direct, ayant sous-titré une variété d'émissions en direct, comprenant les nouvelles, les bulletins météo et les divertissements, des sports rapides (comme le hockey sur glace et le soccer), des sports plus lents (comme le golf), des talk-shows sportifs, des talk-shows diffusés dans la journée ou la soirée, et d'autres types d'émissions sous-titrées en direct (telles que des réunions gouvernementales ou d'affaires) et des événements religieux en direct. Enfin, 23 des 30 participantes ont utilisé la sténographie pour le sous-titrage (CC) des émissions en direct, et sept, l'assistance d'un logiciel de reconnaissance vocale (respeaking). Cinq participants ont accepté de participer à la séance de suivi.

### *Les résultats*

Nous avons attribué une note sur 100 pour la charge de travail mentale subjective (SMW), calculée à l'aide de la moyenne des classements pondérés. Les cotes pondérées ajustées pour les six aspects ont été calculées en multipliant les classements de la catégorie par les pondérations correspondantes.

Le nombre total de la charge de travail mentale subjective (SMW) par participant variait de 16,7 à 91,5 ( $M = 62,87$ ,  $ET = 24,04$ ). La figure 5 montre les moyennes et les écarts-types entre les six sous-échelles qui composent l'ensemble de la charge de travail mentale subjective (SMW).

Les sous-titres des émissions en direct considèrent la charge de travail élevée par rapport aux normes (Grier, 2015). La charge de travail mentale subjective (SMW) pour les sous-titres en direct se classe dans le top 25 dans un large éventail d'emplois incluant le domaine médical. Impactant la charge de travail mentale, la performance venait en tête, suivi de l'effort et la nécessité de respecter les délais, la demande physique venant en dernier. Le faible impact de la demande physique se traduit par le fait qu'elle est considérée être inclus dans la tâche, comme l'utilisation du clavier et l'assistance d'un logiciel de reconnaissance vocale, et devrait devenir plus routinières et automatiques avec la pratique. L'importance de la performance en tant que facteur clé de la charge mentale est en grande partie dû aux normes strictes établies par les réglementations gouvernementales, les clients des radiodiffuseurs et les règlements de l'entreprise. Il est difficile d'atteindre ces normes rapidement surtout pour les sous-titres qui

voient leur niveau de stress augmenter considérablement, en particulier doublé avec la pression d'une date limite.

*Personne 1: "La responsabilité de l'erreur incombe davantage aux sténographes (...) et j'essaie d'être correct à 100%."*

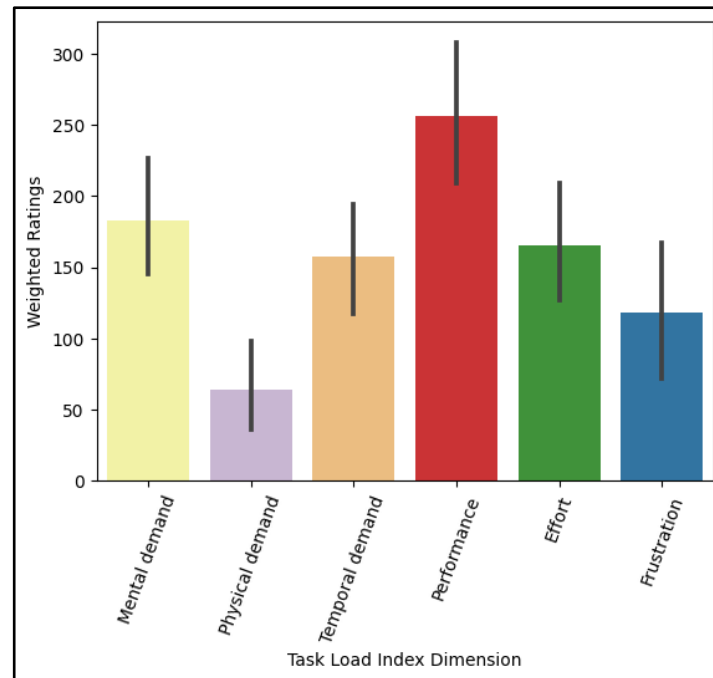
*Personne 2: "Si vous avez commis une erreur, vous ne pouvez pas la corriger"*

Au cours des entrevues, les sous-titres ont indiqué que leur stress mental résidait principalement de devoir fournir des sous-titres (CC) précis, du mot à mot. Ils étaient conscients que la qualité de leur sous-titrage (CC) jouait un rôle crucial dans l'accès des personnes sourdes ou malentendantes (D/HoH) aux médias audiovisuels et que tout manquement à la qualité pouvait aboutir à des plaintes. Ces plaintes sont envoyées aux chaînes de télévision et peuvent avoir une incidence sur leur conformité aux normes gouvernementales, pouvant nuire à la réputation du sous-titreur et à son obtention d'autres contrats.

Une autre analyse croisée a été effectuée pour déterminer l'association entre les catégories de la charge de travail mentale (SMW) et le statut d'emploi des sous-titres. Nous avons trouvé une association statistiquement élevée entre la charge de travail mentale subjective (SMW) et les conditions d'emploi ;  $\chi^2(4, 30) = 10,72$ ,  $p < 0,05$ ,  $\phi = 0,423$ . Un test de suivi de Kruskal-Wallis a révélé des différences significatives dans les scores de la charge de travail mentale subjective (SMW) totaux pour chaque catégorie d'emploi,  $\chi^2(2) = 7,015$ ,  $p < 0,05$  et le test de Dunn utilisant la correction de Bonferroni a révélé une différence significative dans les scores de la charge de travail mentale subjective (SMW) entre les pigistes et les employés en entreprises : pigistes/travailleurs autonomes ( $M = 48,59$ , ÉCART-TYPE = 26,17), employés en entreprise ( $M = 68,61$ , ET = 22,31) et sous-titres qui étaient pigistes et employés par une entreprise ( $M = 76,18$ , ET = 10,65).

