Volume 9, Issue 2, February – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

Characterization and Influence of Video Content on the Viewer

Lokossou Bonaventure
Doctor, Teacher at the National School of Technical Sciences of Information and Communication of the University of Abomey-Calavi (UAC/ENSTIC) in Audiovisual Communication Sciences, Specialty: Audiovisual Communication Techniques and Applied Image

Abstract:-

- ***Study of the Influence of Characterization of Audiovisual on Spectator***

In a highly competitive environment, a primary concern for those offering audiovisual services is ensuring an optimal Quality of Experience (QOE) for the viewer. Presently, QOE tends to be confined to evaluating the perceived audiovisual quality (AVQ) delivered by the system. This assessment typically involves testers rating quality levels on scales after viewing and listening to AV sequences processed through various technologies to be assessed. These subjective tests adhere to protocols recommended by the International Telecommunication Union. However, the actual experience, encompassing factors like fatigue or effort, isn't entirely captured by these quality scores. A more comprehensive method that evaluates not just the received AV quality but also considers the broader quality of experience could better depict how sound and image quality impact the viewer. This study focuses on exploring an alternative approach to current multimedia quality assessment methods in the context of viewing/listening to 2D or 3D AV content. The proposed method delves into QOE by analyzing subjective indicators alongside physiological (electrodermal activity, heart rate, peripheral cutaneous temperature, blood volume pulse) and ocular indicators (PERCLOS, blink duration/frequency, saccadic eye movements, pupil diameter). Physiological and ocular measurements offer advantages by bypassing the biases inherent in subjective measures (such as representativeness and scales) and by revealing phenomena like fatigue or mental effort, possibly triggered by audio and/or video degradations, which significantly impact QOE. Two experimental protocols were implemented to examine the viability of this approach. Findings indicated that AV quality variations influence subjective measures, exposing the inadequacy of quality ratings to accurately represent this impact. However, the influence of quality on physiological and ocular measurements was less straightforward. Specific factors related to certain attributes of test content, such as dynamics or brightness, may have obscured or diminished the observed effects of quality on these measurements. Nonetheless, two physiological indicators reacted to the presence of audio and/or video degradations, particularly when compounded with other factors (like 3D video or test duration effects).

Keywords:- *Audiovisual Quality, Quality of Experience, Subjective Measures, Physiological Measures, Ocular Measures, Mental Fatigue, Mental Effort.*

I. INTRODUCTION AND OBJECTIVES

Hands (2004) cited by Doctor Julie LASSALLE of the University of Brittany emphasized the importance of the influence of content on the perceptual evaluation of quality and the need to offer different types of characterized test content, for example, the level of movements present in the video or the relationship between audio and video media (lyrics or comments). The results of experiment A confirmed this observation by highlighting an influence of content on both subjective and psychophysiological measures. The objective of the experiments presented in this chapter is to propose a set of descriptors making it possible to characterize, in the most complete way possible, the test contents used. The impact of content on the perception of quality and more generally on the viewer's quality of experience can therefore be studied using the specified descriptors. The characterization must allow, on the one hand, to better understand the way in which the content, described by a certain number of criteria, influences the perception of quality and more broadly the quality of experience of the spectator and on the other hand, to facilitate the interpretation of psycho physiological measures. The ITU-T P.912 standard (ITU, 1999) provides a certain number of criteria for describing audiovisual test sequences. All of these criteria are presented in Table 7.1 below. video 7.1. Categories proposed by the ITU-T P.912 standard to describe the audio and video contents of an audiovisual sequence. The proposed description and classification considers audio and video separately, without taking into account the link between sound and image. Generally speaking, the ITU-T P.912 method does not take into account the semantic (dominant modality), technical (change of shots or scenes, movement, etc.) or hedonic (valence and arousal) aspects of the audiovisual content. However, various studies have drawn attention to the influence of these factors such as that of the dynamics and the dominant modality (Hands, 2004), the level of interest (Palhais et al., 2012) or even the presence movements and changes of shots or scenes (Lang A. et al., 2000; Simon et al., 1999) cited by Doctor Julie LASSALLE on the evaluation of perceived quality and on psychophysiological measures. As indicated by Hands (2004), a given modality, audio or video, may contribute more significantly to the audiovisual quality score due to its

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

dominant semantic contribution. The presence of degradations on the dominant modality would then be all the more annoying. The perception of quality is therefore based on different content criteria on which the viewer's judgment will depend. Audiovisual sequences must be described in such a way that a more precise interpretation can be obtained of the quality rating assigned to the audio and/or video signals and of the influences of quality on the quality of experience studied from complementary subjective measures (i.e. (i.e. other than just the quality rating), physiological and ocular. The final objective is to identify the main criteria contributing to the perception of quality and more broadly to the quality of the viewer's experience. The characterization of test content took place in several phases. First, an exhaustive database of content descriptors was developed with the help of an expert in the audiovisual field (professional audiovisual technician – Digipictoris company, Brest). Secondly, each test content was divided into meaningful units close to a plan by plan analysis. Each unit of each content was then characterized based on the previously developed repertoire of descriptors. A final phase consisted of identifying the key descriptors that should constitute the final repertoire. The corpus of test content has been enriched to offer a greater number of audiovisual contexts, namely the contents:

- **Dance :** Extract from the ballet Balé de Rua (14 min 21),
- **Documentary** : Entire documentary on Jean-Marc Mormeck (15 min 25),
- **Opera :** Extract from an adaptation of Don Giovanni (15 min 36),
- **Sport :** Extract from the final of Rolland Garros 2011 (15 min),
- **Theater :** Extract from an adaptation of Les Fourberies de Scapin (10 min 29). Following this expert characterization, sequences of a few seconds were extracted from each content and presented to a panel of participants. Their task was to in turn characterize the proposed sequences on the basis, among other things, of descriptors used by the expert (experiment B1). This step had to fulfill two objectives:

✓ Check the relevance of a set of descriptors considered more “perceptive”, in order to be able to use expert characterization for all content,
✓ Study the relevance of additional descriptors more linked to the quality of spectator experience (pleasure or interest for example).

