Toward a Mathematical Model for Quality of Experience Evaluation of Haptic Applications

Abdelwahab Hamam, *Student Member, IEEE*, and Abdulmotaleb El Saddik, *Fellow, IEEE*

***Abstract*— There is rapid progress in the advancement of user interfaces. One such advancement is enabling the sense of touch, or haptics, as part of the interface. Haptic devices are seeing growth in many types of applications such as gaming and medical simulation. Assessing the quality of experience (QoE) of the user is necessary to evaluate how the user perceives such interfaces. The QoE is a user-centric parameter that shifts the paradigm of evaluation from the technology itself to the user. This paper proposes a mathematical-based QoE evaluation of haptic-based applications. A mathematical model that is able to quantify the QoE of the user is described. By conducting a user study in which users evaluate a haptic-based game application, we were able to test and validate the mathematical model. There are several approaches in determining the weights to be used with the mathematical model. This paper presents and compares different approaches for weight determination, namely even weight distribution, correlation-based weights, even weights–correlation combination, linear regression analysis, and principal component analysis (PCA). Our results show that PCA weight determination performs slightly better than the rest of the approaches.**

***Index Terms*— Game applications, haptic interface, quality of experience (QoE), weighted average.**

I. Introduction

HAPTIC interfaces are instruments that are being incorporated into numerous and diverse range of applications such as gaming and medical simulations. Each application is usually designed to adapt to a certain type of haptic devices [1]. The addition of the sense of touch into those applications promises excitement, realism, and a more natural feel for the users [2]. As this new haptic medium is advancing, the need for its assessment increases.

Haptic-enabled applications are usually assessed by user feedback in the form of user testing, interviews, and questionnaires [3]–[5]. This form of evaluation is resource-expensive and relies on the subjectivity of the user. Thus, there is a need for a more desirable assessment method that converts subjective user ratings into concrete results formed and validated by objective means [6].

The former ideology in the multimedia community was the Quality of Service (QoS). QoS is based on specific parameters that undermine the subjectivity of the user in the assessment process. The newer ideology is the quality of experience (QoE) which focuses more on the user satisfaction and perception. The shift in ideology is in accordance with the user-centric design and evaluation which became a necessary recipe for a product's success or demise [7].

The novelty of the QoE paradigm brings along certain challenges. Especially in virtual reality (VR) and haptic domains, QoE assessment is still in its infant stage. There is still some modular understanding on what constitutes the experience of the user, and how to capture that in a unified systematic model.

This paper describes a mathematical model that is based on weighted averages to quantify the QoE of users. The model takes into account different categories of parameters and it merges QoS parameters with user experience (UX) parameters. A user study was conducted at the DISCOVER Lab of the University of Ottawa to assess a haptic-enabled balance ball game. This paper describes the user study and how the results were adapted to be formalized by the mathematical model including how the weights of the parameters were selected and applied.

The paper's focus is on utilizing different methodologies in weight determination and comparing those methodologies in terms of statistical metrics such as percent and relative error. We adapt five techniques for calculating the weights of the mathematical model among them correlation, regression analysis and principal component analysis (PCA). The quantified QoE based on the calculated weight is confirmed by preliminary users' evaluations that could be further validated with more experiments in the future.

The rest of the paper is organized as follows. Section II reviews the related work done previously in haptics and QoE. In Section III, we describe the user study conducted to assess a haptic-enabled game. The section describes the game application, the parameters selected for testing, the questionnaire results, and the standardization of data. The description of the mathematical model is given in Section IV. Section V lists and explains the weight determination approaches that are considered. In Section VI we discuss and validate the results obtained. Finally, we conclude this paper in Section VII.

II. Related Work

There has been a tacit consensus within the multimedia community that the focus in multimedia evaluation should center on the user. Some authors articulate that by stressing the need to shift the attention from the well-established QoS to QoE [7]–[9]. In [7], the author lists the challenges that arise from that shift, given that QoS-dominated multimedia evaluation for a long time.

Manuscript received November 24, 2012; revised March 8, 2013; accepted March 11, 2013. Date of publication August 7, 2013; date of current version November 6, 2013. The Associate Editor coordinating the review process was Dr. Kurt Barbe.

A. Hamam is with the Discover Laboratory, University of Ottawa, Ottawa ON K1N 6N5, Canada (e-mail: ahamam@site.uottawa.ca).

A. El Saddik is with the MCRLab and Discover Laboratory, University of Ottawa, Ottawa ON K1N 6N5 Canada (e-mail: elsaddik@uottawa.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2013.2272859

In the VR domain, Whalen *et al.* list the common evaluation methods and the challenges encountered when assessing QoE in a virtual environment (VE). Traditionally, there are three recognized methods for assessing the user's feedback and responses in a VE: subjective, performance-based, and physiological measures [3]. Each method enables the collection of a specific type of information regarding the user's responses to the application. For instance, subjective measures evaluate the user's satisfaction, fatigue, intuitiveness, and preferences. (collected via surveys). Performance measures evaluate the user's behavior when performing a task with the VR application. Finally, the physiological measures evaluate nonvoluntary responses of the human body during and immediately after the test session.

