Impact of the Content on Subjective Evaluation of Audiovisual Quality: What dimensions influence our perception?

Julie Lassalle, Laetitia Gros, Thierry Morineau, Gilles Coppin

Impact of the Content on Subjective Evaluation of Audiovisual Quality: What dimensions influence our perception?

Julie Lassalle, *Student Member, IEEE*, Laetitia Gros, Thierry Morineau and Gilles Coppin

***Abstract*—Several studies have shown a test material dependency on perceived quality. Consequently, methods for evaluating AV quality should take account of influence of contents corpus (regarding influence of audio, video and relationship between audio and video) to avoid uncontrolled effects. In this paper, technical, semantic and hedonic dimensions (including different descriptors) were used to characterize a corpus of AV contents. Results have shown significant effects of modality, dynamic and interest descriptors on the viewer AV quality perception. It also seems that verbal sequences and diegesis interact with asynchrony perception. As a result, these descriptors should be considered when selecting test sequences.**

***Index Terms*— Audiovisual Quality, Content-dependency, Evaluation, Perception**

I. Context

DUE to the increasing number of high quality audiovisual technologies and services (HD and 3D images, spatial sound) and the variety of uses such as residential (TV, PC), embedded (mobile) and communicative (tele-conferences), it is necessary to assess how the overall quality of these services is perceived by the user, as the QoE (Quality of Experience) is a key aspect in a competitive market.

Currently, assessing the perceived quality of audio and video signals reproduced/displayed by an audiovisual (AV) service is often done by considering the two media separately. But, the overall perceived quality of an audiovisual signal is the result of interaction between qualities of each medium [1], [2]. In addition, it is highly likely that semantic properties of the content impact the perceived overall quality, not only semantic properties of *individual* audio (A) and video (V) signals but also the semantic relationship between audio and video (diegesis [sound-in/off or off-screen], dominant modality, *etc*.). For example, [2]-[4] showed that perceived quality is content-dependent. Particularly, there is a dominant modality impact e.g. video quality dominates AV perceived quality for a commercial clip and the reverse is observed for a "head and shoulder" content. Thus, modality influence seems to refer to a semantic of content notion. For example, user attention could shift to A or V modality depending on the modality conveying the main information. In consequence, AV perceived quality could be dependent on semantic information conveyed either by video (also including dynamic, motion camera criteria) or by audio (i.e. speech). These semantic properties could also interact with usual transmission conditions. Works of [5], [6] reinforce this assumption by showing that asynchrony sensitivity presents a test material dependency depending on whether a clip is a non-verbal or a verbal sequence. In addition, asynchrony between sound and image may be more or less perceived according to the diegesis of the content.

It is also likely that the user hedonic experience (pleasure/displeasure, interest, *etc.*) related to the content may impact the quality evaluation. It seems important to characterize test sequences, not only with technical criteria but also in terms of semantics (objective and subjective i.e. as perceived by viewer) indeed hedonism to understand how the content influences the conscious perception of the audiovisual quality.

But, today, there are only two methods standardized by the International Telecommunications Union which consider the audiovisual signal as a whole (ITU-T P.911 [7], P.920 [8]). However, in a non-interactive context (P.911), only an overall quality judgment is asked to participants, therefore it is not possible to understand the interaction between audio quality and video quality. In addition, P.911 methods supply some recommendations for the characterization of test sequences regarding audio and video content separately. But it does not provide guidelines regarding the characterization of the audiovisual content in terms of semantic relationship between sound and image (sound-in/off or off-screen for example).

As methods for evaluating AV perceived quality should take AV content effect into account, the objective of this paper is to draw up a list of audio, video and audiovisual descriptors to consider when selecting test sequences. It could avoid uncontrolled content impact on the audiovisual quality evaluation. Section 2 described used descriptors for content characterization. Validation of these descriptors is presented in section 3. Section 4 exposes interactions between descriptors and perceived AV quality. Section 5 discusses the results and concludes with perspectives.

