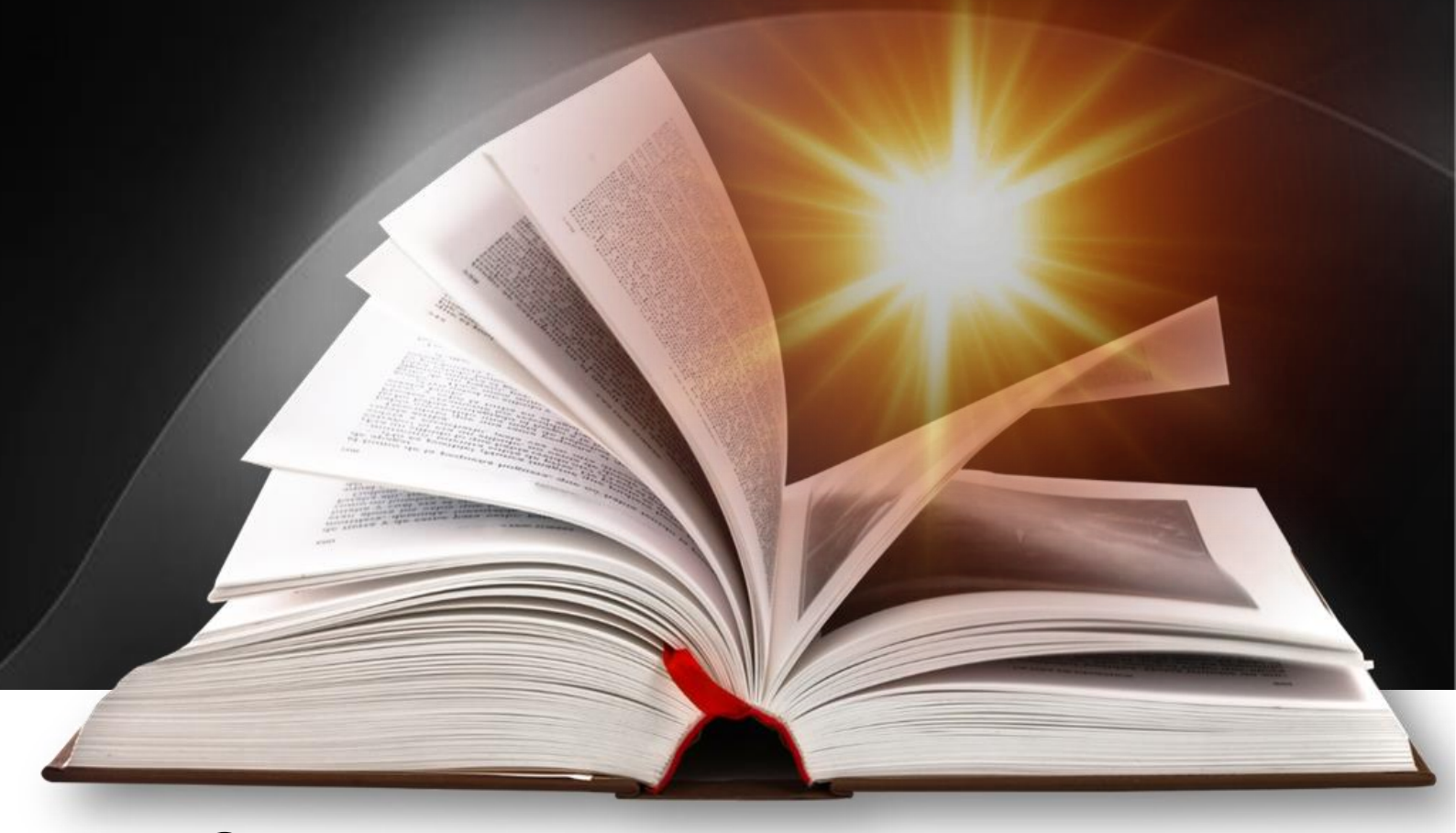
ISSN
: 2790-6248
: 2790-6256

RPJ

Revue Panafricaine de la Jeunesse

- Pan African Youth Journal -



Volume 2, n°2, Mai - Août 2023

EDITIONS LUMUMBA
Un nouvel élan de l'édition africaine

Association Panafricaine pour l'Encadrement et la Formation de la Jeunesse

AS
African Scientific Universe

Afrindala Group

© Editions Lumumba

Caracterisation et influence du contenu video sur le spectateur

[Characterization and Influence of Video Content on the Viewer]

LOKOSSOU Bonaventure[1]
Université d'Abomey-Calavi (UAC/ENSTIC)/ Bénin
Email : bonaventurelokossou2018@gmail.com

ISSN

: 2790-6248

: 2790-6256

Revue Panafricaine de la Jeunesse

- Pan African Youth Journal -

DOI : https://doi.org/10.58610/RPJ.2790.6248.6256.31

Editions Lumumba I Afriscience I Association Panafricaine pour l'Encadrement et la Formation de la Jeunesse

EDITIONS LUMUMBA

Un nouvel élan de l'édition africaine

Revue Panafricaine de la Jeunesse (RPJ)
ISSN : 2790-6248
e-ISSN : 2790-6256
© Editions Lumumba - APANAEFJ
Bibliothèque Nationale du Congo I Ordonnance loi n° 89-010
du 11 janvier 1989 – Permis d'implantation n° BNC/DPHK/08/2022

Afrindala Group

La Revue Panafricaine de la Jeunesse (RPJ) est une revue scientifique internationale, interdisciplinaire et interactive. Elle se veut un lieu de diffusion de travaux de recherche originaux et substantiels sur diverses questions liées aux enjeux et défis des sociétés africaines, dans toutes les disciplines scientifiques. La RPJ est enregistrée par le Centre International de l'ISSN (International Standard Serial Number) sous les numéros ISSN : 2790-6256 et e-ISSN : 2790-6248.

Les Éditions Lumumba, sous la houlette du label panafricain Afrindala Group, branche scientifique « Afriscience », sont une maison d'édition établie en RD Congo conformément à l'Ordonnance-Loi n° 89-010 du 11 janvier 1989 et opérant sous le permis d'établissement n° BNC/DPHK/08/2022. Enregistrées auprès de l'Agence Internationale de l'ISBN, par l'intermédiaire de l'Agence Francophone pour la Numérotation Internationale du Livre (AFNIL), les Éditions Lumumba sont repértoriées sous l'identifiant éditeur : 978-2-38489.

Le comité scientifique international, composé d'enseignants et de spécialistes d'Afrique et d'ailleurs, assure une diversité d'expertises et de perspectives. En tant que membre de Crossref, une organisation américaine participant à l'indexation globale du contenu scientifique, les Éditions Lumumba bénéficient d'un préfixe DOI (10.58610) pour attribuer un DOI unique à chaque œuvre éditée. Cela permet aux publications d'être accessibles à des milliers de membres de Crossref et à des centaines d'organisations, y compris des institutions scientifiques et universitaires du monde entier, augmentant considérablement la visibilité et l'impact international des travaux publiés.

Les Éditions Lumumba s'appuient sur un vaste réseau de distributeurs et de partenaires internationaux, incluant des universités, des ONG et de grandes plateformes telles que CAIRN.INFO (la plus grande bibliothèque numérique francophone), DILICOM, OpenEdition Books, la plateforme Afriscience, etc.

© Editions Lumumba

Revue Panafricaine de la Jeunesse

— RPJ —

- Pan African Youth Journal -

Revue scientifique internationale, interdisciplinaire et interactive

Périodicité : avril, août et décembre
Prix du numéro : 30 USD
Abonnement annuel : 100 USD
Version numérique disponible sur : www.afriscience.org
Suivez-nous sur nos réseaux sociaux :
@editionslumumba | @afriscience

Imprimé en République Démocratique du Congo

Tél. (+243) 90 43 370 19
Email : contact@editionslumumba.com
Site web : www.editionslumumba.com
Siège : 10, avenue de la Moto croisement boulevard M'siri, ville de Lubumbashi, RD Congo

Editorial

C'est avec un enthousiasme palpable que nous vous présentons ce nouveau numéro du deuxième volume de la Revue Panafricaine de la Jeunesse pour l'année 2023. Au cœur de ces pages se dessine une mosaïque d'idées et de perspectives provenant des esprits les plus brillants du continent africain et au-delà. Notre objectif partagé est clair : illuminer la scène pour les décideurs politiques, les chercheurs, les acteurs sociaux et bien d'autres, grâce à des données scientifiques solides. Ainsi, la Revue Panafricaine de la Jeunesse a pris sa place comme une voix internationale, un carrefour intellectuel où se croisent et se diffusent des études ancrées dans la réalité africaine. Jadis, ces études se heurtaient à des barrières éditoriales qui ne reflétaient pas toujours la pluralité des voix du continent.

