

Mesure de la perception de la qualité audiovisuelle par analyse conjointe de signaux physiologiques

JULIE LASSALLE¹, JEROME DANIEL¹, RONAN LE PAGE², JEAN-MARC GOUJON², LAETITIA GROS¹

¹ Orange Labs

2 Avenue Pierre Marzin, 22307 Lannion Cedex, France

² Laboratoire Foton ENSSAT

6 rue Kérampont, BP 80518, 22305 Lannion, France

¹julielassalle84@gmail.com, jerome.daniel@orange.com, laetitia.gros@orange.com

²ronan.le-page@enssat.fr, jean-marc.goujon@enssat.fr

Résumé - L'influence de la qualité audiovisuelle (AV) sur l'utilisateur a été étudiée à partir de l'analyse de mesures physiologiques complétant les mesures subjectives habituellement utilisées. Le présent papier propose une méthode d'analyse de ces signaux basée sur l'extraction d'indicateurs et la définition d'un modèle empirique de détection automatique (par fusion de données hétérogènes) des modifications éventuelles de l'activité physiologique en réaction à la présence de dégradations de qualité et plus globalement, à la présentation de stimuli AV.

Abstract - The influence of the video quality (AV) assessment was studied upon the basis of physiological measurements in addition to subjective measurements usually used. In this paper, we propose new indicators extracted from these signals and an empirical method (with heterogenous data fusion) for automatic detection of physiological reaction due to quality degradation or more generally in response of stimuli.

1 Contexte

Dans un contexte fortement concurrentiel, un des principaux enjeux pour les acteurs du domaine de l'offre de services audiovisuels (AV) est de garantir une qualité d'expérience (QoE) optimale notamment en termes de qualité du signal audio/vidéo restitué à l'utilisateur (QAV). Actuellement, la QAV est principalement évaluée à partir de la collecte de notes données par des testeurs naïfs sur des échelles de qualité, après visualisation/écoute de séquences AV traitées par le service ou la technologie à évaluer. Ces tests subjectifs suivent des procédures recommandées par l'Union Internationale des Télécommunications (UIT). Cependant, la QoE ne peut se réduire à la seule évaluation de la qualité du signal AV restitué mais doit également tenir compte des différents facteurs capables de l'influencer tels que la fatigue ou l'effort mental induits par le niveau de qualité. Une méthode alternative pour pallier les faiblesses des méthodes subjectives actuelles pour évaluer la qualité pourrait consister à compléter les mesures subjectives par des mesures physiologiques reflétant l'activité du système nerveux autonome [1, 2, 3].

1.1 Procédure expérimentale

Trente-trois participants ont visualisé et écouté un total de cinq contenus audiovisuels (Opéra, Sport, Théâtre, Ballet et Documentaire) présentés en format 3D/stéréo. La durée des contenus de test était comprise entre 10 et 14 min. Durant la visualisation, des dégradations audio (10% de perte de paquets), vidéo (réduction du débit, variable entre 94 et 550 Kbps selon

le contenu) et audiovisuelles (combinaison des dégradations audio et vidéo et désynchronisation image/son avec 1500 ms d'avance du son) étaient appliquées. Chaque dégradation audio (A), vidéo, (V), audio-vidéo combinée (AV) et désynchronisation (D) était insérée pendant une minute. Chaque contenu présentait un pattern de dégradations identique pour tous les participants (voir Figure 1 ci-dessous).

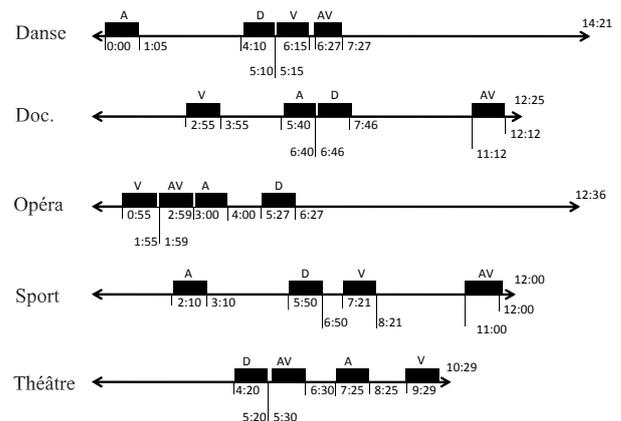


Figure 1 : Pattern d'introduction des dégradations A, V, AV et D pour chacun des cinq contenus de test

