Testing and evaluation of new platforms for delivery of distributed gaming and multimedia

Audrius Jurgelionis
University of Genoa
Biophysical and Electronic Engineering Department
Via Opera Pia 11a, 16145 Genoa, Italy
+39 010 3532780
jurge@elios.unige.it

Francesco Bellotti
University of Genoa
Biophysical and Electronic Engineering Department
Via Opera Pia 11a, 16145 Genoa, Italy
+39 010 3532780
franz@elios.unige.it

Peter Eisert
Fraunhofer Institute for Telecommunications
Einsteinufer 37,
10587 Berlin, Germany
+49 30 31002 614
eisert@hhi.fraunhofer.de

Jukka-Pekka Laulajainen
VTT Technical Research Centre of Finland
P.O. Box 1100, FI-90571 Oulu, Finland
+358 20 722 111
jukka-pekka.laulajainen@vtt.fi

ABSTRACT
This paper presents Games@Large - a new system for pervasive gaming and multimedia and describes its components, which are relevant to the testing and evaluation of the system design. Complex gaming and multimedia frameworks comprise multitude of elements: heterogeneous end user devices, wireless and wired network technologies, and various multimedia and gaming services. They enable an easy access to the offered services through the use of a versatile system architecture. Testing and evaluating the design of the system takes central stage in the project through an iterative process which has a great impact on overall design course and the final product in terms of assuring a proper Quality of Service and usability of the system. The analysis of system elements and their testing and evaluation methods will facilitate the development of the methodology for the purpose of verification of the different components and the entertainment system as an integrated whole. The subsequent step of this analysis is an adaptation of the metrics, used for video game and multimedia evaluation, which will develop a comprehensive testing and evaluation methodology for newly evolving entertainment systems.

Keywords
Testing, evaluation, usability, user experience, multimedia, video games, pervasive entertainment.

1. INTRODUCTION
The future home is an always-on connected digital home. By 2010, there will be over 420 million broadband households worldwide [2], [3]. With the standard set for super-high speed, always-on connection, the way people view entertainment has fundamentally changed and created a new standard for consumption. Consumers no longer expect their internet access to be only from a desktop PC - now they want it through the TV in the living room or in the palm of their hand, inside the house and on the go.

The presented scenario [4] bundles video gaming capabilities into consumer electronics devices, such as Set-Top Boxes (STBs), Digital Video Recorders (DVRs), home entertainment systems, TVs, handhelds and other devices that are not considered, today, as real gaming devices since they lack the necessary CPU power and graphical performance.

In this paper we present a new system for pervasive gaming and multimedia, which is being developed under the EU FP6 project, Games@Large (GaL) [1]. The paper is dedicated to the design testing concept elaboration, in order to base the approach for the development of evaluation and testing methodologies. Testing and verification process is part of the iterative, spiral-life workflow model (user-centred design and incremental improvement based on feedback from user and expert evaluation of prototypes).

In section 2 we present the new framework for pervasive gaming and multimedia, and QoS evaluation methods that are relevant to the design testing and evaluation. In section 3 we analyse relevant GaL system components, properties and possible testing methods, which are based on a literature review. Analysis information includes specifications for testing of different GaL elements, i.e. GUI, network impairment effects on user experience in video games, multimedia tests, etc.

2. SYSTEM FOR PERVASIVE GAMING AND MULTIMEDIA
The increasing number of broadband users, and the demand for service quality and diversity, especially in the entertainment area, drives to the development of new pervasive entertainment systems. Services should be easy accessible without reference to the time and location. A typical system will likely be characterized by:

- Pervasive game access on devices that typically do not posses a full set of technical requirements to run video games, such as hand-held devices or set-top boxes
- Cost-effective infrastructure for running games simultaneously on a server infrastructure. Single server utilization to serve multiple end-devices
- Decoupling of input devices (Wireless Game Pad) and output devices (TV screen or handheld device) from the execution on a home server

A major success factor for ubiquitous entertainment systems will be given by the possibility of hosting existing, legacy games and applications, without need for either re-coding or compromising any game features. The system will be responsible for automatically transcoding/adapting any game or interactive multimedia application to the end-device on which it is rendered.