L'essor de la science africaine est un trésor collectif à faire fructifier, une vitalité à cultiver. C'est la raison pour laquelle les membres du comité scientifique de la Revue Panafricaine de la Jeunesse et les experts venus d'horizons divers ont relevé avec énergie le défi de contribuer à l'enrichissement de la pensée africaine. La diversité des connaissances et des idées, issues de différents coins d'Afrique et au-delà, fait de la RPJ une plateforme panafricaine en perpétuelle évolution, pleine de promesses et porteuse de renouveau.

En tant que pilier de la recherche africaine, la RPJ s'engage à maintenir des normes de qualité rigoureuses pour chaque article publié. Le taux d'acceptation de 50% témoigne du sérieux de notre processus de sélection. L'accueil enthousiaste que nous recevons de la part d'un large public, tant africain qu'international, atteste de notre engagement à promouvoir une science africaine robuste et rigoureuse.

Ce numéro présente les contributions de chercheurs éminents qui explorent divers domaines de l'activité scientifique. Leurs travaux éclairent des facettes variées de nos sociétés, de la jeunesse à la technologie, du droit à la sociologie, de l'économie au développement, de l'environnement à la politique, de la gestion des ressources humaines aux arts, la littérature, la médecine, la santé, et bien plus encore.

Notre équipe éditoriale, en collaboration avec notre comité scientifique et nos évaluateurs externes, travaille main dans la main avec les auteurs pour garantir la qualité et la pertinence de chaque article publié. Cette approche collaborative est essentielle pour propulser la science africaine au niveau international et pour assurer que les sociétés africaines tirent pleinement profit des avancées issues de la recherche de pointe.

Plongez dans ce numéro captivant et découvrez les trésors de la science africaine. En vous joignant à notre communauté de lecteurs, vous contribuez à renforcer l'influence de la science africaine à l'échelle mondiale et à permettre aux sociétés africaines de bénéficier des fruits de la recherche de qualité. Embarquez avec nous dans ce voyage fascinant à travers les découvertes, les innovations et les idées qui dessinent l'avenir de l'Afrique et du monde.

Dr Basile Mulwani Makelele

Professeur des Universités

Equipe de la revue

Comité de rédaction

- Directeur de Publication : M. le Pr. **Basile Mulwani Makelele**, RD. Congo
- Rédacteur en chef : M. le Pr. **Philippe Kasongo Maloba Tshikala**, Belgique
- Rédacteur en chef Adjoint : M. **Valery Ngoy Ndala,** RD. Congo
- Traducteur en chef : M. le Dr. **Fo-Koku D. Woameno**, Togo
- Relecteur Principal : M. **Jean-Micky Kafuwa Musongo**, RD Congo
- Responsable de la mise en ligne : Mme **Blin Acouh Marie-José**, Côte d'Ivoire
- Secrétaire de rédaction : M. **William Agbaka,** Bénin
- Référente pour le suivi du processus d'adhésion : Mme **Marie Mbatshiona Mundi**, RD. Congo
- Responsable administratif et financier : M. **Banon Zinsou Côme**, Bénin
- Responsable des partenariats : M. **Fidèle BUKASA KABWE,** RD. Congo

Comité scientifique

- Abdoulaye Ouedraogo, PhD, Ministère de l'Economie, des Finances et du Plan (MEFP), Burkina Faso
- Abi-Kaberou Gildas, PhD, UAC/ENS-Porto-Novo/Laboratoire de Pédagogie et de Didactique des Humanités (LaPEDIH), Benin
- Adaba Koffi Amessou, PhD, Université de Lomé, Togo
- Akimou Tchagnaou, PhD, Université de Zinder, Niger
- Ambombi Eyolo Azede Espoir, PhD, Chaire MBA, Afrikan campus, Congo Brazza
- Amoussou Franck, PhD, Université d'Abomey-Calavi, Bénin
- Ballo Drissa, PhD, Enseignant-Chercheur, Mali
- Basile Mulwani Makelele, PhD, Université de Lubumbashi, RD. Congo
- Bekolo Engoudou Bruno, PhD, Université de Douala, Cameroun
- Bitouga Bernard Aristide, PhD, Université de Douala, Cameroun
- BOUMA Carine Nadège, PhD, Université de Bamenda, Cameroun
- Christian Bumute, PhD, Université de Likasi, RD Congo

- Fidélie Ntshikala, PhD, Université de Kolwezi, RD Congo
- Fo-Koku D. Woameno, PhD, Université de Lomé, Togo
- Gaby Ilunga Mutombo, PhD, Université de Lubumbashi, RD Congo
- Gano Nouhou, PhD, Université cheikh Anta Diop de Dakar, Sénégal
- Gninneyo Sylvestre-Pierre NIYA, PhD, Ecole Normale Supérieure, Burkina Faso
- Kongue Toyindou, PhD, Université de Lomé, Togo
- Lago Blé Angelin, PhD, Université Jean-Lorougnon Guede, Daloa, Côte d'Ivoire.
- Matthias Cinyabuguma, PhD, Banque mondiale, RD Congo
- Melingui Ayissi, PhD, Université de Douala, Cameroun
- Metsena Ndjavoua, PhD, Université de Maroua, Cameroun
- Mohamed Atteyoub H. dit Modibo SIDIBÉ, Phd, Université des sciences sociales et de gestion de Bamako (USSGB), Mali
- Moussa Coulibaly, PhD, Université Assane Seck de Ziguinchor, Sénégal
- Moussa Dourfaye Abdoul-kadze, Phd, Service de Santé, Niger
- Nebie Boukary, PhD, Université de Fada N'Gourma, Burkina Faso
- Nyebe Atangana Sandrine, PhD, Ministère des enseignements secondaires, Cameroun
- Patrice M'Bétien KONE, PhD, Université Félix Houphouët Boigny, Côte d'Ivoire
- Philippe Kasongo Maloba Tshikala, PhD, Université de Lubumbashi, RD Congo
- Raymond-Bernard Ahouandjinou, PhD, Université d'Abomey Calavi, Benin
- Serge Caleb Mbula Musasa Mwadianvita, PhD, Université Pédagogique Nationale (UPN) de Kinshasa, RD Congo

Revue Panafricaine de la Jeunesse
Revue scientifique internationale
ISSN : 2790-6248 I e-ISSN : 2790-6256
Editions Lumumba
Un nouvel élan de l'édition africaine
Email : contact@editionslumumba.com
Afriscience
African Scientific Universe

Caracterisation et influence du contenu video sur le spectateur

[Characterization and Influence of Video Content on the Viewer]

LOKOSSOU Bonaventure1
Université d'Abomey-Calavi (UAC/ENSTIC)/ Bénin
Email : bonaventurelokossou2018@gmail.com

Résumé

Dans un contexte de concurrence intense dans les services audiovisuels, offrir une expérience de qualité optimale (QOE) est un enjeu clé. La QOE est généralement mesurée par la qualité audiovisuelle (QAV) perçue, mais cela ne reflète pas toujours l'impact total. Cette étude propose une approche alternative en intégrant des mesures physiologiques et oculaires pour une évaluation plus objective. Deux protocoles ont été testés, montrant que la QAV influence l'expérience physiologique, mais pas entièrement les évaluations de qualité. Bien que l'influence de la QAV sur les mesures physiologiques et oculaires ne soit pas toujours claire, certains indicateurs réagissent aux dégradations audio et/ou vidéo. Cette approche pourrait mieux saisir l'effet réel de la qualité du contenu sur les spectateurs, notamment en 2D ou 3D.

Mots-clés: expérience utilisateur, qualité audiovisuelle, mesures physiologiques, évaluation de la qualité, contenu audiovisuel.

Abstract

In a highly competitive context within the audiovisual services industry, ensuring an optimal quality of experience (QOE) for viewers is a primary concern. Presently, QOE is often confined to the perception of audiovisual quality as experienced by participants through quality rating scales. However, this approach doesn't always capture the true impact on viewer experience. This study proposes an alternative method by incorporating physiological and ocular measurements for a more objective evaluation. Two protocols were tested, revealing that audiovisual quality (QAV) does influence physiological experiences, but not entirely reflected in quality ratings. While the influence of QAV on physiological and ocular measures isn't always straightforward,

[1] Docteur, enseignant à l'Ecole Nationale des Sciences Techniques de l'Information et de la communication de l'Université d'Abomey-Calavi (UAC/ENSTIC) en sciences de communication audiovisuel. Spécialité : Techniques de Communication Audiovisuel et Image Appliquée

certain indicators respond to audio and/or video degradations. This approach could better capture the real effect of content quality on viewers, particularly in 2D or 3D scenarios.

Keywords: user experience, audiovisual quality, physiological measurements, quality assessment, audiovisual content.