Cette constatation a révélé l'association entre le statut d'emploi et la charge de travail mentale subjective (SMW), où les pigistes semblaient avoir des scores (pour la charge de travail mentale subjective (SMW) inférieurs à ceux des sous-titres employés par des entreprises ou des agences. Nous pouvons interpréter cette information de la manière suivante: les sous-titres contrôlant leurs horaires semblent avoir une charge de travail moins stressante que ceux qui sont obligés de compléter des tâches assignées. Parmi les autres facteurs influençant le stress des sous-titres, nous pouvons citer la recherche d'emploi, la répartition des tâches, la planification et la préparation des émissions en direct, y compris la recherche des noms et du vocabulaire spécialisé dans le dictionnaire.

**Figure 5:** Moyenne et écart-type des classements pondérés pour chacune des critères de NASA-TLX. Le pourcentage de chaque critère a été conçu à partir du questionnaire de compromis et a été multiplié par le classement de la rangée correspondante (allant de 0 à 100).



### *Paraphrase des sous-titres*

Nous basant sur le questionnaire, nous avons demandé aux participantes si leur politique de travail leur permettait de paraphraser. La plupart des participantes ont indiqué que leur règlement de travail permettait la paraphrase (21/30), quelques -uns effectuaient la paraphrase au besoin même si ce n'était pas permis (6/30), et les autres participantes (3/30) n'y étaient pas autorisés. Parmi ceux qui paraphrasaient, la majorité a dit que c'est "uniquement lorsque c'était nécessaire" pour tous les types d'émission, tels que les sports rapides, les talk -shows, les prévisions météorologiques et les nouvelles. Certains participants ont affirmé qu'ils "paraphrasaient toujours" pour tous les types d'émission, à l'exception des sports lents. Pour les sports lents, 14 participantes n'avaient "jamais paraphrasé" tandis que les 13 autres ont déclaré qu'ils "paraphrasaient quand c'était nécessaire".

Enfin, une majorité de participants ont déclaré qu'ils croyaient que la paraphrase améliorerait "un peu" la qualité du sous -titrage CC (12/27) ou "beaucoup" (4/27). Sept participants ont déclaré que le changement causé par la paraphrase était "minime" (8/27) ou rendait la compréhension "un peu plus difficile" (entraînant une qualité inférieure du sous -titrage CC) (3/27). Une autre question était "Pour quelle raison omettez-vous des mots dans les sous-titres

des émissions en direct?”, à laquelle une majorité des participants ont répondu “pour suivre la vitesse “(24), “pour réduire les discours redondants “(11) et “lors de la paraphrase “(9).

Nous avons effectué une analyse de tableau à double entrée, utilisant le test de Chi-square sur les critères d'indépendance du sous-titre, pour déterminer l'association entre un niveau élevé, moyen et bas de la charge de travail mentale (SMW) et l'utilisation de la paraphrase pour les différents types d'émission. Nous avons trouvé une association statistiquement importante entre la fréquence de la paraphrase dans les sous-titres (CC) de sports à rythme rapide et la catégorie de la charge de travail mentale subjective (SMW);  $\chi^2(4, 27) = 11,25$ ,  $p < 0,05$ ,  $\phi = 0,456$ . Nous n'avons trouvé aucun autre lien statistique entre les autres types d'émission et la fréquence de l'utilisation de la paraphrase, avec la charge de travail mentale subjective (SMW).

Les participantes à l'étude justifiaient l'utilisation de la paraphrase par le trop grand débit des commentaires. Les réponses de l'enquête sur la fréquence des paraphrases des sous-titres (CC) par type d'émission, nous montre que plus le débit des commentaires est rapide, plus il y a de paraphrases. Une autre raison courante de paraphraser mentionnée par les participantes résidait dans les commentaires simultanés de plusieurs locuteurs. Les sous-titres (CC) en direct sont généralement affichés ligne par ligne au fur et à mesure de la prise de parole (roll-up), et parfois affichés par apparition sur l'écran (pop-on) selon le diffuseur.

Lorsque plus d'une personne parle à la fois, les sous-titres doivent décider quels mots mentionner dans le sous-titre et comment paraphraser afin de pouvoir sous-titrer l'ensemble des commentaires. Il s'agit non seulement d'une tâche auditive difficile, mais aussi d'une nouvelle tâche de repérage; la charge mentale du sous-titre augmente encore lorsqu'il doit discerner celui qui parle. Les sous-titres doivent réduire le nombre de mots du sous-titrage (CC) quand ils ne peuvent pas suivre le discours, quels que soient le règlement ou les politiques de l'entreprise. De plus, il semblait difficile pour les sous-titres d'écouter et de se souvenir en même temps. Nous pourrions conclure que la tâche de paraphraser ajoute du stress et une charge mentale aux sous-titres.

*Personne 3: “S'il s'agit d'une vitesse raisonnable, je pourrais tout retranscrire, mais les gens parlent trop vite et en même temps.*

*“Il est impossible de sous-titrer sans paraphraser”*

Cette recherche a révélé la présence des différents facteurs de charge mentale, qui influencent la charge de travail mentale subjective (SMW) des sous-titres pour les émissions en direct. Les résultats ont fourni des preuves sur la nécessité de reconsidérer les tâches des sous-titres (CC) pour les émissions en direct en raison de l'inadéquation entre les préférences du public, les exigences réglementaires et la charge mentale élevée qu'ils subissent.

## Partie 2. Conception et développement d'un outil de sous-titrage automatique

### *PAVOCAT*

Afin d'atténuer certaines des difficultés liées au sous-titrage (CC) en direct, un outil de sous-titrage automatique, PAVOCAT, a été développé pour aider les sous-titres à produire des sous-titres (CC) en direct ou en post-production avec des algorithmes d'apprentissage automatique. Il s'agit d'un outil logiciel de sous-titrage (CC) en direct qui utilise la technologie de l'intelligence artificielle pour produire des sous-titres (CC), surveiller les erreurs et suggérer des corrections. Le sous-titre est alors chargé de superviser PAVOCAT et de sélectionner les corrections. Si le sous-titre détermine que les corrections suggérées sont insuffisantes, il peut faire les corrections à l'aide d'un clavier sténo, d'un clavier QWERTY ou de l'assistance de reconnaissance vocale.