Finally, the interactions between content and perceived quality were studied in the light of these descriptors (experiment B2).

II. STUDY METHOD

A. *Experiment A: Characterization of Contents*

➢ *Selection of Descriptors*

Content can be described using different categories of descriptors, for example, technical descriptors relating to the choice of production such as the number of shot changes, camera dynamics (zooms, tracking shots, etc.) or semantic descriptors such as as the dominant modality, the level of understanding or even the quantity of information perceived. The expert characterization was carried out using twenty-eight descriptors that can be grouped into two main categories: technical descriptors and semantic descriptors. An example of the support used by the expert to describe a given sequence is provided in Appendix 7-A.

Certain nomenclatures exist to describe audiovisual content. In particular, the MPEG7 standard (ISO/IEC, 2004) offers a standard description of multimedia content in the context of extended search applications for archived documents. In particular, it provides a set of so-called low-level abstraction descriptors such as camera movement (fixed, panning - horizontal rotation, tracking shot - horizontal transverse movement, zoom, etc.), texture (level of detail) , color temperature or even dynamics, defined as the intuitive notion of the intensity or rhythm of the action in a video sequence. Shot descriptors were also proposed by Amiar (1995): duration, angle of view (high angle, low angle), camera movements, framing (close-up, long shot, etc.), depth of field (blur , short, tall, etc.). The choice of technical descriptors was based on all of these specifications.

In total, thirteen technical descriptors were retained: level of detail (low-moderate-high), color temperature (warm, daytime, cool), brightness (low-moderate-strong) and camera characteristics (general, mobility, angle shooting -horizontal and vertical-, framing, number of cuts, zoom, camera rotation, depth of field, angle of view, for more details see appendix 7-A). The selection of semantic descriptors was carried out on the basis of the descriptors proposed by the MPEG7 standard as well as those suggested by Amiar (1995). This author notably proposes story parameters (interior/exterior, day/night, visual/dialogue, action tension/inaction-immobility, number of characters, intimate/collective/public), audio characteristics (speech, noise, music) or even the qualification of image/sound relationships (diegetic sound: sound in or out of frame; extra-diegetic sound: sound off, these relationships will be better defined below).

B. *Experiment A: Characterization and Influence of Content*

Furthermore, according to Zettl (1991, cited by Simons et al.), the movement (motion) of a film or television content can be described both as the movement of an object present in the image (tennis ball for example), the movement of the cameras (traveling, zooming, panning, tilting, etc.) and the movement of the sequence (changing shots by using cut or any other means of transition). In total, fifteen semantic descriptors were retained: dominant modality (audio, video, audiovisual: based on the results of experiment A and the findings of Hands, 2004), presence of movements, presence of textual information, dynamics of content (weak-moderate-strong), camera dynamics (weak-moderate-strong), sound expression (speech, music, noise), type of speech (dialogue-monologue, comments, singing), image/sound relationships (sound in, off or off-camera) as well as all the script criteria proposed by Amiar: interior/exterior, day/night, light-dark, visual/dialogue, intimate/collective/public, number of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

characters, action/inaction. The twenty-eight semantic and technical descriptors were used by the expert to characterize all of the contents of the corpus. To enable this process, each content was segmented by the expert into different time sequences (close to a shot-by-shot analysis). Each of these sequences can be considered, as defined by Goliot-Lété and Vanoye (1993, p. 28, cited by Amiar), as a unit of meaning, that is to say a series of scenes which do not take place necessarily in the same setting, but which forms a whole with its own meaning. The expert characterization allowed nine main semantic and technical descriptors to emerge. All of the descriptors were not retained due to the redundancy of certain information or the sometimes too fine granularity of certain descriptors (for example the type of framing, see appendix 7-A). The technical descriptors that were selected are:

- **The Luminosity**(weak, moderate, strong),
- **Color Temperature**(warm -orange-, day -white light-, cold -bluish sect. 1.4, chap. I),
- **Camera Dynamics**(weak, moderate, strong: brings together the different camera movements - tracking shots, rotations, zooms, etc. - including cuts/shot changes),
- **The Level of Detail**(weak, moderate, strong). The semantic descriptors that were selected are:
- **The Audiovisual or Diegesis Report** (sound in, off or off-camera),

C. Experiment B: Characterization and Influence of Content

In this document, the notion of diegesis15 will refer to all sounds that can be qualified as in, off or off-camera sound. Two types of in sounds (diegetic sounds i.e. taking place in the same space-time as the action) can be distinguished: in the field or in sound (accompanies the action and heard by the characters in the scene16) and off-camera (off-stage -out of the camera's field and therefore of the spectator- but heard by the characters17). An off sound (extra-diegetic sound) is defined by a sound outside the space-time of the action and which is not heard by the characters in the scene but by the spectator (narration voice-over18 or music off19). In the following studies, all in-sounds (on-camera and off-camera) will be considered diegetic while off-camera sounds will be considered extra-diegetic.