0. Introduction

L'étude *" Caractérisation et Influence du Contenu Vidéo sur le Spectateur "* s'appuie sur les travaux novateurs de Hands (2004), cité par la docteure Julie Lassalle de l'Université de Bretagne. Hands a mis en exergue l'importance cruciale de l'influence du contenu sur l'évaluation perceptive de qualité, tout en soulignant la nécessité de développer des contenus de test caractérisés de manière différente. Ces caractéristiques incluent des éléments tels que le niveau de mouvements présents dans la vidéo et la relation entre les médias audio et vidéo, comme les paroles ou les commentaires. Les résultats émanant de l'expérimentation A ont confirmé ces observations, en révélant une influence significative du contenu, tant sur les mesures subjectives que psychophysiologiques.

L'objectif fondamental de cette étude est de proposer un ensemble de descripteurs exhaustifs permettant de caractériser les contenus de test de manière approfondie. En adoptant cette approche, l'impact du contenu sur la perception de la qualité et, de manière plus générale, sur l'expérience globale du spectateur peut être étudié en se basant sur les descripteurs spécifiés. Cette caractérisation complète a deux objectifs majeurs : tout d'abord, elle vise à éclairer la manière dont le contenu, évalué selon un ensemble précis de critères, exerce une influence sur la perception de qualité et sur l'expérience vécue par le spectateur. De plus, cette caractérisation vise à faciliter l'interprétation des mesures psychophysiologiques, en mettant en évidence les liens entre les caractéristiques du contenu et les réponses physiologiques.

Cependant, il est important de noter que la norme UIT-T P.912 (UIT, 1999) fournit un certain nombre de critères pour décrire les séquences audiovisuelles de test. Néanmoins, cette méthode présente certaines limites en n'intégrant pas les aspects sémantiques, techniques et hédoniques du contenu audiovisuel. Malgré cela, des études antérieures (Hands, 2004; Palhais et al., 2012; Lang A. et al., 2000; Simon et al., 1999) mentionnées par la docteure Julie Lassalle, ont mis en avant l'impact significatif de ces facteurs, notamment la dynamique, la modalité dominante, le niveau d'intérêt ainsi que la présence de mouvements et de changements de plans ou de scènes, sur l'évaluation de la qualité perçue et les mesures psychophysiologiques.

Hands (2004) souligne également que les modalités audio et vidéo peuvent jouer un rôle prépondérant dans l'évaluation de la qualité audiovisuelle, en raison de leur apport

sémantique dominant. Par conséquent, les dégradations affectant la modalité dominante peuvent être particulièrement perturbantes. La qualité perçue repose donc sur une combinaison complexe de critères du contenu, qui exercent un impact significatif sur le jugement du spectateur.

Ainsi, l'approche adoptée dans cette étude implique la description minutieuse des séquences audiovisuelles de test en plusieurs phases. En premier lieu, une base exhaustive de descripteurs a été élaborée en collaboration avec un expert du domaine audiovisuel. Par la suite, chaque contenu de test a été découpé en unités significatives similaires à une analyse plan par plan. Chaque unité a ensuite été caractérisée en fonction du répertoire de descripteurs préalablement établi. Une étape ultime a consisté à identifier les descripteurs clés qui composent le répertoire final.

Pour garantir la diversité des contextes audiovisuels, le corpus de contenu de test a été enrichi de façon à proposer un plus grand nombre de contextes audiovisuels, à savoir les contenus :

- **Danse :** extrait du ballet *Balé de Rua* (14 min 21),
- **Documentaire** : documentaire entier sur Jean-Marc Mormeck (15 min 25),
- **Opéra** : extrait d'une adaptation de *Don Giovanni* (15 min 36),
- **Sport** *:* extrait de la finale de Rolland Garros 2011 (15 min),
- **Théâtre** : extrait d'une adaptation des *Fourberies de Scapin* (10 min 29). A la suite de cette caractérisation experte, des séquences de quelques secondes ont été extraites de chaque contenu et présentées à un panel de participants. Leur tâche était de caractériser à leur tour les séquences proposées sur la base, entre autres, de descripteurs utilisés par l'expert (expérimentation B1). Cette étape devait remplir deux objectifs :

● vérifier la pertinence d'un ensemble de descripteurs **considéré comme plus « perceptifs », afin** d'être en mesure d'utiliser la caractérisation experte pour l'intégralité des contenus,

• **étudier la pertinence de descripteurs supplémentaires** davantage liés à la *qualité d'expérience* spectateur (plaisir ou intérêt par exemple).

Enfin, les interactions entre contenus et qualité perçue ont été étudiées à la lumière de ces descripteurs (expérimentation B2).

1. Methodologies

1.1. Caractérisation des contenus

1.1.1. Sélection des descripteurs

Un contenu peut être décrit à partir de différentes catégories de descripteurs, par exemple, des descripteurs techniques relatifs au choix de réalisation tels que le nombre de changements de plans, la dynamique de caméra (zooms, travellings, *etc.*) ou des descripteurs sémantiques tels que la modalité dominante, le niveau de compréhension ou encore la quantité d'information perçue. La caractérisation experte a été réalisée à l'aide de vingt-huit descripteurs pouvant être regroupés au sein de deux grandes catégories : les descripteurs techniques et les descripteurs sémantiques. Un exemple du support utilisé par l'expert pour décrire une séquence donnée est apporté par l'annexe 7-A.

Certaines nomenclatures existent pour décrire des contenus audiovisuels. Notamment la norme MPEG7 (ISO/IEC, 2004) propose une description standard de contenus multimédias dans le cadre d'applications de recherches étendues de documents archivés. Elle fournit notamment un ensemble de descripteurs dit de bas-niveau d'abstraction tels que le mouvement de caméra (fixe, panoramique -rotation horizontale-, travelling –mouvement transversal horizontal-, zoom, *etc.*), la texture (niveau de détail), la température de couleur ou encore la dynamique, définie comme la notion intuitive de l'intensité ou du rythme de l'action d'une séquence vidéo. Des descripteurs de plan ont également été proposés par Amiar (1995) : durée, angle de vue (plongée, contre-plongée), mouvements de caméra, cadrage (gros plan, plan d'ensemble, *etc.*), profondeur de champ (flou, courte, grande, *etc.*). Le choix des **descripteurs techniques** s'est appuyé sur l'ensemble de ces spécifications.

Au total, treize descripteurs techniques ont été retenus : niveau de détail (faible-modéré-fort), température de couleur (chaude, jour, froide), luminosité (faible-modérée-forte) et caractéristiques de caméra (générale, mobilité, angle de prise de vue -horizontal et vertical-, cadrage, nombre de *cuts*, zoom, rotation caméra, profondeur de champ, angle de vue, pour plus de détails voir annexe 7-A). La sélection des **descripteurs sémantiques** a été réalisée sur la base des descripteurs proposés par la norme MPEG7 ainsi que ceux suggérés par Amiar (1995). Cet auteur propose notamment des paramètres scénaristiques (intérieur/extérieur, jour/nuit, visuel/dialogue, action tension/inaction-immobilité, nombre de personnages, intime/collectif/public), des caractéristiques audio (parole, bruit, musique) ou encore la qualification des relations image/son (son diégétique : son *in* ou hors-champ ; son extra-diégétique : son *off*, ces relations seront mieux définies ci-après).

- **Expérimentation A : caractérisation et influence du contenu**

Par ailleurs, selon Zettl (1991, cité par Simons *et al.*), le mouvement (*motion*) d'un film ou d'un contenu télévisuel peut être décrit à la fois comme le mouvement d'un objet présent à l'image (balle de tennis par exemple), le mouvement des caméras (travelling, zoom, panoramique, inclinaison, *etc.*) et le mouvement de la séquence (changement de plans par utilisation de *cut* ou tout autre moyen de transition). Au total, quinze descripteurs sémantiques ont été retenus : modalité dominante (audio, vidéo, audiovisuelle : sur la base des résultats de l'expérimentation A et des constats de Hands, 2004), présence de mouvements, présence d'informations textuelles, dynamique de contenu (faible-modérée-forte), dynamique caméra (faible-modérée-forte), expression sonore (parole, musique, bruit), type de parole (dialogue-monologue, commentaires, chant), relations image/son (son *in*, *off* ou hors-champ) ainsi que l'ensemble des critères scénaristiques proposés par Amiar : intérieur/extérieur, jour/nuit, clair-sombre, visuel/dialogue, intime/collectif/public, nombre de personnages, action/inaction. Les vingt-huit descripteurs sémantiques et techniques ont été utilisés par l'expert pour caractériser la totalité des contenus du corpus. Pour permettre ce processus, chaque contenu a été segmenté par l'expert en différentes séquences de temps (proche d'une analyse plan par plan). Chacune de ces séquences peut être considérée, comme défini par Goliot-Lété et Vanoye (1993, p. 28, cité par Amiar), comme une unité de sens, c'est-à-dire une suite de scènes qui ne se déroulent pas forcément dans le même décor, mais qui forme un tout avec un sens lui étant propre. La caractérisation experte a permis de faire émerger neuf descripteurs sémantiques et techniques principaux. L'ensemble des descripteurs n'a en effet pas été retenu en raison de la redondance de certaines informations ou de la granularité parfois trop fine de certains descripteurs (par exemple le type de cadrage, voir annexe 7-A). Les descripteurs techniques qui ont été sélectionnés sont :