Specific to haptic-based applications, subjective measures dominated the form of user evaluation. Usually the user feedback is expressed as a mean opinion score (MOS). In [10] and [11], the authors use MOS to determine the overall QoE. Their main idea is to test network jitter and delay effect on haptic quality reflected on the user. Using different settings and jitter parameter, they test different scenarios using a networked haptic application. In one scenario, they test a hockey game, while in another they test a networked writing application. From the MOS values, the authors use multiple regression analysis to link QoE parameters with application-level parameters.

Previously, we have created a taxonomy for the possible parameters that can be used to evaluate a haptic audio virtual environment (HAVE). Moreover, the taxonomy's higher level organization was divided into QoS and UX, because we reckoned that QoS is an important part of QoE along with the UX. The UX is in turn divided into four subcategories: perception measures, rendering quality, psychological measures, and physiological measures. Perception measures is a user-center subcategory that mirrors how the user perceives the application. The rendering quality relates to the quality of the three major modalities in VR application, namely graphics, audio, and haptics. Psychological measures and physiological measures are complementary sets that indicate the user's state. Psychological measures subcategory reflects the state of the user through observation and user feedback, while physiological measures subcategory uses direct biological measurements, such as heart rate, to reflect the state of the user. Each category groups related parameters together forming an organizational taxonomy. Evaluators could choose the categories and parameters to include in their evaluation. For a full list of parameters please consult [12]–[14].

Moreover, in [12], we have introduced the mathematical model along with the preliminary evaluation of two haptic-enabled applications: haptic learning system and a haptic Unified Modeling Language (UML) case tool. The weights were arbitrary chosen based on expert opinion. In subsequent work, we slightly modified the mathematical model and applied it to the evaluation of a haptic game entitled Balance Ball game. We used a systematic approach in normalizing the results, determining the weights, and validating the model [15].

This paper extends our previous work by considering various techniques in determining the weights to be used in the mathematical model. The results based on those weights are analyzed to observe the optimal technique for the evaluation of QoE of haptic-based applications.

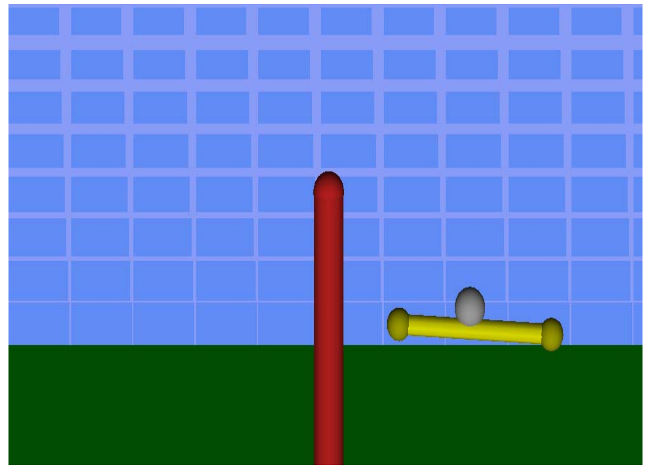

Fig. 1. Screenshot of the Balance Ball haptic game.

III. User Study Description

A. Application Description and Experimental Setup

The application we used to test the proposed model is the Balance Ball game [16]. A screenshot of the game is shown in Fig. 1. A ball is placed on a long wooden board that is held by two players from each side. The game involves the two users collaborating in maintaining the balance of a virtual ball on a board using remote haptic devices. Each player holds one end of the board with his/her haptic device and raises it slowly over a virtual pole to a predefined end mark. The challenge is to collaborate in an attempt to keep the board horizontally balanced as much as possible from the initial location to the destination. Any variation in the horizontal balance will cause the ball to roll away toward one side thus penalizing both the players. The players should remedy that by using the force feedback and the 3-D graphics to apply their judgment in balancing the board again. The score consists of the task completion time and the variations of the ball's position from the middle of the board.

The experiment took place at the haptic laboratory of the DISCOVER Lab at the University of Ottawa. Twenty-two users participated in the experiment. The collaborative application was ran on two computers. The computers were running WinXP SP3 on a 2× Intel Xeon 2.8 GHz with 2 GB of RAM and an Nvidia QuadroFX 2000XGL 128 MB video card. Each computer had a Phantom Desktop haptic device attached to it. The Phantom Desktop is a six degrees of freedom (DOF) positioning and sensing haptic device developed and marketed by SensAble Technologies, Inc. It has a compact design and provides three DOF feedback capabilities. A snapshot of the experiment setup is shown in Fig. 2.