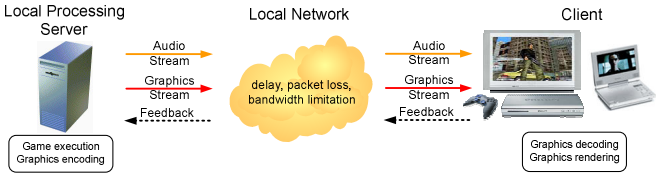

Figure 1. Pre-Rendering (3D graphics streaming).

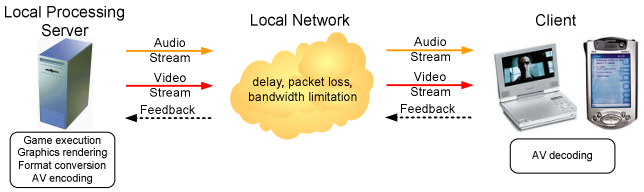

Figure 2. Videostreaming.

Specifically, GaL is not developing any of the offered content, either video games or multimedia; it just integrates and supplies them to the end user. Provision is achieved by the employment of networking technologies and integration of heterogeneous end user devices into GaL infrastructure. The versatile GaL system architecture and effective data encoding/decoding (pre-rendering or video streaming approaches, Fig. 1. and Fig. 2.) techniques facilitate the system's pervasiveness in terms of location, served users and hardware configurations.

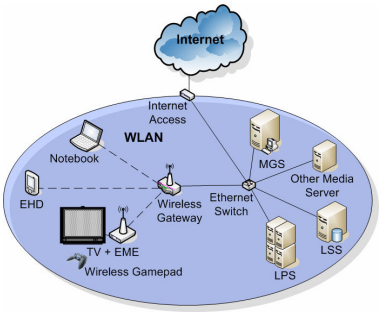

Figure 3. Games@Large framework.

The GaL framework (Local Processing Server – LPS, Local Storage Server – LSS, Management Serve – MGS) as illustrated in Fig. 3. has several particularities, as in the following:

- Heterogeneous end devices: Enhanced Handheld Device (EHD), Set-Top-Box (STB), notebook, Personal Computer (PC)
- Various display devices: TV screen, EHD display, PC monitor, notebook display
- Various control devices: gamepad, keyboard (keyboard control conversion into control from the gamepad is one of the key challenges in terms of usability and enjoyment)
- Distributed computing structure
- Centralized client-server architecture, where servers (Local Processing Server – LPS, Local Storage Server – LSS, Local Management Server - LMS) are performing most of the functions, and end devices serve as control and display units
- Wireless and wired network parts
- Various provided services: (video games, Video-on-Demand (VoD), DVR, Internet TV (IPTV), music, etc.)
- Specific graphical user interface: game selection, billing, parental control, device discovery, content adaptation to the end device capabilities
- Specific data transfer protocols (another key challenge to employ graphics and/or video streaming protocols for visual output transmission from the server to any of the heterogeneous devices)
- Various target environments (house, internet cafe, hotel, elderly house, airplane, cruise ship, train)
- Specific use cases and business scenarios

The specialities defined above pinpoint GaL as a multilayered complex system. The GaL user-centred design is focused on the QoS/bandwidth/latency requirements of each multimedia service. An analysis of system components and their testing methods is necessary in order to build an efficient testing and evaluation methodology. In the next section we introduce the methodology of perceived quality assessment in GaL and in section 3 we analyse the GaL system elements and methods for their evaluation and testing.

2.1 QoS in multimedia networks

In the domain of traditional telecommunications, Quality of Service (QoS) has always been focused on network parameters, looking for different ways of keeping particular sets of these parameters within certain limits to assure the user reasonable quality levels. The problem with this approach is that in the today's Internet, the heterogeneous features of current services make it difficult, sometimes even impossible, to clearly identify the relevant set of performance parameters for each case. Even more, the quality experienced by a user of new multimedia services not only depends on network parameters, but also on higher layers' characteristics (coding and compression of the multimedia, recovery algorithms, nature of the content, etc...). In this sense, a final user might experience acceptable quality levels even in the presence of serious network problems. These observations show that rating the quality of new multimedia services (the Games@Large service suite) from the network side may no longer be effective. The user Perceived Quality of Service

(PQoS) field addresses this problem, assessing the quality of a service as perceived by end users. This seems to be in fact the most coherent approach: after all, the user is the one who pays for the service, and QoS will reflect what he/she understands as such, independently of the state of the network that transports the service. [5]

The assessment of perceived quality in multimedia services can be achieved by two different kinds of methodologies, either subjective or objective [5].