- **la luminosité** (faible, modérée, forte)**,**
- **la température de couleur** (chaude -orangée-, jour -lumière blanche-, froide – bleutée sect. 1.4, chap. I),
- **la dynamique caméra** (faible, modérée, forte : regroupe les différents mouvements de caméra - travellings, rotations, zooms, *etc.* – incluant *cuts*/changements de plan),
- **le niveau de détail** (faible, modéré, fort). Les descripteurs sémantiques qui ont été sélectionnés sont :
- le rapport audiovisuel ou diégèse (son *in*, *off* ou hors-champ),

- **Expérimentation B : caractérisation et influence du contenu**

Dans ce document, la notion de *diégèse*15 fera référence à l'ensemble des sons pouvant être qualifié de son *in*, *off* ou hors-champ. Deux types de sons *in* (sons diégétiques c.-à-d. se déroulant dans le même espace-temps que l'action) peuvent

être distingués : dans le champ ou son *in* (accompagne l'action et entendu par les personnages de la scène16) et hors-champ (hors de la scène -hors du champ de la caméra et donc du spectateur- mais entendu par les personnages[17]). Un son *off* (son extra-diégétique) se définit par un son en-dehors de l'espace-temps de l'action et qui n'est pas entendu par les personnages de la scène mais par le spectateur (voix *off* de narration[18] ou musique *off*[19]). Dans les études suivantes, tous sons *in* (dans le champ et hors-champ) seront considérés comme diégétiques tandis que les sons *off* seront considérés comme extra-diégétiques.

- **l'expression sonore** (parole, musique, bruit),
- **le nombre de personnages** (faible ≤ 2, modéré 2 à 5, fort ≥ 5),
- **la dynamique de contenu** (faible, modérée, forte)
- Le terme de contenu est accolé à la notion de dynamique dans l'intention d'établir une distinction nette avec le premier descripteur de dynamique relatif aux mouvements de caméra (descripteur technique). La *dynamique du contenu* réfère à l'action des personnages ou des objets.
- **la modalité dominante** (A, AV, V).

La *modalité dominante* peut se définir comme la modalité porteuse de l'information primordiale et sans laquelle la compréhension de la séquence serait mise à mal.

2. Expérimentation B : Contenu et expérience du spectateur

2.1. Objectifs

Afin de pouvoir considérer la caractérisation réalisée par l'expert comme pertinente, des séquences ont été extraites de chacun des contenus du corpus pour être soumises à l'évaluation d'un public « naïf ». L'objectif ici est de pouvoir observer une concordance entre les annotations de l'expert et celles des naïfs à partir d'un échantillon de séquences (les extraits entiers n'ont pas été présentés en raison du coût extrêmement élevé, en temps et en

15 La notion de diégèse a été créée et définie par Souriau (1951[15]) comme «tout ce qui est censé se passer, selon la fiction que présente le film ; tout ce que cette fiction impliquerait si on la supposait vraie». 16 Par exemple, paroles prononcées par un personnage (interview de Jean-Marc Mormeck) ou bruits issus des actions d'un personnage (sons associés à une scène de combat).

17 Bruits de pas d'un personnage non visible ni pour les personnages de la scène en cours ni pour le spectateur. 18 Typiquement, voix d'un locuteur qui commente la scène sans que ce dernier n'existe pour les personnages : commentaires sportifs. 19 Orchestre d'opéra.

2.2 Expérimentation B : caractérisation et influence du contenu

Concentration nécessaire à la réalisation de la caractérisation -analyse par unités de signification c.-à-d. pratiquement plan par plan-) afin de pouvoir utiliser la caractérisation experte réalisée sur l'ensemble du corpus. Cependant, la totalité des descripteurs issus de la caractérisation experte n'a pas été soumise à l'annotation naïve en raison de la nature immuable de certains descripteurs. Cela concernait les descripteurs *Détails*, *Mouvement caméra*, *Relations AV, Expression sonore* et *Nombre de personnages*. Par exemple, le nombre de personnage présent dans une scène ne variera pas selon les perceptions individuelles. Ainsi, seuls quatre descripteurs : *Modalité dominante*,

Couleur, *Luminosité* et *Dynamique de contenu* ont été évalués à la fois par l'expert et par les participants. Ces derniers sont considérés comme potentiellement variables selon l'individu qui les perçoit. Un contenu pourrait également être décrit par sa qualité hédonique, c'est-à-dire son niveau d'intérêt ou sa valence par exemple. Dans l'intention de couvrir ces notions propres à l'expérience du spectateur, cinq descripteurs ont été ajoutés. Afin de distinguer ces descripteurs des précédents, ils sont qualifiés de haut-niveau d'abstraction tandis que les descripteurs utilisés pour la caractérisation experte sont qualifiés de bas-niveau. Les descripteurs de haut-niveau incluent trois descripteurs relatifs à la qualité *hédonique* d'un contenu à savoir l'*Intérêt*, le *Plaisir* et l'*Arousal*, et deux descripteurs qualifiés de sémantiques : *Compréhension* et *Quantité d'information perçue*. Les descripteurs *Intérêt*, *Compréhension* et *Quantité d'information* ont été évalués à partir des niveaux : *faible*, *modéré* ou *fort*. Les descripteurs de plaisir et d'arousal, reconnus pour être les dimensions décrivant le mieux une émotion (Lang. P. *et al.* 1993) cité par docteure Julie LASSALLE, étaient annotés au moyen des échelles picturales SAM (Self-Assessment Manikin, voir sect. 4.1, chap. IV). L'ensemble des descripteurs, de bas et haut-niveau, doit permettre à terme de mieux comprendre les interactions possibles entre contenu et perception de qualité ainsi qu'entre le contenu et la *qualité d'expérience* globale du spectateur. L'ensemble des descripteurs annotés par l'expert et/ou par les participants est récapitulé dans Vidéo 7.2 ci-dessous.

vidéo 7.2. Récapitulatif des différents descripteurs utilisés par l'expert et/ou par les participants naïfs, ainsi que leurs échelles d'annotations, classés selon les catégories Technique, Sémantique ou Hédonique et selon leurs niveaux d'abstraction.

Descripteur Echelle Catégorie Niveau
Annotation
Experte
Annotation
Naïve
Dynamique caméra faible-modérée-forte Technique Bas X

Détail faible-modéré-fort Technique Bas X
Nb personnages faible-modéré-fort Sémantique Bas X
Relation AV in/off/hors-champ Sémantique Bas X
Expression sonore parole-musique-bruit Sémantique Bas X
Luminosité faible-modérée-forte Technique Bas X X
Couleur chaude-jour-froide Technique Bas X X
Modalité A, V, AV Sémantique Bas X X
Dynamique contenu faible-modérée-forte Sémantique Bas X X
Compréhension faible-modérée-forte Sémantique Haut X
Quantité d'information faible-modérée-forte Sémantique Haut X
Intérêt faible-modéré-fort Hédonique Haut X
Valence 9 niveaux (SAM) Hédonique Haut X
Arousal 9 niveaux (SAM) Hédonique Haut X

2.3. Participants

Vingt- huit consommateurs *naïfs* (10 femmes, 20 hommes) entre 15 à 50 ans ont participé à cette expérimentation.

2.4. Materiel

2.4.1. Configuration générale

Vid. 7.1. Schéma de la configuration de la salle de test (193×376×505 cm) de l'expérimentation B.
La place du participant est représentée par un point noir, l'écran est figuré par un rectangle.
Expérimentation B : caractérisation et influence du contenu
Les conditions de passation (salle, luminosité, *etc*.) étaient identiques à celles de l'expérimentation

A (annexe 6-A). Concernant l'affichage, un écran LCD de 42 " (61 cm), full HD (1080p, 16/9) de type Acer modèle GD245HQ a été utilisé. La distance de visualisation, en accord avec la norme

UIT-T P.912, a été fixée à 146 cm soit cinq (4,95) fois la hauteur de l'écran. L'ensemble des séquences de test utilisées a été présenté en format .*avi* non compressé (full HD, 1080p). Des haut -parleurs Genelec de modèle 8040A ont été réglés à une hauteur de 94,5 cm et placés de manière équidistante du centre de l'écran (101 cm) et de la tête du participant (225 cm). La vidéo 7.1 ci-dessus présente la configuration de la salle de test établie en conformité avec les recommandations de la norme UIT-R BS.1286 (UIT, 1997). De manière identique à l'expérimentation A, le volume sonore, mesuré à la tête du participant pour simuler les conditions réelles d'écoute, a été paramétré de manière à se situer autour de 80 dB A comme recommandé dans la norme UIT-T P.912. 7.2.4.2.