Users were selected randomly from the School of Electrical Engineering and Computer Science department at the University of Ottawa (8 females and 14 males). Fifteen users were familiar with haptic devices (among them 11 users with

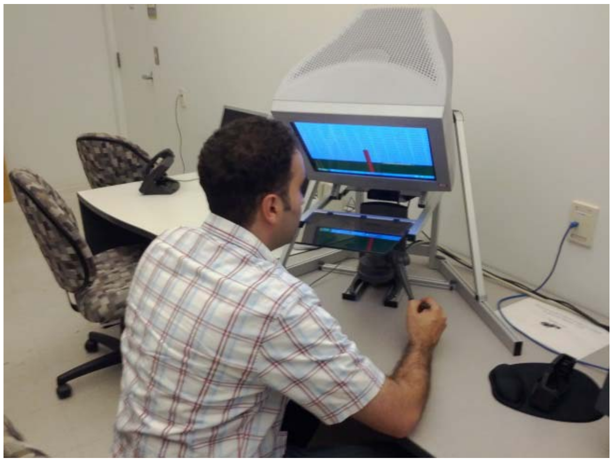

Fig. 2. Experimental setup of the user study conducted. The figure displays the station setup of one of the users during his game play.

previous working haptic experience), while the other seven users were new to the haptic notion. In either case, the user was given a general background about the application, how to handle and hold the haptic device, and what are the goals of the experiment. Users were reminded that the purpose of the experiment is to evaluate the application and not the users' abilities.

Users were divided into eleven teams randomly. Each team consisted of two users, and the experiment lasted on average $\sim$ 15 mins, which included playing the game twice (the first time was a trial run of the application, while the second one was the actual game). In some instances, users performed more than one trial of the application until they got comfortable with the hardware. After the team finished playing the virtual game collaboratively by reaching their destination in the actual game, they were asked to fill out a questionnaire with general questions about the virtual game, past haptic experience, and specific questions that reflect elements of their experience which are described precisely in the following section. Moreover, the questionnaire presented the opportunity for the user to give an overall QoE evaluation of the application which would be used to validate the mathematical model results (the computed QoE from the mathematical model). Each user filled a separate questionnaire.

B. Parameter Selection

The taxonomy in [12] groups the QoE parameters into certain categories for organizational purposes. Five parameters from the taxonomy were selected that are relevant to any haptically rendered 3-D collaborative game application such as the Balance Ball game that we tested. They are listed below along with a description of each parameter as well as the reason it was selected. The category from which the parameter is selected is written in parenthesis. If the parameter is not from the QoS category, then it will be from a subcategory of the UX.

1) Media synchronization (QoS parameter). There are usually three media modals in an HAVE application. Any miss-synchronization between the audio, video, and haptics can cause a drastic loss of perception of both the media that are miss-synchronized. Therefore, media synchronization is necessary for players to maximize their perception and enjoy the game. In this particular case, we focus on the subjective aspect of media synchronization from the user's point of view (even though it can be analyzed through equations, our focus is on user's perspective and experience).
2) Fatigue (perception measures parameter). Research has shown that fatigue, which is caused by muscle exhaustion, is linearly distributed as a function of time [17]. Fatigue is a crucial parameter because the haptic application needs users to interact with the virtual environments by exerting force and it induces fatigue easily compared to audio-visual feedback. Depending on the specifics of an application and on the haptic device used, rapid fatigue can hinder users and limit their rapport with the application. On the other hand, if the application minimized the users' fatigue, then their experience will be more positive.
3) Haptic rendering (rendering quality parameter). Haptic rendering quality remains the same until we reach a threshold [that is usually referred to as the just noticeable difference (JND)] after which the quality starts decaying [18]. For any haptic application, we want the quality to remain above that threshold, otherwise any instability, low resolution, or low haptic fidelity will render the application virtually unrealistic from the user's point of view.
4) Degree of immersion (psychological measures parameter). Even though the degree of immersion will cause a difference in quality, this difference is still not quite understood [19]. However, immersion in gaming application is of importance, because the more the users are immersed in the game, the more they are involved and experiencing enjoyment [20].
5) User intuitiveness (perception measures parameter). User intuitiveness is an important phenomenon that has been considered in disciplines other than human–computer interaction, such as nursing [21]. Although the factors that contribute to intuitiveness are less known, it can be observed through swift and determined actions of the user. It can be determined through user feedback as well.

The selection of the parameters is tailored toward the new medium experience and game experience holistic theory. The new medium (haptic) is emphasized by the two parameters: media (graphic and haptic) synchronization and the haptic rendering parameter.

Game experience psychologists have divided the experience for users of digital games into two categories; immersion and flow [22]. Under flow, there are certain characteristics that model an acceptable level of enjoyment for the user. Most importantly, the interface should not be too cumbersome and it should be responsive to the user (intuitiveness). In addition, the lack of fatigue will increase the flow and enjoyment level among users [23].

TABLE I
SUMMARY OF THE QUESTIONNAIRE RESULTS

Parameter	Mean	Standard Deviation
Media Synchronization	3.95	0.88
Fatigue	1.77	0.90
Haptic Rendering	4.18	0.89
Degree of Immersion	4.09	0.79
User Intuitivenss	4.09	1.04
Overall Rating (%)	84.64	13.00

All the parameters are out of 5, except the overall rating which is a percentage. The table presents the average values of the 22 participants in the experiment along with the standard deviation.

