A QoE Model for Mulsemedia TV in a Smart Home Environment

Lana Jalal, Roberto Puddu, *Student Member, IEEE*, Maria Martini, *Senior Member, IEEE,* Vlad Popescu, *Member, IEEE*, and Maurizio Murroni, *Senior Member*, *IEEE*

***Abstract*—The provision to the users of realistic media contents is one of the main goals of future media services. The sense of reality perceived by the user can be enhanced by adding various sensorial effects to the conventional audio-visual content, through the stimulation of the five senses stimulation (sight, hearing, touch, smell and taste), the so-called multi-sensorial media (mulsemedia). To deliver the additional effects within a smart home (SH) environment, custom devices (e.g., air conditioning, lights) providing opportune smart features, are preferred to ad-hoc devices, often deployed in a specific context such as for example in gaming consoles. In the present study, a prototype for a mulsemedia TV application, implemented in a real smart home scenario, allowed the authors to assess the user's Quality of Experience (QoE) through test measurement campaign. The impact of specific sensory effects (i.e., light, airflow, vibration) on the user experience regarding the enhancement of sense of reality, annoyance, and intensity of the effects was investigated through subjective assessment.**

The need for multi sensorial QoE models is an important challenge for future research in this field, considering the time and cost of subjective quality assessments. Therefore, based on the subjective assessment results, this paper instantiates and validates a parametric QoE model for multi-sensorial TV in a SH scenario which indicates the relationship between the quality of audiovisual contents and user-perceived QoE for sensory effects applications.

Keywords: Quality of Experience, Sensory Effects, Multi sensorial media, Particle Swarm Optimization, smart home.

I. INTRODUCTION

PROVIDING to the users multimedia content together with sensory effects is one of the main features of future media services, called multi-sensorial media [1][2]. The user's viewing experience is enhanced by adding various sensorial effects to traditional media contents, triggering all human senses (i.e., smell, taste, hearing, touch, and sight). The Moving Picture Experts Group (MPEG) created a standard in 2007 to integrate user experiences with sensory effects, which include video and audio with effects like airflow, light, vibration, and temperature. The standard, called RoSE (Representation of Sensory Effects), was merged later in 2008 with the MPEG-V standard to allow the annotation of audio-visual content with sensory effects [3]. The multimedia render is used to reproduce audiovisual contents, the sensory effects renders are employed to enable the stimulation of the remaining human senses [4]-[7]. For instance, the vibration feature of a mobile phone and ventilation, heating, cooling systems can be employed to trigger the haptic sensation. The olfactory system can be triggered by vaporizer devices [5], and finally lighting fixtures can be used to stimulate the visual system. Enhancing the sense of reality and the strength of emotions can be seen as the primary reasons of adding effects to audio-visual contents, to achieve better viewing experience [8]-[11]. Most of the rendering devices for sensory effects in mulsemedia applications are using short range wireless communication standards (i.e., Bluetooth, WiFi, ZigBee,) to connect and render the effects, having as main disadvantage the used architecture. In most of the cases, this architecture is developed only for specific devices, a serious disadvantage in terms of manageability, scalability, and inclusion with a home entertainment system. To overcome these limitations, the authors proposed a different method to connect the sensory effects rendering devices relying on a typical Internet of Things (IoT) architecture [24].

In recent years, the ability of sensors and actuators to communicate and make a global cyber-physical world, gained extensive interest from the Information and Communication Technology (ICT) community. The Smart Home (SH) and Smart Living concepts have expanded, including devices already existing in the home environment [12], [13], [15]. The development of SH applications has been mainly based on the IoT paradigm which was an important part of this scenario's evolution, at pace with the constant evolution of short-range wireless communication.

Adding sensory effects to the multimedia content brings new challenges to the assessment of the Quality of Experience (QoE) for audio-visual contents. The assessment of the users' QoE can be performed via subjective tests, to obtain a Mean Opinion Score (MOS), requiring firm procedures to certify the validity of the results. Sensory effects have a fundamental importance on user's QoE in multi sensorial media applications.

In state of the art of mulsemedia assessments, ad-hoc systems were used rendering the effects. In [4], [9] and [11], a gaming console equipped with two fans and a vibrating bar was used to render haptic effects and the wind flow. The video was displayed on a 24" PC monitor and two loudspeakers with integrated RGB LED lights delivered the visual effects.

L. Jalal, R. Puddu and M. Murroni with the Dept. of Electrical and Electronic Engineering - DIEE/UdR CNIT, University of Cagliari, (e mail: lana.jalal, roberto.puddu, murroni@unica.it).

Maria Martini is with the Department of of Networks and Digital Media, Faculty of Science, Engineering and Computing, Kingston University London, KT1 2EE Penrhyn Road, Kingston upon Thames (e-mail: m.martini@kingston.ac.uk).

Vlad Popescu is with the Department of Electronics and Computers, Transilvania University of Brașov, Romania (e-mail: vlad.popescu@unitbv.ro)

Yuan et al. deployed in [8] and [10] a software-controlled USB fan for airflow effect rendering, a scent dispenser with four types of scents as the olfaction renderer, and a vibrating vest from TN Games [16] for haptic effect.

The drawback of this kind of approach is the need for complete communication architecture to control the rendering devices, with obvious challenges regarding scalability and manageability. The main current approach for QoE evaluation is to map the Quality of Service (QoS) to QoE [17], [18] or to derive the QoE from the audio-visual services [19], [20], not taking into consideration any of the additional sensory effects. Another approach is based on the exponential interdependency between QoS and QoE (IQX hypothesis) [21]. This hypothesis is formulated with QoE and QoS parameters as an exponential function: if the satisfaction level decreases, the disturbance level increases. An exponential function was used because even a small disturbance can drastically decrease the satisfaction. In [22], another QoE model is introduced using a triple user characterization model, taking three dimensions into account: first, the sensorial quality represented by the content quality (i.e., sharpness, brightness, blurriness, number of artifacts, etc.); second is the perceptual quality that characterizes the amount of knowledge a user may acquire; finally, the emotional quality depicts the satisfaction in terms of emotional experience. This model mainly addresses adaptation and presentation issues and does not deal with the sensory effects.

The Pseudo Subjective Quality Assessment (PSQA) [23], is a hybrid approach between subjective and objective evaluations. The subjective assessment results are used to train a learning tool that computes the relationship between the parameters that are generating the video sequences distortion and the perceived quality.

The need for multi-sensorial media QoE models is an important challenge for the current and future research, since evaluating multi-sensorial media QoE using subjective methods is a task consuming significant time and generating supplementary costs. In our previous work [24] we examined the feasibility of an approach based on the IoT concept to render multi-sensorial media sequences in a real SH environment. A cloud-based IoT architecture was implemented using home custom devices.

This paper proposes a novel non-linear parametric model suitable for the estimation of QoE for mulsemedia TV applications. The proposed model can be used to predict the enhancement produced by adding sensory effects to conventional services and plan for mulsemedia delivery as new advanced service. Non-Linear Regression (NLR) [25] is used for the model validation. NLR models are usually employed when the relation between predictor and response has a particular functional form, depending on one or more unknown model parameters. The proposed model for the parameter estimation is based on the Particle Swarm Optimization (PSO) method, proved to be efficient for the optimization of non-linear problems [26], [27].

The remainder of the paper is structured as follows: Section II gives information on implementation and provides details on the experimental set-up used to assess the QoE. The QoE model is presented in section III. The parameters estimation is discussed in section IV, while the obtained results are presented in section V. Final conclusions are drawn in section VI.