En revanche, l'ordre de présentation des contenus était aléatoire. Après chaque visualisation, il était demandé au participant de juger les niveaux de qualité audiovisuelle, vidéo et audio sur une échelle catégorielle en neuf points et cinq items (Excellent-Bon-Satisfaisant-Médiocre-Mauvais). Au cours de l'expérience, quatre indices physiologiques ont été mesurés en continu : la

fréquence cardiaque (FC), l'activité électrodermale (AED, mesurée à partir de la conductance électrodermale), le volume sanguin périphérique (VSP) et la température cutanée périphérique (TCP). Ces mesures ont été recueillies à l'aide de l'outil *Procomp Infiniti* (Thought Technologies™).

1.2 Approches pour l'analyse des données

Il était attendu que la présence de dégradations A et/ou V soit reflétée par les mesures physiologiques recueillies. Une première étape a consisté à réduire les données par la moyenne de chaque signal physiologique pour chaque période : dégradée (1 min) ou non dégradée (durées variables). Une série d'analyse de la variance (ANOVA) a ensuite été réalisée afin d'observer d'éventuelles différences entre les moyennes des différentes périodes. Les résultats n'ont globalement pas révélé d'effets des conditions expérimentales (c.-à-d. des fluctuations de qualité) sur les données obtenues. Pour améliorer la démarche, une investigation s'est développée suivant deux axes : l'extraction de nouveaux indices (section 2) et la proposition d'un modèle empirique de détection automatique (section 3) pour observer les éventuelles réactions physiologiques liées à la présence de dégradations.

2 Traitement des signaux et extraction d'indicateurs

Dans le cadre de cette étude, des techniques d'analyse des signaux bruts (FC, AED, TCP, VSP) ont été développées pour permettre de dégager plus facilement des indicateurs de changement d'état ou « d'événements ». Par ailleurs, l'objectif d'une procédure aussi automatisée que possible amène à élaborer des algorithmes s'adaptant à de fortes variabilités interindividuelles et même intra-individuelles (sur la durée de l'expérience), là où les algorithmes standards demandent en général un ajustement manuel au cas par cas (de seuils, etc.) ou bien manquent de robustesse.

Dans l'exemple illustré en Figure 2, un détecteur a permis de transformer la courbe continue de l'AED en représentation parcimonieuse de réactions dont l'amorce est localisée temporellement et l'intensité quantifiée. Ce détecteur s'appuie sur l'identification de constantes individuelles de pentes en période de relaxation, estimées sur le long terme (méthode différente de [4] bien que d'esprit similaire).

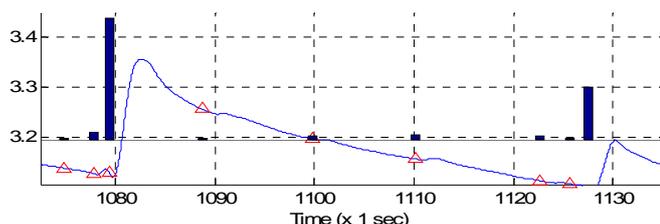


Figure 2 : Conductance électrodermale (courbe bleue) et indicateurs d'événements extraits (triangle : amorce réaction, barre verticale : intensité réaction)

La FC quant à elle est classiquement exploitée à travers sa variabilité temporelle et son observation en bandes de fréquences (hautes, basses et très basses fréquences). Ces indicateurs sont à relativement long terme au regard des événements que l'on cherche à identifier. C'est pourquoi des caractéristiques de l'onde de pouls (VSP), plus en amont, sont extraites pour être prises en compte dans un détecteur tel que décrit en section 3. La Figure 3 permet d'observer des variations de l'enveloppe d'amplitude (simple à évaluer) et de la forme d'onde de chaque battement (qui traduit l'élasticité des artères et reflète un effet du stress [5,6]). Ces caractéristiques apparaissent pour partie corrélées à d'autres signaux comme l'AED et la TCP et doivent permettre à la fois de consolider et affiner le détecteur.