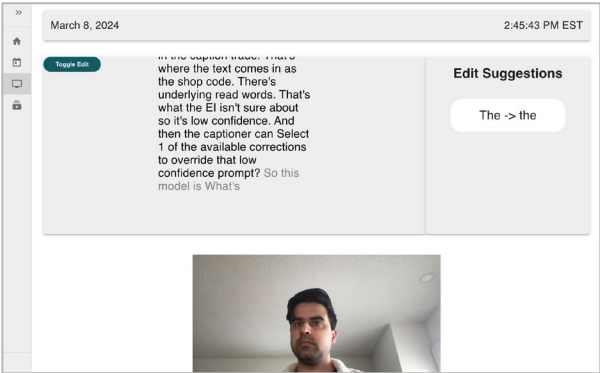
Le système est composé de trois éléments principaux :

- 1) L'interface utilisateur et la fonctionnalité pour les sous-titres et les radiodiffuseurs permet la programmation d'émissions nécessitant un sous-titrage, l'attribution et l'acceptation des tâches de sous-titrage, la communication entre les télédiffuseurs et les sous-titres, le suivi de l'état du système et le contrôle de l'intelligence artificielle avec la correction des erreurs (voir les figures 6a et b). PAVOCAT produit le sous-titrage (CC) initial et une liste d'erreurs potentielles. Il fournit également deux recommandations possibles pour la correction de chaque erreur en temps réel. Les sous-titres peuvent télécharger les dictionnaires pour le vocabulaire spécialisé dans le cadre de différents programmes, vocabulaire qui sera conservé dans la base de données Carl;
- 2) Carl est l'interface de programmation de l'application (API) serveur, qui reçoit les audios du télédiffuseur, les transforme en texte, mentionne les erreurs et fournit des suggestions, pour envoyer ensuite le résultat à PAVOCAT pour l'affichage; et
- 3) PAVOCAT est le serveur de signalisation qui établit et gère la communication entre les sous-titres et les diffuseurs.

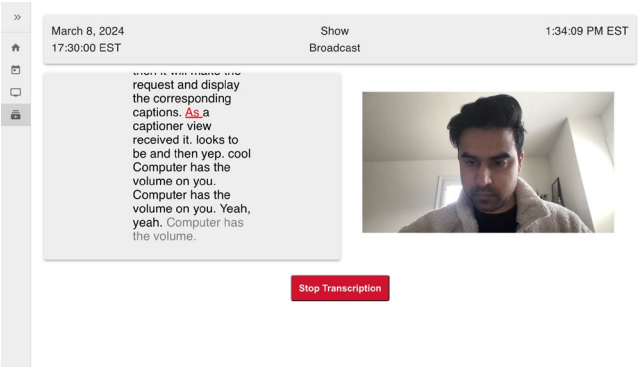
La figure 7 présente le schéma du système PAVOCAT avec les trois éléments mentionnés et leur relation.



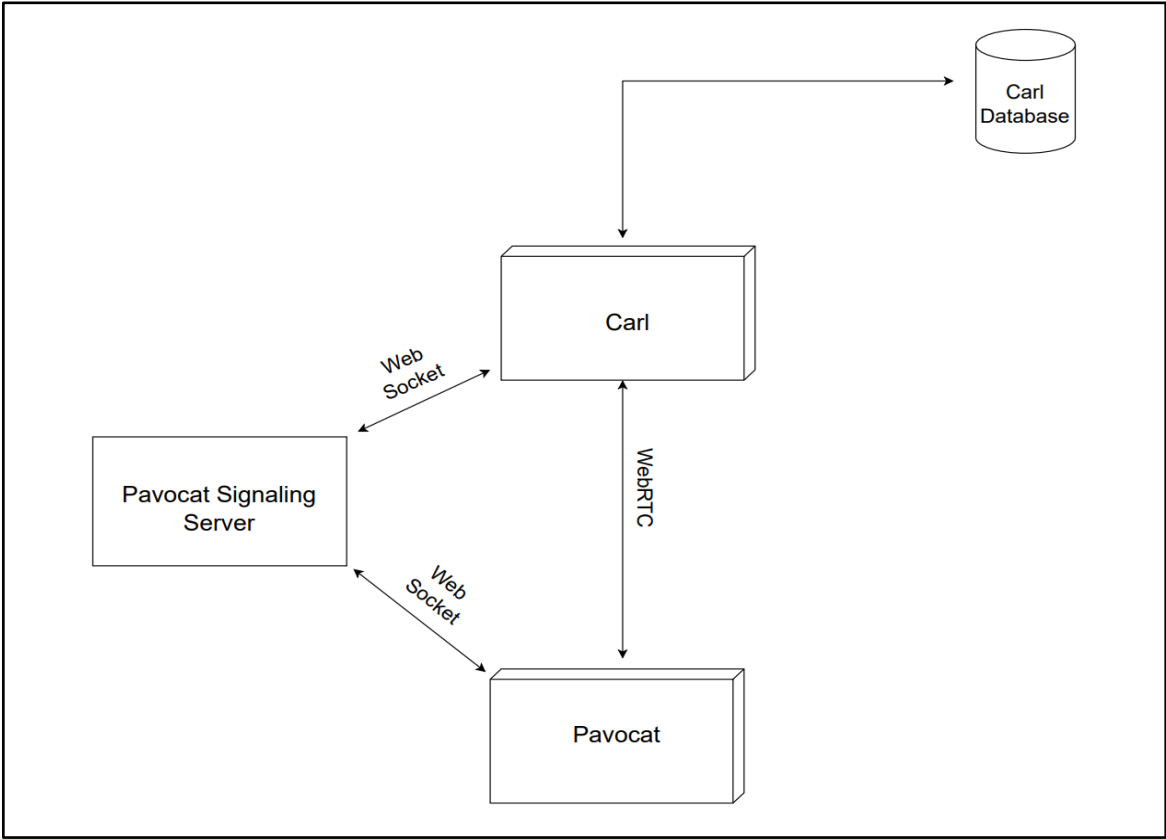
**Figure 6a :** La perspective du sous-titreur avec l'interface PROVOCAT



**Figure 6b :** La perspective du télédiffuseur avec l'interface PROVOCAT



**Figure 7:** les éléments de PROVOCAT et le processus de traitement du signal.



### *Etude auprès des utilisateurs*

À la suite de l'approbation de l'éthique de l'université, une étude auprès des sous-titres novices et expérimentés a été conçue et menée avec le premier prototype entièrement fonctionnel de PAVOCAT.