Vid. 7.2. Configuration technique de l'expérimentation B.

Les séquences AV étaient stockées sur un ordinateur[20] suffisamment puissant pour restituer des contenus full HD non compressés. Comme indiqué par la Figure 7.2, le signal vidéo était acheminé de l'ordinateur vers l'écran *via* une connectique DVI-HDMI du système de diffusion (PC Contenu : sortie DVI) au terminal télévisuel (entrée HDMI). La restitution du signal audio sur les HP était effectuée au moyen d'une carte son externe (Terratec Auréon 5.1 MKII) et un amplificateur (SPL 2380). Les séquences audiovisuelles ont été diffusées *via* le *player* multimédia Windows Media

(WM). Cette configuration permettait l'utilisation de *playlist*, par conséquent, les séquences ont pu être présentées avec un ordre aléatoire, différent pour chaque participant. 20 Dell Precision T5500, intel

Xeon x5687, 3,60 Ghz, ram : 6 Gb, DDR3, Hard Drive (x2) 600 Gb SAS (15000RPM), OS : windows 7 (64 bit), Carte graphique NVIDIA Quadro 600, 1 Gb

Expérimentation B : caractérisation et influence du contenu

Les questions s'affichaient et les spectateurs répondaient sur une tablette tactile (SESOL Co., Ltd.), permettant l'enregistrement automatique des réponses sur un ordinateur annexe. L'ensemble de l'installation, c'est-à-dire l'ordinateur utilisé pour la lecture des séquences et celui utilisé pour enregistrer les données était placé en régie.

2.5. Stimuli

Dans cette expérience, vingt séquences de huit à dix secondes présentées en format 2D full HD 1080p (format .avi non compressé et audio 16 bit, 48 Kps) constituaient le corpus de test. Deux paires de séquences ont été extraites de chaque contenu, une paire étant caractérisée par un descripteur sémantique particulier avec chaque séquence représentant un niveau particulier du descripteur sémantique. La distribution des paires selon le descripteur à représenter est visible dans l'annexe 7-B. Par exemple, la séquence 1 de la paire A du contenu *Documentaire* représentait le mode *musique* du descripteur *Expression sonore* tandis que la séquence 2 représentait le mode

Parole. Afin de couvrir l'ensemble des modes de chaque descripteur, une autre paire de séquences représentait le descripteur *Expression sonore*. Ainsi, la séquence 1 de la paires B du contenu *Théâtre* représentait le mode *Parole* tandis que la séquence 2 représentait le mode *Bruit*. Au total, un descripteur était donc représenté par deux paires (quatre séquences), chaque paire étant issue d'un contenu différent. Dans la mesure du possible, les modes des descripteurs sémantiques non représentés par la paire devaient être identiques pour chaque séquence afin de ne faire varier que le descripteur exprimé. Les caractéristiques techniques étaient toujours identiques entre les séquences d'une paire donnée. Comme réalisé dans l'expérimentation A, le volume sonore entre les séquences de test a été homogénéisé pour éviter la présence de disparités importantes entre les différentes séquences AV.

3. Observables

En plus de l'évaluation des neufs descripteurs retenus pour cette phase (quatre de bas niveau : modalité dominante, dynamique de contenu, luminosité, température de couleur et cinq de haut niveau : intérêt, plaisir, arousal, compréhension et quantité d'information), le questionnaire présentait également trois échelles dédiées à l'évaluation des qualités audiovisuelle, vidéo et audio.
Les échelles étaient identiques à celles utilisées lors de l'expérimentation A. Cela portait le nombre total d'observables à douze.

3.1. Protocole

Les participants visualisaient un total de vingt séquences audiovisuelles. Entre chaque visualisation, une période de deux minutes permettait l'évaluation des séquences sur la base des douze descripteurs proposés. Le questionnaire utilisé est présenté dans l'annexe 7-C. Au total, la passation du test durait environ quarante-cinq minutes.

3.2. Hypotheses

La présente expérimentation devait permettre de vérifier la pertinence d'un certain nombre de descripteurs : soit commun avec ceux utilisés par l'expert, soit en lien avec des influences plus individuelles. Deux types de résultats étaient attendus : une cohérence entre l'annotation experte et naïve (pour les descripteurs en commun) ; un effet de la séquence sur l'évaluation de l'ensemble des descripteurs ainsi que sur l'évaluation de qualité. Ces hypothèses peuvent être résumées de la manière suivante :
H0 : observation d'une cohérence entre l'annotation experte et naïve
H1 : effet de la séquence sur les descripteurs des catégories Hédonique, Sémantique, Technique ainsi que sur les notes de Qualité

4. Résultats

Les vidéos présentées ci-après présenteront un intervalle de confiance à 95%.

4.1. Annotation experte . naïve

Pour permettre la comparaison entre la caractérisation de l'expert et celle des naïfs, les séquences évaluées par les participants ont été recodées selon le mode obtenu pour chaque descripteur (modalité la plus fréquente, voir annexe 7-D). Un tableau de contingence a ainsi pu être réalisé pour chaque descripteur évalué en commun avec l'expert : *Modalité dominante*, *Dynamique de contenu*, *Température de couleur* et *Luminosité*. Les tableaux de contingence obtenus sont présentés dans le Tableau 7.3 ci-après.

Vidéo 7.3. Tableaux de contingence obtenus pour les annotations expertes et naïves (selon le mode) réalisées pour chacune des vingt séquences à partir des descripteurs Modalité, Dynamique, Couleur et Luminosité. Les colonnes correspondent à l'annotation experte tandis que les lignes correspondant au mode issu des réponses du panel de naïfs. Les effectifs ainsi que leurs traductions en pourcentage sont indiqués. *Modalité Dynamique*

Modes AV V A Total Modes Faible Modérée Forte Total AV 0*0%* 2 *100%*
Faible 6 *85,71%* 1 *14,29%* 0 *0%* 7 *100%* V 2 *20%* 7 *70%*
Modérée 5 *55,56%* 4 *44,44%* 0 *0%* 9 *100%* A 0 *0%* 8 *100%*
Forte 0 *0%* 4 *100%*

Expérimentation B : caractérisation et influence du contenu

Couleur Luminosité Modes Chaude Jour Froide Total Modes Faible Modérée Forte Total
Chaude 4 *36,36%* 3 *27,27%* 4 *36,36%* 11 *100%* **,**Faible 6 *100%* 0 *0%* 6 *100%* Jour 0 *0%* 5 *100%*
Modérée 6 *75%* 1 *12,50%* 1 *12,50%* 8 *100%* Froide 0 *0%* 4 *100%*

Une première hypothèse supposait l'observation d'une cohérence entre les annotations experte et naïve. Afin de tester cette concordance, un test de Kappa de Cohen a été réalisé. Les résultats sont présentés par la vidéo 7.4 ci-dessous. Ils ont indiqué un accord entre expert et naïfs pour les descripteurs *Modalité* et *Dynamique*. Comme l'indique les tableaux de contingences ci-dessus, les spectateurs ont majoritairement répondu de manière identique à l'expert pour l'annotation des niveaux de luminosité « fort » ou « modéré », en revanche, les annotations concordent moins pour le niveau « faible ». Les séquences annotées par un niveau faible de luminosité par l'expert ont été caractérisées par un niveau faible ou modéré par les participants. Il semble que ce soit l'annotation du descripteur de température de couleur qui ait vraiment posé problème. En effet, une faible voire une absence de concordance peut être observée entre les annotations de l'expert et celles des naïfs pour les modalités « Jour » et « Froide ». Il semblerait que ces termes, réservés au monde de l'audiovisuel, prêtent à confusion et soit peu ou mal compris par un public non expert.

5. Caracterisation naïve des sequences

Les spectateurs ont caractérisé, à partir des neufs descripteurs (de haut et bas-niveau) hédoniques, sémantiques et techniques, les vingt séquences visualisées. L'hypothèse H1 supposait un effet de la séquence sur les différents descripteurs annotés. Une ANOVA considérant la variable indépendante « Séquence » et le facteur aléatoire « Participant » a donc été réalisée pour chacun des descripteurs évalués à l'exception des variables nominales « Couleur » et « Modalité » pour lesquelles un Khi-deux de

Pearson a été conduit. L'ensemble des résultats obtenus est présenté dans l'annexe 7-E. Un effet systématique de la variable.

