Stevens' Power Law in 3D Tele-immersion

Towards Subjective Modeling of Multimodal Cyber Interaction

Sabrina Schulte
RWTH Aachen University
sabrina.schulte@rwth-aachen.de

Shannon Chen and Klara Nahrstedt
University of Illinois at Urbana-Champaign
{cchen116, klara}@illinois.edu

ABSTRACT

In this paper we verify the insufficiency of Stevens' power law to describe the relationship between QoS and QoE factors. User studies that target different types of application scenarios of 3D Tele-immersion (3DTI) are conducted and the results show no significant power trend in the relationship between packet loss and perceptual quality metrics. We further verify that activity characteristics, activity objectives, and users' roles in the 3DTI session also have profound effects on the service quality aside to the QoS level. Thus, simple one-factor psychophysical laws are inadequate of serving as a QoS-QoE mapping model.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: Models and Principles – *human factors*

General Terms

Experimentation, Human Factors, Measurement

Keywords

3D Tele-Immersion; QoE; Stevens' Power Law

1. INTRODUCTION

Quality assessment of distributed interactive multimedia services has shifted from traditional quality of service (QoS) to quality of experience (QoE). However, the time and human effort for compiling user feedbacks under controlled environment makes QoE-based system adaptation infeasible for it cannot keep up with the network dynamics. Thus, mapping models that connect real-time measureable QoS metrics with QoE are essential for efficient utilization of networking resources.

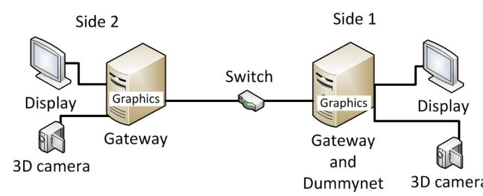
Since more than a decade ago, the thriving of VoIP services (e.g., Skype [2]) urges the investigation of such mapping models to aid the adaptation of bandwidth usage. Power based [3] and log based [6] perceptual models for VoIP scenarios are proved to be practical foundations for network adaptation modules. While elegant closed-form models well describe simple one-dimensional VoIP services, we would like to ask, with the complication of multimodality and application level semantic: *does the elegant psychophysical model still hold in complex multimedia applications such as 3D Tele-immersion?*

The potential of 3D Tele-immersive (3DTI) services has gained its attention from both academia and industry. The development of

MM '14, November 3 – 7, 2014, Orlando, FL, USA

ACM 978-1-4503-3063-3/14/11…$15.00.
http://dx.doi.org/10.1145/2647868.2654998

(a) 3DTI activity



(b) System model

Figure 1. 3DTI system

3DTI systems is currently aiming towards multi-purpose, multi-sites, and multi-modal [7, 9] platforms in order to enable a variety of user activities. In its system view, an efficient adaptation module is essential for the inevitably high bandwidth demand to support its interactive characteristic and complex 3D rendering [8]. Thus, mapping model of QoS-QoE again plays an important role. This paper answers a crucial question on the journey of model building: can (or to what extent can) psychophysical law help us?

Psychophysical laws are results of quantitative psychology [5], a study with long history that well serves the purpose of bridging interdisciplinary studies. The idea is to construct mathematical formula that addresses the relationship between physical stimulus (Φ) and quantified perception (Ψ). Some well-verified laws include the Weber-Fechner Logarithmic Law [5] and the Stevens' Power Law [5]. However, these formula merely focus on simple one-factor stimuli and single human sensory system. For example, the relationship between voltage of the electric shock and level of pain, or weight of an object and the stress perceived when lifting it.

In this paper, we are interested in much more complicated stimulus-perception pair: the packet loss rate versus the perceptual quality of different 3DTI activity sessions. We set the stage with two 3DTI sites with two sets of experiment that target two different application level scenarios. The first one targets fast activities, where users play against each other in a virtual sword fight; while the second experiment targets slow activity, where users play the roles of doctor and patient in a remote physiotherapy session.

