Volume 3, No. 4, July- August 2012

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

QoE Based Modeling and Analysis of 3D Audio Teleconferencing Service

Mansoor Hyder*, Mukhtiar Memon
Information Technology Center,
Sindh Agriculture University
Tandojam, Sindh, Pakistan
(mansoor.hyder,mukhtiar.memon)@sau.edu.pk

Akhtar Ali Jalbani, Gordhan Das Menghwar
Information Technology Center,
Sindh Agriculture University
Tandojam, Sindh, Pakistan
(akjalbani,gdas)@sau.edu.pk

Khalil ur Rehman Laghari
Institute Telecom Sud Paris
Paris, France
Khalil.laghari@it-sudparis.eu

Abstract: Quality of Experience "QoE" represents human centric approach to assess quality of any service or product to understand user needs, feelings, performance and overall user experience. In this article, we propose a generalized QoE Interaction Model which comprises of three domains, QoE-contextual-, QoE-technological- and QoE-business-domain. Furthermore, we particularly investigate the impact of contextual domain over QoE domain. As a case study we investigate "3D Telephony" on the basis of QoE requirements of users/customers. 3D Telephony is an open source VoIP based 3D audio telephone and teleconference service. User studies were conducted to capture QoE data. Benchmarking of QoE terms, their analysis and validation in three different test scenarios were done using statistical empirical approaches. This whole study brought interesting results and QoE findings.

Keywords: QoE Modeling; 3D Audio, Multiparty Conferencing, Telephony, VoIP

I. INTRODUCTION

Humans meter the service quality and Quality of Experience (QoE) represents human centric quality assessment. QoE provides assessment of human expectations, feelings, perceptions and cognition with respect to a particular product/service/application [1,2]. In this era of competition, as the realm of multimedia services expands, service providers in multimedia service markets place their emphasis on Quality of Experience enabled multimedia services to ensure customer satisfaction instead of only considering the pure network resource provisioning. Teleconferencing service provides many advantages like it saves time and budget by avoiding unnecessary trips; importantly it saves energy and reduces CO2 emissions [3]. But it has not yet achieved a wide spread acceptance and success due to various quality issues such as (a) Product/System aspects, (b) Technical network aspects, (c) Business aspects (d) Contextual aspects.

In order to understand QoE requirements for audio teleconferencing service based on VoIP and 3D Audio system, we propose high level QoE interaction model which is multi-domain and multi-disciplinary model. It brings together all the important domains of service life cycle to understand their impact upon QoE. User studies were conducted to evaluate QoE for 3D Telephony [4] system. 3D Telephony is 3D audio phone and teleconference system that helps participants of conference call to spatially separate each other, locate concurrent talkers in space and understand speech with clarity. Furthermore it enhances speech quality, interactivity and brings the feeling of naturalness in communication. Since, it is clear from research studies that overall audio quality of teleconferencing systems in terms of speaker/talker localization and speech intelligibility enhances with incorporation of 3D audio in teleconferencing systems [5,6] but further optimizations would still be required at various levels particularly at virtual acoustic environment, which is part of most 3D audio simulations.

Virtual acoustic environment gives teleconferencing participants a level of freedom to modify specifications of virtual environment like room size, table size and place talkers at specific distance and direction as per their own requirements and ease. In our current work, we study and analyze the important research questions such as, what are possible user/customer QoE requirements in 3D multi-party teleconferencing particularly relating to virtual acoustic environment? How to model QoE requirements? How 3D audio virtual acoustic environment influences QoE factors and vice versa? How simultaneous talkers and their voice type (gender) influence QoE? Does distance between talker and listener in virtual acoustic environment have any impact on QoE? How varying virtual room size can impact QoE? What is overall quality of 3D virtual acoustic environment as perceived by participants?

The reminder of the paper is structured as follows: In section II, we present related work on QoE modeling and QoE issues in 3D audio multiparty conference system. In Section III, at first we present architecture of our solution 3D Telephony system and then detailed discussion is done on our proposed QoE interaction Model. In Section IV, Methodology and Design set up is presented. In Section V, we present test results and discuss our findings. Finally we conclude our work.