Enfin, il importe que la FC elle-même et ses indicateurs dérivés participent efficacement à la détection de réactions aux stimuli et conditions de l'expérience. Or, les mesures recueillies pour certains sujets ont montré des anomalies fréquentes (battements manquants et/ou fausses détections). Un algorithme exploitant des statistiques locales de l'onde de pouls a donc été élaboré afin de fiabiliser la détection des battements cardiaques. Il permet par la même occasion d'identifier d'éventuels accidents (battements vraiment manquants, précoces ou bien tardifs), susceptibles d'être expliqués par des facteurs individuels et non expérimentaux, ou *a minima* d'être isolés. Comme illustré dans la Figure 3, l'algorithme fournit une détection robuste des pulsations cardiaques, complétée par le marquage et la caractérisation d'anomalies effectives. On peut alors choisir de les intégrer telles quelles ou bien de les « régulariser » avant calcul de la FC et indicateurs dérivés.

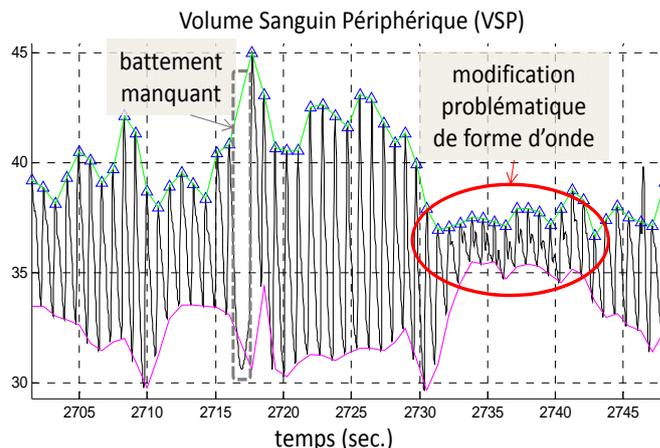


Figure 3 : Exemple d'onde de pouls (en noir) avec marqueurs de battements détectés (triangles bleus), enveloppe d'amplitude (min : magenta ; max : verte). La modification d'enveloppe et de forme d'onde cerclée de rouge est sujette à fausses détections ou détections manquantes avec des algorithmes standards

3 Détection automatique de la présence de dégradations

En contexte d'analyse de signaux multiples et de prise de décision, un modèle empirique de détection de l'influence de dégradations sur le pattern physiologique de l'individu a ensuite été élaboré en combinant

plusieurs signaux au sein d'un même modèle de Markov caché (HMM Hidden Markov Model) [7, 8].

Les paramètres d'entrée du modèle sont des vecteurs de signaux physiologiques : AED et indicateurs d'AED (section 2), TCP, FC (calculée en battements par minute) et/ou autres indicateurs associés (section 2). La phase d'apprentissage a permis d'estimer les densités de probabilité du modèle individuellement à partir de quatre contenus. Le test a été réalisé sur l'ensemble des signaux physiologiques mesurés durant la visualisation des cinq contenus de test par dix-sept individus (choisis pour la qualité de leurs mesures). La Figure 4 présente les résultats obtenus pour deux individus. Elle permet d'observer un certain nombre d'occurrences de détection pendant les périodes de dégradation mais aussi de nombreuses fausses alarmes.