We manipulate the condition of the network that connects the two sites and recruit real human users to participate and evaluate the quality of the experiment sessions. Comprehensive user feedbacks including general service quality, sense of control, interactive-ness, win/loss ratio (for gaming activity), sense of achievement, and sense of time are compiled during the experiment. Analysis shows that the oversimplified psychophysical laws do not describe the relationship of multimodal, interactive services as we experience in 3DTI. With the semantic and sensory complication of the application, we need more sophisticated model building framework to capture the intrinsic relationship of QoS-QoE that involves environmental and social confound factors [10].

The rest of the paper is organized as follows. In Section 2 and 3 we introduce the system model and the psychological model under verification. In Section 4 we describe the targeted 3DTI activities and the experiment flow. In Section 5 we present and analyze the experiment results. Finally, in Section 6 we conclude.

2. SYSTEM MODEL

The purpose of 3DTI system is to allow two physically dispersed parties to interact with each other in a virtual space. As shown in Figure 1 the system includes two sites, each consists of a 3D Kinect camera, a gateway with a graphic engine, and a display.

The camera records the user in front of it. This data stream is sent to the local gateway and passed on to the other gateway through the network. Streams received from the local camera and the remote gateway are merged in the graphics engine in the local gateway to render an in-synced virtual space. Artificial elements can also be added into the virtual space via the graphics engine.

In our experiment setting, the two gateways are connected via a LAN switch so that we have full control of the network conditions using *dummynet* [1]. The network allows a maximum packet size of 1500 bytes and has a bandwidth of 100 Mbps.

The image of user in a site will be captured by the camera and presented as point cloud in its stream. The final virtual space is shown on the displays with resolution of 480 x 640 with frame rate around 20 fps.

3. PSYCHOPHYSICAL MODEL

In this section we discuss the psychological background for this paper. We are going to introduce two different psychophysical laws called Weber-Fechner Law [5] and Stevens' Power Law [5]. Both laws describe the relation between a physical stimulus and the sensational magnitude.

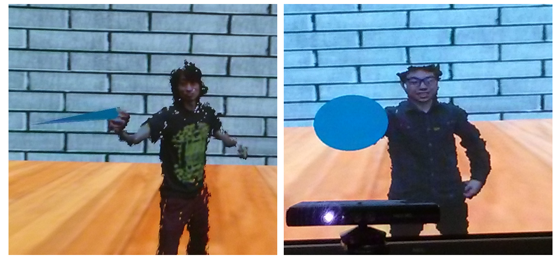
The Weber-Fechner Law describes the correlation between a physical stimulus and the sensation magnitude as a logarithmic function. Given a stimulus Φ and its responding quantified perception Ψ, the relationship between Φ and Ψ is described as

$$d\Psi = \mathrm{k} \times \frac{d\Phi}{\Phi}$$

where $d\Psi$ and $d\Phi$ are the difference of stimulus and the difference of perception, respectively, and *k* is a constant scalar. By integrating both sides of the equation

$$\Psi = \mathrm{k} \times \ln \Phi + c$$

where c is the constant introduced by integration. Hence, the Weber-Fechner law claims that the relationship between a relative stimulus-perception pair is logarithmic, which is proved to be true for most cases including loudness, brightness, sense of touch and heat, and even temporal, spatial, and numerical cognitions [5].



(a) Fast activity (b) Slow activity

Figure 3. Experiment activities

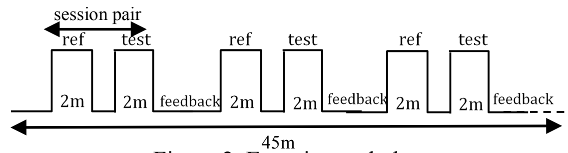


Figure 2. Experimental phase

Later, Stevens' Power Law is developed because Weber-Fechner Law only covers sub-linear relationships between stimulus and perception. Stimuli like electric shock and low temperature have exponential effects on sense of pain and cold [5].

Thus, to cover super-linear, linear and sub-linear relationships, Stevens' Power Law is defined as

$$\Psi = k\Phi^{\Theta}$$

where Θ is a power exponent that determines the relationship. When $\Theta > 1$ the relationship becomes exponential, and when $\Theta < 1$ the relationship is sub-linear. In view of the wider coverage of Stevens' Law, in the rest of the paper we focus on investigating the feasibility of using Stevens' Law to describe the QoS-QoE relationship of 3DTI activities.