L'objectif de l'étude était de comprendre les expériences des sous-titres lorsqu'ils utilisent l'intelligence artificielle pour les aider dans leur sous-titrage (CC).

Les questions de recherche de cette étude sont les suivantes :

1. Comment les sous-titres novices et experts travaillent-ils avec un nouveau logiciel d'intelligence artificielle en mode "superviseur" ?
2. Quelle est l'expérience des sous-titres travaillant avec le nouveau logiciel d'intelligence artificielle PAVOCAT ?

Sur une période de 5 jours, les participantes ont visionné trois vidéos de sports à rythme rapide d'une durée de 5 minutes, des Jeux olympiques d'hiver de 2022 au cours desquels PAVOCAT a automatiquement généré le sous-titrage (CC) en direct, a surveillé les erreurs, puis a fourni des options de correction parmi lesquelles le sous-titre pouvait choisir. Aucune saisie n'était autorisée. Notre hypothèse était que le rôle des sous-titres pour les émissions en direct pourrait être modifié pour superviser le sous-titrage et l'intelligence artificielle, plutôt qu'effectuer les opérations de sous-titrage (sténographes et assistance de reconnaissance vocale).

### *Planification de l'expérience*

Une expérience chronologique a été menée auprès de 21 participants. Dix sous-titres novices (dont 1 femme) avec moins de six mois d'expérience en tant que sous-titre, et 11 personnes expérimentées (dont 5 femmes) avec plus de six mois d'expérience. Les participantes ont rempli préalablement un questionnaire avec 19 questions et ont participé à une séance de formation et d'entraînement pour apprendre à utiliser PAVOCAT. Ils/elles ont utilisé le logiciel PAVOCAT pour sous-titrer trois extraits de 5 minutes (tâches réalisées une fois par jour pendant cinq jours) des Jeux olympiques. Après avoir terminé chaque séance de sous-titrage de dix minutes, les participantes ont rempli un questionnaire de 25 questions composé d'une combinaison de quatre logiciels approuvés. Le logiciel TLX de la NASA (Hart et Staveland, 1988) a été utilisé pour évaluer la charge de travail mentale subjective (SMW).

L'Échelle d'Utilisabilité du Système (the System Usability Scale (SUS) se compose de 10 questions de type Likert (Likert-style), mesurant la facilité d'utilisation, la facilité d'apprentissage, la correction des erreurs et la satisfaction de l'utilisateur (Brooke, 1995). L'échelle de confiance dans l'automatisation (the Trust in Automation scale, un questionnaire, Jian et al., 2000) sur une échelle de 7, avec un questionnaire de 12 questions, évalue le taux de confiance des participantes envers le système (SUS). L'échelle de satisfaction comporte une question demandant explicitement si les participantes sont satisfaites du système (à noter que SUS ne mesure pas explicitement la satisfaction).

À l'issue de l'étude, chaque participant e a réalisé un entretien de type conversationnel de 30 minutes dans lequel il/elle a donné son opinion et a partagé son expérience avec le logiciel PAVOCAT. Il/elle a aussi partagé ses sentiments à propos de l'impact croissant de l'intelligence artificielle et de la reconnaissance automatique de la parole (RSA) dans le domaine du sous-titrage.

**Tableau 2:** Etude sur les thèmes et modificateurs du langage avec le logiciel PROVOCAT.

Themes	Modifiers
<b>Performance and Expectations</b> Performance and quality of PAVOCAT's automated captions	<b>Positive</b> Captioning was accurate and/or the system's performance or the participant's familiarity with the system improved over time  <b>Negative</b> Captioning was not accurate and/or the system's performance or the participant's familiarity with the system did not improve over time
<b>Other (Video Content and Technical)</b> Content chosen for captioning, including speed of captioning provided, proper names, languages, accents or other linguistic features affect AI captioning accuracy Technical issues with captioning in general as well as with the system	
<b>User Interface Modifications</b> Suggestions and omissions of the interface such as control over audio/visual elements including, layout, organization and other adjustments	
<b>Physical and Mental Workload</b> Describing the stress and ability to stay aware and in control of the PAVOCAT environment compared to their current stenography/respeaking work Note: If physical workload was positive and mental workload was negative then that comment should be coded in BOTH positive and negative modifiers	<b>Positive</b> PAVOCAT was not physically or mentally stressful to operate  <b>Neutral/No Change</b> Physical and mental demands of PAVOCAT were as stressful to operate  <b>Negative</b> PAVOCAT was more physically or mentally stressful to operate
<b>Comparison with Other Explicitly Named AI/ASR Software</b> Comparisons of PAVOCAT with other AI/ASR software	
<b>Willingness to use AI</b> Participants' trust in AI and any changes in trust during the study	<b>Positive</b> Participant's trust in AI to delegation of captioning tasks was either high or improved  <b>Neutral/No Change</b> Participant's trust in AI or opportunity to delegate captioning tasks is neutral and/or did not change as a result of the study  <b>Negative</b> Participants' trust in AI or delegate captioning tasks is low and/or reduced as a result of the study OR concerns about losing employment to AI
<b>Human AI Interaction</b> How humans and AI will or will not work together	<b>Override Typing</b> Participants would like to option to override the AI's captioning and respeak/type instead  <b>Override Edit Suggestions</b> Participants would like the option to provide Edit Suggestions back to the AI or force the AI to remember or dismiss Edit Suggestions

## Analyses des données

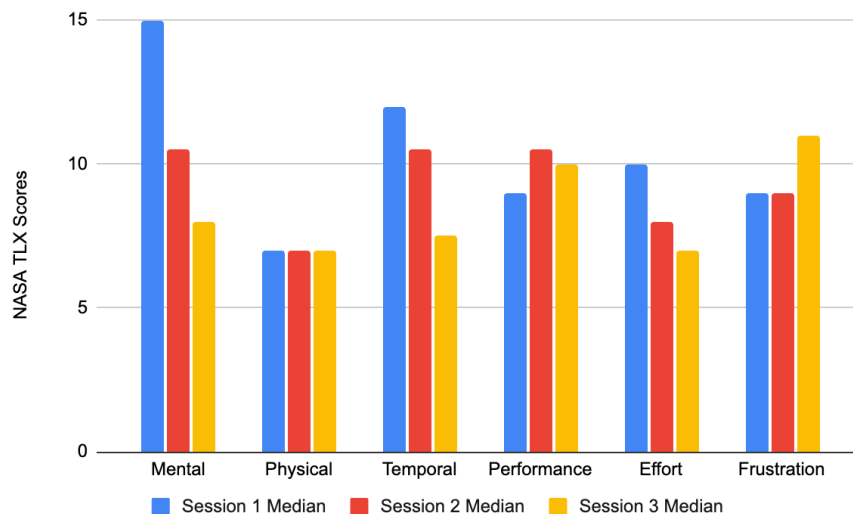
Les entrevues post -étude ont été analysées thématiquement par Noldus Observer (Noldus, 2024). L'intégralité de l'étude a été enregistrée pour pouvoir analyser les observations et les enregistrements vidéos et audios.