The GaL testing is mainly based on subjective methods, as it implies the user-centred design process. Subjective methods define the most accurate metric as they present a direct connection with the user experience. This metric is suitable for evaluating and testing such a complex system, which comprises various elements and devices.

Objective non-intrusive methods can also be applied in parallel with subjective methods, as they are providing useful quantitative data.

Next chapter is dedicated to the analysis of the GaL system and its components in the context of the whole system testing and evaluation. The approach is based on subjective methods, although objective as well as functionality tests are also considered for some of the GaL components.

3. TESTING CONCEPT

This section describes the general concept of the GaL system testing. In order to better evaluate each system component and its importance in the whole system testing, relevant GaL framework properties are differentiated into 3 prime groups: system properties, user properties, and application properties. These groups reflect the key properties of different GaL elements, and show their impact to the system development and testing.

3.1 System properties

System properties or technical properties can be categorized in 3 classes: software testing, hardware testing, and system testing.

3.1.1 Software testing

Software in the GaL system consists of modules (video streaming, pre-rendering, and GUI modules) made by different partners. Before integrating the modules each should be tested separately. After the integration testing methods for these modules are described in application properties section.

3.1.2 Hardware testing

During the project also hardware is produced and must be tested. This hardware consists of Enhanced Set-Top Boxes and Wi-Fi network equipment. Other hardware used does not have to be tested as such, but tests should be made to conclude whether the used hardware is suitable for GaL. This list includes local processing servers, storage servers, handheld devices, and network components like routers.

The WiFi Quality Assurance team will perform a series of well-defined tests to verify the design on software and hardware level, and provide a product able to get Wi-Fi Alliance's certificate, as well as pass all regulatory domain tests for certification (i.e. FCC).

3.1.3 System testing

The individual software and hardware components will be tested separately as described above. In addition to testing individual components, system-level testing for the whole GaL system is foreseen. The aims of system testing include finding out problems and properties of the entire system rather than of its separate components. From the user's point of view, the overall system performance is crucial, as the user is not using its components one by one but as a single service. Thus, the user's perceived quality of the system can be assessed with the results of system testing.

There are many technical properties that affect the system performance. The properties can be categorized in user-related, server-related, end-device-related, and network-related classes. The most important properties to be considered are listed below.

3.1.3.1 User related:

- Number of concurrent game users
- Users using other services through the same network
- User's performance related to delay of controls and graphics output (Recommended/acceptable input delay, recommended/acceptable graphics delay)

3.1.3.2 Server related:

- Processing capacity: The ability of the Local Processing Server (LPS) to execute and serve multiple game sessions (available CPU power)
- The ability of the Local Storage Server (LSS) storage to supply requests from multiple LPS servers to run multiple game sessions in the network (available disk space, available data transfer speed)
- Other programs running on the same computer (required disk space, required CPU power)
- 3D graphics streaming properties: The ability of the protocol to serve games and the compatibility of the protocol with graphics libraries (OpenGL, DirectX)
- Used video codec (coding efficiency, computational complexity, encoder delay)
- Used transport protocol

3.1.3.3 End device related:

The Games@Large software is designed to work on several different devices like Enhanced Set-Top Boxes and handheld devices. All possible combinations should be tested to see that the GaL framework actually works as planned. At least the following parameters should be changeable while testing client devices:

- Processing capacity (by changing devices or simulating lower capacity)
- Clients: The ability of the end device to process the graphics/video input and controls input. The delay should be minimized to achieve minimum delay and maximal user experience
- Screen size, form factor (240 x 320, 1080p, etc), may influence the user experience for each device

3.1.3.4 Network related:

As multimedia applications develop, communications systems are expected to make provision for two of their aspects, namely Quality of Service (QoS) support and group communications services [6].

This section describes general communication characteristics independent of any layer in communication architecture.

- *Available bandwidth (downlink/uplink)*

Playing 3D graphics or video streaming based games sets a requirement for bandwidth in the order of several megabits per second depending on target resolution and frame rate for video and scene complexity for graphics streaming. If the available bandwidth for gaming is not sufficient, the game experience will be deteriorated. In the case of connection-oriented protocol, congestion in either direction causes problems to both game stream and control channel. If the protocol is connectionless, downlink bandwidth is critical for game streaming and uplink for control channel.