4. 3DTI USER STUDY

4.1 Experimental Setup

We recruit eighteen users to participate in our user study. Four of them are female and fourteen are male. All users are affiliate to the *anonymous institute*. We use two separated 3DTI sites to simulate a distributed environment. Both sites are in the same lab but connected via *dummynet* and separated by a visual cover.

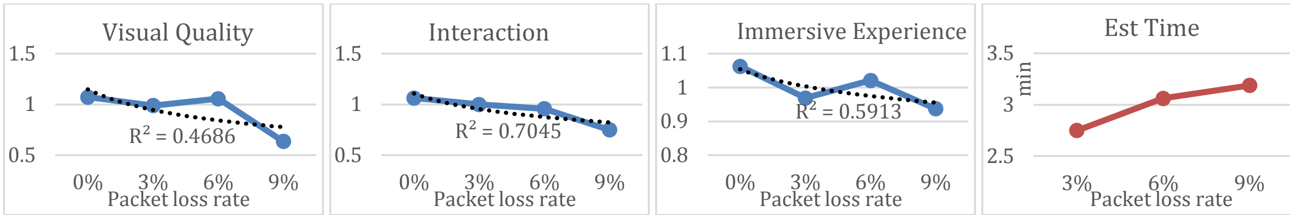
All users participate pairwise. At the beginning of each experiment we ask the user to read the description of the experiment and fill out the user information form. After that follows a training phase, which helps the user to get familiar with the activity and the system. We do not limit the time for the training phase, since it is important that the user understands how to interact through the system.

The experimental phase lasts 45 minutes and consists of four session pairs (Figure 2). Each session lasts two minutes. The first session of a pair is the reference session and the second session is the test session. After each session pair, users are asked to fill out a questionnaire. At the end of the whole experiment we ask them to fill out a final questionnaire which refers to the overall experience.

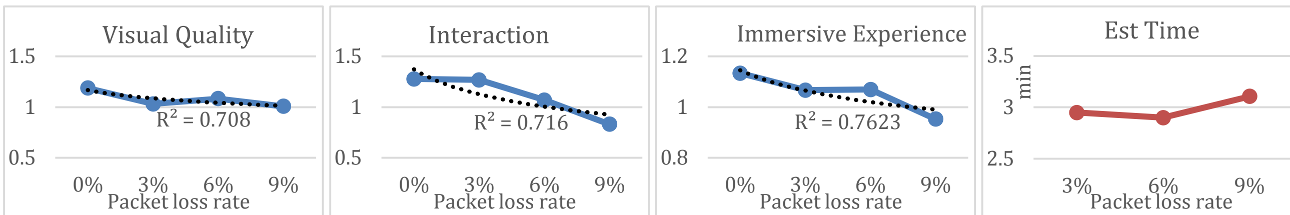
During the reference session we set the *dummynet* to simulate QoS of 4.5% packet loss rate in the link that connects the two sites. During the test session, 0%, 3%, 6%, and 9% packet loss rate are simulated in random order in the four session pairs.

4.2 3DTI User Activities

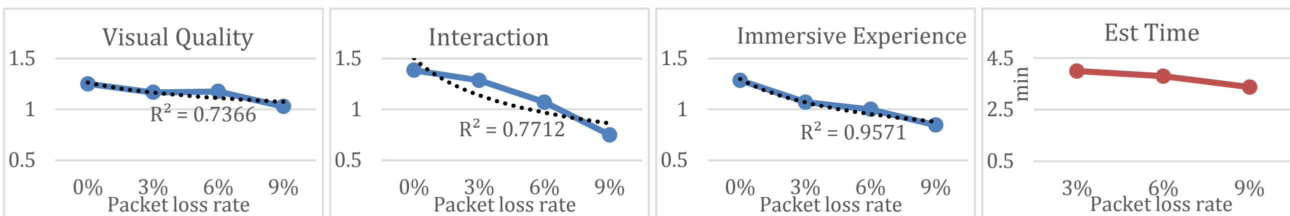
Based on the activity objectives and characteristics, users' tolerance towards quality degradation can be very different [7]. One



(a) Fast Activity



(b) Slow Activity



(c) Slow Activity: leader's feedback

(d) Slow Activity: follower's feedback

Figure 4. Experiment results

important characteristic of the activities that significantly affects the QoS-QoE relationship is the speed of movement and user interaction. Thus, we study two signature user activities in 3DTI scenarios: the virtual sword fight game, which represents fast activity; and the remote physiotherapy, which represents slow activity.