- **Sound Expression**(speech, music, noise),
- **The Number of Characters**(weak ≤ 2, moderate 2 to 5, strong ≥ 5),
- **Content Dynamics**(weak, moderate, strong)
- The term content is attached to the notion of dynamics with the intention of establishing a clear distinction with the first dynamic descriptor relating to camera movements (technical descriptor). Content dynamics refers to the action of characters or objects.
- **The Dominant Modality**(A, AV, V).
- The dominant modality can be defined as the modality carrying the primordial information and without which the understanding of the sequence would be undermined.

D. Experiment B: Content and Viewer Experience

➢ *Goals*

In order to be able to consider the characterization carried out by the expert as relevant, sequences were extracted from each of the contents of the corpus to be submitted to the evaluation of a "naive" public. The objective here is to be able to observe a concordance between the annotations of the expert and those of the naïve from a sample of sequences (the entire extracts were not presented due to the extremely high cost, in terms of time and effort).

15 The notion of diegesis was created and defined by Souriau (195115) as "everything that is supposed to happen, according to the fiction that the film presents; everything that this fiction would imply if we supposed it to be true. 16 For example, words spoken by a character (interview with Jean-Marc Mormeck) or noises coming from a character's actions (sounds associated with a fight scene).

17 Sounds of footsteps of a character not visible either to the characters in the current scene or to the spectator. 18 Typically, the voice of a speaker who comments on the scene without the latter existing for the characters: sports comments. 19 Opera Orchestra

E. Experiment B: Characterization and Influence of Content

Concentration necessary to carry out the characterization -analysis by units of meaning i.e. practically plan by plan-) in order to be able to use the expert characterization carried out on the entire corpus. However, all of the descriptors resulting from the expert characterization were not subjected to naive annotation due to the immutable nature of certain descriptors. This concerned the Details, Camera Movement, AV Relations, Sound Expression and Number of Characters descriptors. For example, the number of characters present in a scene will not vary according to individual perceptions. Thus, only four descriptors: Dominant modality,

Color, Brightness and Content Dynamics were evaluated by both the expert and the participants. The latter are considered potentially variable depending on the individual who perceives them. Content could also be described by its hedonic quality, that is to say its level of interest or its valence for example. In order to cover these notions specific to the spectator experience, five descriptors have been added. In order to distinguish these descriptors from the previous ones, they are qualified as high-level abstraction while the descriptors used for expert characterization are qualified as low-level. The high-level descriptors include three descriptors relating to the hedonic quality of content, namely Interest, Pleasure and Arousal, and two descriptors qualified as semantic: Comprehension and Quantity of information perceived. The descriptors Interest, Understanding and Quantity of information were evaluated based on the levels: weak, moderate or strong. The descriptors of pleasure and arousal, recognized to be the dimensions best describing an emotion (Lang. P. et al. 1993) cited by Doctor Julie LASSALLE, were annotated using the SAM pictorial scales (Self-Assessment Manikin, see section

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

4.1, chapter IV). All of the descriptors, low and high-level, should ultimately provide a better understanding of the possible interactions between content and perception of quality as well as between content and the quality of the viewer's overall experience. All of the descriptors annotated by the expert and/or by the participants are summarized in Video 7.2 below.

Video 7.2. Summary of the different descriptors used by the expert and/or by the naive participants, as well as their annotation scales, classified according to the Technical, Semantic or Hedonic categories and according to their levels of abstraction.

- Descriptor Scale Category Level
- Annotation
- Expert
- Annotation
- Naive
- Camera dynamicsweak-moderate-strong Technique Low
- Detailweak-moderate-strong Technical Low
- Number of charactersweak-moderate-strong Semantic Low
- AV relationshipin/off/off-screen Semantics Low
- Sound expressionspeech-music-noise Semantic Low
- Brightnessweak-moderate-strong Technical Low XX
- Colorhot-day-cold Technical Low XX
- ModalityA, V, AV Semantic Low XX
- Dynamic contentweak-moderate-strong Semantics Low XX
- Understandingweak-moderate-strong Semantic High
- Amount of informationweak-moderate-strong Semantic High
- Interestweak-moderate-strong Hedonic High
- Valence9 levels (SAM) Hedonic High
- Arousal9 levels (SAM) Hedonic High

➢ *Participants*

Twenty-eight naïve consumers (10 women, 20 men) between 15 and 50 years old participated in this experiment.

➢ *Material*

- *General Configuration*

Vid. 7.1. Schematic of the configuration of the test room (193×376×505 cm) of experiment B.

The participant's place is represented by a black dot, the screen is represented by a rectangle.

F. Experiment B: Characterization and Influence of Content

The testing conditions (room, lighting, etc.) were identical to those of the experiment.