II. RELATED WORK

Mostly, audio telephony services such as VoIP services are assessed based on Quality of Service (QoS) parameters [7,8]. QoE is considered as an extension to QoS concept, in [9] the relationship between QoE and QoS is investigated and authors propose logarithmic dependencies between QoS and QoE in order to understand more deeply the quantitative relationships and causality issues between these two quality concepts. In [10] QoE Model for VoIP is presented to debug and tune VoIP issues which could negatively impact QoE. The paper [11] focuses on the QoE based evaluation of VoIP services over Best Effort UMTS networks. They have analyzed the most relevant configuration parameters in order to evaluate the performance of VoIP communications in different conditions. However QoS may not be considered as only influencing domain over QoE, because there are other aspects of service life cycle as well such as contextual and business factors which could also influence user experience.

For 3D Audio teleconferencing services, virtual acoustic environment plays important part and its influence over user experience is worth investigating. But there is very less literature available in the field of virtual acoustic environment which covers QoE modeling aspects of localization performance, localization easiness and spatial audio quality. 3D audio teleconferencing systems under development are far from mass market usage as their quality of experience does not fulfill all user demands yet. Consequently, it is very important to measure the quality of existing systems to understand how to improve them.

A lot of research has been carried on teleconferencing topic but there is very less literature available on QoE aspects of teleconferencing solution, since most of the work is focused on, studying sound localization in multi-party conference [12, 13, 14], recognition of unfamiliar voice with the help of specialization and visual representation of voice location [15, 16], specialized audio and video multi-way conferencing [17] and cocktail party effect [18], because of a lack of research work in QoE domain of teleconferencing we see it as an important task to study and model QoE requirements of 3D Teleconferencing service and system.

In the speech intelligibility area, in their work [5] study the influence of stereo audio coding by testing subjective quality of localized speech at various azimuth on horizontal plane, specifically on sound quality and Japanese word intelligibility. Their aim was to use sound localization to separate the main speaker speech from other speakers in a multi-party 3D audio conferencing environment utilizing voiscape [19]. Also in their work [13, 14] studied the effect of competing noise source on the intelligibility of target speech by particularly focusing acoustical aspects of conference system in which participants from stand-alone PCs share a common virtual space. However, in our study we have incorporated human speech as competing sound sources which have been thought of as a direct application to the problem of a multi-party teleconferencing system. Additionally, relative and absolute differences in concurrent talkers voice type such as two male, two female and two mixed gender talker scenarios were also tested. Work in [20] investigated the speech intelligibility of English phonetically balanced words with competing speech, tests were performed having fixed distances in relation to the listener. However, we study three different listeners to talker distances to better map the user/customer QoE requirements.

III. QOE MODEL AND 3D TELEPHONY ARCHITECTURE

A. 3D Teleconferencing Service-Architecture and Implementation:

The 3D Telephony [4] setup is based on a centralized conference bridge and virtual reality server. Each user connects to the conference bridge using a communication device of his or her choice, either a VoIP (soft) phone, a PSTN or ISDN phone, or a mobile communication device. All call control is left to the bridge, and audio streams are forwarded to the rendering engine and rendered individually for each user before being transmitted back to the bridge for mixing and transcoding. Head-tracking is achieved by a separate and direct connection between each user and the virtual environment.

Implemented system is based on the open-source VoIP soft-phone Ekiga, which has been enhanced by a plug-in to control the virtual environment in order to support QoE requirements. As a rendering engine we utilized Uni-Verse [21] acoustic simulation framework. The Asterisk telephony toolkit was employed as a conference bridge and enhanced by a dial-plan application that connects to the rendering front-end. The system will be provided to the scientific community as an entirely open-source application. The current prototype system can be installed on any desktop computer or laptop running an Ubuntu/Debian based operating system.