## Resultats

Des tests de Friedman non paramétriques ont été effectués pour tous les facteurs (NASA-TLX, SUS, Trust in Automation et Satisfaction), en utilisant une analyse croisée pour les classifications novices/experts. La charge mentale a connu une réduction statistiquement significative entre les trois séances,  $M = 10,33$  ( $ET = 5,46$ ) pour la séance 1,  $M = 9,25$  ( $ET = 4,78$ ) pour la séance 2 et  $M = 8,42$  ( $ET = 5,17$ ) pour la séance 3. Les autres facteurs de NASA-TLX n'ont pas été jugés statistiquement significatifs, mais ont montré des améliorations au cours de la durée de l'étude. Les scores globaux de l'USS ont donné une médiane de  $M = 57,5$  ( $ET = 18,75$ ) pour la session 1, de  $M = 55$  ( $ET = 17,30$ ) pour la session 2 et de  $M = 52,5$  ( $ET = 12,93$ ) pour la session 3. L'échelle de confiance dans l'automatisation a donné une médiane globale de  $M = 32$  ( $ET = 11,75$ ) pour la session 1,  $M = 33$  ( $ET = 11,39$ ) pour la session 2 et  $M = 39$  ( $ET = 10,28$ ) pour la session 3. L'échelle de satisfaction a donné une médiane globale de  $M$

= 5 (ET = 1,34) pour la session 1, M = 4 (ET = 1,56) pour la session 2 et M = 4,5 (ET = 1,54) pour la session 3. L'analyse thématique est toujours en cours et fait l'objet d'une vérification de fiabilité entre les évaluateurs.

**Figure 8:** Les résultats des sous-facteurs de NASA-TLX.



## Objective 4. Formuler des recommandations

**Formuler des recommandations avec des preuves empiriques appropriées et des exemples pour soutenir les normes de sous-titrage (CC) et l'amélioration de la réglementation en tant que note de synthèse. Cela inclura également l'impact sur diverses technologies et processus de distribution.**

Nous basant sur les résultats de l'étude, nous avons élaboré une série préliminaire de recommandations afin de comprendre les points de vue des sous-titres et des téléspectateurs. Les résultats ont révélé des preuves et des solutions potentielles pour améliorer les normes et les règlements de sous-titrage (CC). Ces recommandations ont été consignées dans des rapports techniques et universitaires en vue de leur diffusion. L'étape suivante serait de présenter les résultats de cette recherche pour la prochaine réunion sur les normes du sous-titrage au Conseil de la Radiodiffusion et des Télécommunications Canadiennes (CRTC).

D'après l'étude menée auprès des utilisatrices de sous-titre intégral (PBP), nous recommandons:

- 1) L'utilisatrice des sous-titres pour les émissions sportives à rythme rapide et en direct pourrait apprécier davantage le visionnement avec moins de mots dans les sous-titres (CC). Une étude longitudinale est nécessaire pour confirmer ces résultats.

**Recommandation provisoire :** Sous-titrer les commentaires des commentateurs (CO) et ne pas mentionner les commentaires de l'action en intégral (PBP) pendant le match d'un sport à rythme rapide.

Nous basant sur les résultats de l'étude sur la charge de travail des sous-titres, de l'étude sur les utilisatrices de PAVOCAT et des principes d'ergonomie et d'ingénierie industrielle visant à réduire la charge de travail subjective des sous-titres pour les émissions en direct, à parfaire leur satisfaction professionnelle et leur rendement au travail, les recommandations sont les suivantes:

1. **Encourager la collaboration:** Les sous-titres travaillant en binôme peuvent se compléter et prendre des pauses. Une pratique courante pour les services de traduction des émissions en direct, y compris les interprètes en langue des signes, consiste à travailler en binôme (comme des partenaires) pour se soutenir mutuellement, détecter les erreurs et prendre des pauses (Syahputra et al., 2017). Pour aider à réduire la charge de travail mentale subjective (SMW), il est préférable d'avoir un groupe de deux sous-titres (ou plus) pour les émissions en direct. Une raison venant à l'encontre de cette proposition est d'abord le coût (deux personnes au lieu d'une) et puis le partage de travail. En effet, les sous-titres ont tendance à être assez mal payés et sont souvent réticents à partager leur travail, en particulier en tant que pigistes. De plus, il n'existe pas d'organisation professionnelle qui se consacre à la défense des droits des sous-titres ou à l'élaboration de normes en milieu de travail. Pour changer les mœurs dans le cadre du travail des sous-titres des émissions en direct, une prise de conscience et plus de leadership sont nécessaires.
2. **Prévoir des pauses:** Avoir des pauses régulières est une autre recommandation courante pour réduire la charge de travail mentale (Boucsein et Thum, 1997; Krueger, 1989). Dans la diffusion des émissions en direct, les pauses publicitaires offrent cette possibilité, car les publicités sont sous-titrées ultérieurement, de sorte que le sous-titre en direct n'est pas tenu de les sous-titrer. Une pause publicitaire typique dure entre 2 et 3 minutes et se répète environ toutes les 10 minutes dans une retransmission (Bollapragada et Garbiras, 2004). La durée des pauses doit être déterminée après l'examen de la relation entre la récupération cognitive, les exigences du travail, les ressources et le stress (Hobfoll, 2001; Meijman et Mulder, 2013). Cependant, pour la diffusion des émissions sportives en direct, le nombre des pauses publicitaires et leur durée peuvent varier en fonction du match.
3. **Optimiser les horaires de travail:** Planifier le travail de façon à tenir compte de la charge cognitive et des exigences de traitement des sous-titres des émissions en direct. Éviter de planifier des séances consécutives et essayer de fournir un temps de récupération adéquat entre les séances. Nous savons pertinemment que le défi de l'attribution du nombre d'heures et leur partage entre les employés ont un impact direct sur la charge de travail mentale subjective (SMW) (Trumball, 1966; Martens et al., 1999). À l'instar d'autres types de travail, la charge de travail mentale subjective (SMW) des sous-titres

en direct pourrait être atténué en modulant les horaires et en répartissant de manière équilibrée les heures de travail.