The parameters selected represent all the categories in the taxonomy (considering that both the psychological and physiological measures represent the user state). This was convenient to the evaluation process because we wanted to conduct an overall QoE evaluation. An overall QoE evaluation would present the general mental association of the user to the quality of the application. Moreover, because the parameters selected stems from game theory, they are relevant to any haptically rendered 3-D collaborative game application such as the Balance Ball game that we tested.

C. Questionnaire Results

The results of the questionnaire are summarized in Table I. The questionnaire presented to the users was a Likert-like scale developed for this experiment. Essentially, most questions are followed by a five-point scale in which the users are required to circle the point that is closest to their level of agreement. Each extreme poles of the scale are marked by opposing descriptive labels based on the content of the questions. The descriptive labels help the user reflect on the question as he/she is completing the questionnaire. A sample question presented to the user was the following:

To what extent do you think the haptic feedback was useful?

Not useful Completely useful

1 2 3 4 5

There was mostly one-to-one mapping between the user preference and the evaluation assigned to the parameter, such that one question corresponded to the evaluation of the parameter by the user. The focus was on the preference of the user, while the performance was used in another research area [16]. Our goal here was to quantify the subjective evaluation of the user without including the performance metrics.

The users rated the overall QoE as a percentage to get a more precise value for several reasons. The value is important as it is the ultimate goal of the evaluation. A precise value would ease the validation of the model and enhance error calculations. Hence given the users' percentage rating of the application, we will have two QoE values to compare: one from the user and another from the mathematical model described in Section IV.

D. Standardizing Data

To calculate a weighted average of the QoE, two modifications need to be performed for the data to adhere to two rules that we have specified. The first rule is that all the values selected by the users in the questionnaire should be converted to a normalized number between zero and one. This will facilitate the calculation of the QoE value in a percentage format. The second rule dictates that all the values should be in ascending order. That is, the higher end of the value of the parameter indicates better rating, while the lower end indicate worse evaluation.

We applied the two modifications necessary to the results. To normalize the numbers, we applied the following formula (x-min)/range, where x is the Likert-scale value selected by the user and range = max−min (max is the maximum value that can be selected by the user and min is the minimum value).

Looking at the results in Table I, all the parameters present follow rule two except fatigue. Higher fatigue ratings in the questionnaire indicate that the application causes higher fatigue. In this case, fatigue is undesirable and higher subjective fatigue values will degrade the user's satisfaction. Modification of the fatigue Likert values was done by subtracting the normalized fatigue value from one. We applied 1 (norm. fatigue value) according to [24]. We renamed the variable as comfort.

IV. MATHEMATICAL MODEL DESCRIPTION

This section describes the mathematical model where the QoE is computed as the weighted linear combination of the QoS and UX for a particular haptic user interface. In turn, the QoS is computed as a weighted linear combination of the parameters in the QoS category. For the UX, each subcategory is adjusted by weights of its own. Moreover, the subcategories are treated as a weighted average of their own parameters.

The mathematical model equations are as follows:

$$\mathrm{QoE} = \zeta \times \mathrm{QoS} + (1 - \zeta) \times \mathrm{UX} \tag{1}$$

where

$$\mathrm{QoS} = \frac{\sum_l \eta_l S_l}{\sum_l \eta_l} \tag{2}$$

and

$$\mathrm{UX} = A \frac{\sum_i \alpha_i P_i}{\sum_i \alpha_i} + B \frac{\sum_j \beta_j R_j}{\sum_j \beta_j} + C \frac{\sum_k \gamma_k U_k}{\sum_k \gamma_k}. \tag{3}$$

The symbols are defined as follows:

1) ζ controls the relative weight given to the QoS parameters compared to the user experience parameters.
2) S_l, P_i, R_j, U_k represent the quality values given to the individual parameters of QoS measures (S_l), perception measures (P_i), rendering quality measures (R_i), and user state measures (U_k).
3) A, B, C are empirically determined weighing constants for the respective perception measures, rendering quality measures, and user state measures.

4) η_l, α_i, β_j, γ_k are weighing factors which depend on the relative quality value of individual user experience parameters underneath QoS measures, perception measures, rendering quality measures, and user state measures, respectively.

If the quality factors are restricted between 0 and 1, then the overall QoE will also have a value between 0 and 1 (i.e., $0 \leq$ QoE ≤ 1). To achieve this condition, the constant coefficients A, B, and C in (3) should satisfy the constraint

$$A + B + C = 1. \tag{4}$$

In this mathematical model, we have combined both the psychological and physiological categories from the taxonomy presented in [12] into one category called user state. If physiological parameters are included for physical validation in the future, it can be combined with psychological measures, because both categories reflect the user state.