B. QoE Model for 3D Telephony:

QoE Interaction Model produces blue print of various domains across the life cycle of 3D Telephony service which could impact quality of user/customer experience. A generalized QoE Interaction Model for 3D audio teleconferencing service is presented in (Figure 1). The model represents the main actors QoE domain, Contextual domain, Technological and Business domain. All domains are modeled to understand the process of formation of QoE requirements by end user and/or customer. In our current work, we primarily focus upon impact of Contextual domain aspects over QoE domain, however in our future work; we intend to include technological domain and business domain parameters as well.

a) ***QoE Domain:*** QoE Factors are formed based on objective human factors and subjective human factors. Objective human factors are quantitative in nature and are related to human performance and cognition. These factors produce quantitative data based on various human factors such as human memory, audio visual capacity, human reaction time etc. Quantitative data answers the questions like “how much”, ”how many” and “where” (cooper, 2007). These factors can be gathered and evaluated both through subjective testing or quantitative research.

a) ***Localization Performance (LP):*** LP is an objective human factor because it produces quantitative data on the basis of human performance of localizing talkers correctly in virtual environment. Subjective human

factors are qualitative parameters which reflect customer/user perceptions, feelings and intentions. These factors are normally obtained through surveys, customer interviews, and ethnographic field studies (cooper, 2007). We define three QoE factors (i) *Localization Easiness* (ii) *Spatial Audio Quality* (iii) *Overall Audio Quality*

b) Localization Easiness (LE): It represents human perception of easiness with respect to localizing talkers in Virtual Acoustic environment VAE. Subjects were asked to give MOS score on how easy they feel localizing simultaneous talkers in VAE?

c) Spatial Audio Quality (SOQ): Describes how well the participant could perceive that talkers were spatially separated and their speech quality is also intelligible. Subjects were asked to give MOS score on spatial audio quality.

d) Overall Audio Quality (OAQ): Referred to the total audio quality and listening environment as experienced by subjects. Subjective and Objective Human Factors can be further analyzed to verify their interdependence or any relationship. In our work, we define one such parameter Localization Efficiency (LEF) to study the relationship between LE and LP as defined below.

e) Localization Efficiency (LEF): It describes relationship between human perception factor LE and human performance factor LP. Smaller the gap between human perception of easiness (LE) and human performance factor (LP) in localizing talkers, greater the sense of being there and increased localization efficiency.

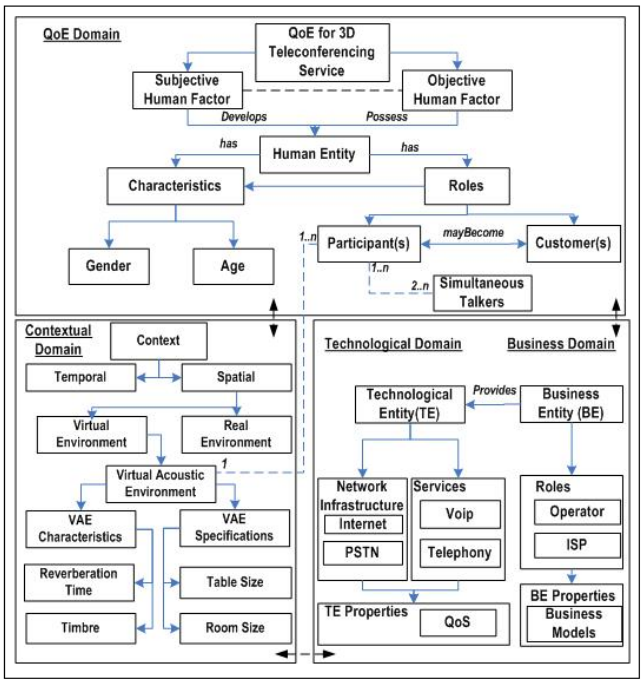

Figure: 1 QoE Model 3D Telephony

In addition to subjective and objective human factors, human entity represents various other aspects such as roles and characteristics. Roles are broadly categorized into two important roles Participants (i.e. Teleconferencing Service users) and Customers. Customer is one who subscribes to service and is legal owner of the service, while the participant(s) is one who actually uses the service. The line between participant and customer boxes shows the possibility of interchanging roles of the two. The cardinality between participants and Virtual Acoustic Environment is also shown, it means the n number of participants (1..n) can be presented in VAE for multi-party teleconferencing. While Simultaneous Talkers refer to number of concurrent talkers during multiparty teleconferencing and cardinality of (2..n) suggests that number of simultaneous talkers can be (2,3,4..n).