A (Annex 6-A). Concerning the display, a 42 "(61 cm), full HD (1080p, 16/9) LCD screen of the Acer model GD245HQ was used. The viewing distance, in accordance with the standard

ITU-T P.912, was set at 146 cm or five (4.95) times the height of the screen. All of the test sequences used were presented in uncompressed .avi format (full HD, 1080p). Genelec Model 8040A speakers were set at a height of 94.5 cm and placed equidistant from the center of the screen (101 cm) and the participant's head (225 cm). Video 7.1 above shows the test room configuration established in accordance with the recommendations of the ITU-R BS.1286 standard (ITU, 1997). Identical to experiment A, the sound volume, measured at the participant's head to simulate real listening conditions, was set to be around 80 dB A as recommended in the ITU-T P standard. .912. 7.2.4.2.

Vid. 7.2. Technical configuration of experiment B.

The AV sequences were stored on a computer20 powerful enough to render uncompressed full HD content. As shown in Figure 7.2, the video signal was routed from the computer to the screen via a DVI-HDMI connector from the broadcast system (PC Content: DVI output) to the television terminal (HDMI input). The restitution of the audio signal on the HPs was carried out using an external sound card (Terratec Auréon 5.1 MKII) and an amplifier (SPL 2380). The audiovisual sequences were broadcast via the Windows Media multimedia player

(WM). This configuration allowed the use of playlists, therefore, the sequences could be presented with a random order, different for each participant. 20 Dell Precision T5500, intel Xeons

G. Experiment B: Characterization and Influence of Content

Questions were displayed and viewers responded on a touchscreen tablet (SESOL Co., Ltd.), allowing automatic recording of responses on an annex computer. The entire installation, that is to say the computer used to play back the sequences and the one used to record the data, was placed under control.

➢ *Stimuli*

In this experiment, twenty sequences of eight to ten seconds presented in 2D full HD 1080p format (uncompressed .avi format and 16 bit audio, 48 Kps) constituted the test corpus. Two pairs of sequences were extracted from each content, one pair being characterized by a particular semantic descriptor with each sequence representing a particular level of the semantic descriptor. The distribution of pairs according to the descriptor to be represented can be seen in Appendix 7-B. For example, sequence 1 of pair A of the Documentary content represented the music mode of the Sound Expression descriptor while sequence 2 represented the mode Word.

In order to cover all the modes of each descriptor, another pair of sequences represented the Sound Expression descriptor. Thus, sequence 1 of pair B of Theater content represented Speech mode while sequence 2 represented Noise mode. In total, a descriptor was therefore represented by two pairs (four sequences), each pair coming from a different content. As far as possible, the modes of the semantic descriptors not represented by the pair had to be identical for

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

each sequence in order to vary only the expressed descriptor. The technical characteristics were always identical between the sequences of a given pair. As carried out in experiment A, the sound volume between the test sequences was homogenized to avoid the presence of significant disparities between the different AV sequences.

III. OBSERVABLES

In addition to the evaluation of the nine descriptors retained for this phase (four low level: dominant modality, content dynamics, brightness, color temperature and five high level: interest, pleasure, arousal, comprehension and quantity of information), the questionnaire also presented three scales dedicated to the evaluation of audiovisual, video and audio qualities.

The scales were identical to those used in Experiment A. This brought the total number of observables to twelve.

A. Protocol

Participants viewed a total of twenty audiovisual sequences. Between each visualization, a period of two minutes allowed the evaluation of the sequences on the basis of the twelve descriptors proposed. The questionnaire used is presented in Annex 7-C. In total, the test took approximately forty-five minutes.

B. Hypotheses

The present experiment was intended to make it possible to verify the relevance of a certain number of descriptors: either common with those used by the expert, or linked to more individual influences. Two types of results were expected: consistency between expert and naive annotation (for common descriptors); an effect of the sequence on the evaluation of all the descriptors as well as on the quality evaluation. These hypotheses can be summarized as follows:

- **H0**: Observation of consistency between expert and naive annotation
- **H1**: Effect of the sequence on the descriptors of the Hedonic, Semantic, Technical categories as well as on the Quality scores

IV. RESULTS

The videos presented below will present a 95% confidence interval.

A. Expert Annotation. Naive

To allow comparison between the characterization of the expert and that of the naïve, the sequences evaluated by the participants were recoded according to the mode obtained for each descriptor (the most frequent modality, see appendix 7-D). A contingency table was thus produced for each descriptor evaluated jointly with the expert: Dominant modality, Content dynamics, Color temperature and Brightness. The contingency tables obtained are presented in Table 7.3 below.

Video 7.3. Contingency tables obtained for the expert and naive annotations (depending on the mode) carried out for each of the twenty sequences based on the Modality, Dynamics, Color and Luminosity descriptors. The columns correspond to the expert annotation while the rows correspond to the mode taken from the responses of the naive panel. The numbers and their translations into percentages are indicated. Dynamic Modality.