II. Content-based descriptors

"References [9] and [10] define several objective descriptors to describe an AV segment." MPEG-7 standard provides audio and video information and distinguishes two abstraction levels: low-level descriptors such as color, texture, motion, melody or speech and high-level with objective semantic concepts (events, objects, localization, *etc*.). Another

Manuscript received November 30, 2011.

J. Lassalle is with Orange Labs, Lannion, France and with UBS, Vannes, France and Telecom Bretagne, France
L. Gros is with Orange Labs, Lannion, France
T. Morineau is with the CRPCC Department, UBS, Vannes, France
G. Coppin is with Lussi department, Telecom Bretagne, Brest, France and with Lab-STICC, France

approach with different elements for film analysis had been proposed by [10]. It especially concerns component of the plan (angle of view, camera motion, framing, depth of field, *etc*.), narrative script (outdoor/indoor, day/night, visual/dialogues, action-tension/ inaction-immobility, intimate/collective/public, number of characters), audio characteristics (speech, sound effects, music) or again image and sound relationship description (diegetic sound i.e. sound-in or off-screen and non-diegetic sound i.e. sound-off).

A list of 32 technical and objective semantic criteria was established on the basis of [9], [10] and proposals of an audiovisual expert. These criteria could constitute an objective content description with firstly, video low-level descriptors (camera motion, brightness, details -texture-, temperature color, framing and angle of view) and secondly, elementary semantic descriptors:

--For audio: speech, sound effects, music

--For video: narrative composition (indoor/outdoor, day/night, intimate/collective, action nature, numbers of characters)

--For audiovisual: dominant modality and sequence's AV diegesis

This objective description material was subjected and validated by an audiovisual expert (professional audiovisual technician). This list of descriptors was used to characterize five 2D/stereo contents from 10 to 14 minutes: opera (*Don Giovanni* extract (*DG*)), theater (*The Tricks of Scapin* extract (*ToS*)), ballet (*Bale de rua* extract (*BdR*), sport (final of *Roland Garros* 2011 extract (*RG*)) and documentary (report on the French boxer *Mormeck* (*MM*)). These contents were chosen to provide various AV realistic contexts. Each content was cut into different time-sequences for analysis process. Each of these time-sequences constitutes a meaningful sequence i.e. a series of scenes which do not take place necessarily in the same film set but which form a whole with a proper sense, as defined by [11].

This expert characterization step has led to the emergence of nine descriptors given in table 1, defining *technical*, *semantic* (low-level and high-level) and *hedonic* dimensions.

TABLE 1
FINAL DESCRIPTION MATERIAL

Descriptors	Modes	Dim.	Level	Used Ex.
Brightness	low, moderate, high	Tech	Low	X
Color temperature	hot, moderate, cool	Tech	Low	X
Details	low, moderate, high	Tech	Low	X
Camera dynamic	low, moderate, high	Tech	Low	X
AV diegesis	Sound-in/off, off screen	Sem.	Low	X
Sound type	speech, music, sound effects	Sem.	Low	X
Content dynamic	low, moderate, high	Sem.	Low	X
Dominant modality	A, V or AV	Sem.	Low	X
Nb of characters	low, moderate, high	Sem.	Low	X
Comprehension	low, moderate, high	Sem.	High	
Quantity of information	low, moderate, high	Sem.	High	
Interest	low, moderate, high	Hed.	High	
Valence	9-points scale	Hed.	High	
Arousal	9-points scale	Hed.	high	

List of chosen descriptors and their modes according to their technical (Tech), semantic (Sem.) or hedonic (Hed) dimension and their abstraction level (low or high) belonging. Last column indicates descriptors used for expert characterization.

III. EXPERIMENT 1

A. Objectives

The objectives of this first experiment were to validate low-level expert characterization, to check the relevance of the descriptors and also to obtain a corpus of audiovisual sequences representing these different descriptors. The totality of objective criteria (low-level description material) was not subjected to naive annotation because of their unchanging nature as for example *level of details* (texture), *motion camera*, *AV relationship* or *sound type* descriptors. Only two low-level semantic and two technical descriptors (*dominant modality, color, brightness* and *content dynamic*) considered as potentially variable according to individual perception, were evaluated by both expert and naïve participants (naïve low-level material).