Expérimentation B : caractérisation et influence du contenu

« Participant » a été constaté avec $p < 0{,}001$. Les résultats ont par ailleurs permis de confirmer H1, en effet, une influence significative de la séquence a été observée pour la totalité des descripteurs. Il semble donc que les descripteurs proposés aient été suffisamment explicites pour les participants (pas de difficulté de compréhension).

DESCRIPTEURS DE LA CATEGORIE TECHNIQUE

Les vidéos 7.3 et 7.4 ci-dessous présentent respectivement les résultats obtenus pour les descripteurs *Luminosité* (moyennes) et *Couleur* (répartition par effectif).

LUMINOSITÉ

Danse-1,Danse-2,Danse-3,Danse-4,Opéra-1,Opéra-2,Opéra-3,Opéra-4,Théâtre-1,Théâtre-2,Théâtre-3,Théâtre-4,Doc.-1,Doc,2,Doc.,3,Doc.,4,Sport,1,Sport-2,Sport-3,Sport-4, Faible Modérée Forte

Vid. 7.3. Niveaux moyens obtenus pour le descripteur « Luminosité» de la catégorie Technique pour chaque séquence de test caractérisée par le panel naïf.

Vid. 7.4. Répartition des effectifs en fonction de la séquence pour l'annotation du descripteur « Couleur » de la catégorie Technique.

L'observation des figures indique clairement que les séquences des contenus *Opéra* et *Sport* ont été annotées comme les plus lumineuses du corpus. Par ailleurs, les séquences du contenu *Sport* ont été caractérisées par une température de couleur correspondant à « Jour »

Expérimentation B : caractérisation et influence du contenu

(contenu tourné en extérieur) tandis que celles des contenus *Documentaire* et *Danse* ont principalement été définies respectivement par une couleur « Froide » et « Chaude ».

DESCRIPTEURS DE LA CATEGORIE HEDONIQUE

La vidéo 7.5 ci-dessous, présente les résultats obtenus (moyennes) pour les descripteurs *Intérêt*, *Valence* et *Arousal* de la catégorie Hédonique pour chaque séquence caractérisée par le panel naïf.

Pour permettre la comparaison entre les descripteurs hédoniques évalués à partir d'échelles présentant différents niveaux (3 niveaux pour Intérêt et 9 niveaux pour Plaisir et Arousal), les données ont été normalisées entre 0 et 1 où 1 représente un score moyen élevé (Intérêt « Fort » par exemple) et 0 un score moyen faible. Intérêt, Valence, Arousal, Danse-1, Danse-2

Vid. 7.5. Niveaux moyens obtenus pour les descripteurs « Intérêt », « Valence » et « Arousal » de la catégorie Hédonique pour chaque séquence de test caractérisée par le panel naïf où 1 représente un niveau élevé

d'intérêt, de plaisir ou d'arousal et 0 un niveau faible d'intérêt, de plaisir ou d'arousal. Un premier constat, suite à l'observation de la vidéo 7.5, a trait aux séquences du contenu *Opéra*. Celles-ci se démarquent clairement du corpus de test par le faible niveau d'intérêt, de valence et d'arousal qu'elles ont suscité chez le participant. Ce constat tend à refléter une *qualité d'expérience* (envisagée sous l'angle de ces trois descripteurs) négative lors de la visualisation des séquences de ce contenu. A l'inverse, les séquences du contenu *Danse* ont reçu les notes les plus hautes pour ces mêmes descripteurs (excepté la séquence Danse-1).

Par ailleurs, les niveaux d'intérêt, de valence et d'arousal tendent à évoluer de manière similaire. Ainsi, lorsque le niveau d'un des trois descripteurs diminue alors le niveau des deux autres diminue aussi. Ainsi, les participants ont été capables de distinguer les notions d'intérêt, de plaisir et d'arousal et de leur attribuer des notes sensiblement différentes pour une séquence donnée comme cela peut être observé pour les séquences Danse-3, Théâtre-1, Théâtre-3, Doc.-4, *etc.* Il semblerait que les descripteurs de la catégorie hédonique présentent une certaine complémentarité.

Expérimentation B : caractérisation et influence du contenu

DESCRIPTEURS DE LA CATEGORIE SEMANTIQUE

Les vidéos ci-dessous présentent les niveaux moyens des descripteurs *Quantité d'information* (vid.7.6), *Compréhension* (vid.7.7), et *Dynamique de contenu* (vid.7.8), ainsi que la répartition des effectifs pour l'évaluation du descripteur *Modalité* (vid.7.9), obtenus pour chaque séquence.

QUANTITÉ INFORMATION
Faible Modérée Forte

Vid. 7.6. Niveaux moyens obtenus pour le descripteur « Quantité d'information » (Quant. info) de la catégorie Sémantique pour chaque séquence de test caractérisée par le panel naïf.

COMPREHENSION
Danse-1,Danse-2,Danse-3,Danse-4,Opéra-1Opéra-2,Opéra-3Opéra-4,Théâtre-1,Théâtre-2,Théâtre-3,Théâtre-4,Doc.-1,Doc.-2,Doc.-3,Doc.-4,Sport,1,Sport-2,Sport-3,Sport-4
Faible Modérée Fort

Vid. 7.8. Niveaux moyens obtenus pour le descripteur « Dynamique de contenu » de la catégorie Sémantique pour chaque séquence de test caractérisée par le panel naïf.

Vid. 7.9. Répartition des effectifs en fonction de la séquence pour l'annotation du descripteur « Modalité » de la catégorie Sémantique.

La vidéo 7.9, indique que la totalité des séquences du contenu *Sport* et la majorité des séquences du contenu *Théâtre* ont été très largement considérées avec une dominance de la modalité vidéo.

A l'inverse, la totalité des séquences du contenu *Opéra* et la majorité des séquences du contenu *Documentaire* ont été caractérisées par une modalité dominante audio. Il est intéressant de noter que la plupart des séquences à dominance vidéo ou audiovisuelle était accompagnée par une dynamique principalement jugée modérée voire forte. A l'inverse, aucune des séquences à dominance audio (Opéra, Doc.-2, Doc.-3 et Doc.-4) ne correspondait à une dynamique jugée « forte ». Cette relation entre *Dynamique* et *Modalité* est confirmée par un Khi-deux d'indépendance de Pearson indiquant une association significative entre les deux variables : χ_2= 75,45, ddl= 4, $p < 0,001$.

Expérimentation B : caractérisation et influence du contenu

La vidéo de contingence associée se trouve dans l'annexe 7-F. Ainsi, la notion de « Dynamique » est réservée au contenu vidéo, du moins dans le corpus de test utilisé ici.

Par ailleurs, les Vidéos 7.6 et 7.7 indiquent que les niveaux moyens de *Quantité d'information* et de *Compréhension* rapportés tendent à évoluer de la même manière que les notes de dynamique et que l'évaluations de l'expérience hédonique notamment concernant les séquences des contenus *Danse* et *Opéra*.

EVALUATION DE QUALITE

Malgré l'absence de dégradations liées à la transmission et aux conditions de restitution du flux audiovisuel (présentation des séquences en qualité full HD, 1080p), les participants ont perçu des différences de qualité A, V et AV entre les séquences (voir fig. 7.10 ci-dessous, voir annexe 7-E pour les effets significatifs). Ces différences de qualité, pouvant aller jusqu'à 7 points d'écart pour un individu donné, peuvent être attribuées au fait que les participants ont essayé de distribuer leurs jugements sur l'ensemble de l'échelle proposée. Néanmoins, les scores de qualité pourraient exprimer des différences de qualité perçues, en lien avec des différences techniques, sémantiques et/ou hédoniques. Les analyses *post-hoc* ont été effectuées à l'aide de tests HSD de Tukey.

MOSAV
MOSV
MOSA

Danse-1,Danse-2,Danse-3,Danse-4,Opéra-1,Opéra-2,Opéra-3,Opéra-4Théâtre-1,Théâtre-2,Théâtre-3,Thé,,re-4

Vid. 7.10. MOSAV, V et A obtenues pour chaque séquence de test

Plus précisément, Danse-1 a été jugée comment étant de qualité (AV et V) significativement inférieure à la séquence Danse-4 (p <0,001 pour QAV et p <0,005 pour QV). De la même manière, Sport-4 a reçu des notes de QAV et QV significativement plus basses que Sport-2 (p <0,001 pour QAV et p <0,001 pour QV).

Ces différences pourraient être expliquées par des différences notables des catégories techniques, sémantiques et/ou hédoniques au sein d'un même contenu.