- AV Modes VA Total Modes Low Moderate Strong Total AV 0*0%* 2 100%
- Low 6 85.71% 1 14.29% 0 0% 7 100% V 2 20% 7 70%
- Moderate 5 55.56% 4 44.44% 0 0% 9 100% A 0 0% 8 100%
- Strong 0 0% 4 100%

B. Experiment B: Characterization and Influence of Content

➢ *Color Brightness*

- Modes Warm Day Cold Total Modes Low Moderate Strong Total
- Warm 4 36.36% 3 27.27% 4 36.36% 11 100%, Low 6 100% 0 0% 6 100% Day 0 0% 5 100%
- Moderate 6 75% 1 12.50% 1 12.50% 8 100% Cold 0 0% 4 100%

A first hypothesis assumed the observation of consistency between expert and naive annotations. To test this agreement, a Cohen's Kappa test was performed. The results are presented in video 7.4 below. They indicated agreement between expert and naïve for the Modality and Dynamic descriptors. As indicated in the contingency tables above, the majority of spectators responded identically to the expert for the annotation of "strong" or "moderate" brightness levels, however, the annotations agreed less for the level " weak ". The sequences annotated by a low level of brightness by the expert were characterized by a low or moderate level by the participants. It seems to be the annotation of the color temperature descriptor that was really the problem. Indeed, a weak or even absence of concordance can be observed between the annotations of the expert and those of the naïve for the "Day" and "Cold" modalities. It would seem that these terms, reserved for the audiovisual world, are confusing and are little or poorly understood by a non-expert public.

V. NAIVE CHARACTERIZATION OF SEQUENCES

The spectators characterized, based on the nine descriptors (high and low level) hedonic, semantic and technical, the twenty sequences visualized. Hypothesis H1 assumed an effect of the sequence on the different annotated descriptors. An ANOVA considering the independent variable "Sequence" and the random factor "Participant" was therefore carried out for each of the descriptors evaluated with the exception of the nominal variables "Color" and "Modality" for which a Pearson chi-square was conducted. . All the results obtained are presented in Appendix 7-E.A systematic effect of the variable.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

A. Experiment B: Characterization and Influence of Content

"Participating" was found with $p < 0.001$. The results also made it possible to confirm H1, in fact, a significant influence of the sequence was observed for all of the descriptors. It therefore seems that the descriptors proposed were sufficiently explicit for the participants (no difficulty in understanding).

➢ *Technical Category Descriptors*

Videos 7.3 and 7.4 below respectively present the results obtained for the Brightness (average) and Color (distribution by number) descriptors.

➢ *Brightness*

Dance-1, Dance-2, Dance-3, Dance-4, Opera-1, Opera-2, Opera-3, Opera-4, Theater-1, Theater-2, Theater-3, Theater-4, Doc. -1,Doc,2,Doc.,3,Doc.,4,Sport,1,Sport-2,Sport-3, Sport-4, Weak Moderate Strong

Vid. 7.3. Average levels obtained for the "Brightness" descriptor of the Technical category for each test sequence characterized by the naive panel.

Vid. 7.4. Distribution of staff according to the sequence for the annotation of the "Color" descriptor of the Technical category.

Observation of the figures clearly indicates that the Opera and Sport content sequences were annotated as the brightest in the corpus. Furthermore, the Sports content sequences were characterized by a color temperature corresponding to "Day"

B. Experiment B: Characterization and Influence of Content

(content shot on location) while those of Documentary and Dance content were mainly defined by a "Cold" and "Warm" color respectively.

➢ *Hedonic Category Descriptors*

Video 7.5 below presents the results obtained (averages) for the Interest, Valence and Arousal descriptors of the Hedonic category for each sequence characterized by the naive panel.

To allow comparison between hedonic descriptors assessed from scales presenting different levels (3 levels for Interest and 9 levels for Pleasure and Arousal), the data were normalized between 0 and 1 where 1 represents a high average score ("Interest" Strong" for example) and 0 a low average score. Interest, Valencia, Arousal, Dance-1, Dance-2

Vid. 7.5. Average levels obtained for the descriptors "Interest", "Valence" and "Arousal" of the Hedonic category for each test sequence characterized by the naïve panel where 1 represents a high level of interest, pleasure or arousal and 0 represents low level of interest, pleasure or arousal. A first observation, following the observation of video 7.5, relates to the sequences of Opera content. These clearly stand out from the test corpus by the low level of interest, valence and arousal that they aroused in the participant. This observation tends to reflect a negative quality of experience (considered from the angle of these three descriptors) when viewing sequences of this content. Conversely, the Dance content sequences received the highest scores for these same descriptors (except the Dance-1 sequence).

Furthermore, the levels of interest, valence and arousal tend to evolve in a similar way. Thus, when the level of one of the three descriptors decreases then the level of the other two also decreases. Thus, the participants were able to distinguish the notions of interest, pleasure and arousal and to assign them significantly different notes for a given sequence as can be observed for the sequences Dance-3, Theater-1, Theater -3, Doc.-4, etc. It would seem that the descriptors of the hedonic category present a certain complementarity.

C. Experiment B: Characterization and Influence of Content

➢ *Semantic Category Descriptors*

The videos below present the average levels of the descriptors Quantity of information (vid.7.6), Comprehension (vid.7.7), and Content dynamics (vid.7.8), as well as the distribution of staff for the evaluation of the descriptor Modality (vid.7.9), obtained for each sequence.

➢ *Quantity Information*

- *Weak Moderate Strong*

Vid. 7.6. Average levels obtained for the "Quantity of information" descriptor (Quant. info) of the Semantics category for each test sequence characterized by the naive panel.

➢ *Understanding*

Dance-1,Dance-2,Dance-3,Dance-4,Opera-1Opera-2,Opera-3Opera-4,Theater-1,Theater-2,Theater-3,Theater-4,Doc.-1,Doc .-2,Doc.-3,Doc.-4,Sport,1,Sport-2,Sport-3, Sport-4

- *Weak Moderate Strong*

Vid. 7.8. Average levels obtained for the "Content Dynamics" descriptor of the Semantics category for each test sequence characterized by the naive panel.