B. Stimuli

For this experiment, four sequences, from 8 to 10-s, were taken from each of five AV contents characterized by expert. Sequence 1 to 4 were extracted from *BdR* content, 5 to 8 from *DG*, 9 to 12 from *ToS*, 13 to 16 from *MM* and 17 to 20 from *RG*. Table 2 shows sequences characterization according to low-level descriptors, its gives the characteristics of each sequence (Seq) extracted from the five contents (Ct), according to the low-level semantic descriptors: *Modality* (Mod), *Diegesis* (Dieg, with In for sound-in, Off for sound-off and Off-S for off-screen sound), *Sound type* (Snd type, with MuS for music, Sp. for speech and Sd-E for sound effects), *number of characters* (Ch. nb, with L for "low," M for "moderate" and H for "high" level), *Content dynamic* (Ct *D,* see Ch. nb for abbreviations), *Brightness* (Bs, see Ch. nb for abbreviations), *camera dynamic* (Cam *D,* see Ch. nb for abbreviations*), details* (Det, see Ch. nb abbreviations), and *Color* (Col, with H for hot, M for moderate and C for cool). For sequences extracted from the same content, technical criteria did not vary.

TABLE 2
SEQUENCES LOW-LEVEL SEMANTIC CHARACTERISATION

Seq	Ct	Mod	Dieg	Snd type	Ch. *nb*	Ct *D*	Bs	Det	Cam *D*	Col
1	BdR	V	In	MuS	H	L	L	M	L	H
2	BdR	V	In	MuS	H	H	L	M	L	H
3	BdR	V	In	MuS	H	H	L	M	L	H
4	BdR	V	In	MuS	M	H	L	M	L	H
5	DG	A	Off-S	Sp.	H	L	M	M	L	M

6	DG	A	In	Sp.	H	L	M	M	L	M
7	DG	A	In	Sp.	H	L	M	M	L	M
8	DG	A	In	Sp.	L	L	M	M	L	M
9	ToS	A	In	Sp.	L	L	L	M	L	C
10	ToS	V	In	Sd-E	L	L	L	M	L	C
11	ToS	V	Off	MuS	M	M	L	M	L	C
12	ToS	AV	In	Sp.	M	M	L	M	L	C
13	MM	V	Off	MuS	L	M	L	M	L	M
14	MM	A	Off	Sp.	L	M	L	M	L	M
15	MM	A	In	Sp.	L	L	L	M	L	M
16	MM	A	Off	Sp.	L	L	L	M	L	M
17	RG	aV	Off	Sp.	H	L	H	H	H	M
18	RG	A	Off	Sp.	H	L	H	H	H	M
19	RG	V	Off	Sp.	H	M	H	H	H	M
20	RG	V	Off	Sp.	H	H	H	H	H	M

Table 2: characteristics of the twenty chosen sequence (Seq) extracted from the five contents (Ct), according to Modality (Mod), Diegesis (Dieg) Sound type (Snd type), number of characters (Ch. *nb*), Content dynamic (Ct *D*), Brightness (Bs), Color (Col), Camera dynamic (Cam *D*) and details (Det)

C. Method

The twenty AV sequences were presented in random order to 28 non expert participants. They were asked to evaluate each sequence according to 12 criteria given in table 3. In addition to technical, semantic and hedonic descriptors, A, V and AV quality judgments had been asked in order to have some information on intrinsic quality of sequences. Visualization conditions complied with the recommendations proposed by the ITU-T P.911.