II. EXPERIMENTAL SET-UP

The experimental set-up was based on the Recommendation of ITU-T P.913 [28]. The tests for the subjective assessment were conducted in the QoE lab of the University of Cagliari [29]. The lab, which is $4 \times 4 \times 2.70$ m ($l \times w \times h$), was furnished with a large sofa and parquet floor as shown in Fig. 2(a) to simulate as close as possible the feeling of a typical living room. For the implementation of the whole SH scenario, we employed a cloud IoT platform called Lysis [30], where the effects rendering devices are implemented as Virtual Objects (VOs) [31]. The overall architecture of the platform, organized on four layers [24], is presented in Fig. 1.

The bottom layer is the physical one, built of objects with Internet access. For this specific application scenario, these objects are electronic devices with processing capabilities and integrated peripherals (i.e., smartphones), or single board computers (SBCs) with switching capabilities (i.e., Arduino, or Raspberry Pi), able to manage the renderers. The physical layer links with the upper layers by means of standard communication methods (i.e., wireless or wired) and data protocols (i.e., HTTP and MQTT).

The Virtualization Layer is based on the concept of the virtual object (VO), as the digital counterpart of the real-world entity (RWOs). The VOs represent the RWOs in terms of semantic description and functionalities and have two interfaces, allowing on one side the VO to communicate with the aggregation layer, while on the other side representing the access point to the lower layer RWOs. For this specific architecture, the virtualization layer is implemented as a software driver installed on the RWOs.

The Aggregation Layer is responsible for combining data generated by one or more VOs, ensuring a high level of re-usability. The uppermost layer of the adopted architecture is the Application Layer, where user applications are used for the final data processing and presentation.

A. Test equipment

The hardware deployed to implement the system consists of the following components, each one in charge for a specific sensory effect. The hardware was explained in detail in [24], here are briefly the main components and features of the set-up.

- An air conditioner (AC) wall split [32] employed as airflow effects renderer, controlled by an Arduino MEGA 2560 fitted out with an IR transmitter;
- A smart light system [33] (Philips Hue) placed behind the TV to emulate their integration with TV frame. The light

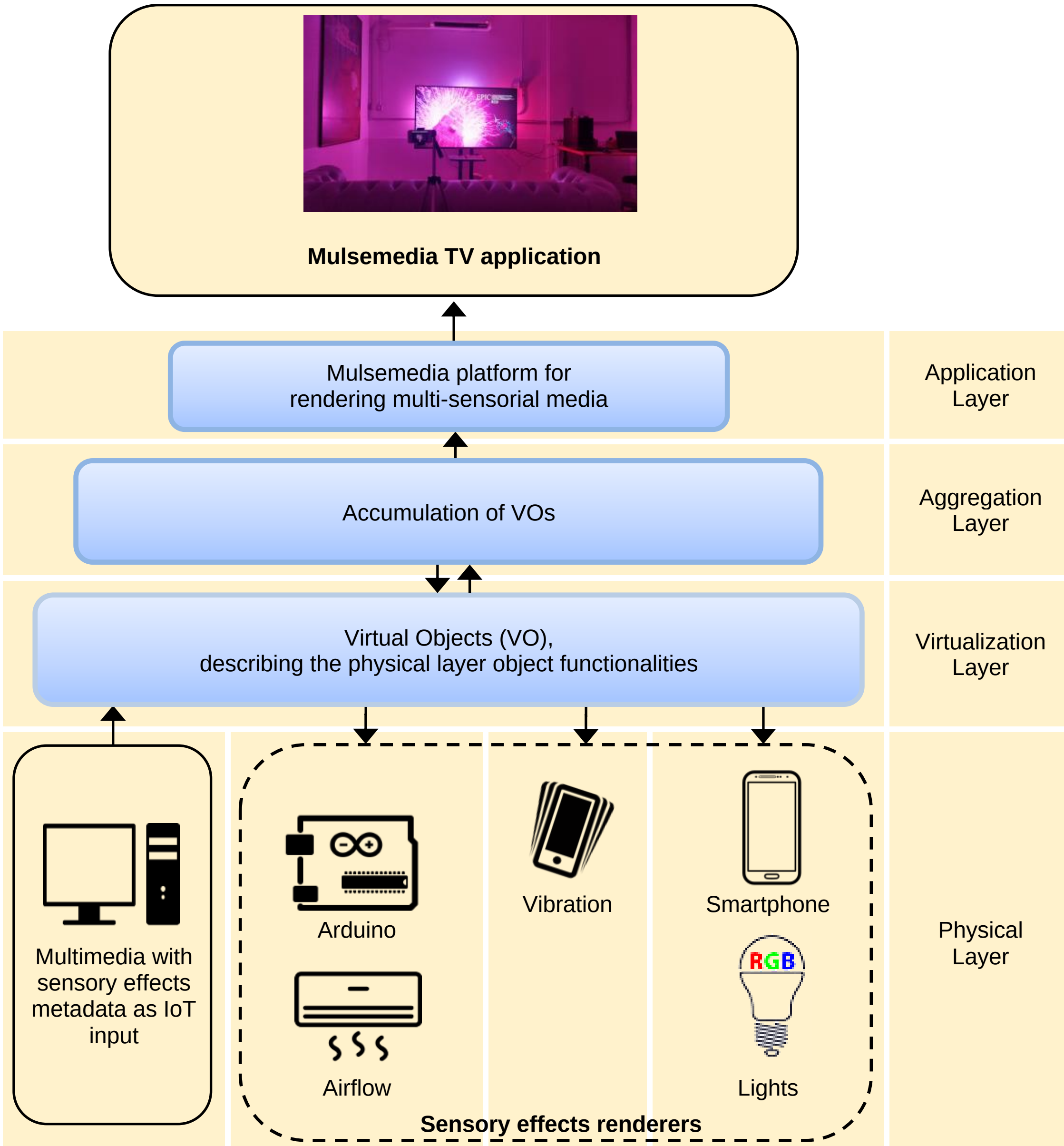
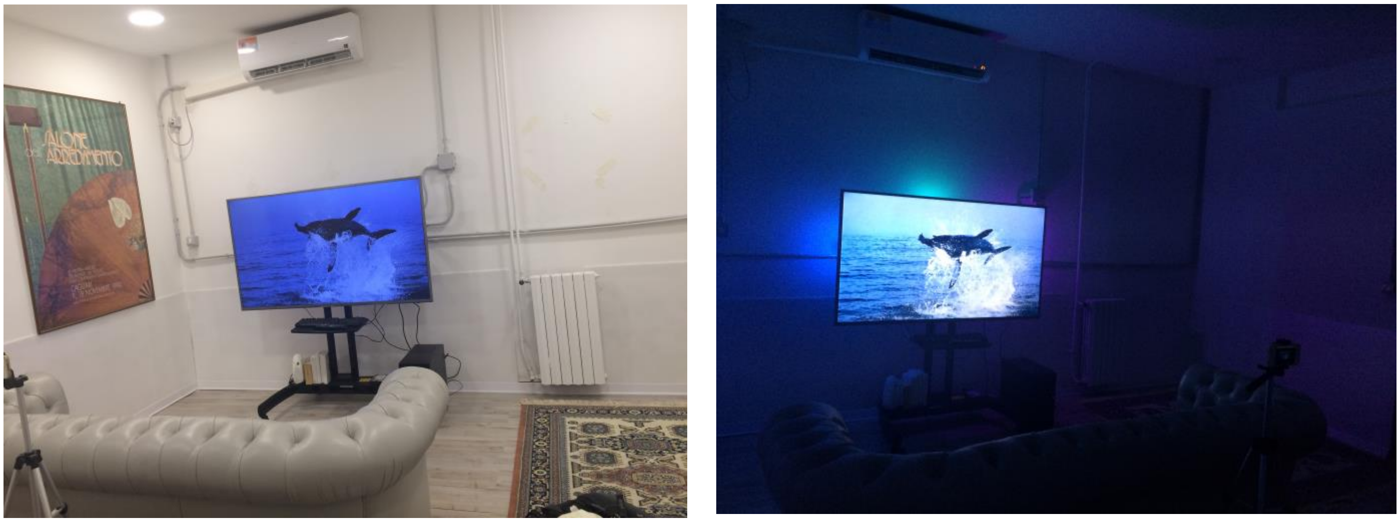


Fig. 1. Mulsemedia IoT architecture

effects are automatically extracted using an application running on a smartphone wirelessly connected to the lights. The smartphone is placed in front of the smart TV, performing real-time light effects extraction. The room was completely darkened during the tests;

- The display is a 60″ SAMSUNG TV UHD 4K Flat Smart TV. The distance between sofa and display is 1,86 m, representing 2.5 times the screen’s height;
- The haptic effect is generated by the vibration feature of a smartphone. The test cohort held the phone in their hand, in their pockets, or placed beside them on the sofa;
- The media renderer is a multimedia PC connected through a HDMI cable to the TV, running the video content annotated with the sensory effects.