4. Assurer la flexibilité: Encourager la flexibilité dans les modalités de travail afin de permettre aux sous-titres de trouver un équilibre entre leur travail et leurs engagements personnels. Ils/elles pourraient avoir des horaires de travail flexibles, l'option de travailler à distance et des formules proposant le partage de travail (Russell et coll., 2009).
5. Fournir des ressources adéquates: S'assurer que les sous-titres disposent des ressources nécessaires pour effectuer leur travail efficacement. Les diffuseurs pourraient fournir les informations pertinentes au sujet de l'émission avant sa diffusion (par exemple, des informations sur l'équipe d'un match en direct de football, des informations sur les invités des talk-shows, etc.), les logiciels de sous-titrage avancés ou encore du matériel (tels que des écouteurs de haute qualité), des chaises et des bureaux ergonomiques et un éclairage approprié.
6. Formation et soutien: Fournir une formation et un soutien adéquats aux sous-titres pour les émissions en direct ayant pour but de les aider à gérer les exigences cognitives de leur travail. Il pourrait s'agir d'une formation sur les raccourcis clavier et d'autres outils de productivité, ainsi que d'une formation sur la gestion du stress et les techniques de relaxation.
7. Rétroaction et amélioration continue: Encourager les retours des sous-titres pour les émissions en direct et utiliser ces commentaires pour améliorer de façon continue l'environnement et les procédures de travail. Les radiodiffuseurs, les pigistes et les entreprises de sous-titrage (CC) devraient examiner régulièrement les horaires de travail, les exigences du travail et les ressources disponibles pour s'assurer que leurs employés travaillent dans des conditions optimales.
8. La combinaison du système de la synthèse vocale, de l'intelligence artificielle et du sous-titre en tant que superviseur pourrait réduire le décalage, la charge de travail mentale subjective (SMW) et ainsi améliorer la précision pour les émissions de télévision en direct au rythme rapide. Cependant, des travaux de recherche et de développement supplémentaires sont nécessaires pour non seulement fournir des preuves suffisantes à l'appui de cette affirmation, mais aussi mettre en place des outils entièrement fonctionnels et de soutien pour obtenir un sous-titrage (CC) en direct de haute qualité.

## **Objectif 5. Diffusion des résultats du projet**

**Les résultats du projet ont été publiés dans deux revues ou conférences avec comité de lecture (une soumission évaluée par des pairs sera publiée dans une revue à accès libre sans devoir faire une demande à une bibliothèque ; les revues à accès libre ont un coût de publication pour les auteurs qui ont été répertoriés dans le budget du projet), une publication spécialisée et un communiqué de presse.**

Les résultats de l'étude sur la charge de travail des sous-titres ont été publiés dans une revue et un programme de conférence. A noter que les élèves sont indiqués en gras.

1. Nam, S., Karam, M., Christelis, C., **Bhargav, H .**, Fels, D.I. (2023). Assessing Subjective Workload for Live Captioners. *Journal of Applied Ergonomics*. 113, 104094.
2. Karam, M., Christelis, C., Hibbard, E., **Leung, J., Kumarasamy, T .**, Whitfield, M., Fels, D.I. (2022, July). Workload evaluations for closed captioners. *ICCHP 2022*.

L'article sur la méthodologie de compréhension a été soumis à une conférence.

1. **Kumarasamy, T .**, Nam, S., Fels., D.I. (2024) Using a Novel Conversational Method for Measuring Comprehension of D/HoH Viewers Consuming Captioning Content for Entertainment Purposes. *ICCHP 2024*.

Les résultats de l'étude sur le sous-titrage intégral (PBP) ont été publiés dans une publication spécialisée et soumis à une revue:

1. Nam, S. Karam, M., **Kumarasamy, T .**, Whitfield, M., Fels, D.I. (April 2023). Colour commentary versus gameplay captions of live fast-paced sports for Deaf and Hard of Hearing television viewers. Live Subtitling and Accessibility Symposium.
2. Nam, S. **Kumarasamy, T .**, Karam, M., Whitfield, M., Hibbard, E., **Leung, J .**, Fels, DI. (2024 submitted). Eye gaze behaviour and comprehension of color commentary and gameplay captions of live fast-paced sports for Deaf and Hard of Hearing television viewers. *Behavior and Information Technology*.

Nous prévoyons que les étudiants de maîtrise publient leurs travaux de thèse dans des revues académiques appropriées. Les rapports de recherche sont disponibles sur le site Web de Sous-titrage en direct Canada: <https://www.livecaptioningcanada.ca/>

## Les limites du projet

### Limites de l'étude comparative

Cette étude a produit des résultats concis pouvant être pris en compte pour l'avenir du sous-titrage (CC) dans le cadre des retransmissions sportives rapides en direct. Cependant, nous pouvons noter quelques restrictions. La première réside dans le nombre limité des personnes évaluées pouvant donner lieu à des données insuffisantes pour montrer des différences concluantes entre les conditions et/ou les groupes de participants. Il y avait également quelques problèmes avec le dispositif de suivi oculaire et les données qui en résultaient. Par exemple, le dispositif de suivi oculaire perdait facilement son calibrage en raison des mouvements de la tête ou du corps, ce qui aurait nécessité que l'étude soit mise en pause et que le système soit

recalibré. De plus, le système ne saisissait pas systématiquement la même quantité de données des suivis oculaires pour chaque participant.e. Bien que certaines de ces limitations puissent être corrigées par la normalisation, d'autres ne peuvent pas l'être. Nous n'avons pas étudié plus en détail les données de suivi oculaire dans le cadre des entrevues pour pouvoir déduire le processus cognitif par le biais d'auto-évaluations, ou pour expliquer ce que les participantes regardaient ou ressentaient. Le système de suivi oculaire pouvait enregistrer plusieurs types de données (par exemple, sur le diamètre de la pupille) et nous n'en avons utilisé que deux: la fréquence et la durée des visites. Les autres données de suivi oculaire pourraient fournir des indications plus détaillées sur certains des comportements de visionnement (comme le degré d'attention accordé lors de l'observation d'une zone spécifique qui peut être indiquée à l'aide du diamètre de la pupille). D'autres études enregistrant et utilisant ces autres données en association avec des techniques complémentaires, telles que des entrevues approfondies, pourraient être nécessaires pour comprendre plus en profondeur le comportement de visionnement. Les participants, n'ayant visionné qu'une seule mi-temps du match et surpris par les nouveaux sous-titres (transcription des commentaires CO) n'ont peut-être pas pu assimiler correctement ces derniers.