TABLE 3
EVALUATION CRITERIA

Abbrev.	Descriptors	Used Scales
AVQ, VQ, AQ	A, V, AV quality	9-poins likert scales (5 items) *excellent-good-fair-poor-bad*
Int.	Interest	3-points scales *low-moderate-high*
Compr.	Comprehension	3-points scales *low-moderate-high*
Quant.Inf	Quantity of information	3-points scales *low-moderate-high*
Mod.	Dominant modality	3-points scales *A-V-AV*
Dyn.	Content Dynamic	3-points scales *low-moderate-high*
Col.	Color	3-points scales *hot-moderate-cool*
Lum.	Luminosity	3-points scales *low-moderate-high*
Val.	Valence	Self-Assessment Manikin-Scales [12]
Ar.	Arousal	Self-Assessment Manikin-Scales [12]

D. Results

Expert vs. naïve approach

Results showed an agreement between expert and naïve annotation: *modality* (X^2 (4) = 17.33, $p < 0.05$), *color* (X^2 (4) = 10.91, $p < 0.05$), *dynamic* (X^2 (4) = 22.08, $p < 0.001$), and *luminosity* (X^2 (4) = 17.75, $p < 0.05$). That is, when a sequence was tagged by expert as video in terms of dominant modality, viewers have significantly responded similarly. Thus expert characterization allows the naïve one to be predicted. This result validates the expert qualification and, consequently, could be generalized to the entire corpus of audiovisual content. Hence, expert characterization of the AV contents corpus meets an ecological validity.

General effects

A series of Analysis of Variance (ANOVA) was performed with 95%-confidence intervals on individual "Quality" scores (AQ, VQ and AVQ) and on scores corresponding to the different descriptors (see Table 3) considering fixed independent variable "Sequence."

Firstly, results revealed that all semantic, technical and hedonic descriptors significantly depend on the sequence and more generally on the content (for example, *DG* content has been largely considered as the least interesting, pleasant and comprehensive whereas *BdR* and *RG* received high scores for these same descriptors. Mitigated scores were attributed to *ToS* and *MM*). However, sequences from a same content could be judged differently according to some descriptors (for example, RG sequences significantly vary according to *dynamic*). Fig. 1 presents content characterization on the basis of all descriptors.

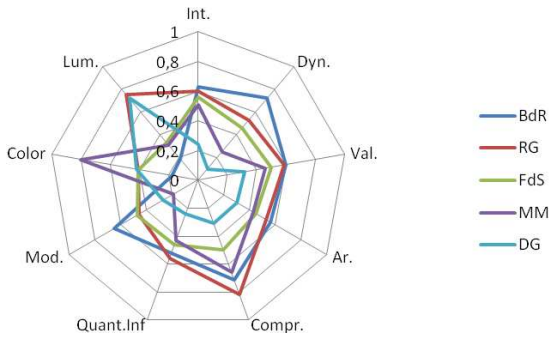

Fig. 1. Result of the naïve characterization for each content (mean obtained for each annotated descriptors). For comparison, data were normalized between 0 and 1. Values close to 1 represent descriptors "high" level, values close to 0 represent descriptors "low" levels. *Color* descriptor must read as follow: value close to 0=hot, 0.5=moderate, 1=cool and *modality* descriptor 0=A, 0.5=V and 1=AV

Secondly, results revealed a significant influence of sequences on the AVQ scores, $F(19, 513)=4.42$, $p<0.001$, VQ scores, $F(19, 513)=5.76$, $p<0.001$ and AQ scores, $F(19, 513)=1.82$, $p<0.05$, as illustrated by Fig. 2.

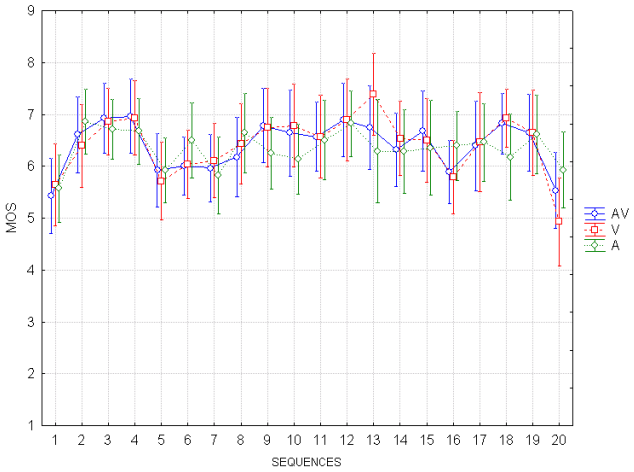

Fig. 2. Impact of "Sequence" condition on the mean opinion scores (MOS) for AV, A and V qualities with 95% confidence interval

