



ÉTUDE DE L'IMPACT DE LA QUALITÉ AUDIOS ET VISUELLES (QAV) SUR LE SPECTATEUR ADULTE

[Étapes de traitement de l'article]

Date de soumission : 20-05-2025 / Date de retour d'instruction : 05-06-2025 / Date de publication : 15-07-2025

Bonaventure LOKOSSOU

Université d'Abomey-Calavi

École Nationale des Sciences Techniques de l'Information et de la Communication
(ENSTIC/UAC) Bénin

bonaventurelokossou2018@gmail.com

&

Dorothée Elavagnon DOGNON

Institut National des Métiers d'Arts, d'Archéologie et de la culture de l'Université
d'Abomey-Calavi (INMAAC /UAC)

delavagnon@gmail.com

Résumé : Dans un domaine très concurrentiel, l'un des principaux enjeux pour les acteurs de l'offre de services audiovisuels (AV) est de garantir au spectateur une qualité d'expérience optimale. Aujourd'hui, la (QOE) est souvent restreinte à la perception de la qualité audiovisuelle restituée par le système. Elle est principalement mesurée à travers la collecte de données par des participants sur des échelles de qualité, après visualisation et écoute de séquences audiovisuelles (AV) traitées par le ou les technologies à évaluer. Ces tests perceptifs suivent des procédures recommandées par l'Union Internationale des Télécommunications. Cependant, la qualité restituée peut affecter la QOE (fatigue, effort...) sans être reflétée par les notes de qualité. Une méthode considérant l'évaluation non plus de la qualité audiovisuelle (QAV) perçue seule mais de la *qualité d'expérience*, plus largement considérée, pourrait permettre de mieux rendre compte de l'influence de la qualité du son et de l'image sur le consommateur adulte. Alors quel est donc l'impact de la qualité audiovisuelle sur un spectateur adulte ? L'objectif de cette recherche est d'identifier l'impact de la dégradation des AV sur un adulte. L'hypothèse qui soutient cet objectif est que la fatigue et la malcompréhension des contenus audiovisuels par les adultes sont les résultats de la dégradation des AV. Une expérience pratique a été faite sur 45 personnes (25 femmes et 20 hommes). Les informations sont recueillies et traitées en Excel. Au terme de cette recherche on note que 20% des femmes n'ont pu cerner les contenus audiovisuels et 25% des hommes non plus. Le niveau de compréhension est considérablement réduit dans une proportion de 100% chez tous les participants. Le présent travail est centré sur la recherche d'une méthode alternative aux méthodes actuelles de l'examen de qualité pour applications multimédias dans un contexte de visualisation et d'éco

Mots clés : Audio, Audiovisuel, Dégradation vidéo, Impact audiovisuel, QAV.

STUDY OF THE IMPACT OF AUDIO AND VISUAL QUALITY (AVQ) ON THE ADULT VIEWER

Abstract : In a strongly competitive context, one of the main stakes for the actors of the audiovisual services offer is to guarantee to the spectator an optimal Quality of Experience (QOE). Nowadays, QOE is often limited to the perception of the audiovisual quality (AVQ) received by the system. It is mainly measured through the collection of rates given by testers onto quality scales, after visualization and listening of the AV sequences treated by one or several technologies to be evaluated. These subjective tests are following procedures recommended by the International Telecommunication Union. However, the restored quality can affect some factors of QOE (fatigue, effort, etc.) which are not reflected by the quality scores. A method considering the evaluation either of the AV quality only received but of the

quality of experience, widely considered, could allow to report better the influence of the sound and image quality on the adult spectator. So what is the impact of audiovisual quality on an adult spectator ? the objective of this research is to identify the impact of the deterioration of AV on adult. The hypothesis which supports this objective is that fatigue and the malcompressing of audiovisual content by adults are the results of the degradation of AV. A practical experience was made on 45 people. The information is collected and processed in excel . At the end of this research, we note 20% of women could not identify audiovisual content an 25% of the either. The level of understanding is considerably reduced in a proportion of 100% in all participants.

Keywords: audio, audiovisual, video degradation, audiovisual impact, QAV.