Vid. 7.9. Distribution of numbers according to the sequence for the annotation of the "Modality" descriptor of the Semantics category.

Video 7.9 indicates that all of the sequences of Sport content and the majority of sequences of Theater content were very largely considered with a dominance of the video modality.

Conversely, all of the Opera content sequences and the majority of the Documentary content sequences were characterized by a dominant audio modality. It is interesting to note that most of the predominantly video or audiovisual sequences were accompanied by dynamics mainly considered moderate or even strong. Conversely, none of the predominantly audio sequences (Opera, Doc.-2, Doc.-3 and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

Doc.-4) corresponded to a dynamic considered "strong". This relationship between Dynamics and Modality is confirmed by a Pearson chi-square of independence indicating a significant association between the two variables: $\chi 2 = 75.45$, $df = 4$, $p < 0.001$.

D. Experiment B: Characterization and Influence of Content

The associated contingency video can be found in Appendix 7-F. Thus, the notion of "Dynamic" is reserved for video content, at least in the test corpus used here.

Furthermore, Videos 7.6 and 7.7 indicate that the average levels of Quantity of information and Comprehension reported tend to evolve in the same way as the dynamic scores and the evaluations of the hedonic experience, particularly regarding the sequences of content. Dance and Opera.

➢ *Quality Assessment*

Despite the absence of degradation linked to the transmission and restitution conditions of the audiovisual stream (presentation of the sequences in full HD, 1080p quality), the participants perceived differences in A, V and AV quality between the sequences (see fig. 7.10 below, see Annex 7-E for significant effects). These differences in quality, which can be up to 7 points apart for a given individual, can be attributed to the fact that the participants tried to distribute their judgments across the entire proposed scale. However, quality scores could express perceived quality differences, linked to technical, semantic and/or hedonic differences. Post hoc analyzes were performed using Tukey's HSD tests.

- MOSAV
- MOSV
- MOSA

Dance-1,Dance-2,Dance-3,Dance-4,Opera-1,Opera-2,Opera-3,Opera-4Theater-1,Theater-2,Theater-3,Tea,,re-4

Vid. 7.10. MOSAV, V and A obtained for each test sequence.

More precisely, Dance-1 was judged to be of significantly lower quality (AV and V) than the Dance-4 sequence ($p < 0.001$ for QAV and $p < 0.005$ for QV). Similarly, Sport-4 received significantly lower QAV and QOL scores than Sport-2 ($p < 0.001$ for QAV and $p < 0.001$ for QOL). These differences could be explained by notable differences in technical, semantic and/or hedonic categories within the same content.

E. Experiment B: Characterization and Influence of Content

The level of dynamics could help explain the differences in perceived quality observed. Indeed, Dance-1 like Sport-4 stood out from the other sequences of their original content by their levels of dynamics. Remember that the technical descriptors (brightness, color, camera dynamics and detail) did not vary between the sequences selected for a given content. Sport-4 corresponded to the only sequence of the entire test corpus, combining a high level of Camera Dynamics (expert characterization) and Content Dynamics. This sequence therefore offered the participant highly dynamic content, that is to say informationally rich (large amount of visual information). It is likely that the accumulation of dynamics (camera and content) is the cause of a reduction in video quality (considered dominant for these sequences), the audio not presenting a significant difference with the scores obtained for the sequences. other content sequences. Nevertheless, ratings of hedonic descriptors were high for Sport-4. This sequence indeed obtained the highest averages, all sequences combined, for the descriptors Interest (after the Dance-2 sequence), Valence and Arousal reflecting a hedonic experience that could be described as strongly positive. Conversely, Dance-1 was characterized by a video modality and a low level of content dynamics. It was also qualified by low or moderate levels for all of the high-level descriptors Interest (moderate), Pleasure (3), Arousal (5), Amount of information (low) and Understanding (low). Dance-1 therefore corresponds to a sequence that is not very dynamic (especially compared to the other sequences of the content, which are highly dynamic), poor in both auditory and visual information and the origin of a rather negative spectator experience. It seems that the more the dynamic increases, the more the hedonic experience is positive and vice versa. The link between dynamics and "hedonic" experience tends to be confirmed by the results obtained for all the Opera content sequences. Indeed, the latter were all characterized by a low level of Content Dynamics (see vid. 7.8 above) and judged by the participants with a low level of interest, arousal and valence (fig. 7.5) . Thus, the perception of the level of dynamics tends to explain the differences in quality perceived between the sequences of Dance content and content.

Sport. In the case of Danse-1, the absence of dynamics would result in lower quality scores. Conversely, too high dynamics (cumulative camera and content dynamics) as is the case for Sport-4, would reduce the video and audiovisual quality levels as reflected by the quality scores.

Furthermore, a low level of dynamism was associated with low levels of interest, pleasure and arousal (negative hedonic experience). Conversely, high dynamics were responsible for a positive hedonic experience. Thus the notion of "Dynamic" would strongly participate in the "hedonic" experience of the spectator.