Despite the absence of transmission degradations (network degradations, coding-decoding, compression), participants perceived a difference in quality between sequences (up to 7-points gap for a given participant). These perceived differences may be attributed to the fact that participants try to use the whole of the scales range. Nevertheless, quality scores reflect intrinsic quality differences perceived between sequences. These quality differences could be explained by technical, semantic and/or hedonic differences (interest, understanding, dynamic, *etc.*). For example, S1 has been judged significantly lower than S4 (for VQ and AVQ). Regarding low-level descriptors, only *dynamic* ("low" for S1 and "high" for S4) and *number of characters* ("high" for S1, "moderate" for S4) differed between these two sequences. For high-level descriptors, S1 received very low scores regarding *interest*, *pleasure* and *arousal* descriptors whereas S4 received high scores for the same dimension. In brief, S1 represented a poor informative sequence (both on audio and video media) probably at the origin of a negative experience for participants.

Another example is the significant difference between S20 and S18 (for VQ and AVQ) of RG content. Distinction between these sequences is based on *dynamic* ("high" for S20, "low" for S18) and *modality* ("V" for S20, "A" for S18). Additionally, S20 was the only sequence, of the whole test corpus, with technical (i.e. camera motion) and semantic dynamic (i.e. viewer perceived dynamic) annotated with a "high" level simultaneously. S20 could be considered as a too wealthy informative sequence (A and V) in terms of technical and low-level semantic dimensions. It could be interesting to note that S20 received significant highest scores for *dynamic* and *arousal* (highest scores of annotated corpus) than S18. In this way, quality perception differences, attributable to technical, semantic or hedonic factors, could exist inside a same content.

In other words, these preliminary results highlighted that *dynamic, modality* or *number of characters* descriptors interacts with our perception of quality. These findings could lead to discussion about possible interactions between annotated dimensions and perceived AV quality, in a degraded context.

IV. Experiment 2

A. Objectives

The objective of this experiment was to study the potential impacts of low-level semantic and technical dimensions (low-level description material) on the overall perceived AV quality, in interaction with various degradations. This step should also allow a catalog of interactions between degradation and some semantic and/or technical descriptors (generalization of degradations effects on individual perception according to some of studied criteria).

B. Method

35 participants viewed 20 audiovisual sequences (corpus of experiment 1). Each of these sequences was presented under 10 different quality conditions. Degradations were chosen on the one hand, to represent current impairments of AV quality (i.e. could really occur when viewing audiovisual content) and on the other hand, as being strong enough to be perceptible. Degradations were:

--**Reference** (no degradation): uncompressed .avi with audio at 48 Khz, 16 bit

--(**AsynC**.): asynchrony between audio and video where audio was presented with a 1500ms delay

--(**AbV**): audio bitrates variations with AC3 codec at 64Kbps/8Khz

--(**VbV**): video bitrates variations with MPEG4-AVC codec between 93 and 1600 Kbps (depending on sequence detail level)

--(**VfrZ**): five periods of 18 frozen frames (random position)

--(**ApL**): 10% audio packets loss

--A and V degradations combination: **(ApL+VfrZ); (ApL+VbV); (AbV+VbV); (AbV+VfrZ)**

In all, each participant viewed and listened to 200 sequences, presented in a different random order for each participant. In compliance with P.911 standard, participants were asked to rate only AV quality with a five items (excellent-good-fair-poor-bad) 9-points scale after each visualized sequence. Participants had a few seconds break, between each sequences, for scale completion. As for the previous experiment, conditions of visualization respected P.911 recommendations.