Introduction

Dans un contexte concurrentiel, la qualité d'expérience du consommateur (QOE : Quality Of Experience) est une des préoccupations majeures des acteurs du domaine de ces services audiovisuels (télévisuels, visioconférences, productions audiovisuelles). Actuellement, l'évaluation de la QOE se réalise généralement à travers l'évaluation de la qualité, telle que perçue par les consommateurs, des signaux audio et/ou vidéo restitués. Les méthodes d'évaluation utilisées sont recommandées par l'Union Internationale des Télécommunications. Ces approches reposent sur des mesures subjectives dont l'interprétation et la validité sont limitées par un certain nombre d'interface. Elles ne permettent pas non plus de rendre compte fidèlement de l'influence de la qualité audiovisuelle sur la *qualité d'expérience* du consommateur. Par exemple, ces méthodes n'apportent pas d'informations sur le coût pour l'utilisateur, du point de vue de la fatigue ou de l'effort mental, induit par des dégradations du signal et pouvant à terme conduire à un rejet du système ou de la technologie de restitution. Le *coût* peut être mesuré à partir d'indices de l'activité physiologique et oculaire de l'individu. Ce type de mesures présente l'avantage de ne pas être soumis aux biais des mesures subjectives, capables de diminuer la fiabilité des observations (problème de représentativité, interprétation des items, références implicites, etc.). Wilson G. M. et Sasse (1000a, 2000b) cité par docteur Julie LASSALLE ont montré que des fluctuations importantes de qualité audio ou vidéo peuvent ne pas être consciemment perçues par les participants et pour autant être reflétées par l'activité physiologique. L'objectif de notre recherche présenté dans ce document est de contribuer à une méthode transversale et quantitative pour l'évaluation de la qualité audiovisuelle pour des applications multimédias. La démarche a consisté à proposer, au sein d'une même méthode, l'étude conjointe à la fois de mesures subjectives, avec un questionnaire enrichi (c.-à-d. évaluation de critères autres que les seules notes de qualité) et de mesures de l'activité physiologique et oculaire du spectateur.



1. Méthodologie

1.1. Type et nature de l'étude

Il s'agit d'une étude transversale, de type descriptive à visée analytique, à la fois quantitative et qualitative, qui s'est déroulée sur une période de 02 mois.

1.2. Population et méthode d'étude

1.2.1. Population d'étude

Elle porte sur les adultes spectateurs et fans des contenus audios et visuels, âgés de 18 à 45 ans qui ont accueilli une expérimentation de nature exploratoire, qui propose un corpus de trois contenus de test différents, d'une durée de cinq minutes chacun et présentant différents types de dégradations audio, vidéos sans son et audiovisuelles (dégradation simultanée sur l'audio et la vidéo). L'objectif est d'étudier la manière dont les mesures subjectives classiques d'évaluation de la qualité peuvent être complétées par les mesures physiologiques et oculaires des adultes de 18 à 45 ans.

1.2.2. Échantillonnage

L'échantillonnage exploité dans le cadre de notre étude est aléatoire stratifié. La base de sondage est constituée à partir de quarante-cinq participants adultes (25 femmes, 20 hommes), entre 18 et 45 ans provenant de différents départements, avec une audition et une vision normales ou corrigées à la normale (déclaratif), ont participé à l'expérience. Ils étaient rémunérés pour leur participation.

1.3. Matériel

Configuration générale

Vid. 6.1. Schéma de la configuration de la salle de test (520×370×285 cm) de l'expérimentation. La place du participant est représentée par un point noir, l'écran est figuré par un rectangle.

Le test se déroulait dans une pièce insonorisée où les conditions de visualisation et d'écoute recommandées par la norme UIT-T P.912 étaient respectées à l'exception de la chroma citée (pour plus de détails voir vid exp A). Le niveau de luminosité de la salle était identique pour tous les participants (≤ 18 lux -lx-). Les paramètres de brillance et de contraste de l'écran remplissaient les conditions de la norme UIT-R BT.814-2 (UIT, 2007). Un écran LCD Hyundai S465D de 46" (117 cm) full HD (1080p, 16:9) a été utilisé pour afficher les contenus de test en format full HD (1920×1080p). En accord avec la norme UIT-T P.912, la distance de visualisation était fixée à six fois la hauteur de l'écran (250 cm). Les haut-parleurs (HP) de marque Genelec (modèle 190A) étaient disposés selon une configuration triangulaire, c'est-à-dire qu'ils étaient à la fois équidistants du centre de l'écran (1 m) et équidistants de la tête du participant. La configuration de la salle de test est présentée par la vid 6.1 de la vidéo témoin, celle-ci respectait les recommandations de la norme UIT-R BS.1286 (UIT, 1998). Le niveau d'écoute respectait le seuil de 80 dBA recommandé par la norme UIT-T P.912 (mesuré à l'aide d'un bruit blanc et d'un sonomètre placé au point d'écoute).

1.4. Configuration technique

Vid. 6.2. Configuration technique de l'expérimentation.

La configuration technique de l'expérience A est présentée par la vid 6.2 ci-dessus. Le stockage et la restitution des contenus AV de test a été effectuée au moyen d'un magnétoscope numérique (Digital Vidéo System -DVS- Pronto2K). Les signaux audio et vidéo étaient ensuite acheminés séparément vers le matériel de restitution vidéo (écran) et audio (HP externes) *via* un amplificateur (SPL 2380). Avant cela, chaque signal était converti au bon format d'entrée, c'est-à-dire d'un signal numérique (signal de sortie du DVS) à un signal analogique (signal d'entrée HP) pour l'audio (convertisseur Apogée Rosetta 800) et d'un signal SDI (signal de sortie du DVS) à DVI (signal de sortie du convertisseur, SDM- 875p HD-SDI to DVI) puis HDMI (signal d'entrée écran) pour la vidéo. L'utilisation du DVS présente l'avantage de pouvoir restituer les signaux audio et vidéo en format full HD (1920×1080) non compressé, c'est-à-dire sans aucune perte de qualité (absence de dégradations liées aux phénomènes de codage/décodage). Cependant, l'utilisation de ce matériel, dans la présente configuration technique, ne permettait pas l'utilisation de différentes listes de lecture des séquences AV de test (fichier appelant les séquences AV selon un ordre déterminé - playlist -). En conséquence, l'ordre de présentation était fixé et identique pour l'ensemble des participants. Pour obtenir un ordre de présentation randomisé, il aurait été nécessaire de disposer d'autant de trames vidéo (contenant les neuf extraits AV à présenter) que de participants. A titre indicatif, une trame correspondait à environ à 550 Gigabits (Gb), il était impossible, dans le cadre de la configuration mise en place, d'envisager cette solution. Un autre inconvénient résidait dans l'impossibilité de synchroniser le magnétoscope avec les logiciels de mesures physiologiques et oculaires. Par conséquent, le lancement de la vidéo et des différents logiciels de mesures était effectué manuellement. Le magnétoscope ainsi que les écrans de contrôle de la salle étaient installés dans une régie séparée de la salle de test.