This three-tier organization of the mathematical model allows the evaluator to retain a greater control of the equations from the higher level to the lower one. As an example, supposedly the evaluator wants to focus all his/her study on UX, while maintaining the effects of the QoS constant. In this case, ζ would be set to zero then the full weight would go to UX in the higher level equation. Moreover in (3), supposedly the effect of the user state is negligible, then C could be set to zero and (4) becomes A + B = 1 and the weights then can be distributed between A and B.

V. Determining Weights

We have conducted five approaches in determining the weights to be used within the mathematical model: even weight distribution, correlation values, even weight/correlation amalgamation, regression analysis, and PCA. This section describes these approaches as well as the results obtained.

A. Even-Weight Distribution

Even-weight determination is used when there is no preference on which is the most dominant parameter involved [24]. In this approach, the weights are determined based on the number of the parameters involved, to maintain an equal distribution of weights to parameters ratio [24]. For the top-level equation, we have one QoS parameter and four UX parameters (the parameters that are presented in Table I). Therefore $\zeta = 1/5$ and the top-level equation becomes 0.2 QoS + 0.8 UX.

The same process applies to (3). The A, B, and C weights are determined by the number of parameters involved. In our case, we have two parameters from the perception measures category weighted by A, one parameter from rendering quality category weighted by B, and one parameter from the user state category weighted by C. Because we have a constraint in (4) that the sum of A, B, and C should be equal to one, we can calculate the weights values as follows: A = 0.2, B = C = 0.4.

The designation of B = C and B = 2A come from the fact that because A weighs two parameters, then it should be half of the weight that controls only one parameter. This way each category contributes equally to the UX calculation [24].

The weights associated with each individual parameter (η_l, α_i, β_j, γ_k) can be set to one which allows the single parameters in a category to have full weight, while if a category has multiple parameters each one of those parameters can be equally weighted.

B. Weight Based On Correlation

Our next approach calculates the weight of a parameter based on the correlation of that parameter with the overall rating of the application. The higher the correlation value, the more weight is the parameter encompasses. The correlation formula used is

$$r = \frac{\sum[(x - Mx)(y - My)]}{\sqrt{(SSx)(SSy)}} \tag{5}$$

where X is the overall QoE rating and Y is the intended parameter (with means of Mx, My and sum of squared deviation of SSx, SSy respectively).

To determine A, B, and C of (4), which define the weight of each parameter category of the user experience, we calculated an aggregate correlation value of that parameter category (e.g., perception measures) and the overall QoE rating, using (5). The aggregate correlations for categories with single parameters are equivalent to the correlation of the parameter. For categories with multiple parameters, the aggregate correlation was an overall correlation value (instead of averaged correlation) according to [25]. The values were normalized to satisfy (4). For weight parameters of (1), we used aggregate correlation values normalized as well to add up to one.

C. Combination of Even Distribution and Correlation

This approach was performed as an amalgamation of the previous two approaches. We set the weights of the categories [in (1) and (4)] as even-distributed weights, while we varied the weights of multiple parameters in the same category. The way we treated those multiple parameters is to set their weight according to the correlation value between the parameter and the overall user evaluation which can be inferred from the questionnaire results.

D. Linear Regression Analysis

Regression analysis models the relationship between different variables. Linear regression model is a type of regression analysis where there is a regressor x and a response y related through a straight line, represented by (6) [26]

$$y = \beta_o + \beta_1 x + \varepsilon \tag{6}$$

where β_o, β_1, and ε are the intercept, slope, and random error component, respectively.

The coefficient of determination (R^2), associated with the linear regression model, describes how well the line of the linear regression fits the data.

Studying the behavior of the data, we assumed a linear regression fit between the input (parameter evaluation) and the QoE output of the user. Human behavior is not linear, but this is the best fit that the data provided (with certain acceptable

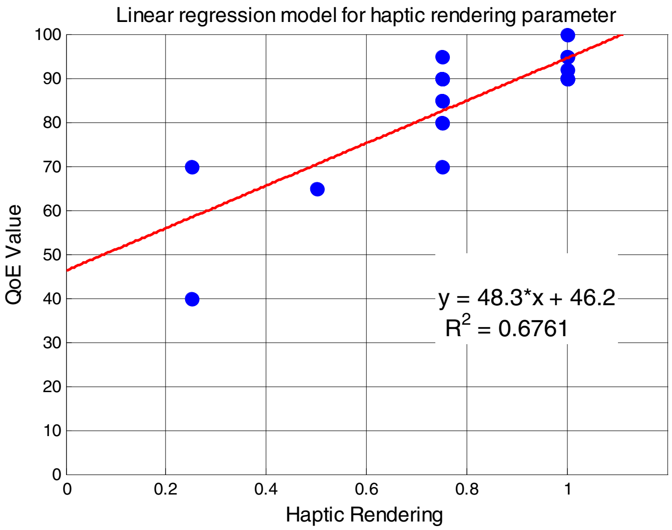

Fig. 3. Linear regression graph for haptic rendering parameter. The equation is also displayed. The haptic rendering values are the Likert value for each user normalized between 0 and 1.