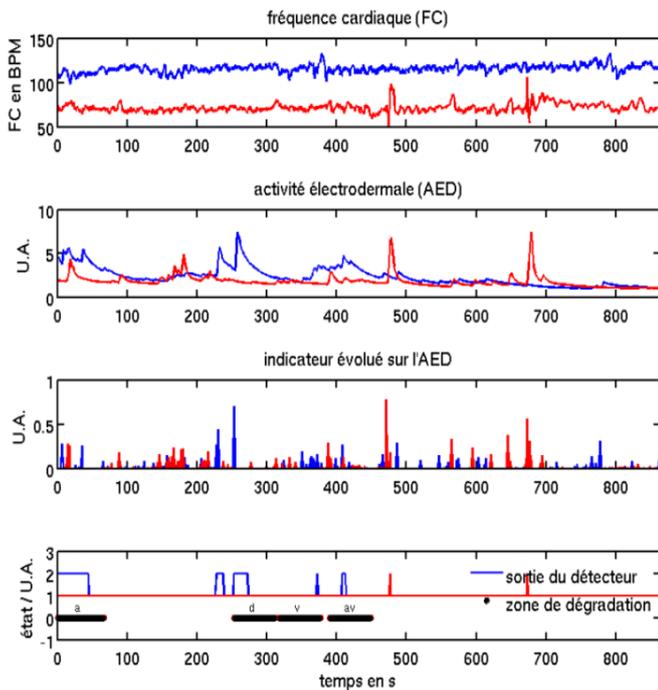


Figure 4 : Signaux d'entrée du modèle, état détecté en sortie du modèle et localisation des zones de dégradations

La présence de fausses alarmes pourrait être expliquée par une influence du contenu de la séquence (changement de plan, dynamique, intérêt, émotion induite, etc., caractérisés lors d'un questionnaire). Des images représentatives des différents niveaux de dynamique possibles au sein d'un même contenu sont apportées dans la Figure 5.



Figure 5 : Images extraites du contenu Documentaire. L'image de gauche illustre une dynamique faible (repos), celle de droite une dynamique élevée (combat)

La Figure 6 présente les variations du niveau de dynamique (en lien avec les mouvements de caméra) en

parallèle des détections d'évènements obtenus sur l'ensemble des individus.

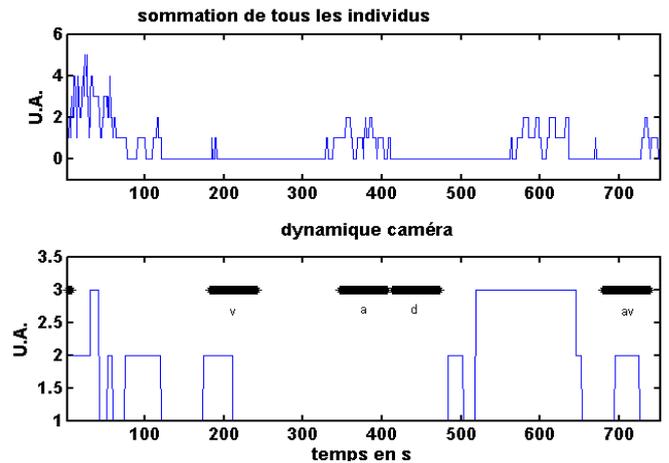


Figure 6 : Etat détecté et variation de la dynamique caméra (1 correspondant à une dynamique faible et 3 à une dynamique forte) pour le contenu Documentaire

La Figure 6 permet également de constater un effet du début du contenu pouvant être expliqué par un impact du changement d'activité (repos/visualisation), de découverte, de surprise ou encore d'anticipation (le début de chaque nouveau contenu était prévenu quelques secondes à l'avance). Cet effet lié au protocole a été observé pour chacun des cinq contenus de test.

Au-delà de l'influence du contenu ou du protocole, des facteurs propres à l'individu pourraient également être pris en compte comme des patterns spécifiques de réponse physiologique (individus labiles vs. stables) et éventuellement des facteurs psychologiques comme l'humeur (état émotionnel du moment), la fatigue, le stress, etc. Les différents facteurs d'influences, étudiés ou supposés, sur les mesures physiologiques sont schématisés dans la Figure 7 ci-après.