We took into consideration that the air flow effect strongly depends on the distance between the AC fan and the sofa. Literature studies revealed that haptic media can be presented with a delay up to 1 second behind the video content to be acceptable for most of the users [38]. We calculated the delay in the activation time of the AC fan in order not to impair the overall QoE. and released the airflow 5 seconds before the video content [39] to achieve an acceptable QoE level.



(a) (b)

Fig. 2. Test environment (a) QoE lab (b) Test environment during the assessment

B. Subjects

The assessment cohort was based on 40 participants from various backgrounds (30 males and 10 females) with ages between 22 and 50 years (45% between 20 and 30 years, 35% between 30 and 40 years, 20% between 40 and 50 years), and an average age of 31. Only one assessor participated in a similar assessment. Information regarding to participants education, age, gender, and occupation was collected.

C. Methodology

The Absolute Category Rating (ACR) method [28] was adopted for this assessment. We based the rating scale used is on the Likert scale as defined by the ITU-T Rec. P.913, with 5 levels labelled as: bad, poor, fair, good, and excellent. Each participant was seated on the sofa in front of the AC situated on the wall above the smart TV (Fig. 2a). The video test sequences, with a duration between 21 and 32 seconds, were shown in a random sequence, interleaved with a sequence of 5 seconds of grey screen, to allow the assessors to rate the previous sequences. Furthermore, in our assessment the duration of the sequences was increased to more than the 10 seconds defined by the ACR method, in order to accommodate multiple sensory effects.

In the first part of the assessment, the participant was orally briefed on the assessment procedures and the rating scale. To reduce the total number of sessions, two participants per session were. Fig. 2(b), shows the test environment used during the subjective assessment. The entire subjective evaluation had a duration of approximatively 20 minutes.

D. Mulsemedia video sequences

The participants had to watch 40 multi-sensorial video sequences from the sensory effect dataset [35] downloaded from Sensory Experience Lab [36], coupled with three additional sensory effects: light (*L*), airflow (*A*) and vibration (*V*). Effects were also combined, creating seven different test cases: *L*, *A*, *V*, *L*+*A*, *L*+*V*, *A*+*V*, *L*+*A*+*V*. Each video sequence was also presented without any sensory effects.

The *L* effect can be extracted in an automated manner directly from the video content as previously described. The *V* and *A* effects cannot be extracted in real-time, so they had to be a-priori annotated to each video sequence using the SEVino video annotation tool [37], as illustrated in Fig. 3. The PlaySEM SE and SER software packages [38] are deployed for video playback and effects rendering.

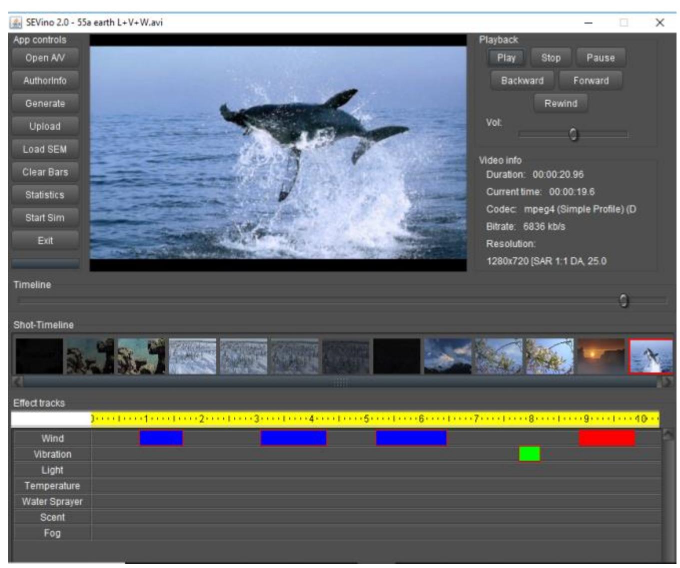


Fig. 3. Video annotation tool (SEVino)

TABLE I. VIDEO SEQUENCES LIST

Video Sequence	**Resolution**	**Bit-rate (Kbit/sec)**	**Category**	**Duration (sec)**	**Effect**	**Scenario**
2012	1280x720	2186	Action	30	L,A,V	earthquake, tornado
Earth	1280x720	4116	Documentary	21	L,A,V	wind, animal jump
Berrecloth	1280x720	3552	Sport	32	L,A,V	downhill cycling on rocks
Bridgestone	1280x720	2421	Commercial	30	L,A,V	windy weather, car moving
Pastranas	1280x720	2619	Sport	32	L,A,V	rally

Fig. 4 shows a snapshot from each video sequence (2012, Berrecloth, Bridgestone, Earth, and Pastranas). The video sequences were chosen from more categories such as sports, action, documentary and commercial. Table I gives an overview for each mulsemedia sequence in terms of resolution, bitrates, duration, category, and the video scenario.

E. Questionnaire

The participants were asked to describe their multi-sensorial experience after viewing all the multi-sensorial sequences. The questionnaire comprised questions related to the participants' perception of mulsemedia. Participants were asked to rate their response, to investigate the impact of multi-sensorial media in terms of:

- improvement of the overall sense of reality;
- the impact of annoyance of each sensory effect on the user experience;
- the impact of the intensity of each sensory effect on the user experience;
- the impact of sensory effects on user enjoyment.

III. QoE MODEL FOR MULSEMEDIA APPLICATIONS

As stated in the introduction, the topic of mulsemedia QoE models is currently a challenging topic, due to the time and cost of subjective quality assessments.

We propose here the following model for QoE for mulsemedia TV applications in a smart home environment:

$$Q_oE_{Muls} = Q_oE_{Mult} * \delta * exp\left(\sum w_i * b_i\right) \tag{1}$$

where Q_OE_{Mult} is the multimedia QoE, (i.e., the audio and video content quality), w_i is the weighting factor for each sensory effect (i.e., light, vibration, or airflow), b_i is a boolean variable used to indicate the presence or not of the sensory effect, and δ is for tuning. The model is validated using the subjective assessment data (MOS). In our assessment, three sensory effects were used: light (*L*), airflow (*A*) and vibration (*V*); the effects were also combined. This model appears as a good compromise between accuracy and complexity.

The specificity of this work is that the effects renderers are deployed in a SH environment using only custom devices. These SH devices are connected to the IoT hence are accessible and can be added to the smart TV to enable a multi-sensorial experience.

(a) (b) (c) (d) (e)

Fig. 4. Snapshot from each video sequence (a) 2012 (b) Berrecloth (c) Bridgestone (d) Earth (e) Pastranas

IV. MODEL PARAMETER ESTIMATION USING PARTICLE SWARM OPTIMIZATION

This section describes the implementation of PSO to estimate the optimal values of the parameters in the proposed model. The PSO is a self-adaptive search optimization technique, consisting of a set of solutions (particles) called population. Each solution is characterised by a set of parameters and can be represented by a point in a multidimensional space [26].