F. Experiment B: Characterization and Influence of Content

➢ *Quality of Experience*

The descriptors of the Hedonic, Semantic and Technical categories evaluated in this study constitute factors potentially capable of influencing the participant's quality of experience. In order to study the descriptors determining a positive experience, a simultaneous multiple regression analysis was conducted from the naive characterization obtained (namely the means obtained for each descriptor and each sequence) by considering the dependent variable "Valence" and the explanatory variables: "Interest", "Arousal", "Comprehension", "Dynamic", "Quantity of information", "Brightness" and "Quality" (QAV, QV, QA)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

(the "Modality" and "Color" descriptors could not be integrated due to their nominal nature). The analysis revealed the participation of three descriptors: Interest, Understanding and Dynamic. The valence of the experience can be modeled, within the framework of this experiment, by the following equation (R=0.99, R2 = 0.98): Valence of the experience = 1.37×Interest + 0, 56×Comprehension + 0.40×Dynamic - 0.76 This result indicates that increasing levels of interest, understanding and dynamics contribute to the positive valence of the viewer's experience. The quality of the viewer's experience, considered from the angle of its valence, therefore seems to be able to be predicted by a subset of indicators.

- *Final Characterization of the Sequences*

The previous paragraphs presented the naive characterization of the sequences (appendix 7-D). However, the description of content using low-level descriptors will always be carried out using expert characterization. Table 7.5 below presents the description obtained for each of the twenty sequences of experiment B1, namely the high-level descriptors of the Hedonic and Semantic categories (naïve characterization) and the low-level descriptors of the Semantic and Technical categories (expert characterization).

➢ *Experiment B: Characterization and Influence of Content*

Video7.5. video presenting the characterization of the sequences carried out by the expert and by the panel of "naive" spectators.

C. Expert C. Naïve

Seq. Mod R.AV EX.S Nb P D.Ct Lum Det D.Cm Col Int Val Ar Info Comp

Dance-1V In MuS F Fa Fa M Fa CM 3 5 Fa Fa
Dance-2V In MuS FF Fa M Fa CF 7 7 MF
Dance-3V In MuS FF Fa M Fa CF 7 6 MF
Dance-4V In MuS MF Fa M Fa CF 7 7 MF
Opera-1A HC PF Fa MM Fa J Fa 3 2 Fa Fa
Opera-2A In PF Fa MM Fa J Fa 4 4 MM
Opera-3A In PF Fa MM Fa J Fa 3 3 Fa M
Opera-4A In P Fa Fa MM Fa J Fa 4 3 Fa Fa
Theater-1A In P Fa Fa Fa M Fa FM 6 5 MM
Theater-2V In B Fa Fa Fa M Fa FM 5 5 MF
Theater-3V Off MuS MM Fa M Fa FM 5 3 Fa M
Theater-4AV In BMM Fa M Fa FM 5 5 MM
Doc-1V Off MuS Fa M Fa M Fa JM 3 5 Fa Fa
Doc-2A Off P Fa M Fa M Fa JM 5 3 MF
Doc-3A In P Fa Fa Fa M Fa JM 5 5 MF
Doc-4A Off P Fa Fa Fa M Fa JM 5 5 MF
Sports-1AV Off PF Fa FFFJF 7 5 FF
Sport-2A Off PF Fa FFFJF 7 6 MF
Sports-3V Off PFMFFFJF 6 5 FF Sport-4

Characterization of sequences (Séq.) carried out by the expert (C.Expert) using low-level descriptors Semantics: Modality (Mod), AV Relationship (R.AV), Sound Expression (EX.S), Number of Characters (Nb P), Content Dynamics (D.Ct) and Technique: Brightness (Lum), Detail (Det), Camera Dynamics (D.Cm) and Color (Col); according to the levels Low (Fa), Moderate (M), Strong (F) or Audio (A), Video (V), AudioVisual (AV) or Hot (C), Day (J), Cold or Music (Mus), Speech (P), Noise (B) and by the panel of "naïve" spectators (C.Naïve) according to the calculation of the mode carried out for each sequence and each high-level descriptor of the Hedonic categories: Interest (Int) Valence (Val) and Arousal (Ar) and Semantics: Quantity of information (Info) and Comprehension (Comp).

VI. CONCLUSIONS

One of the main objectives of experiment B was to characterize the test contents with the intention of better understanding the relationship between content and perceived quality. This characterization was carried out by an expert and made it possible to have a set of contents for which the low-level technical and semantic characteristics are known. In the following studies, each sequence extracted from one of these five contents can therefore be defined on the basis of these descriptors.

Secondly, the relevance of a subset of descriptors (considered to be the most subjective) used by the expert was verified with naïve participants.