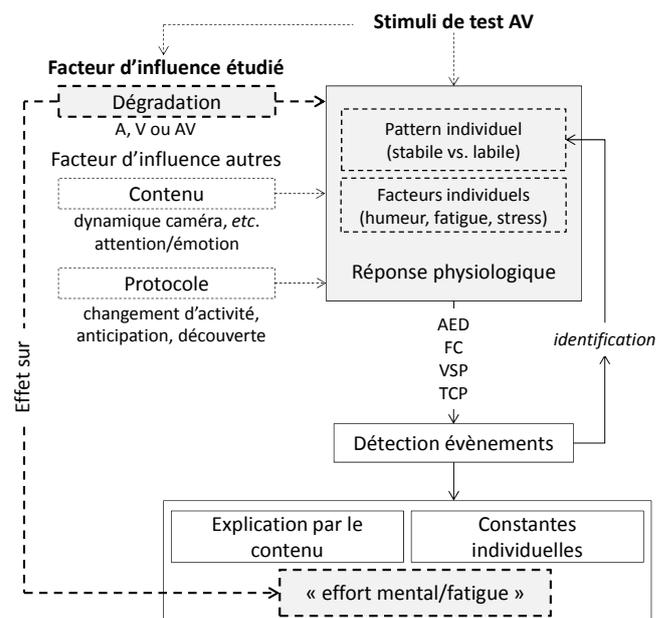


Figure 7 : Schéma présentant les différents facteurs d'influences des mesures physiologiques recueillies

4 Conclusions et perspectives

Dans le cadre d'une proposition de méthode alternative aux méthodes subjectives actuelles pour évaluer l'influence de la qualité audiovisuelle sur l'expérience du spectateur, un ensemble d'indicateurs physiologiques a été étudié. Différents traitements ont été réalisés dans l'optique de pouvoir détecter de manière automatique l'impact de dégradation de la qualité audiovisuelle à partir des signaux recueillis. Les résultats obtenus n'ont pas permis d'observer des réactions systématiques en présence de dégradation. La qualité de la détection effectuée doit être améliorée en envisageant par exemple un meilleur apprentissage et en tenant compte des différents facteurs d'influences de ces mesures (contenu, protocole, individu).

5 Références

- [1] Wilson, G., & Sasse, A. (2000). Do Users Always Know What's Good For Them ? Utilising Physiological Responses to Assess Media Quality. Dans *Proceedings of HCI 2000 : People and Computers XIV - Usability or Else!* (p. 327-339), Sunderland, UK : Springer.
- [2] Wilson, G., & Sasse, A. (2000). Investigating the impact of audio degradations on users : Subjective vs. Objective assessment methods. Dans *Proceedings of OZCHI 2000 : Interfacing Reality in the New Millennium*, 135-142.
- [3] Lassalle, J., Gros, L., & Coppin, G. (2011). Combination of physiological and subjective measures to assess quality of experience for audiovisual technologies. Dans *Proceedings of the Third International Workshop on Quality of Multimedia Experience (QoMEX)*, 13-18.
- [4] Clarion, A. (2009). *Recherche d'indicateurs électrodermiques pour l'analyse de la charge mentale en conduit automobile* (Thèse de Doctorat, Université Lyon I, France). Récupérée du site thèse en ligne : <http://tel.archives-ouvertes.fr>.
- [5] Kageyama, Y. Odagaki, M., & Hosaka, H. (2007). Wavelet Analysis for Quantification of Mental Stress Stage by Finger-tip Photo-plethysmography. Dans *Proceedings of IEEE, Engineering in Medicine and Biology Society (EMBS)*, 1846-1849.
- [6] Kim, K.H., Bang, S.W., & Kim, S.R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing.*, 42(3), 419–427.
- [7] Rabiner, L. R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. Dans *Proceedings of the IEEE*, 77(2), 257–286.
- [8] Le Page, R. (2003). *Détection et analyse de l'onde P d'un électrocardiogramme : application au dépistage de la fibrillation auriculaire* (Thèse de doctorat, Université de Bretagne Occidentale, Brest, France).