De plus, de nombreux participants étaient non seulement habitué au match, mais ils étaient aussi fans; ils savaient tout des équipes, des joueurs, des arbitres et des règles, ce qui a peut-être, à son tour, influencé leurs connaissances, leur intérêt à regarder le match et leur volonté de ne pas tenir compte des erreurs ou des difficultés dans le sous-titrage. D'autres recherches devraient envisager des études longitudinales avec la diffusion de tout le match et d'autres sports, moins connus, comme le hockey féminin ou le rugby. Enfin, il y avait des limites concernant la traduction de la langue des signes (ASL) en anglais parlé, certaines traductions n'ayant peut-être pas reflété fidèlement le sens du message que les traducteurs essayaient de transmettre aux personnes sourdes et malentendantes (D/HoH). Encourager les chercheuses sourdes à participer à cette recherche pourrait aider à réduire les problèmes potentiels de traduction.

## **Limites de l'étude sur le sous -titreur**

Notre étude a révélé des résultats instructifs pour l'environnement de travail, le débit de travail et la charge de travail des sous -titreurs des émissions en direct, avec cependant quelques limitations. La première réside dans la taille de l'échantillon, qui limite la généralisation des résultats. Les sous-titreur.s n'ont été invités à évaluer leur charge de travail mentale subjective (SMW) que lors de leur dernier travail de sous-titrage (CC) en direct, avec des disparités dans la durée, le genre et le type de tâche. Dans le cadre d'une future étude, les sous-titreur.s seraient évalués dans les mêmes circonstances et avec le même contenu. Toutefois, même en contrôlant le type et la longueur du contenu, nous trouverons des inégalités entre les participant.es du point de vue de l'expérience, de la familiarité et de l'aisance avec le sujet. Les mesures subjectives, telles que la charge de travail mentale subjective (SMW), peuvent être affectées par des différences individuelles et des biais cognitifs et sociaux de désirabilité. L'une des conclusions de notre étude porte sur l'altérité entre différents domaines en rapport avec la fréquence



d'utilisation. Compte tenu de ces preuves, il serait utile d'étudier les niveaux de stress mesurés dans chaque domaine. De plus, alors que le premier NASA-TLX vise à mesurer une seule tâche, de nombreuses études antérieures ont utilisé le NASA-TLX pour des activités multitâches. Le résultat des entrevues a prouvé que le sous-titrage (CC) des émissions en direct demande des compétences multitâches et dans ce contexte, d'autres études sur les sous-tâches du sous-titrage (CC) des émissions en direct pourraient être menées ultérieurement. Enfin, les études ne peuvent fournir qu'un instantané de l'expérience des participants à un moment précis. Dans notre étude, nous avons demandé aux participantes de se souvenir de la dernière émission qu'ils/elles avaient sous-titré (CC). Par conséquent, les résultats de l'étude n'incluraient pas toute la variété de la charge de travail mentale subjective (SMW), auxquelles les sous-titres des émissions en direct sont confrontés dans leur travail quotidien. Pour mieux comprendre les points de vue et les facteurs de charge de travail des sous-titres, il est nécessaire d'effectuer des recherches, des observations et des discussions approfondies avec les sous-titres.

## **Limites de l'étude sur les moyens d'usage du logiciel de sous-titrage utilisant l'intelligence artificielle**

Dans cette étude, certaines limites sont similaires à celles des autres études, à savoir la faible taille de l'échantillon, la variété limitée du contenu et l'expérience du sous-titre. De plus, nous avons été confrontés à d'autres limites, spécifiques à cette étude, y compris son déroulement en mode virtuelle, entraînant des problèmes techniques (tel que la connectivité Internet) qui ont nui à la vitesse et à la synchronisation du système de synthèse vocale PAVOCAT et ont entraîné des incohérences dans la performance du logiciel. Les participants auraient ressenti de la frustration face à ces problèmes techniques, se concrétisant par une moins bonne accessibilité et des cotes de confiance plus faibles. Nous pensons que des études en face à face pourraient résoudre certains de ces problèmes de connectivité, en réduisant toutefois le bassin de participant.e.s.

Le système PAVOCAT a présenté également une certaine instabilité et a cessé de fonctionner après environ 10 minutes, il a donc dû être redémarré pour poursuivre l'étude. L'équipe de développement de PAVOCAT est en train d'apporter des révisions au logiciel pour répondre aux problèmes soulevés par cette étude. Nous espérons ainsi que les pannes du système ne se reproduisent plus dans les études futures.