TABLE II
CORRELATION MATRIX OF THE PARAMETERS

Parameter	A	B	C	D	E	F
A	1	0.59	0.42	0.4	0.4	0.68
B	0.59	1	0.4	0.42	0.22	0.53
C	0.42	0.4	1	0.3	0.33	0.82
D	0.4	0.42	0.3	1	0.49	0.58
E	0.4	0.22	0.33	0.49	1	0.58
F	0.68	0.53	0.82	0.58	0.58	1

Parameters: A is Media Synchronization, B is Comfort (Fatigue inversed), C is Haptic Rendering, D is Degree of Immersion, E is User Intuitiveness, and F is overall QoE rating. Correlation is significant ($p < 0.05$) if it is greater than 0.36.

error range as described in the analysis section), and the most convenient given the linear relationship of the weighted average methodology. Other evaluation methodologies which assume nonlinear relationship also exist [13].

The linear regression model for the haptic rendering parameter is provided in Fig. 3 as an example. The QoE value of the user is the response to the regressor which is the haptic rendering parameter in this case. The weight of the parameter would be equal to the strength of the linear regression model represented by the slope of the line multiplied by R^2.

We have also attempted a linear regression model aggregated for user experience parameters and QoS parameters to compute (1). Moreover, for each category of user experience parameters, we calculated an aggregate linear regression model to satisfy (4). The values obtained for the linear regression model after optimization are $\zeta = 0.57$, $\alpha_1 = 047$, $\alpha_2 = 0.53$, A $= 0.17$, B $= 0.62$, C $= 0.21$, where ζ is used for (1), α_1 and α_2 are for multiple perception measures parameters (comfort and intuitiveness) and A, B, and C satisfy (4).

E. Principal Component Analysis

PCA is a statistical technique that has been used in various fields such as finite data representation [27] and sound recognition [28]. PCA transforms the original set of data into a simplified set of data by removing any redundancy that is present in the data [29].

Table II displays the correlation matrix of all the parameters including the overall QoE rated by the user. The table shows that most of the parameters correlate significantly with one another. This suggests that there is redundant information present in the data. By performing PCA, we remove this redundancy and maintain a set of uncorrelated variables.

The advantages of PCA are threefold.

1) Minimizing random error that may arise from one of the measures taken.
2) Eliminating redundant data from the variables.
3) Determining how much each parameter weighs in the model and which one has the highest weight.

Our PCA results produced similar number of uncorrelated variables but some of those variables had really low weight. Most notably is the perception measures variables had a weight of 0.02 (A $= 0.02$). The weight of QoS variables could also be eliminated with PCA ($\zeta = 0.01$).

VI. RESULT ANALYSIS AND DISCUSSION

The weights derived in each approach were inserted in the mathematical model and a quantified QoE was calculated. For each user, there are two sets of results: one derived by the mathematical model and one provided by the user as a subjective evaluation of the overall system. The QoE value quantified by the mathematical model varied by the weight approach. The values produced by the mathematical model were validated against the users' own evaluation of the application. Table III summarizes the results obtained.

The average calculated QoE value represents the overall QoE of the application using the mathematical model and given a certain weight approach. The values along with the standard deviation vary only slightly within the different approaches. Actually, the similarity between some approaches can be noticed. For example, the even-weight distribution and the weight–correlation combination are almost identical in all the attributes. This is because of the fact that the parameters correlation values are very close in magnitude as well. We had one category with multiple parameters. The perception measures under the UX had two parameters. The correlation values for both the parameters with the overall QoE were 0.53 and 0.58 for comfort and user intuitiveness, respectively. No matter what the magnitudes of the weights are, if they are similar in value, then the result of the weighted average would be the same (or close to the equal weights result).

On the other hand, just by considering correlation values as weights (approach 2), then the results differ from approaches 1 and 3 described earlier. Correlation weights produce higher percent error and higher average of differences. The average of differences is defined as the average value of all the differences between the user ratings and the QoE ratings obtained from the given mathematical model under a specific approach.

TABLE III
SUMMARY OF THE RESULTS OBTAINED BY DIFFERENT WEIGHT APPROACHES

Attribute/Approach	Even weight distribution	Correlation weights	Weight - correlation combination	Linear regression analysis	PCA
Average calculated QoE value	77.59 ± 15.91	76.18 ± 17.22	77.58 ± 15.92	76.053 ± 17.74	78.71 ± 16.22
Average of differences	7.77	9.51	7.76	10.07	7.47
Range of differences	0 - 18	0 – 34.2	0 - 18	0 - 35.9	0-17
Percent Error (%)	9.95	12.04	9.95	12.78	9.51
Correlation (p<0.005)	0.92	0.86	0.92	0.86	0.89

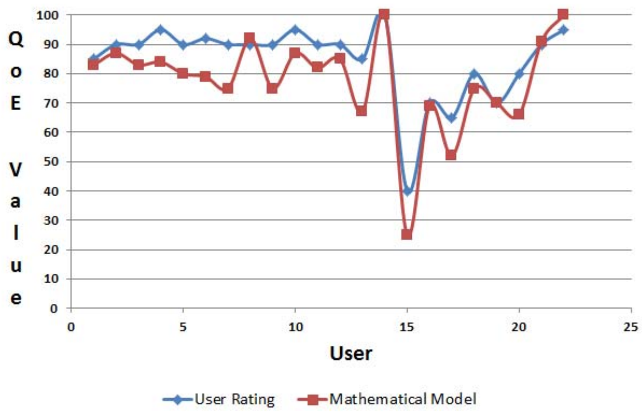

Fig. 4. Users' ratings versus even-weight distribution results. The mathematical model results are based on the even-weight distribution.