4.2.1 Fast Activity

In the fast activity we simulate a sword fight (Figure 3a). When the users come into the site and raise their right hand they see a sword attached to it in the virtual world. The sword follows the movement of their hands. In the top corners of the display there are the "life points" of the users - one for the user herself and one for her opponent. Each experiment session consists of multiple rounds with the users fighting against each other.

4.2.2 Slow Activity

For slow activity we design an application for remote physiotherapy session, in which one user plays the role of a leader (trainer) and the other plays the role of a follower (patient). When the leader comes into the site and raises his right hand in front of the camera, there will be a circle attached to his hand (Figure 3b). The leader's task is to move the circle to a certain position and then instruct the follower to touch the circle in the virtual world with a specified body part. As soon as the follower touches the circle, its color changes to signify a successful touch. The leader then moves it to a new position.

5. RESULTS AND ANALYSIS

In this section we show the results of the experiments and further show that Stevens' Power Law alone is insufficient to cover the QoS-QoE relationship in 3DTI. In the following discussions, we use the normalized MOS (Mean Opinion Score) value to represent the perceptual quality. For each experiment session, the users evaluate the quality by 5-point MOS scale [11]. The normalized MOS is obtained by normalizing the score of a test session by the score of the reference session. Thus, a normalized score greater than 1 indicates an improved quality from the reference while a less than 1 score indicates a degraded quality. Reflecting to Stevens' power law, the stimulus (Φ) is the packet loss rate and the quantified perception (Ψ) is the normalized MOS.

For fast and slow activities, we ask three subjective questions about the visual quality, the interaction, and the overall immersive experience. In addition, we ask the users to estimate the time lapse of an experiment session. This estimated time serves as an indicator of the user's *psychological flow* [4]. The flow measures "the holistic sensation that people feel when they act with total involvement", which is the main intrinsic motivation for people to perform activities that provide no discernible extrinsic rewards [4]. Losing track of time is one explicit trait of flow. We argue that a shorter-than-actual estimated time indicates full concentration supported by an acceptable service quality, while a longer estimated time indicates boredom and reluctance, which can be incurred by the unacceptable service.

5.1 General Analysis

In Figure 4 we show the normalized MOS given by users (y-axis) plotted against the packet loss rate (x-axis). Not surprisingly there is a positive correlation between QoS and QoE degradations, which incurs monotonically decreasing plots in the graphs. However, none of them show a significant trend that complies with the power law. Along with the normalized MOS (solid line) we also show the fitted power-based regression curve (dotted line) on each graph with its goodness-of-fit index (R-square value). As we can see, the small R-square values indicate that the Stevens' power law poorly describes the relationships between the degradation of packet loss (QoS) and the three QoE metrics (visual quality, interaction, immersion) we target in the experiment questions.

Comparing the effect of packet loss to different QoE metrics (columns 1, 2, 3 of Figure 4), we see that interaction metric is the most affected. Packet loss during content transmission results in screen freezes and decrease of frame rate. Therefore, for fast paced interaction (fast activity) and closely collaborative tasks (slow activity), the effect of packet loss is highly aware and causes interruption of the flow of both activities.

On the other hand, the effect of packet loss on the immersive experience is less predictable (column 3 of Figure 4). Comparing to the other two metrics, the immersive experience is a compound result of various variables that reflects many confound factors [10]. Thus, a simplified one-factor QoS-QoE psychological mapping like Stevens' law does not serve the purpose. While frameworks of identifying confound factors in delivery of internet video [10] and VoIP [6] services has been proposed, for intrinsic multimodal service as 3DTI, a comprehensive mapping that covers wide range of human, environmental, and social factors is yet to be seen to our best knowledge.