- *Delay (downlink/uplink)*

Games are very interactive and require minimal delay from the network. Both downlink and uplink delays should be minimized to enable the fast response to user commands [13]. This significantly differs from other multimedia applications like IPTV or music streaming, where delays can be tolerated to some extent. Minimizing the delay for the transmission of video and graphics is one of the most challenging aspects in Games@Large.

- *Jitter (downlink/uplink)*

Unlike the case of traditional video streaming, there is no possibility for using long jitter buffering in the case of game applications. Thus, jitter should be minimized. Too high jitter degrades the game experience as the response time of the game varies.

- *Packet loss* (downlink/uplink)

The effect of packet loss to user experience depends on the 3D streaming protocol. At least some of the 3D data are such that loss cannot be accepted and lost parts have to be recovered by retransmissions or forward error correction. In the case of video streaming-based gaming, some packet losses may be accepted if proper error concealment or resiliency techniques are used.

- *QoS enabled/disabled*

If the network supports QoS, the game sessions can be assigned higher priority in the network. This enables the smooth operation of game applications regardless of the other traffic. If QoS is not supported by the network, the game traffic competes with the other traffic with the same priority. This means that the game experience is affected if there is too much other traffic in the same network.

- *Wireless properties*

Presence of RF interference. RF interference may deteriorate the network performance, which in turn, degrades the user experience of the game user.

- Distance from the access point.

If the game users reside too far away from the access point or there are obstacles between the stations and the access point, the signal strength may decrease. This causes the usable bitrate to decrease, which affects the maximum possible number of users.

3.2 User and stakeholder characteristics and impact

GaL targets at four basic scenarios: house, internet cafe, hotel, elderly house. User demographics plays an important role in the whole testing process. It should be noticed that stakeholder or environmental properties, and their implications to user demographics have to be considered when selecting test participants. In early project stages a detailed analysis and predefinition of possible choices of test participants has to be performed. When performing tests the test participants have to be chosen according to the specifications for each environment. This characterizes the users that should be selected and defines the criteria for the selection. With respect to appropriate use cases at different GaL focus areas, it will facilitate testing process procedures and enable to obtain useful and objective test results.

Environmental analysis is performed in terms of contextual analysis, which is differentiated in three categories of characteristics that bear direct relevance to gaming and game experience: the social, physical, and service context.

Finally, stakeholder properties, such as infrastructure capabilities, user demographics differences, offered services and requirements to the system will none the less impact testing and evaluation. And in particular, in test participants' selection, i.e. elderly houses have a different distribution of the age of users and different requirements to the GaL system, than hotels or internet cafes.

3.3 Application properties and components

The description below presents Games@Large application components and relevant properties. They represent GUI, games, video, and other services and interfaces that users are interacting with. Therefore, testing those components from the user's point of view (subjective method) implies testing the whole system and its performance. As it was motioned previously, objective tests of separate system components are useful for obtaining quantitative evaluation.

Literature analysis on possible testing methods which can be applied in GaL testing process is conducted in order to base the development of GaL testing methodology.

3.3.1 Graphical user interface

According to given specifications, the GUI will be a web-based user interface, which will be transferred to various end devices, such as TVs, EHDs and laptops.

Functionality testing enables evaluating how the application and the user interact. This includes how the application handles keyboard/gamepad and mouse input and how it displays screen text, images, buttons, menus, dialog boxes, icons, toolbars and more. Functionality testing is commonly done by human testers, or by automatic tools. Given their increased importance, testing GUIs for correctness can enhance the entire system's safety, robustness, and usability. GUI designers during the development

process will test its functionality. In the GUI testing, it is important to evaluate how it meets its written specifications.

An iterative user-centered design (UCD) design process is used during the development of a graphical user interface (GUI) for the GaL system. The fourth phase in an iterative process is the evaluation phase during which feedback about the interface is obtained through the functionality and usability testing [7].

Usability testing is defined as: "In System Testing, testing which attempts to find any human-factor problems". A better description is "testing the software from a user's point of view". Essentially it means testing software to prove/ensure that it is 'user-friendly', as distinct from testing the functionality of the software. In practical terms it includes ergonomic considerations, screen design, standardization etc [7].