b. Contextual Domain: It represents all aspects of environment in which a user can use a service. Contextual domain is broadly classified into two categories (i) *Temporal* (ii) *Spatial.* (i) Spatial is further classified into two more categories *physical* and *virtual*. Virtual environment can be acoustic or visual or both. As we analyze in our work the impact of virtual 3D audio service for teleconferencing in virtual acoustic environment, we put more focus on virtual concept of context domain. Virtual Acoustic Environment (VAE) is further sub-classified into two categories, first VAE Specifications which specify the various actors which constitute virtual acoustic environment such as dimensions of virtual room, table size, types of room etc. Second VAE Properties, which is set of specific parameters that defines properties of virtual environment such as echo, reverberation and timbre etc.

The Quality of Experience in virtual acoustic environment depends upon specifications of virtual acoustic environment such as room size, table size, and characteristics of virtual acoustic environment such as reverberation in room and timbre. Furthermore, QoE in virtual acoustic environment varies also on the basis of characteristics of participants such as voice type and number of concurrent talker etc.

c. Technological and Business Domain: These are two important domains in QoE Interaction Model. However in our current work, we mainly investigate VAE and QoE aspects while in our future work, technological and business domain may be explored further.

a) Business Entity: The Business Entity represents Service Provider, Network Operator, and Device Vendor etc. Customers establish interaction with service provider and/or network operator to subscribe a service that fulfills their intended goals. This interaction between customer and provider can be direct or indirect (on line) but in both cases this interaction experience also develops positive or negative feelings. The Business Entity has some properties such as business model & strategies which defines the direction of its business. There should be alignment between business and technical entities to create an integrated technical and business solution which could guarantee the rich quality of experience.

b) Technological Entity: The Technological Entity represents set of services, network resources and devices offered by Business Entity. The user can use various technological entities to achieve its goals. The

usage experience of these technological entities also influences overall QoE. In order to retain customers, it is obligatory that Business and Technology domain aspects completely fulfill the needs and desires of customers and users.

IV. METHODOLOGY

In order to design, test and implement QoE enhanced 3D Telephony system we conducted formal listening-only tests to study various virtual acoustic teleconferencing scenarios. User experiments were conducted with 31 paid subjects, 13 of them female and 18 of them male, according to ITU-T P.800 recommendations as far as possible. All tests were conducted in a quiet listening room on a computer using a specially designed user interface on Linux operating system. To enable participants to distinguish the different talkers contained in each sample, each talker was represented by a number as well as its spoken text. Each participant was asked a series of questions to be answered for each talker contained within each sample. Localization performance of each test participant was measured separately by presenting him/her a map with possible talker locations. Localization easiness, spatial and overall audio quality were measured using discrete MOS-LQSW (Listening Quality Scale Wide-band) scores with the values 1 (bad), 2 (poor), 3 (fair), 4 (good) and 5 (excellent). All audio samples consisted of anechoic speech samples taken from the ITU-T Rec. P.50 Appendix 1 library. They were prerecorded from and processed by the open-source 3D audio rendering engine Uni-Verse at a sampling rate of 16 kHz. The speech samples were recorded using three different male and three different female voices, each speaking four sentences in American English. Selection of scenarios and sub-scenarios to form QoE Modeling has been taken on the following facts and grounds.