Initially, based on preliminary tests, we defined the swarm size p (i.e., number of particles = 20), and the maximum number of iterations, $t\_max=200$. The independent variables V, L, and A represent vibration, light, and airflow effects, and the dependent variable y represent the Q_OE_{Muls}, whereas the outcome of the PSO will be the MOS. In the proposed model the influence of the independent variables (i.e., effects) on the dependent variable (Q_OE_{Muls}), is represented by the coefficients w_L, w_A, w_V. The position of each particle and its velocity within the swarm are denoted by y_j, and v_j, respectively, where the index j is the particle number. For each particle at the 1st iteration ($k = 1$), the initial values of velocity v_j^1, weights w_{Lj}^1, w_{Aj}^1, w_{Vj}^1 and position y_j^1 are randomly selected. Particles are moved iteratively to find the new position in the search space dimension.For each of the k iterations the fitness value f_j is calculated as the difference between the previous and the current position. Then, the best location visited by each particle ($pbest_j^k$), and the best position in the whole swarm ($gbest^k$) is determined. Therefore, if the value of f_j^k is greater than the biggest f_j ($pbest_j^k$) in history, then it is used as the new $pbest_j^k$, the particle with the best fitness value achieved among all particles in the swarm is selected as the $gbest^k$. Particles update their velocity based on the following equation [43]:

$$v_j^{k+1}=I*v_j^k+c_1*r_1*(pbest_j^k-y_j^k)+c_2*r_2*(gbest^k-y_j^k) \quad (2)$$

where v_j^k and v_j^{k+1} are the current and updated particle's velocity, I is the inertia weight which is a constant=0.02 based on preliminary tests, r_1 and r_2 are random variables in [0,1], c_1 (self-confidence factor) and c_2 (swarm-confidence factor) are constants values equal to 2 as recommended by [43] and [44]. The cognitive component term $pbest_j^k-y_j^k$, represents the best solution found by each particle. The social component term $gbest^k- y_j^k$ is referred to the best solution in the whole swarm.

At the iteration $k+1$, the weights $w_{i,j}^k$ are updated for each particle, where $i \in$ {vibration (V), light (L), and airflow (A)}. The new updated weights values $w_{i,j}^{k+1}$ are calculated using to the following equation:

$$w_{i,j}^{k+1} = w_{i,j}^k + v_j^{k+1} \quad , j=1,\dots,p \quad (3)$$

Then, for each particle the new updated weights $w_{i,j}^{k+1}$ from equation (3) are substituted in (1) to compute the new position y_j^{k+1}. Each individual particle keeps searching for the individual and global best position based on updating the velocities. This process continues until the optimal parameter values of the proposed model are achieved or the maximum iteration number t_max is reached. The process of the PSO algorithm is summarized in Algorithm 1. To ensure the algorithm achieve convergence preliminary tests on the PSO had been run, the final convergence of the model presented in Fig. 5.

Algorithm 1 : PSO Algorithm

Inputs:
  $Dataset((Lm, Am, Vm), y)$, $test\ cases\ (m{:}1 \rightarrow 7)$
Initialize:
  p, t_max, c_1, c_2, $and\ I$
for $j : 1 \rightarrow N$ **do**
  Initialize:
  $y_j^1, v_j^1, w_{Lj}^1, w_{Aj}^1, w_{Vj}^1$
end for
while optimum parameter values or t_max is not attained **do**
  for $j : 1 \rightarrow N$ **do**
    $k : 1 \rightarrow t_{max}$
    compute y_j^k
    compute f_j^k
    if $f_j^k < f_j^{k-1}$ **then**
      p_{bestj}^k is set to the current location
    end if
    $gbest^k$=best particle in the whole swarm
    for $j : 1 \rightarrow N$ **do**
      Compute $v_j^{K+1}, w_{Lj}^{k+1}, w_{Aj}^{k+1}, w_{Vj}^{k+1}$
      Compute y_j^{K+1}
    end for
  end for
end while

The estimated parameters value (the weights w_i) for the proposed model are shown in equation (4).

$$Q_oE_{Muls} = Q_oE_{Mult} * exp\,(0.117 * b_L + 0.049 * b_A + 0.074 * b_V) \quad (4)$$

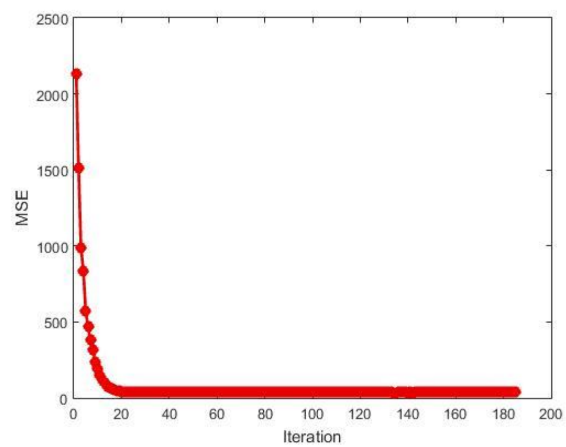


Fig. 5. The convergence of MSE for the PSO algorithm.

In the subjective assessment the Q_OE_{Mult} is assessed by presenting the reference video sequence without sensory effects. In production systems, this may not be possible, so the Q_OE_{Mult} can be assessed from existing QoS models [17], [21], [22], [45] and estimating the Q_OE_{Mult} from QoS parameters [38].

The results of the subjective assessment allowed us to derive a non-linear exponential model for the prediction of the QoE of mulsemedia content for TV applications on smart home scenario.

V. RESULTS

This section analyses the cumulated impact of multi-sensorial media on user experience, by taking into consideration the impact of each sensory effect derived from the post-experiment questions. 8 outliers had been detected from the 40 participants, according to the procedure described in [28] and [29]. Fig. 6 to Fig. 10 show the subjective MOS test results and Confidence Interval CI (95%) for the mulsemedia sequences, with eight different configurations. The following subsections analyze the influence of each of the sensory effects on the perceived experience.

A. *Impact of the combined sensory effects on user experience*

This section presents the combined impact of mulsemedia on user experience through analysis of the impact that each sensory effect (light L, airflow A, vibration V) has on the user, which had to answer to the following statements:

- if the sensory effects (L, A, V) are annoying.
- the sensory effects (L, A, V) enhance the sense of reality.
- if the intensity of each sensory effect (L, A, V) is too weak, weak, fine, strong, too strong.
- if the multi-sensorial media is enjoyable.

Each participant watched 40 video sequences from different categories with different configurations. 4 of these configurations contain light effects (L, L+A, L+V, L+A+V), 4 configurations with airflow, as well for vibration effect as shown in Fig. 6 to Fig. 10. These figures will be subsequently commented in section V B.

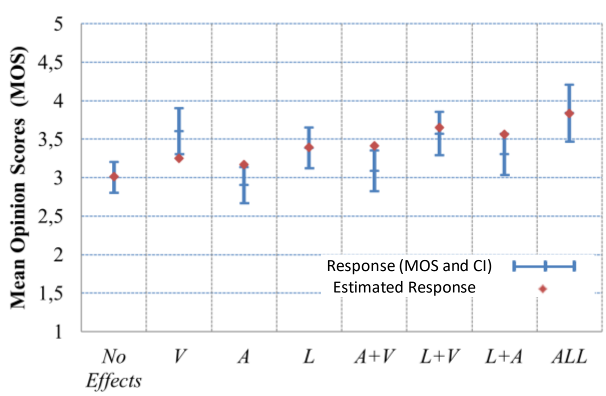


Fig. 6. The model estimated response and the MOS for the action sequence “2012”.