## **Conclusion**

Une grande partie du travail effectué dans le cadre de ce projet de recherche était nouvelle et originale. Les principales conclusions de ce projet se basent sur le compromis entre la vitesse des sous-titres, la précision et le délai de synchronisation qui doivent être pris en compte afin d'améliorer la qualité du sous-titrage (CC) pour les téléspectateurs sourds et malentendants (D/HoH). L'une des principales conclusions est que les sous-titres des émissions en direct connaissent une charge de travail mentale subjective (SMW) élevée lors

d'événements sportifs au rythme rapide. Les facteurs humains et les pratiques psychologiques peuvent fournir des méthodes pour atténuer cette charge de travail, notamment des pauses accrues et un travail en binôme/partagé. Toutefois, d'autres options, telles que la reconnaissance vocale automatique (ASR) et l'intelligence artificielle, peuvent fournir des alternatives. PAVOCAT a été conçu pour exploiter la reconnaissance vocale automatique (ASR) et l'intelligence artificielle pour le sous-titrage (CC) en direct avec des sous-titres en tant que superviseurs pour en surveiller les performances et choisir parmi les suggestions d'erreur proposées par ce logiciel. Les sous-titres se méfiaient de ce système et voulaient avoir la possibilité d'outrepasser les suggestions ou le sous-titrage (CC) proposés par logiciel lorsqu'ils/elles estimaient que les suggestions étaient inadéquates. Cependant, ils/elles estimaient également que l'avènement d'un logiciel tel que PAVOCAT était presque inévitable et se disaient favorables à participer à son développement. Une autre option étudiée dans le cadre de ce projet consistait à réduire la quantité de sous-titrage (CC) pour les sports rapides, tels que le hockey et le basketball, car le délai de synchronisation, les erreurs et le format des sous-titres (CC) les rendaient presque inutiles. Nous basant sur le suivi oculaire et une nouvelle méthode d'évaluation de la compréhension, nous avons mené une étude dans laquelle les téléspectateurs pouvaient visionner des sous-titres (CC) de type transcription des commentaires (et non du type intégral (PBP)). Cette première étude a révélé que les participantes sourdes et malentendantes (D/HoH) préféraient les sous-titres de la transcription des commentaires (CO), mais que d'autres études longitudinales étaient nécessaires pour réduire le potentiel effet de nouveauté. L'entrevue décontractée visant à mesurer la compréhension des participant.e.s sourdes et malentendantes (D/HoH) d'une manière plus naturelle a permis d'améliorer potentiellement les techniques normalisées d'évaluation de la compréhension, même si une comparaison complète reste à effectuer. À la suite du travail innovant et inédit réalisé dans le cadre de ce projet, de nouvelles questions, de nouveaux problèmes et de nouvelles lacunes sont également apparus, nécessitant des recherches plus approfondies.

## Remerciements

Nous tenons à remercier Normes d'accessibilité Canada pour son généreux financement. Nous remercions également tous les participants qui se sont portés volontaires pour toutes les études. Nous tenons également à remercier les membres du conseil consultatif qui ont fourni des conseils, de l'expérience et de la sagesse pour le projet.

## Références

- Bhattacharjee, A. (2001). Understanding information systems continuance: An expectation -confirmation model. *MIS Quarterly*, 351–370.
- Bollapragada, S., & Garbiras, M. (2004). Scheduling Commercials on Broadcast Television. *Operations Research*, 52(3), 337–345. <https://doi.org/10.1287/opre.1030.0083>
- Boucsein, W., & Thum, M. (1997). Design of work/rest schedules for computer work based on psychophysiological recovery measures. *International Journal of Industrial Ergonomics*, 20(1), 51–57.

- Bowers, J. W., & Phillips, W. A. (1967). A note on the generality of source-credibility scales. *Speech Monographs*, 34(2), 185–186. <https://doi.org/10.1080/03637756709375542>
- Brooke, J. (1996). SUS: a “quick and dirty” usability scale. In *Usability evaluation in industry* (pp. 189–194). Taylor Francis.
- Grier, R. A. (2015). How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), 1727–1731. <https://doi.org/10.1177/1541931215591373>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkat (Eds.), *Human Mental Workload* (1–Book, Section, pp. 139–183).
- Hobfoll, S. E. (2001). The Influence of Culture, Community, and the Nested-Self in the Stress Process: Advancing Conservation of Resources Theory. *Applied Psychology*, 50(3), 337–421. <https://doi.org/10.1111/1464-0597.00062>
- Jensema, C. J. (1998). Viewer reaction to different television captioning speeds. *American Annals of the Deaf*, 143(4), Article 4.
- Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4(1), 53–71. [https://doi.org/10.1207/S15327566IJCE0401\\_04](https://doi.org/10.1207/S15327566IJCE0401_04)
- Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work & Stress*, 3(2), 129–141. <https://doi.org/10.1080/02678378908256939>
- Martens, M. F. J., Nijhuis, F. J. N., Van Boxtel, M. P. J., & Knottnerus, J. A. (1999). Flexible work schedules and mental and physical health. A study of a working population with non-traditional working hours. *Journal of Organizational Behavior*, 20(1), 35–46. [https://doi.org/10.1002/\(SICI\)1099-1379\(199901\)20:1<35::AID-JOB879>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1099-1379(199901)20:1<35::AID-JOB879>3.0.CO;2-Z)
- Meijman, T. F., & Mulder, G. (2013). Psychological aspects of workload. In *A handbook of work and organizational psychology* (pp. 5–33). Psychology press. <https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9780203765425-2&type=chapterpdf>
- Noldus. (2024). *Behavioral coding—Event logging software | The Observer XT. Behavioral Coding—Event Logging Software | The Observer XT*. <https://www.noldus.com/observer-xt>
- Ofcom. (2005). *Guidance on Standards for Subtitling*. [http://www.ofcom.org.uk/tv/ifi/guidance/tv\\_access\\_serv/archive/subtitling\\_stnds/](http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/archive/subtitling_stnds/)
- Russell, H., O’Connell, P. J., & McGinnity, F. (2009). The Impact of Flexible Working Arrangements on Work–life Conflict and Work Pressure in Ireland. *Gender, Work & Organization*, 16(1), 73–97. <https://doi.org/10.1111/j.1468-0432.2008.00431.x>
- Simpson, E. K., & Kahler, R. C. (1981). A Scale for source credibility; Validated in the selling context. *Journal of Personal Selling & Sales Management*, 1(1), 17–25.
- Syahputra, B. P., Saragih, A., Lubis, S., & Muchtar, M. (2017). INTERPRETING TECHNIQUES BY A TOUR GUIDE AT THE ANCIENT TOMBS OF RAJA SIDABUTAR. *Researchers World*, 8(1), 151.
- Szarkowska, A. (2013). Auteur Description: From the Director’s Creative Vision to Audio Description. *Journal of Visual Impairment & Blindness*, 107(5), Article 5.

Trumbull, R. (1966). Diurnal Cycles and Work-Rest Scheduling in Unusual Environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 8(5), 385–398.  
<https://doi.org/10.1177/001872086600800502>.