2. Résultats

Dans cette expérimentation, l'influence des fluctuations de qualité sur l'expérience subjective, physiologique et le comportement oculaire était étudiée. Il était attendu que la présence de dégradations diminue les scores moyens obtenus à l'évaluation des niveaux de qualité audio (MOSA : Mean Opinion Score for Audio), vidéo (MOSV) et audiovisuelle (MOSAV). Deux hypothèses ont été testées pour les mesures subjectives :

- **H0s** : MOSAV, MOSV et MOSA obtenues pour la condition sans dégradations devraient être supérieures à celles obtenues pour les conditions dégradées,
- **H1s** : MOSAV devrait être inférieure à MOSA et MOSV obtenues pour la condition de désynchronisation image/son, la désynchronisation ne dégradant pas les signaux audio et vidéo. Par ailleurs, il était attendu que les mesures psychophysiologiques réagissent à la présence de l'ensemble ou d'une partie des dégradations introduites. Le traitement du signal dégradé devrait amener le participant à mobiliser plus de



ressources (dépenses énergétiques/effort). Une activation du système nerveux autonome et particulièrement de sa branche sympathique (SNS) ainsi que des modifications du comportement oculaire étaient alors attendues. La présence d'un effort répétitif ou prolongé en réponse à la présence de dégradations pourrait également conduire à un état de fatigue. Trois hypothèses ont été testées pour les mesures psychophysiologiques :

- **H0p** : la présence de dégradations audio et/ou vidéo (désynchronisation, réduction du débit) pourrait être à l'origine d'un effort mental supplémentaire (pour décoder et interpréter le message dégradé) puis d'un état de fatigue, visibles à travers des modifications des patterns physiologiques et/ou oculaires,
- **H1p** : un effort mental supplémentaire lié à la présence de dégradations audio et/ou vidéo conduirait à une activation majoritaire du système nerveux sympathique traduite par une augmentation des indices AED, FC, DP et une diminution des indices VSP, TCP, Ebfreq et EBdur,
- **H2p** : un état de fatigue consécutif à l'effort mental lié à la présence de dégradations audio et/ou vidéo serait traduit par une augmentation du PERCLOS, de Ebfreq et de EBdur.

2.1. Préparation et réduction des données

Pour chacun des participants sélectionnés, environ une heure trente de données ont été recueillies et ce, pour chaque indice. Pour chaque signal mesuré, la première et la dernière minute (pour lesquelles aucune dégradation n'était présente) a été retirée afin de ne pas tenir compte des « effets de bord » éventuels (surprise, anticipation du début et de la fin du contenu ceux-ci étant d'une durée fixe, réponse d'orientation au début du contenu).

Les mesures physiologiques et oculaires de quatre spectateurs n'ont pu être utilisées en raison de problèmes techniques des outils de mesure. En plus de cette première série d'exclusion, d'autres participants ont dû être rejetés en raison du manque de fiabilité de leurs mesures (données manquantes, valeurs aberrantes) pour chaque jeu de données, physiologiques ou oculaires, considérés et traités séparément. Ainsi, un participant supplémentaire a été retiré du jeu de données physiologiques, portant à quarante-cinq le nombre total de participants dont les mesures ont été utilisées. Pour le jeu de données oculaires, six individus ont dû être exclus, portant à vingt le nombre total de participant retenus. Aucun participant n'a été rejeté pour l'analyse des mesures subjectives. Le nombre de participants dont les mesures ont été retenues pour chaque jeu de données est récapitulé par la vidéo 6.2.