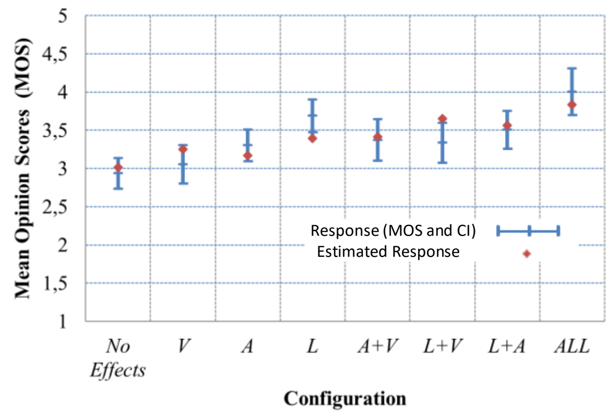


Fig. 7. The model estimated response and the MOS for the documentary sequence “Earth”.

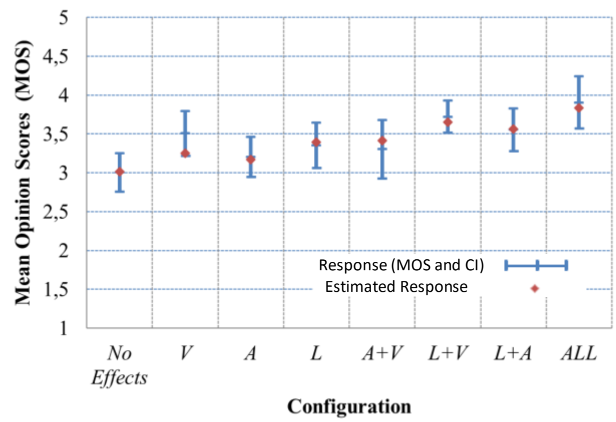


Fig. 8. The model estimated response and the MOS for the sport sequence “Berrecloth”.

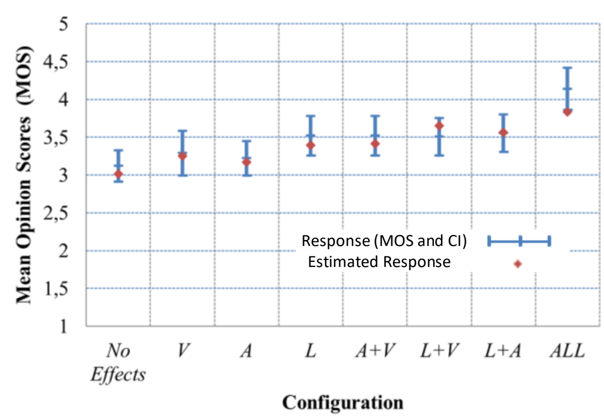


Fig. 9. The model estimated response and the MOS for the commercial sequence “Bridgestone”.

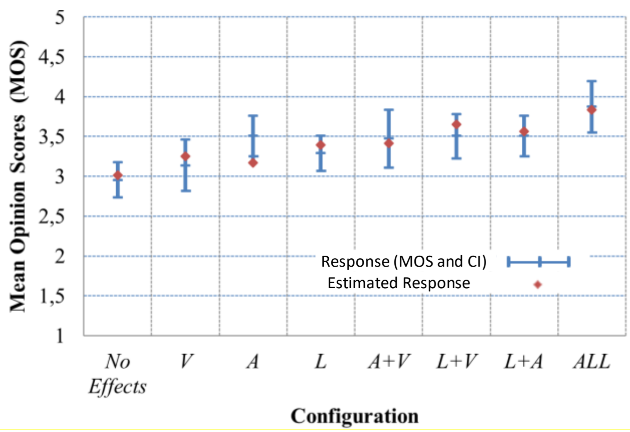


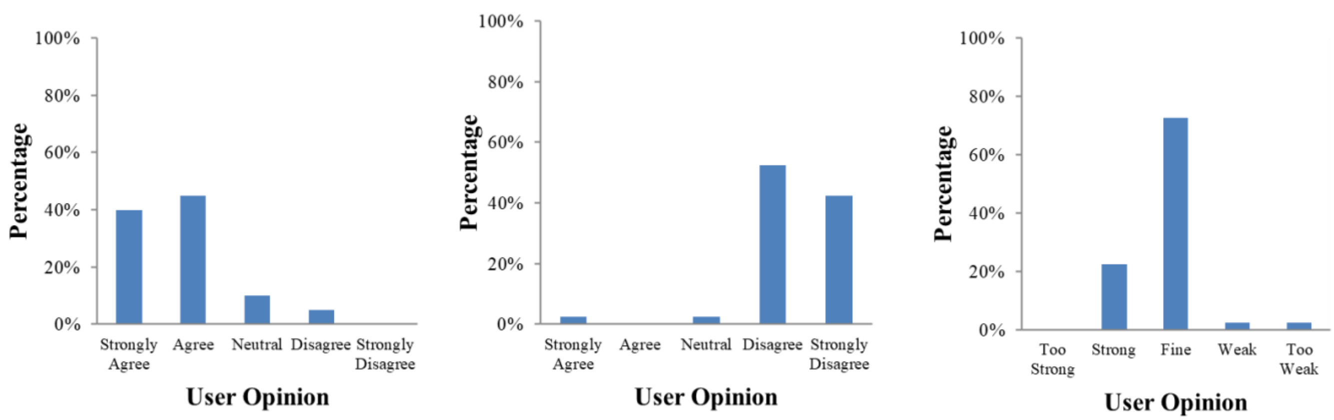
Fig. 10. The model estimated response and the MOS for the sport sequence "Pastranas".

1) Impact of the light effect on user experience

The participants were asked to give their opinion regarding the annoyance of light effect, i.e., if the light enhances the sense of reality, and the intensity of the light effect, as shown in Fig. 11. User opinions provided as answers to the questionnaire regarding to the impact of light effect in the experiment gave the following results: 95% of the participants were not annoyed by the light effect, 2.5% had neutral opinion, and 2.5% were annoyed by the light effect. Regarding to the sense of reality, 85% of the participants agreed and strongly agreed that light effect enhances the sense of reality, 10% had neutral opinion, and 5% felt that light effect did not enhance the sense of reality. Finally, the participants opinion about the light intensity during the assessment was the following: 72.5% found the intensity fine, 22.5% felt it was strong, and 5% found it weak-too weak intensity.

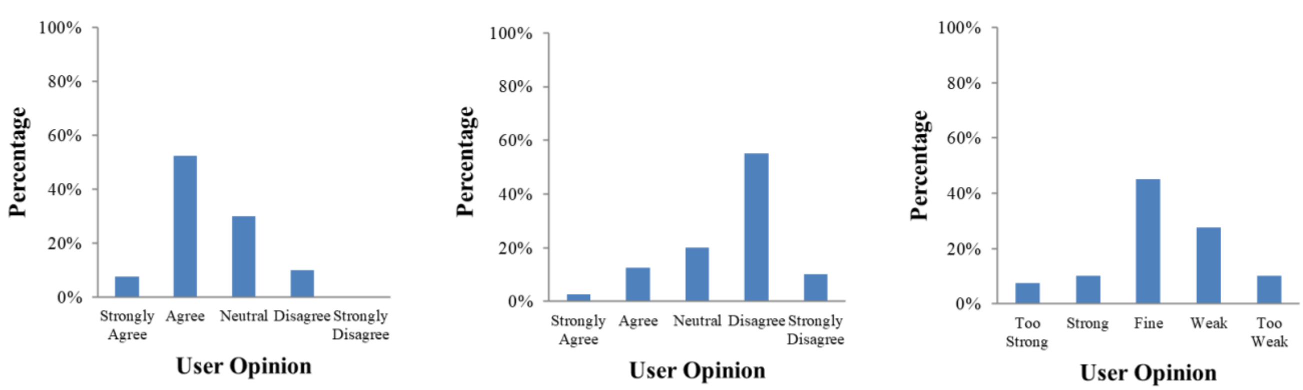
2) Impact of the airflow effect on user experience

The influence of the airflow effect on the perceived sense of reality, annoyance, and intensity were collected, and displayed in Fig. 12. The majority of the participant (60%) felt that the airflow enhances the sense of reality, 30% had neutral opinion, and 10% did not feel the any enhancement. The response to the annoyance of the airflow effect indicates that 65% of the participants were not annoyed, 20% had neutral opinion, and 15% were annoyed by the airflow effect.