C. Results

An ANOVA was performed on "Quality" (i.e. Mean Opinion Scores of audiovisual evaluation –MOSav-) with

"Sequence" and "Degradation" as independent variables.

Results showed "Degradation" ($F(9, 297)=470.77$, $p<0.001$) and "Sequence" ($F(19, 627)=15.91$, $p<0.001$) impacts on MOSav. We also observed a significant interaction of "Sequence" and "Degradation" factors on MOSav ($F(171, 5643)=12.25$, $p<0.001$). This interaction is illustrated by Fig. 3 which gives MOSav for sequences 9, 10, 13, 14 according to degradations. Discomfort for audiovisual asynchrony seems stronger for verbal sequence (S9, tagged with "speech" mode) compared to sequences 13 (tagged "music") and 10 (tagged "sound effects"). Sequence 14 even tagged "speech" seems to be less impacted by asynchrony. Indeed, it had been also tagged as "sound-off," so asynchrony can not be perceived

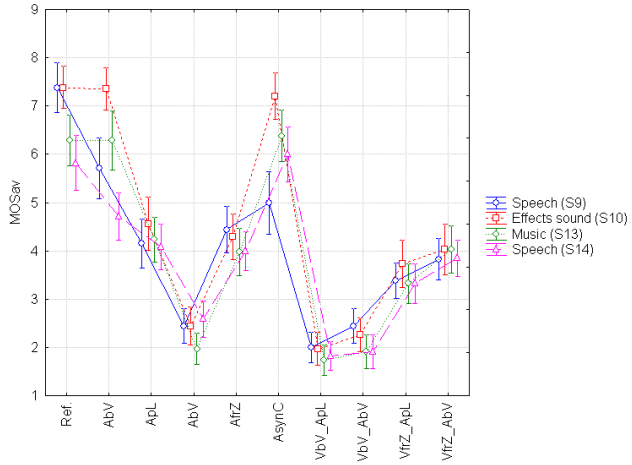

Fig. 3. Illustration of the influence of *Sound Type* descriptor (i.e. speech, music or sound effects) in interaction with "Degradation" variable on MOSav with 95% confidence interval

(diegesis effect).

The results also revealed patterns in perception of some degradation regardless of the type of content. For example, video bitrates degradation has always been evaluated as being more annoying than breaks in the signal continuity (frozen images), regardless of content. For audio, on the contrary, the continuity breaks were perceived as more troublesome than the bitrates degradation.

This result highlighted the fact that perceived degradations are content-dependent. In order to study thoroughly this effect, previous experiment was used to characterize sequences, according to low and high-levels descriptors (see Table 1).

For technical descriptors (brightness, color temperature, details level and camera dynamic), and low-level semantic descriptors (AV diegesis, sound type, content dynamic, dominant modality and number of characters), we considered the value given by the expert. For the other descriptors (high-level semantic descriptor -comprehension and quantity of information- and hedonic descriptors -interest, valence and arousal-), we considered the more frequent value (mode) given by naïve participants.

Then a series of ANOVAs was conducted on MOSav considering each descriptor as an independent factor and in interaction with the factor "Degradation."

Results showed a significant main effect of the descriptor only for *dynamic, modality* and *interest*, respectively $F(2, 170) = 5.92$, $p<0.05$, $F(2, 170) = 4.93$, $p<0.005$ and $F(2, 170) = 4.52$, $p<0.05$. Fig. 4 shows an increase of MOSav for "strong" dynamic sequences compared to "low" and "moderate" dynamic sequences (*HSD Tukey* test).