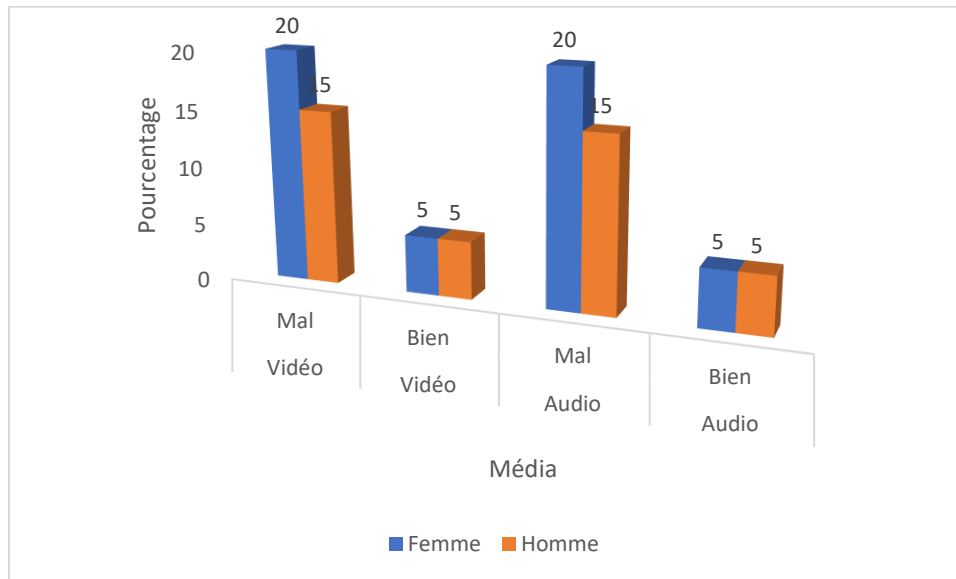


Figure 1 : Histogramme montrant l'effet des dégradés visuels sur l'interprétation des vidéos selon le genre.

Les résultats montrent que l'interprétation des vidéos, comparée à celle des audios, est globalement moins bonne, en particulier chez les femmes. Cette mauvaise interprétation peut s'expliquer par la présence de formes de dégradés dans les images des cinq séquences de vidéos utilisées, dont deux présentaient des dégradés particulièrement marqués, susceptibles d'altérer la perception visuelle et donc la compréhension du message.

vidéo 6.2. Nombre de participants dont les mesures ont été retenues pour l'analyse statistique à partir des données subjectives, physiologiques ou oculaires.

2.2. Type de mesures participants retenus

SUBJECTIVES 20

PHYSIOLOGIQUES 25

OCULAIRES 45

Les scores subjectifs ont été réduits par la moyenne obtenue pour chaque extrait (Opéra, Documentaire et Sport) et chaque condition (C0, C1 et C2) afin d'obtenir une note MOS pour chaque qualité évaluée, AV (MOSAV), audio (MOSA) et vidéo (MOSV). Deux autres jeux de données ont été constitués pour les mesures physiologiques et oculaires. Le premier correspondait à la moyenne temporelle de chaque indice, pour chaque participant et pour chacun des neuf extraits soit sur une fenêtre temporelle de cinq minutes (JD-5min). Ce premier jeu de données devait permettre d'observer les éventuelles différences entre les neuf extraits (séquences) visualisés (effet du contenu, effet global de la condition de qualité). Le second jeu de données a été réalisé de manière identique au premier à la différence que la moyenne temporelle était calculée pour chaque période, dégradée et non dégradée soit une fenêtre temporelle de trente secondes (JD-30s). L'objectif était de pouvoir étudier les éventuelles modifications de l'activité physiologique et/ou oculaire, au sein de chaque



extrait (séquence) et selon la présence, le type et le niveau de dégradation (« V- » par exemple). Comme indiqué dans la section 3.5, les réponses physiologiques VSP, FC, AED et TCP, sont sujettes à une forte variabilité inter-individuelle. Afin de minimiser cet impact, les moyennes calculées pour les différents indicateurs ont été normalisées par la baseline (*bsl*) pour chaque participant et chaque jeu de données à partir de la formule suivante (Calcanis, Callaghan, Gardner et Walker, 2008 ; Lin *et al.*, 2005) :

Moyenne temporelle *bsl*

Moyenne temporelle signal min ou $s = \text{abs}$

Moyenne temporelle $\text{bsl signal} = n$

Les données normalisées seront notées de la manière suivante : VSP_n , FC_n , AED_n et TCP_n . La démarche générale pour l'analyse des données physiologiques et oculaires est schématisée par la vidéo 6.8. L'ensemble des vid présentées dans les vidéos témoins afficheront un intervalle de confiance à 95%.

vid. 6.8. Approche générale pour l'analyse des données physiologiques et oculaires.

2.3. Mesures subjectives

Selon l'hypothèse H_0 s, les notes de qualité obtenues pour la condition sans dégradations devaient être supérieures à celles obtenues pour les conditions dégradées. Une analyse multivariée (MANOVA - Multivariate ANalysis Of Variance) réalisée sur les scores individuels en considérant les variables indépendantes à trois modalités « Contenu » (Opéra, Documentaire et Sport) et « Qualité » (C0, C1 et C2) a confirmé cette hypothèse. Les résultats ont révélé un effet significatif à la fois des facteurs « Contenu » ($F(6, 27) = 73,36$, ($p < 0,0001$)). La normalité des données a été vérifiée pour l'expérimentation A ainsi que pour l'ensemble des expérimentations présentées dans la suite du document. Les résultats obtenus (tests Kolmogorov-Smirnov et Lilliefors) montrent qu'une grande partie des distributions étudiées suivaient la loi normale, d'autres non, notamment en raison d'une influence forte du contenu sur la répartition des données.