A. Experiment B: Characterization and Influence of Content

The concordance between the characterization of the expert and that of the spectators for the Modality and Content Dynamics descriptors confirms the relevance of these criteria. The luminosity descriptor met with weaker agreement between experts and naive people, mainly concerning the qualification of the sequences by the "low" level. Despite the disagreement observed for this level, the term still seems to have been correctly assimilated by the naive. On the other hand, the terms used to qualify color temperature seem reserved for the technical field of audiovisual and not very accessible to a panel of non-expert participants. This descriptor, which is difficult to understand for non-expert participants, will not be integrated into a future evaluation questionnaire. This study also made it possible to have a corpus of short sequences (8 to 10 s) characterized not only by low-level technical and semantic criteria but also by high-level hedonic and semantic criteria. Overall, the results showed that the participants were able to describe the sequences visualized using the proposed descriptors. Each descriptor seems to provide relevant information on technical, semantic or hedonic levels to describe the audiovisual content viewed. The participants' evaluations of the descriptors were therefore influenced by the sequence but also, more broadly, by the content. For example, Opera aroused little interest as well as negative valence and low arousal in the participant. This could be explained by the fact that the chosen sequences were extracted from a more general context, that is to say that they were separated from a global framework carrying meaning. These snippets of events could have cut the participant off from the general meaning of the content. This is all the more true as the Opera content sequences present a foreign language context (the meaning is therefore all the more difficult to extract, limiting the understanding of the extract) for sequences whose dominant modality has been judged to be as being audio. Thus, in order to avoid this

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

disinterest, it would be appropriate to place the participant in the overall narrative framework of the content, for example by offering the reading of synopses before evaluating the test sequences. The results also highlighted a link between the Dynamic and Dominant Modality descriptors. Indeed, dynamics have been associated, within the framework of the corpus of test sequences proposed here, with the video modality. Specifically, a predominance of audio would be mainly associated with low dynamics and a predominance of video would be mainly assimilated to moderate or even strong dynamics. This observation is in line with that made by Hands (2004) assuming that audio quality would be dominant for a weakly dynamic AV context while video quality would be dominant for a highly dynamic AV context. Despite the absence of degradations, differences in ratings could be observed regarding perceived quality. It would seem that the Dynamic descriptor is a good candidate to explain these differences.

B. Experiment B: Characterization and Influence of Content

Indeed, the not very dynamic sequences (Dance-1) and a fortiori audio (Opera content sequences), providing little information to the viewer and causing a negative hedonic experience, were characterized by low levels of perceived quality. Conversely, a highly dynamic sequence and even more so video, providing a lot of information to spectators

(Sport-4) and at the origin of a strongly positive experience (from the point of view of interest, pleasure and level of arousal), resulted in an altered level of perceived quality. This result also indicates that the participants were able, in a context of non-degraded quality, to judge the quality of the audio and/or video signals independently of the positive or negative trend of their experiences.

Finally, this study also made it possible to better understand the link between the content (studied from the descriptors evaluated) and the quality of experience considered from the angle of its valence. Indeed, a subset of descriptors (interest, understanding and dynamics) would make it possible to predict the valence of the experience. These descriptors could be taken into account in the context of methods relating to the evaluation of the quality of the restored audio and video signals.

REFERENCES

[1]. Kohlrausch, A. and van de Par, S. (2005). Audio-visual interaction in the context of multimedia applications. In J. Blauert (ed.), Communication acoustics (pp. 109-138). Berlin, Germany: Springer-Verlag.

[2]. 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.

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

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

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

[6]. Knoche, H., De Meer, HG and Kirsh, D. (1999). Utility curves: Mean opinion scores considered biased. In Proceedings of the 7th International Workshop on the Quality of Service (IWQoS), 12-14.

[7]. Koch, C. (2004). The Quest for Consciousness: A Neurobiological Approach. Englewood, CO: Robert & Company Publishers.

[8]. Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In L. Damos (ed.), Multiple-task performance (pp. 279-328). London, UK: Taylor & Francis.

[9]. Applied Anthropology Laboratory (1996). Establishment of a method for studying pilot fatigue in air transport, phase 1 (Report AA 358/96). Paris, France. Retrieved fromhttp://www.developpementdurable. gouv.fr/IMG/pdf/fatigue1.pdf cited by doctor Julie LASSALLE

[10]. Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In M. Appley and R. Trumbull (Eds.), Psychological stress: Issues in research (pp. 14-42). New York, NY: Appleton century crofts.

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

[12]. Lacey, J.I. and Lacey, B.C. (1970). Some autonomic-central nervous system interrelationships. In P. Black (Ed.), Physiological correlates of emotion (pp. 205-227). New York, NY: Academic Press.

[13]. Lacombe, M. (2009). Lacombe: summary of human anatomy and physiology (vol. 2, 30th ed.). Rueil-Malmaison, France: Lamarre.

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

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

[16]. Lang, A. (1991). Emotion, formal features, and memory for televised political advertisements. *Television and political advertising, 1*, 221-243.

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

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24FEB816

[18]. Lang, A. and Basil, M. (1998). Attention, resource allocation, and communication research: What do secondary task reaction times measure, anyway? In M. Roloff (ed.), Mass Communication yearbook (vol. 21, pp. 443-473). Beverly Hills, CA: Sage.
[19]. Lang, A., Bolls, P., Potter, R. F., & 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.
[20]. Lang, A., Newhagen, J. and Reeves, B. (1996). Negative video as structure: Emotion, attention, capacity, and memory. Journal of Broadcasting and Electronic Media, 40(4), 460-477.
[21]. Lang, A., Zhou, S., Schwartz, N., Bolls, PD and Potter, RF (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.
[22]. Lang, L. and Qi, H. (2008). The study of driver fatigue monitor algorithm combined PERCLOS and AECS. In Proceedings of the International Conference on Computer Science and Software Engineering, 1, 349-352.
[23]. Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In JB Sidowski, JH Johnson, & GA Williams (Eds.), Technology in mental health care delivery systems (pp. 119–137). Norwood, NJ: Ablex.