3.3.2 Multimedia

GaL may work simultaneously with a variety of other multimedia services such as: Video on Demand, VoIP, IPTV, and DVR.

These services will not be developed in the project, but may be integrated into the same network. The testing process is likely to focus on the effects of games on multimedia applications, when both are running concurrently and consumed by the one or several end-users simultaneously.

Testing processes are intended to cover several types of multimedia applications, which will be available in different target environments, in order to comprehensively verify the current system design.

Testing should use different types of multimedia clips, representing different classes of motion contents. Performance should be evaluated in terms of infrastructure utilization, loss rates, number of simultaneous customers served and end-user perceived quality [8].

An additional assessment of GaL performance in terms of network-related and device-related parameters will be done. Subjective tests will be performed in order to determine the real end-users' perceptual assessment of the quality of multimedia clips streamed using the GaL infrastructure.

Since the perceived quality of any multimedia sequence is highly influenced by the associated sound (e.g. audio and video must be appropriately synchronized), all clips should be streamed with their soundtracks.

The verification tests used to asses the performance of different coding standards can be slightly modified and adapted to the GaL testing purposes. Here we will not only consider the encoding efficiency, but rather we will investigate overall system performance, overall delay and user perceived quality, when multiple processes are running on the server (multiple game applications and/or multimedia applications are running concurrently). Certainly, a number of concurrent end users (processes) and application types should be varied. Network simulation tests will be performed in order to investigate and evaluate the influence of network parameters and to obtain data for quantitative evaluation.

The methodology of the AVC compression performance assessing is given in [9], [10], [11].

3.3.3 Video Games

This subsection describes testing of different types of games which will be covered in the GaL design process. Each game type described below may have unique QoS requirements that should be categorized during the project. Multiple games that represent game genres and types will be selected and evaluated.

While designing a network-based entertainment system, it is important to understand how the degradation of network conditions affects the overall user performance.

Tests of first person shooter games (FPS) revealed latency to be a critical issue. While movements are not significantly affected by network conditions, precision shooting is highly affected by high latencies. Moreover, the game becomes less reactive and less enjoyable at higher latencies. While low values of packet loss (0%-4%) might be acceptable, it is important to keep the latency at low values (<100 ms) to ensure the quality of the service [12], [13], [14], [15]. However, it must be noted that all these studies have been focusing on traditional type of networked gaming where the game software runs on each player's computer and the network is used only to synchronize the players' contexts together. Thus, the results may not be exactly applicable to GaL type of gaming.

Further studies may also be performed on different kinds of games, such as strategy and racing games, where network performance decay may not alter the overall quality of the games as much as it does in FPS games.

Tests will be done while simulating a wireless network connection, in which latency and packet loss are not constant and usually higher than in regular cabled connections.

A subjective and objective evaluation of the effects of loss and latency on videogames in the GaL system is therefore useful for making required improvements or modifications in the design [12], [13], [14]. [15].

On the other hand, frame rates and resolutions rendered in a computer game directly impact the player performance, influencing both the overall game playability and the game's enjoyability. Insights into the effects of frame rates and resolutions can inform system designers in their development of new platform. [16], [17] present results of carefully designed user studies that measure the impact of frame rate and frame resolution on user performance in first person shooter games. Contrary to previous results for streaming video, frame rate has a marked impact on both player performance and game enjoyment while resolution has little impact on performance and some impact on enjoyment [17].

There have been numerous studies that examined the effects of frame rate and frame resolution on users passively watching streamed video. These studies have found that a decrease in frame resolution corresponds to a decrease in user satisfaction, while a decrease in frame rate does not decrease user satisfaction as much. However, watching video, even during a video-conference, does not have the same interaction requirements, in terms of the required response time, as do some other interactive media applications [17].

The diversity of game hardware in GaL results in the same game being played at different frame rates and frame resolutions. A

quantitative understanding of the effects of frame rate and resolution on game playability is therefore critical for GaL platform developers, to enable better targeting of platform improvements to aspects of the display that matter [17].

Finally, simultaneously running games and multimedia applications will impact each other's performance and definitely the Perceived Quality of Service. Thus the effects of multimedia on the video games should be considered and evaluated in the testing process.

Concluding this section we also have to mention controllability issues, since heterogeneous control devices and units are used in the GaL system. Special attention should be paid to gamepad control capabilities, as keyboard controls are mapped onto the gamepad. This issue has to be taken into account while considering usability issues in GaL testing process.