Une approche par analyse de la variance (ANOVA) a donc généralement été privilégiée en raison de sa robustesse aux écarts à la normalité et sa capacité à intégrer, en autres, des facteurs aléatoires. Et « Qualité » ($F(6, 27) = 7,43$ ($p < 0,0001$)) ainsi qu'une interaction significative entre ces deux variables ($F(12, 21) = 17,17$, ($p < 0,0001$)). La Figure 6.9 ci-dessous présente les notes MOSAV, MOSV et MOSA obtenues pour chaque contenu et chaque condition de qualité.

MOSAV

MOSV

MOSA

C0:OPERA DOC. SPORT

C1:OPERA DOC. SPORT

C2:OPERA DOC. SPORT

vid. 6.9. MOSAV, MOSV et MOSA obtenues pour les conditions C0, C1 et C2 pour chacun des contenus visualisés : Opéra, Documentaire (Doc.) et Sport.

La vidéo 6.9 indique différents effets. Premièrement, il semble que les notes de QAV correspondent globalement à une moyenne des notes de QA et QV. Ainsi, les participants tiennent compte à la fois de la qualité audio et vidéo perçue pour élaborer leurs notes de jugements de qualité AV. Aucune prédominance de la qualité vidéo ou audio ne semble être mise en avant. Une seconde hypothèse (H1s) supposait que les notes de qualité AV devaient être inférieures aux notes obtenues pour QA et QV en présence de désynchronisation (C1), la désynchronisation ne dégradant pas les signaux audio et vidéo. Cependant, l'observation des effets de C1 sur la qualité perçue permet de constater une diminution non seulement de QAV mais aussi, et principalement, de QA par rapport à la condition sans altérations (C0). La qualité audio (présentée sans aucune dégradation) a en effet reçu les scores les plus bas comme si cette modalité était identifiée comme étant porteuse de l'altération de désynchronisation : le son est en retard ou en avance par rapport à l'image (et non l'image en avance ou en retard par rapport au son).

Par ailleurs, C2 (dégradations A et/ou V) a été perçue comme dégradant plus fortement la qualité que C1, cela est d'autant plus marqué pour les contenus *Documentaire* et *Sport*. L'analyse plus détaillée des effets de C2 sur la qualité perçue a notamment indiqué des scores de QV significativement plus élevés ($p < 0,0001$ d'après un test *post-hoc* HSD de Tukey) pour le contenu *Opéra* que pour les deux autres. Ce constat indique que l'impact de dégradations, pourtant objectivement équivalentes, sur la perception de qualité est dépendant du type de contenu.

3. Discussion

Comme indiqué dans les résultats, les données oculaires étaient capturées par deux optiques (eye tracker faceLAB4™) et acheminées *via firewire* à un ordinateur dédié¹². Les signaux physiologiques étaient recueillis à l'aide de l'encodeur *Procomp Infiniti* (boîtier placé à proximité du participant) et acheminés *via* fibre optique de l'encodeur au logiciel d'enregistrement *Biograph Infiniti* (Thought Technologies™) installé sur un second ordinateur¹³. La Figure 6.3 présente le matériel de recueil des données physiologiques et oculaires. Dans cette expérimentation, le capteur de conductivité électrodermale fourni avec l'outil de mesure *Procomp infiniti* (SC-Flex/Pro, électrodes Ag-AgCl) était utilisé. Celles-ci

¹² Dell Optiplex, Intel Pentium 4 GX620, 3,20 Ghz, ram : 1 Gb, OS: Windows XP (32 bits)

¹³ Dell Precision T5400, Intel Xeon E5420, 2,49 GHz, ram : 3,25 Gb, OS : Windows XP (32 bits)

Expérimentation: exploratoire présentent la particularité d'être insérées (cousues) à l'intérieur de sangles velcro permettant leur maintien sur le site d'accueil. L'utilisation de gel pour ces électrodes est déconseillée par le fournisseur. Pour des raisons techniques, les ordinateurs de recueil de mesures étaient placés dans la salle de test. Toutefois, ceux-ci étaient situés à une distance supérieure à 2 m de la place du participant et dissimulés par un rideau, la salle pouvant être scindée en deux espaces



distincts. Le bruit inhérent au fonctionnement des ordinateurs a été contrôlé et ne dépassait pas le seuil de niveau de bruit de fond recommandé par la norme UIT-T P.912 (≤ 30 dBA). Pour éviter toute interférence avec les mesures physiologiques, les téléphones portables étaient interdits dans la salle de test.

3.1. Stimuli

Dans cette expérience, trois contenus audiovisuels 2D étaient présentés en format full HD :

- **Opéra** : extrait d'une adaptation de *Don Giovanni*,
 - **Documentaire** : extrait d'un documentaire sur le boxeur français Jean-Marc Mormeck,
 - **Sport** : extrait de la finale de Roland Garros 2010 cité par docteur Julie LASSALLE.
- Un aperçu des contenus AV de test est présenté par la Vidéo 6.4 ci-dessous.

Vid. 6.4. Aperçu des contenus de test de gauche à droite : Documentaire, Opéra et Sport.