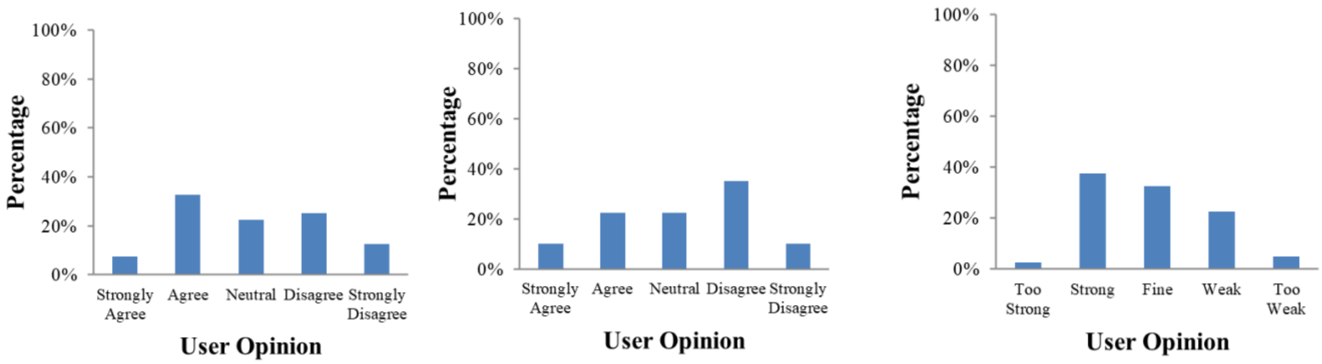
(a) (b) (c)

Fig. 11. User response to the impact of light effect (a) Light effect enhances the sense of reality (b) Light effect is annoying (c) The intensity of light effect



(a) (b) (c)

Fig. 12. User response to the impact of airflow effect
(a) Airflow effect enhances the sense of reality (b) Airflow effect is annoying (c) The intensity of airflow effect



(a) (b) (c)

Fig. 13. User response to the impact of vibration effect
(a) Vibration effect enhances the sense of reality (b) Vibration effect is annoying (c) The intensity of vibration effect

The impact of airflow intensity on user experience was the following: 45% of the participants felt the intensity to be fine, 17.5% found ”strong” as too strong, and 37.5% found the intensity ”weak” as being too weak.

3) Impact of vibration effect on user experience

User opinions regarding the impact of vibration effect are shown in Fig. 13. Most of the participants (40%) agreed and strongly agreed that the vibration effect enhances the sense of reality, 22.5% had neutral opinion, and 37.5% disagreed. Regarding to the annoyance of vibration effect, 45% of the participants did not consider the vibration as annoying, 22.5% expressed a neutral opinion, and 32.5% found it annoying. The users' opinion on the vibration intensity indicates that 40% found ”strong” as too strong, 32.5% felt it as fine, and 27.5% felt it too weak.

4) User enjoyment and satisfaction

User enjoyment is important to indicate the quality of experience. The results from the study presented in [10] showed that 70% of the assessors consider haptic and airflow effects enhance user enjoyment levels. In our assessment, results obtained from the question “I enjoyed multi-sensory experience” indicated that 80% enjoyed the multi-sensorial media, 12.5% had a neutral opinion, and 7.5% did not enjoy the mulsemedia experience (Fig. 14). The impact of the effect type on quality of experience variation is shown in Fig. 15, which can be considered as an improved impact compared to the state-of-the-art study [39] as shown in Fig. 16, in which sensory effects renderers are ad-hoc devices. The main disadvantage of these applications is the architecture used for controlling the renders, developed only for specific devices. The results from our assessment reveal that each effect has a higher impact on user opinion, the same for the combination of all effect together. In an SH environment, the user preferences can be saved by the IoT architecture, and the devices settings can be adjusted accordingly, to improve the user experiences through delivering personalized services and increasing user enjoyment and satisfaction level.

The estimated responses by the model for all the sequences are shown in Fig. 6 to Fig. 10. Most of the responses are inside the confidence interval (95%) of the subjective quality evaluation. The model response is also close to the average of the MOS. The MOS without sensory effects is lower than the MOS with sensory effects in all configurations for all video sequences. The sensory effects impact on the MOS varies based on the category and contents of the sequence. For example, in the “action” category, the vibration effects were the most appreciated effect by users. For “documentary” and “commercial” categories, light was the most preferred effect. For the “sport” category, the vibration effect was the one with the highest impact, specifically for the “Berrecloth” sequence, containing scenes downhill bicycle rides. On the other hand, for the “sport” sequence “Pastranas”, the airflow effect was the preferred effect by users. The impact of the combinations of sensory effects (L+A, L+V, V+W) is either lower than or equal to the impact of an individual sensory effect, though the combination of all effects together has the highest impact on the MOS.

The relatively low MOS score (3) for the case where no effect is added can be motivated by the fact that the order of presentation of the sequences was random, and therefore the sequences with no impairments could have been evaluated after the ones with effects and the evaluators have considered an overall MOS lower, considering positively the multisensory effects to the overall QoE and thus reducing the MOS of the sequences with no effects.

The proposed model performance compared to other mulsemedia QoE models performance [39] [40]. The performance comparison is in terms of means square error (MSE) for two different empirical dataset MOS from two different subjective assessments. Fig.17. Shows the MSE of the proposed model compared to the other models using the subjective assessment data MOS from our assessment, the proposed nonlinear (exponential) model allows obtaining an improvement of 11.83% compared to the power model presented in [40] and 55.27% compared to the linear model proposed in [39].

Finally, the proposed model performance compared to the linear model [39] performance utilizing the empirical dataset MOS from the subjective assessment presented in [38] is shown in Fig. 18.

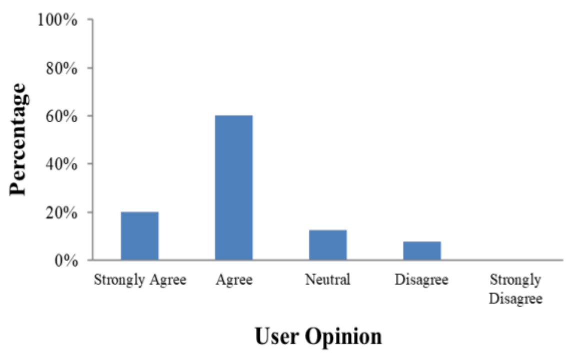


Fig. 14. Percentage of user enjoyment

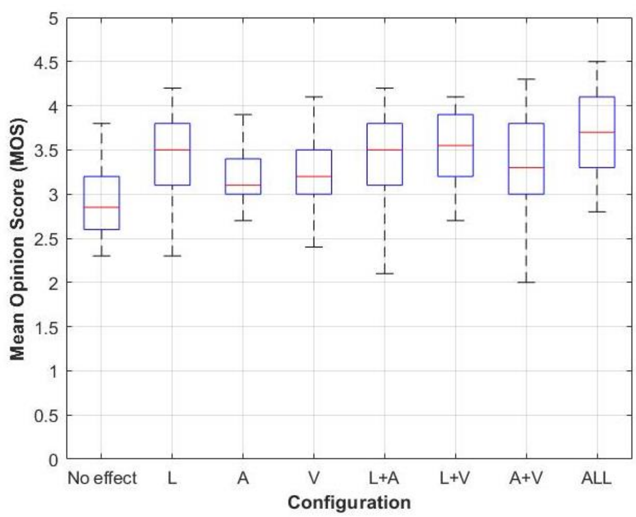


Fig. 15. MOS response and the impact of each effect on the response.