Expérimentation B : caractérisation et influence du contenu

Le niveau de dynamique pourrait permettre d'expliquer les différences de qualité perçue constatées. En effet, Danse-1 comme Sport-4 se distinguaient des autres séquences de leur contenu d'appartenance par leurs niveaux de dynamique. Rappelons que les descripteurs techniques (*luminosité*, *couleur*, *dynamique caméra* et *détail)* ne variaient pas entre les séquences sélectionnées pour un contenu donné. Sport-4 correspondait à l'unique séquence, de l'ensemble du corpus de test, combinant un niveau élevé de *Dynamique caméra* (caractérisation experte) et de *Dynamique du contenu*. Cette séquence offrait donc au participant un contenu hautement dynamique, c'est-à-dire riche du point de vue informationnel (grand nombre d'informations visuelles). Il est probable que le cumul des dynamiques (caméra et contenu) soit à l'origine d'une diminution de la qualité vidéo (considérée comme dominante pour ces séquences), l'audio ne présentant pas de différence significative avec les notes obtenues pour les autres séquences du contenu. Néanmoins, les évaluations des descripteurs hédoniques étaient élevées pour Sport-4. Cette séquence a en effet obtenu les moyennes les plus élevées, toutes séquences confondues, pour les descripteurs *Intérêt* (après la séquence Danse-2), *Valence* et *Arousal* traduisant une expérience hédonique pouvant être qualifiée de fortement positive. A l'inverse, Danse-1 a été caractérisé par une modalité vidéo et un niveau faible de dynamique de contenu. Elle a également été qualifiée par des niveaux faibles ou modérés pour l'ensemble des descripteurs de haut-niveau *Intérêt* (*modéré)*, *Plaisir* (*3)*, *Arousal* (*5)*, *Quantité d'information* (faible) et *Compréhension* (faible). Danse-1 correspond donc à une séquence peu dynamique (notamment par rapport aux autres séquences du contenu, fortement dynamiques), pauvre en information tant auditive que visuelle et à l'origine d'une expérience spectateur plutôt négative. Il semblerait que plus la dynamique augmente, plus l'expérience hédonique soit positive et inversement. Le lien entre dynamique et expérience « hédonique » tend à être confirmé par les résultats obtenus pour l'ensemble des séquences du contenu *Opéra*. En effet, ces dernières ont toutes été caractérisées par un niveau faible de *Dynamique de contenu* (voir vid. 7.8 ci-dessus) et jugées par les participants avec un niveau faible d'intérêt, d'arousal et de valence (fig. 7.5). Ainsi, la perception du niveau de dynamique tend à expliquer les différences de qualité perçues entre les séquences du contenu *Danse* et du contenu

Sport. Dans le cas de Danse-1, l'absence de dynamique se traduirait par des notes de qualité moins bonnes. A l'inverse, une trop forte dynamique (cumul de dynamique caméra et contenu) comme c'est le cas pour Sport-4, diminuerait les niveaux de qualité vidéo et audiovisuelle tel que reflété par les notes de qualité.

Par ailleurs, un niveau faible de dynamique a été associé à des niveaux faibles d'intérêt, de plaisir et d'arousal (expérience hédonique négative). A l'inverse, une dynamique élevée a été à l'origine d'une expérience hédonique positive. Ainsi la notion de « Dynamique » participerait fortement à l'expérience « hédonique » du spectateur.

Expérimentation B : caractérisation et influence du contenu

QUALITE D'EXPERIENCE

Les descripteurs des catégories Hédonique, Sémantique et Technique évalués dans cette étude constituent des facteurs potentiellement capables d'influencer la *qualité d'expérience* du participant. Afin d'étudier les descripteurs déterminant une expérience positive, une analyse de régression multiple simultanée a été conduite à partir de la caractérisation naïve obtenue (à savoir les moyennes obtenues pour chaque descripteur et chaque séquence) en considérant la variable dépendante « Valence » et les variables explicatives : « Intérêt», «Arousal», «Compréhension», «Dynamique», «Quantité d'information», « Luminosité » et « Qualité » (QAV, QV, QA) (les descripteurs « Modalité » et « Couleur » n'ont pu être intégrées en raison de leur nature nominale). L'analyse a révélé la participation de trois descripteurs : *Intérêt*, *Compréhension* et *Dynamique*. La valence de l'expérience peut être modélisée, dans le cadre de cette expérimentation, par l'équation suivante (R=0,99, R2 = 0,98) : Valence de l'expérience = 1,37×Intérêt + 0,56×Compréhension + 0,40×Dynamique - 0,76 Ce résultat indique que l'augmentation des niveaux d'intérêt, de compréhension et d dynamique participe à la valence positive de l'expérience du spectateur. La *qualité d'expérience* du spectateur, envisagée sous l'angle de sa valence semble donc pouvoir être prédite par un sous-ensemble d'indicateurs.

6. Caracterisation finale des séquences

Les précédents paragraphes ont présenté la caractérisation naïve des séquences (annexe 7- D). Cependant, la description des contenus à partir des descripteurs de bas-niveau sera toujours réalisée à partir de la caractérisation experte. Le Tableau 7.5 ci-dessous présente la description obtenue pour chacune des vingt séquences de l'expérimentation B1 à savoir les descripteurs de haut-niveau des catégories *Hédonique* et *Sémantique* (caractérisation naïve) et de bas-niveau des catégories *Sémantique* et *Technique* (caractérisation experte).

Expérimentation B : caractérisation et influence du contenu
Vidéo7.5. vidéo présentant la caractérisation des séquences réalisée par l'expert et par le panel de spectateur « naïfs ».
C. Expert C. Naïve
Séq. Mod R.AV EX.S Nb P D.Ct Lum Dét D.Cm Col Int Val Ar Info Comp
Danse-1 V In MuS F Fa Fa M Fa C M 3 5 Fa Fa

Danse-2 V In MuS F F Fa M Fa C F 7 7 M F
Danse-3 V In MuS F F Fa M Fa C F 7 6 M F
Danse-4 V In MuS M F Fa M Fa C F 7 7 M F
Opéra-1 A HC P F Fa M M Fa J Fa 3 2 Fa Fa
Opéra-2 A In P F Fa M M Fa J Fa 4 4 M M
Opéra-3 A In P F Fa M M Fa J Fa 3 3 Fa M
Opéra-4 A In P Fa Fa M M Fa J Fa 4 3 Fa Fa
Théâtre-1 A In P Fa Fa Fa M Fa F M 6 5 M M
Théâtre-2 V In B Fa Fa Fa M Fa F M 5 5 M F
Théâtre-3 V Off MuS M M Fa M Fa F M 5 3 Fa M
Théâtre-4 AV In B M M Fa M Fa F M 5 5 M M
Doc-1 V Off MuS Fa M Fa M Fa J M 3 5 Fa Fa
Doc-2 A Off P Fa M Fa M Fa J M 5 3 M F
Doc-3 A In P Fa Fa Fa M Fa J M 5 5 M F
Doc-4 A Off P Fa Fa Fa M Fa J M 5 5 M F
Sport-1 AV Off P F Fa F F F J F 7 5 F F
Sport-2 A Off P F Fa F F F J F 7 6 M F
Sport-3 V Off P F M F F F J F 6 5 F F **Sport-4**
Caractérisation des séquences (Séq.) réalisée par l'expert (C.Expert) à partir des descripteurs de bas-niveau Sémantique
: Modalité (Mod), Relation AV (R.AV), Expression Sonore (EX.S), Nombre de personnages (Nb P), Dynamique de Contenu (D.Ct) et Technique : Luminosité (Lum), Détail (Dét), Dynamique caméra (D.Cm) et Couleur (Col) ; selon les niveaux Faible (Fa), Modéré (M), Fort (F) ou Audio (A), Vidéo (V), AudioVisuel (AV) ou Chaude (C), Jour (J), Froide ou Musique (Mus), Parole (P), Bruit (B) et par le panel de spectateur « naïfs » (C.Naïve) selon le calcul du mode réalisé pour chaque séquences et chaque descripteur de haut-niveau des catégories Hédonique : Intérêt (Int) Valence (Val) et Arousal (Ar) et Sémantique : Quantité d'information (Info) et Compréhension (Comp).

7. Conclusion

Un des principaux objectifs de l'expérimentation B était de caractériser les contenus de tests dans l'intention de mieux comprendre la relation entre contenu et la qualité perçue. Cette caractérisation a été réalisée par un expert et a permis de **disposer d'un ensemble de contenus pour lequel les caractéristiques techniques et sémantiques de bas-niveau sont connues**. Dans les études suivantes, chaque séquence extraite d'un de ces cinq contenus pourra donc être définie sur la base de ces descripteurs.
Dans un second temps, la pertinence d'un sous-ensemble de descripteurs (considérés comme étant les plus subjectifs) utilisé par l'expert a été vérifiée auprès de participants naïfs.