Chaque extrait durait sept minutes. Ces trois types de contenus ont été choisis pour couvrir des catégories de programmes présentant des contextes audiovisuels très différents en matière de sémantiques et de choix scénaristiques (dynamique, relation entre les modalités audio et vidéo, changement de plan, *etc.*). Chaque extrait était présenté trois fois, avec trois conditions de qualité différentes : une condition référence (C0) et deux conditions dégradées (C1 et C2). La condition C0 correspondait à l'extrait original présenté en qualité source, c'est-à-dire sans compression audio (48Kps, 16 bit) ou vidéo (.avi non compressé). Une première condition de dégradation (C1) présentait des périodes de désynchronisation (**D**) entre le son et l'image. Chaque période dégradée correspondait à un seuil différent de désynchronisation : - 250 ms, -120 ms, - 40 ms, +220 ms et +350 ms, les valeurs négatives représentant un son en avance par rapport à l'image. Ces valeurs ont été choisies pour se situer au-dessus des seuils d'acceptabilité (§ 2.1.3.3, chap. II), à l'exception du seuil fixé à - 40 ms devant permettre de tester l'influence d'un niveau de désynchronisation à la limite de la perceptibilité. Chaque période dégradée durait trente secondes et était insérée suivant l'ordre noté ci-dessus. Chacune de ces périodes étaient précédées d'une période de trente secondes non dégradées (sans désynchronisation). Outre le délai entre l'image et le son, les qualités des signaux audio et vidéo étaient identiques à celle de la condition de référence. Une deuxième condition de dégradation (C2) présentait des périodes de dégradations du signal A, V ou AV par réduction de débit. Pour rappel, les contenus TVHD sont diffusés avec un minimum de 6 Mbps pour la vidéo et de 128 Kbps pour l'audio. Cinq niveaux de dégradations étaient insérés :

- **A-** : audio compressé (MP3, 16 Kbps)
- **A+** : audio compressé (MP3, 24 Kbps)
- **V+** : vidéo compressée (AVC (H264/ MPEG-4 Part 10) - x264, 1500 Kbps)
- **V-** : vidéo compressée (AVC-x264, 1000 Kbps)

- **AV--** : combinaison des dégradations A- et VComme pour la condition C1, chaque période de dégradation (insérée selon l'ordre donné ci-dessus) durait trente secondes et était précédée par une période de trente secondes non dégradée. La vid 6.5 présente les patterns des dégradations pour les conditions C1 et C2. Après introduction des dégradations, les segments audios étaient normalisés (homogénéisation du volume) afin d'éviter la présence de sons trop forts ou trop faibles intra ou inter contenu (plage de variation de l'énergie, liée à l'amplitude, comprise dans un intervalle de -1 à 1). Enfin, les contenus étaient décompressés pour pouvoir être diffusé au format full HD. vid. 6.5. Pattern des dégradations des conditions C1 et C2 où Ø représente une période sans dégradations.

3.2. Observables

3.2.1. Mesures subjectives

La norme UIT-T P.912 (UIT, 1998) ne proposant pas de question pour évaluer séparément QA et QV, des questions pour l'évaluation séparée de QA et QV ont été extraites de la norme P.920 et ajoutées au protocole d'évaluation. Ainsi, après la visualisation de chaque extrait, les participants devaient évaluer successivement QAV, QV et QA à partir d'une échelle catégorielle en neuf points et cinq items (*Excellent-Bon-Satisfaisant-Médiocre-Mauvais*) comme recommandée par la méthode ACR de la norme P.912 (voir sect. 1.5.1, chap. I) cité par docteur Julie LASSALLE. La Vidéo 6.6 ci-dessous rappelle l'échelle recommandée par la méthode ACR. L'ordre de présentation des questions était conforme à celui proposé par la norme P.920 et était fixe pour l'ensemble des participants (QAV, QV puis QA). L'annexe 6-B présente le questionnaire utilisé. 98 Excellent 76 Bon 54 Satisfaisant 32 Médiocre 1 Mauvais
Vid. 6.6. Echelle d'évaluation de la qualité recommandée par la méthode ACR de la norme P.911.

3.2.2. Mesures physiologiques et oculaires

Dans cette expérimentation, les données oculaires recueillies correspondaient à la fréquence et la durée de fermeture de l'oeil (EBfreq, EBdur), au PERCLOS et au diamètre pupillaire (DP). Pour les indices physiologiques, la conductance cutanée (AED), la volumétrie sanguine périphérique (VSP), la fréquence cardiaque (FC) et la température cutanée périphérique (TCP) étaient mesurées.

Les fréquences d'échantillonnage choisies pour chaque indice correspondaient aux valeurs maximales par défaut proposées par les outils de mesure, à savoir : 256 Hz pour les données d'AED et de TCP, 2048 Hz pour le VSP et la FC (calculée, *a posteriori*, à partir du VSP) et 60 Hz pour l'ensemble des indices oculaires. Les observables subjectifs et psychophysiologiques pour les différents types de mesures et les outils de recueil utilisés sont synthétisés dans la vidéo 6.1 ci-après.

Vidéo 6.1. Synthèse des différents observables et outils de recueil pour chaque type de mesures étudiées.