5.2 Activity Pace: Fast vs. Slow

In this section we focus on analyzing the difference between the two activities (Figure 4a and 4b). The results of ANOVA test shows significant differences regarding different aspects of service quality ($p<0.05$ for interaction and immersive experience). From previous studies of QoS, it is well expected that a visual content with higher level of motion would require a higher frame rate. Packet loss has direct effect on the frame rate perceived at the 3DTI site, thus causing the visual content not being able to keep up with the fast changing movement during the sword fight activity. Users report irritation of not being able to see the movement clearly and motion blur effect during the session, which cause the degradation of MOS feedbacks.

5.3 User Role: Leader vs. Follower

In this section we focus on the slow activity (Figure 4c and 4d). We group the feedbacks by the role a user played in the session to see the effect of activity objectives. From ANOVA of the plots, we see that the leader's QoE is more impacted by packet loss than the follower ($p<0.1$ for visual quality). This is due to the fact that the leader is in charge of the progress of a session. A leader has to give instructions and wait for the follower to react. Thus, QoE of the session depends heavily on whether a leader can accomplish her task smoothly. Frame freeze caused by packet loss confuses the leader about whether her instruction is correctly received or whether the follower has already accomplished the task. The confusion incur lower scores to the interaction and the overall immersive experience at the leader's site.

Comparing the session time lapse estimated by the leaders and the followers (column 4 of Figure 4c and 4d), we find that generally the leader has a longer feeling of time lapse (about double of the time lapse estimated by the followers). From the qualitative feedbacks of the leaders, we find frustration and loss of confidence due to loss of control. The effect of packet loss makes the leading of the session become a difficult and unpleasant task. On the contrary, the task of a follower is merely to receive instruction from the leader and to react accordingly. From the qualitative feedbacks, followers do not even notice the frame freeze sometimes because they think the pauses of the flow are caused by actual pauses of the leader. Thus, there are no specific negative comments towards the interactive-ness at the follower's site.

6. CONCLUSION AND FUTURE WORKS

In this work we verify the insufficiency of Stevens' power law to describe the relationship between QoS and QoE factors. Subjective user studies that target different types of application scenarios of 3DTI are conducted and the results show no significant power trend in the relationship between packet loss and perceptual quality metrics. We further verify that activity characteristics, activity objectives, and users' roles in the 3DTI session also have profound effect on the service quality aside to the QoS level.

To address these confound factors, a comprehensive framework of 3DTI that takes in environmental, social, and psychological factors should be considered. While communities of internet video streaming and VoIP have long involved in the development of similar frameworks, a comparative framework is not seen in 3DTI. The logical next step of 3DTI is to investigate in such direction to aid the delivery of multimodal, interactive multimedia services.

7. REFERENCES

[1] Dummynet. info.iet.unipi.it/~luigi/dummynet/

[2] Skype. www.skype.com

[3] K. Chen, C. Huang, P. Huang et al., "Quantifying Skype User Satisfaction," ACM SIGCOMM, 2006.

[4] M. Csikszentmihalyi. "Finding Flow: The Psychology of Optimal Experience". Basic Books, 1997.

[5] E. B. Goldstein, "Sensation and Perception," 8/e, Wadsworth, 2010.

[6] C. Chen, C. Chu, S. Yeh et al., "Measuring the Perceptual Quality of SKYPE Sources," ACM SIGCOMM CCR, 2012.

[7] S. Chen, P. Xia, and K. Nahrstedt. "Activity-Aware Adaptive Compression: A Morphing-Based Frame Synthesis Application in 3DTI," ACM MM, 2013.

[8] S. Chen, K. Nahrstedt, and I. Gupta, "3DTI Amphitheater: A Manageable 3DTI Environment with Hierarchical Stream Prioritization," ACM MMSys, 2014.

[9] Z. Tang, A. Patel, X. Guoet al., "A Multimodal Virtual Environment for Interacting with 3D Deformable Models," ACM MM, 2010.

[10] A. Balachandran, S. Seshan, V. Sekar et al., "Developing a Predictive Model of Quality of Experience for Internet Video," ACM SIGCOMM, 2013.

[11] ITU-T Recommendation P.800, Methods for Subjective Determination of Transmission Quality, 1996.