A. Voice Type:

In this scenario, The goal was to test the impact of relative and absolute differences in voice types (such as two concurrent male, female or mixed gender talkers) over QoE. Therefore, the three tests within this setup were conducted, Voice Type-1 utilize two simultaneous female talkers with an average signal length of 13:03s, and Voice Type-2 with two mixed gender talkers with an average signal length of 14:42s and Voice Type-3 is for two concurrent male talkers with speech signals of average length of 14:38s, from four possible locations distributed around the table.

B. Virtual Room Size:

In this scenario, we analyze how varying virtual room size and sound source/talker-to-wall distance impact upon QoE factors. How participants' opinions and performance vary with varying room size. To determine the effect of room size and sound source/talker-to-wall distance on all QoE scores, this test uses three different rooms with dimensions of 10^3m^3, 15^3m^3 and 20^3m^3 . The average lengths of the presented stimuli add up to 14:38s, 14:65s,14:43s for the three tests.

C. Virtual Table Size:

This scenario was used to measure the impact of the distance between the individual sound sources and the listener. Varying table size varies listener-to-sound/talker distance, therefore it is to see how this varying table size influences QoE. Virtual conference tables with radii of 2m, 3m and 4m respectively are used in this set up. The average stimuli lengths were 14:38s, 14:42s and 14:47s

V. RESULTS AND DISCUSSION

a. ***Reliability and Validity Testing:*** Before proceeding to results, it's important to verify reliability and internal consistency of QoE factors (LP,LE,SAQ,OAQ) utilized in various scenarios. Cronbach's Alpha test is normally employed to verify reliability and validity of data. QoE factors at each sub scenario as well as at whole scenarios are tested and the results vary from .813 to .893. The cutoff threshold is 0.6 and it is evident from the results that all values are more than 0.6, thus it suggest a high level of reliability of construct variables and underlying measurement items.

b. ***Discussion:*** In this section, we take one by one each QoE factor and evaluate them with respect to different scenarios (e.g., voice type, room and table size).

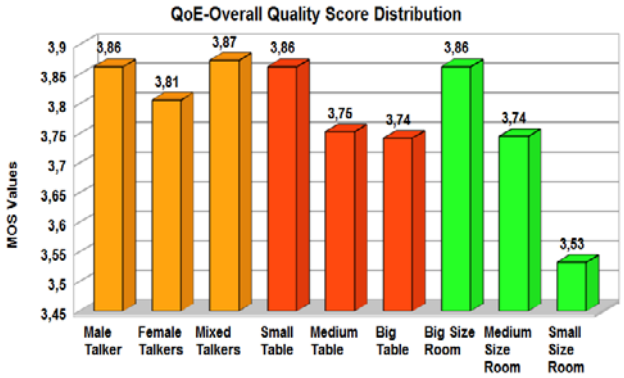

Figure: 2 QoE – Overall Audio Quality.

a) ***Localization Performance (LP):*** The quantitative LP data is presented in percentage to compare LP scores for various scenarios in 3D virtual environment (See Table 1). It's found out that highest LP scores are achieved with mixed gender Voice Type Scenario (76.61 %), i.e., 76.61 percent times subjects correctly localized positions of simultaneous mixed gender talkers. The lowest LP was found to be female simultaneous talker (48.6%) set up. In Table Size scenario, when we change table size, LP score for 4m radius table was found to be the highest (74%) and for medium table, it was about (70%). In Room Size Scenario, LP for small room (72.3%) was almost equal to medium size room (71.7%). The large room and small table have the lowest LP score (63.44%). We select cut off range of LP as (60%) required for good QoE. Mixed Gender Voice type LP score, small size room LP scores, Big Table LP scores produce "good" QoE.