The highest average of differences and percent error stems from the linear regression approach. The linear fit had low coefficient of determination values in some instances which suggest that the data may deviate from the linear regression fit and this could be the reason for the high error rate.

The range of differences attribute lists the minimum difference encountered between user rating and the mathematical model result in a given approach. This attribute does not allow us to infer conclusions because these values are sporadic and occurred mostly once within the data. They were included here for interested readers.

All the approaches correlated significantly with the user ratings. The correlation of even-weight distribution mathematical model and user rating is displayed in Fig. 4. The correlation of each approach additionally validates the mathematical model. In Fig. 4, we have plotted the data of the 22 users. One line shows the users' QoE value and the another line shows the values computed by the mathematical model. The values follow the same pattern and coincide sometimes. This indicates a high correlation pattern. The correlation value was computed to be 0.92, $p < 0.005$ (DOF $= 20$). This is the highest correlation value in all the approaches (the weight-correlation combination approach had similar correlation). This means that the mathematical model results significantly follow the users' ratings of the haptic application.

The PCA approach seems to have a slightly better percent error and average of differences than the rest of the approaches. Removing the redundancy of the variables had a certain effect which could be attributed to the advantages of PCA. The correlation of PCA mathematical model results and user ratings is not the highest but it is significant and relatively high.

VII. CONCLUSION

This paper presented a mathematical model capable of quantifying the QoE of users when utilizing haptic-based applications. This paper also presented the user study conducted and how the data were applied in the mathematical model. It focuses on approaches for weight determination of the mathematical model to investigate which approach produces the more accurate results when utilized.

The results suggest that there was a variation on how the approaches faired with certain degree of similarity. Linear regression analysis and correlation weights produced the highest error rates. PCA and even-weight distribution produced the best results with PCA generating a slightly smaller error rate. A combination of even weights and correlation values for multiple parameters produced similar results as even-weight distribution approach.

We have attempted the even-weight approach previously [15] and our goal was to reduce the percent error by considering different and more advanced approaches such as the linear regression and PCA. Although the PCA approach managed to reduce the error slightly, it is still similar in range with the even-weight approach. Our conclusion is that there will always be some limitations when modeling human behavior through a mathematical model and certain range of error is expected. With PCA and even-weight distribution, this range of error is minimal and acceptable.

In accordance, the five approaches can be used in weight determination, but our recommendation is to use even-weight distribution or PCA if the evaluators do not desire the weights to be even. Our future work includes further testing with different types of applications and experimental setups.

REFERENCES

[1] M. Orozco and A. E. Saddik, "AdHapticA: Adaptive haptic application framework," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 9, pp. 1840–1851, Sep. 2008.

[2] A. E. Saddik, "The potential of haptics technologies," *IEEE Instrum. Meas. Magn.*, vol. 10, no. 1, pp. 10–17, Feb. 2007.

[3] T. E. Whalen, S. Noel, and J. Stewart, "Measuring the human side of virtual reality," in *Proc. IEEE VECIMS*, Jul. 2003, pp. 8–12.

[4] A. Guerraz, C. Loscos, and H. R. Widenfeld, "How to use physical parameters coming from the haptic device itself to enhance the evaluation of haptic benefits in user interface," in *Proc. Eur.*, 2003, pp. 137–145.

[5] R. Chaudhari, E. Steinbach, and S. Hirche, "Towards an objective quality evaluation framework for haptic data reduction," in *Proc. WHC*, 2011, pp. 539–544.

[6] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang, "Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 481–490.

[7] R. Jain, "Quality of experience," *IEEE Multimedia*, vol. 11, no. 1, pp. 96–95, Aug. 2004.

[8] T. Ebrahimi, "Quality of multimedia experience: Past, present and future," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 3–4.

[9] L. De Marez and K. De Moor, "The challenge of user-and QoE-centric research and product development in today's ICT-environment," *Observatorio*, vol. 3, no. 1, pp. 1–22, 2007.

[10] Y. Kusunose, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE assessment in networked air hockey game with haptic media," in *Proc. 9th Annu. Workshop Netw. Syst. Support Games*, 2010, pp. 1–2.

[11] A. Tatematsu, Y. Ishibashi, N. Fukushima, and S. Sugawara, "QoE assessment in haptic media, sound and video transmission: Influences of network latency," in *Proc. IEEE Int. Workshop Tech. Committee CQR*, Jun. 2010, pp. 1–6.