SUBJECTIVE : QAV, QV, QA

Echelles 09 points 06 items (*Excellent-Bon-Mauvais*)

PHYSIOLOGIQUE AED, FC, TCP, VSP, Paire d'électrodes Pléthysmographe
Thermistor

OCULAIRE DP, EBdur, EBfreq, PERCLOS, Eye tracker

3.3. Protocole

A l'arrivée des techniciens, les consignes (présentées dans l'annexe 6-C) ainsi que le questionnaire pour l'évaluation de qualité étaient distribués. Directement après cette étape, les capteurs pour l'enregistrement des mesures physiologiques étaient installés sur la main non dominante des participants (paire d'électrodes, pléthysmographie et thermistor). Le placement des différents capteurs a été réalisé comme présenté dans (électrodes pour mesures d'AED placées sur les phalanges médiales de l'index et du majeur). Les mesures enregistrées étant particulièrement sensibles au mouvement, les consignes demandaient donc au participant d'éviter, dans la mesure du possible, tous mouvements du bras ou de la main sur laquelle étaient disposés les capteurs. Il était toutefois possible de repositionner précautionneusement le bras ou la main durant les périodes d'évaluation. L'installation des capteurs peu de temps après l'arrivée des participants avait pour objectif d'une part de stabiliser les mesures physiologiques (différences de température extérieure/intérieure, effort - escalier-, prise de nicotine, *etc.*) et d'autre part, d'habituer le participant au port des capteurs. Après cette étape, le calibrage et la création du modèle de tête individuel étaient réalisés suivant la procédure décrite dans la section vid 3.6.1. Cette étape durait en moyenne 5 à quinze minutes. Après le calibrage de l'*eye tracker*, une seconde phase correspondait à l'enregistrement de la baseline durant sept minutes afin de récolter les mesures nécessaires pour établir la condition de comparaison. Durant l'enregistrement de la baseline, les participants avaient pour consigne de se détendre. Aucun stimulus n'était présenté durant cette étape. Après les phases d'installation et d'enregistrement de la baseline, le participant visualisait les trois contenus (Opéra, Documentaire, Sport selon cet ordre, fixe pour l'ensemble des participants), chaque contenu étant présenté successivement trois fois, soit neuf séquences d'une durée de sept minutes chacune. Une pause de deux minutes entre chaque séquence permettait aux participants d'évaluer les niveaux de qualité audio, vidéo et audiovisuelle. La durée totale de passation était d'environ une heure quarante-cinq. La vid 6.7 présente le déroulement et le chronogramme de l'expérimentation A.

vid. 6.7. Déroulement et chronogramme de l'expérimentation A. Chaque contenu (5 min), était présenté trois fois, une fois pour chaque condition de qualité (C0, C1 et C2). Après chaque contenu, une pause de 2 min permettait aux participants de compléter le questionnaire.

Conclusion

Les expérimentations A et B présentées dans cette étude ont permis d'étudier l'impact du contenu, d'une part, sur la *qualité d'expérience* du spectateur (en matière d'intérêt, de plaisir) et d'autre part, sur la perception de qualité (B). L'expérience B a permis de disposer d'un corpus de contenus et de séquences de test audiovisuels caractérisés selon des descripteurs sémantiques (de bas et haut-niveau), techniques et hédoniques. A a montré que la caractérisation réalisée par l'expert (descripteurs de bas-niveau seulement) peut être considérée comme pertinente pour une description des contenus compréhensible par des participants naïfs en-dehors du descripteur de température de couleur. D'un point de vue méthodologique, cela autorise à ne mobiliser qu'un expert seul pour l'annotation des séquences ou contenus de test. L'expérience B a permis de mieux comprendre les interactions entre le contenu, défini par les descripteurs proposés, et la perception de qualité. La modalité dominante, la dynamique, la relation audiovisuelle (diégétique) et l'expression sonore influencent notamment la qualité audiovisuelle perçue par le spectateur. Ce constat souligne l'importance de l'étape de sélection des séquences de test ainsi que la description préalable à réaliser. Par exemple, pour être étudiée, la désynchronisation devra être appliquée à des séquences présentant un segment audio à la fois verbal et diégétique. Une seconde campagne de test pourrait être conduite afin de tester le questionnaire utilisé dans l'expérimentation B pour un protocole similaire à l'expérimentation A. Cela permettrait d'étudier plus précisément les effets des dégradations non plus sur la perception de qualité seule mais également sur la *qualité d'expérience* finale du spectateur. Il pourrait également être intéressant d'élargir le cadre de la norme UIT-T P.912 aux évaluations des qualités audio et vidéo pour permettre l'étude du poids de chaque qualité à la perception de qualité audiovisuelle. De plus, dans le cadre du corpus de séquences étudiées, les descripteurs *Modalité* et *Dynamique* étaient fortement liés (dynamique faible associée à une dominance audio). Ce constat va dans le sens de celui de Hands (2004) cité par docteur Julie LASSALLE dans sa thèse de doctorat, cependant, cette association pourrait être l'expression d'une limite liée au questionnaire où le terme de dynamique réfère uniquement à la modalité vidéo. Une étude complémentaire où il serait demandé aux participants d'évaluer à la fois la dynamique audio et vidéo pourrait être envisagée pour approfondir ce point.