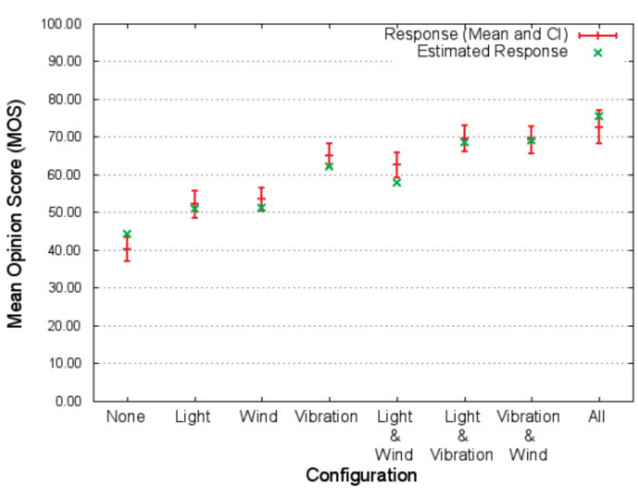


Fig. 16. MOS response and estimated response from user study [39].

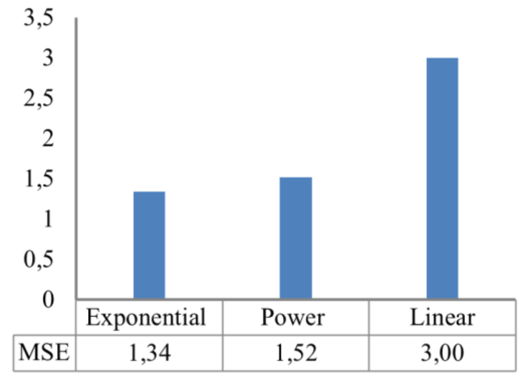


Fig. 17. Models performance comparison in term of MSE using our assessment dataset.

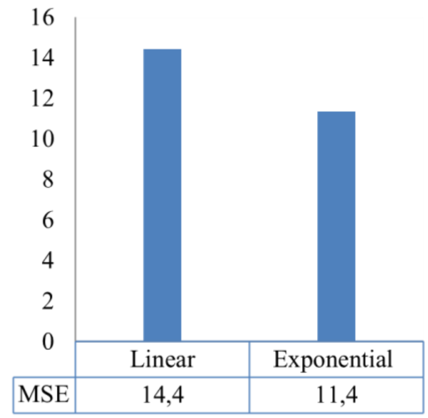


Fig. 18. Model performance comparison in term of MSE using the MOS presented in [39].

B. VALIDATION

Considering the novelty of the proposed QoE model, we repeated the MOS test with a different set of participants in very similar testing conditions at the QoE Laboratory of Transilvania University of Brașov, Romania.

The laboratory has similar dimensions, and, to closely replicate the conditions from the initial training phase, we used the same screen size for the TV set and same distance from the couch to the screen and to the AC fan.

A number of 43 participants have been invited to this assessment (14 females and 29 males), between 20–60 years old, (50% between 20 and 30 years, 35% between 30 and 40 years, and 15% between 40 and 60 years), with the average age 28 years. None of the assessor participated in a similar assessment campaign.

The background of the participants was mixed, being composed in a balanced manner from both technical students and researchers (from the Department of Electronics and Computers) and humanist students and researchers (from the Departments of Social Sciences).

In this case, 6 outliers where identified (<14%). A further analysis, on the outliers in both measurement campaigns show that the majority of them were older people (> 45 years). This is in line with the reduction of the percentage of outliers in this second test assessment procedure, where the percentage of older assessors was reduced. It seems that the multisensorial feature is more shocking for older people that have reacted out of the statistical mean and variance of the rest of the population. The result obtained in this validation phase were very similar in terms of measured MOS, falling into the confidence interval (95% just like in the training phase) as to confirm the effectiveness of the proposed model.

VI. CONCLUSIONS

This paper presents the results of the subjective assessment conducted for mulsemedia TV applications in a SH scenario with real custom devices (e.g., air conditioning, lights, etc.). The impact of light, vibration, and airflow effects on user experience was analyzed in terms of enhancement of the sense of reality, impact of annoyance, user response to the intensity of the effects, and user overall enjoyment. Most of the participants (80%) enjoyed the mulsemedia experience with the three effects. Furthermore, most of the assessors (85% for light effect, 60% for airflow, and 40% for vibration effect) felt that sensory effects enhance the sense of reality. 95% of the participants did not experience any annoyance because of the light effects, 65% for airflow, and 45% for vibration effects. In addition, results show that users globally assess the intensity of the sensory effect during the assessment as fine (72.5% for light effect, 45% for airflow, and 32.5% for vibration).

A parametric QoE non-linear model for mulsemedia in a SH scenario has been instantiated and validated in a real SH environment, through two MOS assessment campaigns. The proposed model considers the number of effects and their impact on the QoE and has been instantiated with three sensory effects (light, airflow, and vibration), with further effects that can be added without changing the model. The subjective assessment results applied to the audiovisual sequences, selected from the category action, sport, documentary, and commercial, will reduce the need to conduct further subjective quality assessments to assess how the model performs in other categories. As for the result achieved, the proposed model can provide satisfactory accuracy in the estimation of the QoE on mulsemedia applications and can enhance the estimation accuracy compared to the models currently presented in the literature.

REFERENCES

[1] G. Ghinea, C. Timmerer, W. Lin, and SR. Gulliver, “Mulsemedia: State of the Art, Perspectives, and Challenges,” ACM Trans. Multimedia Computing, Communications and Applications, vol.11, no.1, 2014.

[2] L. Jalal and M. Murroni, "Enhancing TV broadcasting services: A survey on mulsemedia quality of experience," 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1-7, Cagliari, Italy, 2017. doi: 10.1109/BMSB.2017.7986192

[3] B. Choi, E. Lee, and K.Yoon, “Streaming media with sensory effect,” In: Information Science and Applications (ICISA), pp. 1–6, 2011.

[4] M. Waltl, C. Timmerer, B. Rainer, and H. Hellwagner, “Sensory effects for ambient experiences in the World Wide Web,” Multimedia tools and applications, vol. 70, no. 2, pp. 1141-60, 2014.

[5] J. J. Kaye, “Making Scents: Aromatic Output for HCI,” Interactions, vol. 11. no. 1, pp 48-61, 2004.

[6] Z. Yuan, T. Bi, G. Muntean and G. Ghinea, "Perceived Synchronization of Mulsemedia Services," in IEEE Transactions on Multimedia, vol. 17, no. 7, pp. 957-966, July 2015.

[7] O. A. Ademoye and G. Ghinea, "Synchronization of Olfaction-Enhanced Multimedia," in IEEE Transactions on Multimedia, vol. 11, no. 3, pp. 561-565, April 2009.

[8] Z. Yuan, G. Ghinea and G. Muntean, "Beyond Multimedia Adaptation: Quality of Experience-Aware Multi-Sensorial Media Delivery," in IEEE Transactions on Multimedia, vol. 17, pp. 104-117, Jan. 2015.

[9] M. Waltl, C. Timmerer, and H. Hellwagner, “Increasing the User Experience of Multimedia Presentations with Sensory Effects,” in Proc. 11th International Workshop on Image Analysis for Multimedia Interactive Services, pp. 1-4, 2010.

[10] Z. Yuan, S. Chen, G. Ghinea, and G. Muntean, “User quality of experience of mulsemedia applications,” ACM Trans. Multimedia Comput. Commun, vol. 11, no. 1, pp. 15:1–15:19, Sep. 2014.