[12] A. Hamam, M. Eid, A. E. Saddik, and N. D. Georganas, "A quality of experience model for haptic user interfaces," in *Proc. Ambi-Sys Workshop Haptic User Inter. Ambient Media Syst.*, 2008, pp. 1–6.

[13] A. Hamam, M. Eid, A. E. Saddik, and N. D. Georganas, "A fuzzy logic system for evaluating quality of experience of haptic-based applications," in *Proc. 6th Int. Conf. Haptics, Perception, Devices Scenarios*, 2008, pp. 129–138.

[14] A. Hamam and N. D. Georganas, "A comparison of Mamdani and Sugeno fuzzy inference systems for evaluating the quality of experience of hapto-audio-visual applications," in *Proc. IEEE Int. Workshop HAVE Games*, Oct. 2008, pp. 87–92.

[15] A. Hamam and A. El Saddik, "Evaluating the quality of experience of haptic-based applications through mathematical modeling," in *Proc. IEEE Int. Workshop HAVE Games*, Oct. 2012, pp. 56–61.

[16] H. Al Osman, M. Eid, and A. El Saddik, "Evaluating ALPHAN: A network protocol for haptic interaction," in *Proc. Symp. Haptic Inter. Virtual Environ. Teleoperator Syst.*, Mar. 2008, pp. 361–366.

[17] R. Seroussi, M. H. Krag, P. Wilder, and M. H. Pope, "The design and use of a microcomputerized real-time muscle fatigue monitor based on the medical frequency shift in the electromyographic signal," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 2, pp. 284–286, Feb. 1989.

[18] M. A. Srinivasan and C. Basdogan, "Haptics in virtual environments: Taxonomy, research status, and challenges," *Comput. Graph.*, vol. 21, no. 4, pp. 393–404, 1997.

[19] F. Gutierrez, J. Pierce, V. M. Vergara, and R. Coulter, "The effect of degree of immersion upon learning performance in virtual reality simulations for medical education," *Stud. Health Technol. Inform.*, vol. 125, pp. 155–160, Jan. 2007.

[20] W. IJsselsteijn, Y. de Kort, K. Poels, A. Jurgelionis, and F. Bellotti, "Characterising and measuring user experiences in digital games," in *Proc. Int. Conf. Adv. Comput. Entertainment Technol.*, 2007, pp. 1–3.

[21] V. G. Miller, "Measurement of self-perception of intuitiveness," *Western J. Nursing Res.*, vol. 15, no. 5, pp. 595–606, 1993.

[22] W. IJsselsteijn, Y. de Kort, K. Poels, A. Jurgelionis, and F. Bellotti, "Characterising and measuring user experiences in digital games," in *Proc. Int. Conf. Adv. Comput. Entertainment Technol.*, 2007, pp. 1–3.

[23] P. Sweetser and P. Wyeth, "GameFlow: A model for evaluating player enjoyment in games," *Comput. Entertainment*, vol. 3, no. 3, pp. 1–3, 2005.

[24] T. Tullis and B. Albert, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*, 1st ed. San Mateo, CA, USA: Morgan Kaufmann, 2008.

[25] B. Monin and D. M. Oppenheimer, "Correlated averages versus averaged correlations: Demonstrating the warm glow heuristic beyond aggregation," *Soc. Cognit.*, vol. 23, no. 3, pp. 257–278, 2005.

[26] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 4th ed. New York, NY, USA: Wiley, 2006.

[27] T. FdSzabo and G. Horvath, "Finite word length computational effects of the principal component analysis networks," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 5, pp. 1218–1222, Oct. 1998.

[28] W. Huadong, M. Siegel, and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," *IEEE Trans. Instrum. Meas.*, vol. 48, no. 5, pp. 1005–1009, Oct. 1999.

[29] J. Sauro and E. Kindlund, "A method to standardize usability metrics into a single score," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2005, pp. 401–409.

Abdelwahab Hamam (S'04) is a Ph.D. student in Electrical and Computer Engineering at the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada. He received his M.A.Sc. in Electrical Engineering from the University of Ottawa in 2007. He is a member of the DISCOVER Lab research group supervised by Dr. Abdulmotaleb El Saddik. His research interests include haptic applications, medical simulations, and quality of experience for virtual environments.

Abdulmotaleb El Saddik (F'09) is University Research Chair and Professor in the School of Electrical Engineering and Computer Science at the University of Ottawa. He held regular and visiting positions in Canada, Spain, UAE, Germany and China. He is an internationally-recognized scholar who has made strong contributions to the knowledge and understanding of multimedia computing, communications and applications. He has authored and co-authored four books and more than 400 publications. Chaired more than 30 conferences and workshop and has received research grants and contracts totaling more than $18 Mio. He has supervised more than 100 researchers. He received several international awards among others ACM Distinguished Scientist, Fellow of the Engineering Institute of Canada, and Fellow of the Canadian Academy of Engineers and Fellow of IEEE and IEEE Canada Computer Medal.