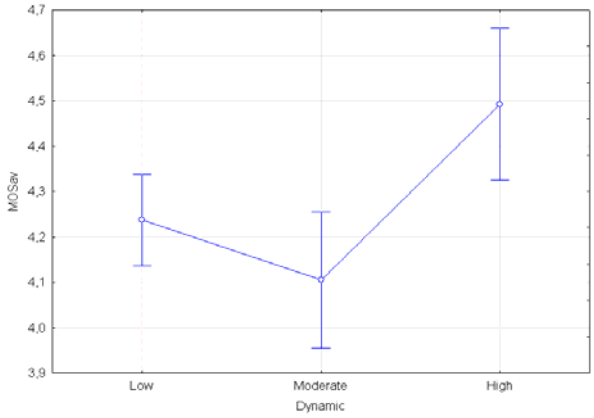

Fig. 4. Impact of *dynamic* on MOSav with 95% confidence interval

The same trend is observed for sequences with dominant video compared to sequences with dominant audio (*HSD Tukey* test) see Fig. 5.

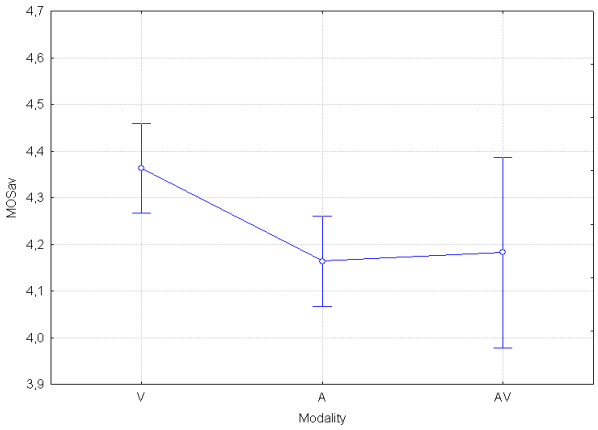

Fig. 5. Impact of *modality* on MOSav with 95% confidence interval

However, the majority of sequences with dominant video was accompanied by a strong or moderate judged dynamic. In return, none of the dominant audio sequences did correspond to a "strong" dynamic. We can assume that, either, there is no sequences presenting a character both audio and dynamic, or dynamic is always associated to video.

In addition to significant main effects, there are significant interactions between these three descriptors and the factor "Degradation." Regarding *modality*, the audio degradations have been rated as more embarrassing for dominant audio sequences, *a fortiori* with low dynamic, than for dominant video sequences as showed by Fig. 6. The same effect was observed for video.

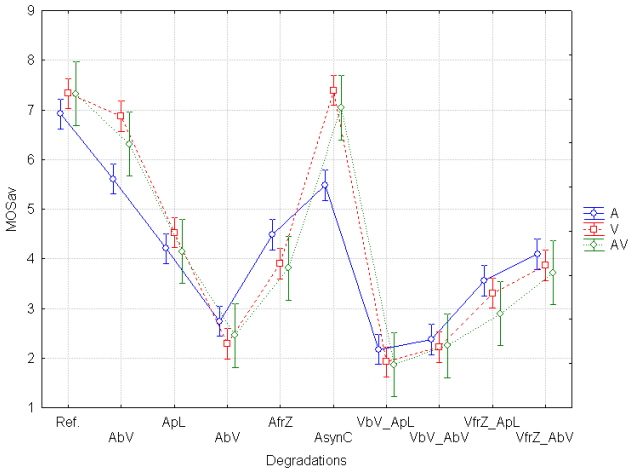

Fig. 6. Impact of *modality* on MOSav in interaction with "Degradation" factor with 95% confidence interval

Notice that *modality*, *interest* and *dynamic* descriptors are strongly linked in presence of degraded quality. This link was confirmed by a Chi^2 analysis which showed dependency between these three descriptors: *modality*/*interest,* X^2 (4) = 12.86, $p < 0.05$, *dynamic*/*interest*, X^2 (4) = 17.46, $p < 0.05$ and *dynamic*/*modality,* X^2 (4) = 17.14, $p < 0.05$.

Dynamic, *modality* and *interest* descriptors significantly influence the quality scores and, therefore, the viewer perception of AV content.