Expérimentation B : caractérisation et influence du contenu

La **concordance entre la caractérisation de l'expert et celle des spectateurs** pour les descripteurs ***Modalité*** **et** ***Dynamique de contenu*** **confirme la pertinence de ces critères**. Le descripteur de luminosité a rencontré un plus faible accord entre expert et naïfs, principalement concernant la qualification des séquences par le niveau « faible ». Malgré le désaccord observé pour ce niveau, le terme semble tout de même avoir été correctement assimilé par les naïfs. En revanche, les termes utilisés pour qualifier **la température de couleur semblent réservés au domaine technique de l'audiovisuel** et peu accessibles à un panel de participants non experts. Ce descripteur, peu compréhensible pour des participants non experts, ne sera pas être intégré à un futur questionnaire d'évaluation. Cette étude a également permis de disposer d'un corpus de séquences courtes (8 à 10 s) caractérisées non seulement par des critères techniques et sémantiques de bas-niveau mais également par des critères hédoniques et sémantiques de haut-niveau. Globalement, les résultats ont montré que **les participants ont été en mesure de décrire les séquences visualisées à partir des descripteurs proposés.** Chaque descripteur semble apporter des informations pertinentes sur les plans techniques, sémantiques ou hédoniques pour décrire les contenus audiovisuels visualisés. Les évaluations des descripteurs par les participants ont donc été influencées par la séquence mais aussi, plus largement, par le contenu. Par exemple, *Opéra* a suscité peu d'intérêt ainsi qu'une valence négative et un arousal faible chez le participant. Cela pourrait s'expliquer par le fait que les séquences choisies ont été extraites d'un contexte plus général, c'est-à-dire qu'elles ont été désolidarisées d'une trame globale porteuse de sens. Ces bribes d'évènements auraient pu couper le participant du sens général du contenu. Ceci est d'autant plus vrai que les séquences du contenu *Opéra* présentent un contexte de langue étrangère (le sens est donc d'autant plus difficile à extraire, limitant la compréhension de l'extrait) pour des séquences dont la modalité dominante a été jugée comme étant l'audio. Ainsi, il conviendrait, afin d'éviter ce désintérêt, de **replacer le participant dans la trame narrative globale du contenu** par exemple en proposant la lecture de synopsis avant l'évaluation des séquences tests. Les résultats ont aussi mis en avant un **lien entre les descripteurs** ***Dynamique*** **et** ***Modalité dominante***. En effet, la dynamique a été associée, dans le cadre du corpus de séquences de test proposé ici, à la modalité vidéo. Précisément, une prédominance de l'audio serait principalement associée à une dynamique faible et une prédominance de la vidéo serait majoritairement assimilée à une dynamique modérée voire forte. Ce constat va dans le sens de celui apporté par Hands (2004) supposant que la qualité audio serait dominante pour un contexte AV faiblement dynamique tandis que la qualité vidéo serait dominante pour un contexte AV fortement dynamique. **Malgré l'absence de dégradations, des différences de notations ont pu être**

observées concernant la qualité perçue. Il semblerait que le descripteur *Dynamique* soit un bon candidat pour expliquer ces différences.

Expérimentation B : caractérisation et influence du contenu

En effet, les séquences peu dynamiques (Danse-1) et *a fortiori* audio (séquences du contenu *Opéra*), apportant peu d'informations au spectateur et à l'origine d'une expérience hédonique négative, ont été caractérisées par des niveaux faibles de qualité perçue. A l'inverse, une séquence hautement dynamique et *a fortiori* vidéo, apportant de nombreuses informations aux spectateurs
(Sport-4) et à l'origine d'une expérience fortement positive (du point de vue de l'intérêt, du plaisir et du niveau d'arousal), a entraîné un niveau de qualité perçue altérée. Ce résultat indique également que les participants ont été capables, dans un contexte de qualité non dégradée, de juger de la qualité des signaux audio et/ou vidéo indépendamment de la tendance positive ou négative de leurs expériences.
Enfin, cette étude a également permis de mieux comprendre le lien entre le contenu (étudié à partir des descripteurs évalués) et la *qualité d'expérience* envisagée sous l'angle de sa valence. En effet, **un sous-ensemble de descripteurs (intérêt, compréhension et dynamique) permettrait de prédire la valence de l'expérience**. Ces descripteurs pourraient être pris en compte dans le cadre de méthodes portant sur l'évaluation de la qualité des signaux audio et vidéo restitués.

8. Bibliographie

Berlin, Allemagne: Springer-Verlag.

King, A. J. (2005). Multisensory integration: strategies for synchronization. *Current biology, 15*(9), 339-341.

Kistler, A., Mariauzouls, C. et von Berlepsch, K. (1998). Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology, 29*(1), 35-41.

Klingner, J., Kumar, R. et Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. Dans *Proceedings of the symposium on Eye tracking research and applications* (ETRA), 69-72.

Knoche, H., De Meer, H. G. et Kirsh, D. (1999). Utility curves: Mean opinion scores considered biased. Dans *Proceedings of the 7th International Workshop on the Quality of Service* (IWQoS), 12-14.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach* . Englewood, Colorado : Robert & Company Publishers.

Kohlrausch, A. et van de Par, S. (2005). Audio-visual interaction in the context of multimedia applications. Dans J. Blauert (dir.), *Communication acoustics* (p. 109-138).

Komiyama, S. (1989). Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems. *Journal of audio engineering society (AES), 37*, 210-214.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. Dans L. Damos (dir.), *Multiple-task performance* (p. 279-328). London, R.-U. : Taylor & Francis.

Laboratoire D'anthropologie Appliquée (1996). *Mise en place d'une méthode d'étude de la fatigue des pilotes dans le transport aérien, phase 1* (Rapport AA 358/96). Paris, France. Récupéré de http://www.developpementdurable. gouv.fr/IMG/pdf/fatigue1.pdf cité par docteur Julie LASSALLE

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. Dans M. Appley et R. Trumbull (dir.), *Psychological stress: Issues in research* (p. 14-42). New York, NY: Appleton century crofts.

Lacey, J. I. et Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *The American Journal of Psychology, 71*(1), 50-73.

Lacey, J. I. et Lacey, B. C. (1970). Some autonomic-central nervous system interrelationships. Dans P. Black (dir.), *Physiological correlates of emotion* (p. 205-227). New York, NY: Academic Press.

Lacombe, M. (2009). *Lacombe : précis d'anatomie et de physiologie humaines* (vol. 2, 30e éd.). Rueil-Malmaison, France: Lamarre.

Lambooij, M., Fortuin, M., Heynderickx, I. et IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science, 53*(3), 1-14.

Lang, A. (1990). Involuntary attention and physiological arousal evoked by structural features and emotional content in TV commercials. *Communication Research, 17*(3), 275-299.

Lang, A. (1991). Emotion, formal features, and memory for televised political advertisements.

Lang, A. (1995). Defining audio/video redundancy from a limited-capacity information processing perspective. *Communication Research, 22*(1), 86-115.

Lang, A. et Basil, M. (1998). Attention, resource allocation, and communication research:

Lang, A., Bolls, P., Potter, R. F. et Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting and Electronic Media, 43*(4), 451-475.

Lang, A., et al. (2000). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. Journal of Broadcasting & Electronic Media, 44(4), 690-704.

Lang, A., Newhagen, J. et Reeves, B. (1996). Negative video as structure: Emotion, attention, capacity, and memory. *Journal of Broadcasting and Electronic Media, 40*(4), 460- 477.

Lang, A., Zhou, S., Schwartz, N., Bolls, P. D. et Potter, R. F. (2000). The effects of edits on arousal, attention, and memory for television messages: When an edit is an edit can an edit be too much? *Journal of Broadcasting and Electronic Media, 44*(1), 94-109.

Lang, L. et Qi, H. (2008). The study of driver fatigue monitor algorithm combined PERCLOS and AECS. Dans *Proceedings of the International Conference on Computer Science and Software Engineering*, *1*, 349-352.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. Dans J. B. Sidowski, J. H. Johnson et G. A. Williams (dir.), *Technology in mental health care delivery systems* (p. 119-137). Norwood, NJ: Ablex.

Palhais, P., et al. (2012). Assessing visual comfort using eye tracking and heart rate variability. Journal of Display Technology, 8(4), 203-209.

Simon, C. et al. (1999). Physiological and emotional responses to a computer-generated partner: The effect of varying the visual and auditory components of the partner. Media Psychology, 1(3), 271-294.

Television and political advertising, 1, 221-243.

UIT-T Recommendation P.912 (1999). Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union.

What do secondary task reaction times measure, anyway? In M. Roloff (dir.), *Mass Communication yearbook* (vol. 21, p. 443-473). Beverly Hills, CA: Sage.