Références bibliographiques

- Wastell, D. G. et Newman, M. (1996). Stress, control and computer system design: a psychophysiological field study. *Behaviour and information technology*, 15(3), 183- 192.
- Welch, R. B., DuttonHurt, L. D. et Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Attention, Perception et Psychophysics*, 39(4), 294-300. Welch, R. B. et Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, 88(3), 638-667.



- Welch, R. B. et Warren, D. H. (1989). Intersensory interaction. Dans K. R. Boff, L. Kaufman et J. P. Thomas (dir.), *Handbook of perception and human performance*. Dordrecht, Pays-Bas: Kluwer academic.
- Widmaier, E. P., Raff, H. et Strang, K. T. (2012). *Vander physiologie humaine : les mécanismes du fonctionnement de l'organisme* (5e éd.). Paris, France.
- Wierwille, W. W., Wreggit, S. S. et Knipling, R. R. (1994). Development of improved algorithms for on-line detection of driver drowsiness. Dans *Proceedings of the International Congress on Transportation Electronics*, 331-340.
- Wilcott, R. (1967). Arousal sweating and electrodermal phenomena. *Psychological bulletin*, 67(1), 58-72.
- Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.
- Wilson, G. F. et Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. *Multiple-task performance*, 329-360.
- Wilson, G. F. et O'Donnell, R.D. (1988). Measurement of operator workload with the neurophysiological workload test battery Dans P.A. Hancock et N. Meshkati (dir.), *Human mental workload* (p. 63-100). Amsterdam, Pays-Bas.
- Wilson, G.M. et Sasse, M.A. (2000a). Do Users Always Know What's Good For Them? Utilizing Physiological Responses to Assess Media Quality. Dans *Proceedings of Human Computer Interaction (HCI) : People and Computers XIV - Usability or Else!* 327-339.
- Wilson, G.M. et Sasse, M.A (2000b). Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. Dans *Proceedings of OZCHI : Interfacing Reality in the New Millennium*, 135-142.
- Winton, W. M., Putnam, L. E. et Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology*, 20(3), 195-216.
- Woodmansee, J. (1967). The pupil reaction as an index of positive and negative affect. Dans *Proceedings of the Annual Convention of the American Psychological Association*, Washington, DC.
- Woodworth, R. S. et Schlosberg, H. (1954). *Experimental psychology*. New York, NY: Holt, Rinehart & Winston.
- Wu, S. et Lin, T. (2011). Exploring the use of physiology in adaptive game design. Dans *Proceedings of the International Conference on the Consumer Electronics, Communications and Networks (CECNet)*, 1280-1283.
- Wundt W (1896): *Gundriss der Psychologie* [Outlines of Psychology]. Leipzig, Germany: Entgelman.
- Yano, S., Ide, S., Mitsuhashi, T. et Thwaites, H. (2002). A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays*, 23(4), 191-201.

- Yerkes, R. M. et Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of comparative neurology and psychology*, 18(5), 459-482.
- Yoon, K., Bolls, P. et Lang, A. (1998). The effects of arousal on liking and believability of commercials. *Journal of Marketing Communications*, 4(2), 101-114.
- You, J., Reiter, U., Hannuksela, M., Gabbouj, M. et Perkis, A. (2010). Perceptual-based quality assessment for audio-visual services: A survey. *Signal Processing: Image Communication*, 25, 482-501.
- Ahlstrom, U. et Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, 36(7), 623-636.
- Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. et Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science*, 213(4504), 220-222.
- Allport, D. A. (1980). Patterns and actions: Cognitive mechanisms are content-specific. In G. Claxton (Ed.), *Cognitive psychology: New directions* (p. 32-59). London, R. -U.: Routledge & Kegan Paul.
- Amiar, Y. (1995). L'analyse du film et de l'image fixe: approche méthodologique *Revue Recherche sur l'information scientifique et technique*, 5(2), 23-28.
- Andreassi, J. L. (2007). *Psychophysiology: Human behavior and physiological response* (5e éd.). Mahwah, New Jersey : Lawrence Erlbaum.
- Andreassi, J.L., Rapisardi, S.C. et Whalen, P.M. (1969). Autonomic responsivity and reaction time under fixed and variable signal schedules. *Psychophysiology*, 6, 58-69.
- Appel, M. L., Berger, R. D., Saul, J. P., Smith, J. M. et Cohen, R. J. (1989). Beat to beat variability in cardiovascular variables: noise or music? *Journal of the American College of Cardiology*, 14(5), 1139-1148.
- Arrighi, R., Alais, D. et Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, 6(3), 260-268.
- Backs, R. W. et Boucsein, W. (2000). *Engineering psychophysiology: issues and applications*. Mahwah, New Jersey: Lawrence Erlbaum Mahwah.
- Baddeley, A. D. (1986). *Working memory*. Oxford, R. -U. : Clarendon.
- Bahill, A. T. et Stark, L. (1975). Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Experimental Neurology*, 48(1), 95-106.