[11] S. F. Langa, M. M. Climent, G. Cernigliaro and D. Rincón Rivera, "Toward Hyper-Realistic and Interactive Social VR Experiences in Live TV Scenarios," in IEEE Transactions on Broadcasting, vol. 68, no. 1, pp. 13-32, March 2022, doi: 10.1109/TBC.2021.3123499.

[12] The State of Traditional TV: Updated With Q2 2017 Data. https://www.marketingcharts.com/featured-24817.

[13] Z. L. Wu and N. Saito, "The Smart Home [Scanning the Issue]," in Proceedings of the IEEE, vol. 101, no. 11, pp. 2319-2321, Nov. 2013.,doi: 10.1109/JPROC.2013.2282668.

[14] L. Atzori, A. Iera, G. Morabito, The Internet of Things: A survey, Computer Networks, vol. 54, no. 15, pp. 2787-2805, 2010. SSN 1389-1286, http://dx.doi.org/10.1016/j.comnet.2010.05.010.

[15] Lee, Edward (January 23, 2008). "Cyber Physical Systems: Design Challenges". University of California, Berkeley Technical Report No.UCB/EECS-2008-8. Retrieved 2008-06-07.

[16] “TN Games” [Online]. Available: http://tngames.com/.

[17] Ullah, S., Thar, K. & Hong, C.S. Management of scalable video streaming in information centric networking. Multimed Tools Appl 76, 21519–21546 (2017). https://doi.org/10.1007/s11042-016-4008-8

[18] S. Kumar, R. Devaraj, A. Sarkar and A. Sur, "Client-Side QoE Management for SVC Video Streaming: An FSM Supported Design Approach," in IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 1113-1126, Sept. 2019, doi: 10.1109/TNSM.2019.2926720.

[19] Mingfu Li, Chien-Lin Yeh, Shao-Yu Lu, "Real-Time QoE Monitoring System for Video Streaming Services with Adaptive Media Playout", International Journal of Digital Multimedia Broadcasting, vol. 2018, Article ID 2619438, 11 pages, 2018. https://doi.org/10.1155/2018/2619438

[20] Kiani Mehr, S., Jogalekar, P. & Medhi, D. Moving QoE for monitoring DASH video streaming: models and a study of multiple mobile clients. J Internet Serv Appl 12, 1 (2021). https://doi.org/10.1186/s13174-021-00133-y.

[21] Hoßfeld, T., Heegaard, P.E., Skorin-Kapov, L. et al. Deriving QoE in systems: from fundamental relationships to a QoE-based Service-level Quality Index. Qual User Exp 5, 7 (2020). https://doi.org/10.1007/s41233-020-00035-0.

[22] X. Ma et al., "QAVA: QoE-Aware Adaptive Video Bitrate Aggregation for HTTP Live Streaming Based on Smart Edge Computing," in IEEE Transactions on Broadcasting, March 2022 (early access), doi: 10.1109/TBC.2022.3171131.

[23] A. Polakovi č, G. Rozinaj and G. -M. Muntean, "User Gaze-Driven Adaptation of Omnidirectional Video Delivery Using Spatial Tiling and Scalable Video Encoding," in IEEE Transactions on Broadcasting, March 2022 (early access), doi: 10.1109/TBC.2022.3157470.

[24] L. Jalal, M. Anedda, V. Popescu, M. Murroni, QoE Assessment for IoT-Based Multi Sensorial Media Broadcasting, IEEE Transactions on Broadcasting, vol. 64, no. 2, pp. 552-560. June 2018.

[25] GAF. Seber, and CJ. Wild “Nonlinear Regression,” Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1989.

[26] V. Özsoy, and H. Örkçü “Estimating the Parameters of Nonlinear Regression Models Through Particle Swarm Optimization,” Gazi University Journal of Science, vol. 29, no. 1, pp. 187-199, 2016.

[27] L. Jalal, V. Popescu and M. Murroni, "Quality-of-experience parameter estimation for multisensorial media using Particle Swarm Optimization," 2017 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) & 2017 Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP), Brasov, 2017, pp. 965-970. doi: 10.1109/OPTIM.2017.7975095

[28] ITU-T Rec. P.913, “Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment”, 2016.

[29] C. Gavrila, V. Popescu, M. Fadda, M. Anedda and M. Murroni, "On the Suitability of HbbTV for Unified Smart Home Experience," in IEEE Transactions on Broadcasting, vol. 67, no. 1, pp. 253-262, March 2021, doi: 10.1109/TBC.2020.2977539

[30] R. Girau, S. Martis, L. Atzori “Lysis: A platform for IoT distributed applications over socially connected objects,” IEEE Internet of Things Journal, 2017.

[31] M. Nitti, V. Pilloni, G. Colistra, and L. Atzori, “The virtual object as a major element of the internet of things: a survey,” IEEE Communications Surveys & Tutorials, vol. 18, pp. 1228–1240, 2015.

[32] “Haier inverter technology” [Online]. Available: http://www.haier.net/en/.

[33] “Philips hue personal wireless lighting” [Online]. Available: https://www2.meethue.com/en-us.

[34] “TV 60” UHD 4K Flat Smart Serie 9 JU6800 [Online]. Available: http://www.samsung.com/it/tvs/uhd-ju6800/UE60JU6800KXZT/.

[35] M. Waltl, C. Timmerer, B. Rainer, H. Hellwagner, “Sensory Effect Dataset and Test Setup,” IEEE Proc. 4th Int. Workshop Quality Multimedia Experience, pp. 115-120, 2012.

[36] Sensory Experience Lab, http://selab.itec.aau.at.

[37] M. Waltl, B. Rainer, C. Timmerer, H. Hellwagner, “An End to-End Tool Chain for Sensory Experience based on MPEG-V,” Signal Processing: Image Communication, vol. 28, no. 2, pp. 136-150, 2013.

[38] E. Saleme, and C. Santos, “PlaySEM: a Platform for Rendering MulSeMedia Compatible with MPEG-V,” WebMedia’15, Manaus, Brazil, pp.145-148, 2015.

[39] A. Covaci, E. B. Saleme, G. Mesfin, N. Hussain, E. Kani-Zabihi and G. Ghinea, "How Do We Experience Crossmodal Correspondent Mulsemedia Content?," in IEEE Transactions on Multimedia, vol. 22, no. 5, pp. 1249-1258, May 2020, doi: 10.1109/TMM.2019.2941274.

[40] L. Jalal, M. Murroni, “A Nonlinear Quality of Experience Model for High Dynamic Spatio-Temporal Mulsemedia” 9th International Conference on Quality of Multimedia Experience (QoMEX), pp. 224-229, Erfurt, Germany 2017.

[41] N. Murray, G. Muntean, Y. Qiao, S. Brennan, and B. Lee, “Modeling User Quality of Experience of Olfaction-Enhanced Multimedia,” IEEE Transactions on Broadcasting, vol. 64, no. 2, pp. 539–552, 2018.

[42] W. Jatmiko, K. Sekiyama, and T. Fukuda “Modified particle swarm robotic for odor source localization in dynamic environment,” International Journal of Intelligent Control and Systems: Special Issue on Swarm Robotic, pp.176-184, 2006.

[43] R. Eberhart, and Y. Shi “Particle swarm optimization: developments, applications and resources,” Proceedings of IEEE International Congress on Evolutionary Computation, pp.81–86, 2001.

[44] Y. Shi, and R. Eberhart “A modified particle swarm optimizer, ” International Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence, pp. 69-73, 1998.

[45] G. Ghinea and J. P. Thomas, "Quality of perception: user quality of service in multimedia presentations," in IEEE Transactions on Multimedia, vol. 7, no. 4, pp. 786-789, Aug. 2005