4. CONCLUSIONS

This paper has introduced a new system and its components for pervasive gaming and multimedia. The aim of the paper was to identify system elements and their parameters, which are relevant to the testing and evaluation of the design.

The GaL system and prospective entertainment systems consists of multiple elements and is intended to serve a number of different users in terms of cultural belonging and environmental dependence. Therefore testing and evaluating of such systems is a complicated task. For the purpose of verification of the different components and the entertainment system as an integrated whole, the GaL framework and its services were differentiated into 3 general groups (system, user and application properties). The differentiation and analysis of the GaL elements, and their parameters and testing methods will facilitate the development of testing methodology intended for comprehensive system design verification. The testing process provides qualitative and quantitative results, which are evaluated and supplied to the system designers, thus advising them and facilitating the correction of the errors and improvement of the design.

5. ACKNOWLEDGMENTS

This article was composed within the process of testing and evaluation methodologies development for EU FP6 project, Games@Large.

6. REFERENCES

[1] http://www.gamesatlarge.eu

[2] http://www.findarticles.com/p/articles/mi_m0EIN/is_2006_Sept_26/ai_n16837715

[3] http://www.emarketer.com/Report.aspx?bband_world_jun06&src=report_summary_reportsell

[4] Y. Tzruya, A. Shani, F. Bellotti, A. Jurgelionis, Games@Large - a new platform for ubiquitous gaming, BroadBand Europe 2006, Geneva, Switzerland, November 2006

[5] P. Casas, D. Guerra, I. Irigaray, User Perceived Quality of Service in Multimedia Networks: a Software Implementation, Joint Research Group of the Electrical Engineering and Mathematics and Statistics Departments, 2006

[6] L. Mathy, C. Edwards, D. Hutchison, Principles of QoS in group communications, Journal Article, Telecommunication Systems, Springer Netherlands, 2004, pp. 59-84

[7] Software Quality Assurance & Usability Testing, http://members.tripod.com/bazman/usability.html?button6=Article+on+Usability+Testing

[8] G. M. Muntean, P. Perry, and L. Murphy, Subjective Assessment of the Quality-Oriented Adaptive Scheme, IEEE Transactions on Broadcasting, vol. 51, no. 3, Sept. 2005

[9] T. Oelbaum, V. Baroncini, T. K. Tan, C. Fenimore, Subjective quality assessment of the emerging AVC/H.264 video coding standard, International Broadcast Conference (IBC), Amsterdam, Netherlands, Sept. 2004

[10] C. Fenimore, V. Baroncini, T. Oelbaum, Tobias, T. K. Tan., Subjective testing methodology in MPEG video verification, Applications of Digital Image Processing XXVII. 2004, pp. 503-511

[11] International organization for standardization, ISO/IEC JTC 1/SC 29/WG 11, Coding of moving pictures and audio, Report of The Formal Verification Tests on AVC (ISO/IEC 14496-10 | ITU-T Rec. H.264), JVT, Test and Video Group, 2003, Waikoloa Janice C. Redish, Joseph S. Dumas, A Practical Guide to Usability Testing, 1999 Intellect Books

[12] M. Claypool, The effect of latency on user performance in Real-Time Strategy games, Int. Journal of Computer and Telecommunications Networking, pp. 52-70, 2005

[13] T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, M. Claypool, The effect of loss and latency on User Performance in Unreal Tournament 2003, Proc. ACM Network and System Support for Games Workshop (NetGames), Portland, USA, Sept 2004

[14] T. Lang, User Experience while playing Halo with network delay or loss, Technical Report CAIA-TR-031205A, Centre for Advanced Internet Architectures, Swinburne University of Technology, Dec 2003

[15] T. Henderson, The effects of relative delay in networked games, Department of Computer Science University College London, February 2003

[16] T. Connor, A. Fiske and R. Kennedy, The Impact of Frame Rate and Resolution on Player Movement in First-Person Shooters, Interactive Qualifying project, Worcester Polytechnic Institute, 2006K

[17] M. Claypool, K. Claypool and F. Damaa, The Effects of Frame Rate and Resolution on Users Playing First Person Shooter Games, Proc. ACM/SPIE Multimedia Computing and Networking (MMCN), San Jose, USA, Jan. 2006