V. Conclusions and Perspectives

To conclude, with experiments 1 and 2, a corpus of short test sequences characterized in terms of technical, semantic and hedonic descriptors has been obtained. This corpus would be available for other quality evaluation framework (e.g. AV quality model study, *etc.*). Moreover, owing to the annotation consistency between the expert and naives, characterization may be used for the entire corpus (long extracts). In addition, experiment 2 reveals different interactions between perceived quality and used descriptors. In effect, it seems that *sound type* (verbal or non verbal sequence) descriptor influence discomfort for audiovisual synchronization. This observation confirms those found by [5], [6]. Beyond the *sound type* influence, *diegesis* should be also considered when selecting test sequences, as well as *dynamic, modality* and *interest.* Indeed, experiment 2 highlights an impact of these two low-levels descriptors (*dynamic, modality*) and the high-level descriptor *interest* on perceived audiovisual quality. Results show that asynchrony is discomfort only for dominant audio sequence. It implies that participants are not perturbed when asynchrony degradation occurs on dominant video or audiovisual modality. This could indicate the necessity to propose a specific question on asynchrony to allow participants to express themselves on this kind of degradation. Indeed, these results do not allow us to know whether asynchrony was perceived or not.

In order to understand interaction between descriptors and audio and video qualities separately evaluated, A and V quality judgment should be added to the single AV evaluation, as recommended by P.920 in an interactive context [8].

Therefore, at least, *dynamic, modality* and *interest* descriptors influence our perception of AV quality. Consequently, these descriptors should be considered in AV quality assessment (for instance, selecting test sequences) because of their non-negligible impact.

Finally, these two experiments are part of a broader quality evaluation study whose objective is to propose a holistic user experience by aggregating subjective, behavioral and physiological indicators [13]. In this context, it was necessary to obtain a test corpus (whose characteristics are known) in order to avoid uncontrolled content effects on the viewer's experience e.g. unwanted impacts of pleasure, interest or dynamic on physiological pattern.

References

[1] J. G. Beerends, and F. E. De Caluwe, "The influence of video quality on perceived audio quality and vice versa," *J. Audio Eng. Soc.*, vol. 47(5), pp. 355-362, May 1999

[2] M. P. Hollier, and R. M. Voelcker, "Towards a multi-modal perceptual model," *BT Technology Journal*, vol. 15(4), pp. 163-172, 1997

[3] ITU-T Contribution COM 12-19-E, *Relation between audio, video and audiovisual qualitys*, KPN, The Netherlands, December 1997

[4] D. H. Hands, "A basic multimedia quality model," *IEEE Trans. Multimedia*, vol.6(6), pp.806-816, December 2004

[5] N. F. Dixon, and L. Spitz, "The diction of auditory visual desynchrony," *Perception*, vol. 9, pp. 719–721, 1980

[6] M. P. Hollier, A. N. Rimell, D.S. Hand, and R.M. Voelcker, "Multi-modal perception," *J. BT Technol.*, vol. 17, pp. 35–46 January 1999

[7] Recommendation ITU P.911, *Subjective audiovisual quality assessment methods for multimedia application*, ITU Telecommunication Standardization Sector, December 1998

[8] Recommendation ITU-T P.920, *Interactive test methods for audiovisual communications*, May 2000

[9] *MPEG-7 Overview V.10* ISO/IEC JTC1/SC29/WG11/M6485, Oct. 2004

[10] Y. Amiar, "L'analyse du film et de l'image fixe : approche méthodologique," *Revue Recherche sur l'information scientifique et technique*, vol.5(2), pp. 23-28,1995

[11] A. Goliot-Lété & F. Vanoye, "Précis d'analyse filmique," Paris: Armand Colin, 2009

[12] P. J. Lang, "Behavioral treatment and bio-behavioral assessment: Computer applications," In J.B. Sidowski, J.H. Johnson, & T.A. Williams (Eds.), *Technology in mental health care delivery*: Norwood, pp. 119-137, 1980

[13] J. Lassalle, L. Gros and G. Coppin, "Combination of physiological and subjective measures to assess quality of experience for audiovisual technologies," *Third international workshop on Quality of multimedia experience*, pp.13-18, Sept. 2011