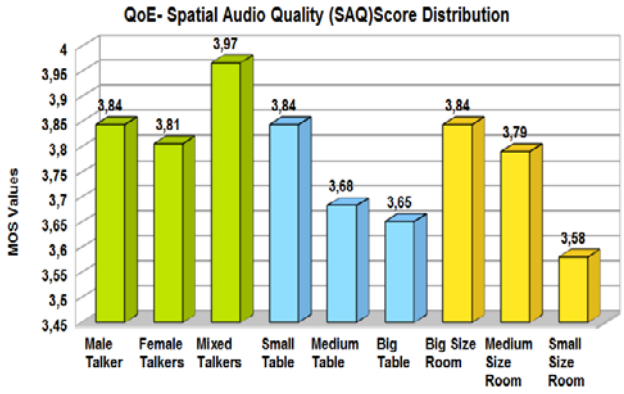

Figure: 3 QoE – Overall Audio Quality.

b) ***Localization Easiness (LE):*** LE is compared across these three scenarios in MOS scale as presented in (See Table 1), it is obvious that human perceives it very easy to locate mixed gender voice (3.8) score in Voice Type scenario, In room size scenario, LE is higher in medium size room (3.76) score than other room sizes and in Table size scenario, medium table has the highest LE scores (3.72).

c) ***Localization Efficiency:*** LE and LP are compared in various scenarios in (Table 1) to understand their nearness. If gap is zero between LP and LE range and correlation between them is also positive, then we consider natural or satisfied relationship between LP and LE thus higher localization efficiency. Positive Correlation means increasing trend between LE and LP i.e subject's LP scores increase with increase in subject's perception of localization easiness. Negative correlation shows negative inverse relationship which is unacceptable. However if gap increases between localization performance and localization easiness as it's the case with scenario voice type Female simultaneous talkers, the LP and LE have one step difference. This suggest slight gap between what participants perceive and perform. While all other scenarios have positive correlation between LP and LE, with minor mismatch between their values. Thus it's safe to say that they have better localization efficiency.

d) ***Spatial Audio Quality (SAQ):*** On whole, SAQ of three scenarios fall into good category (i.e., between 3.1 to 4.0) as perceived by subjects. It means 3D Telephony service provides good spatial audio quality. If we see (Figures 2 & 3) to find precise difference in SAQ score with respect to scenario, then mixed gender has highest SAQ score in Voice Type scenario and Simultaneous female talkers have lowest SAQ. In Virtual room size scenario, subjects give the highest SAQ score to 20m³ room and lowest to 10m³ room size. In Virtual Table size scenario, the highest SAQ score are for 4m radius table and lowest score for 2m radius size table.

e) ***Overall Quality:*** This is one of the important QoE factor which encompasses over all experience and feelings of subjects about quality of 3D teleconferencing service. Like SAQ score, OAQ score also lies in MOS good category and more precisely, only small size room has lowest QoE score (3.58), while all other scenarios and sub scenarios produce better quality impression as shown in graph. It's safe to conclude that 3D Telephony service provides "Good" quality of experience to users.

Table: 1 Localization Performance (LP) and Localization Easiness (LE) Comparison

Subjective Score	LP Score	LE Score	LP	LE
Bad	1-20 %	0.0 to 1.0	—	—
Poor	21-40 %	1.1 to 2.0	—	—
Fair	41-60%	2.1 to 3.0	FxF=48.66%	—
Good	61-80%	3.1 to 4.0	MxM=63.44% MxF=76.61% MT=70% LT=74% MR=71.70% SR=72.3	MxM=3.68 MxF=3.83 MT=3.72 BT=3.67 FxF=3.70 MR=3.76 SR=3.62
Excellent	81-100%	4.1 to 5.0	—	—

VI. CONCLUSION

In this paper, a generalized high level QoE Interaction Model comprised of the main actors such as: Quality of Experience (QoE), Contextual, Technological and Business domain is presented. All domains in QoE model were modeled to understand the formation of QoE requirements by user/customer. Main focus was to understand the relationship between contextual domain and QoE domain. Also QoE terms relating to all domains were benchmarked. Additionally, subjective user studies were conducted and QoE findings for "3D Audio Telephony" service were analyzed and results were presented. Most of QoE results suggests a range of "Good MOS score" when users are using 3D Audio Teleconferencing service based on our solution. In future work, we intend to extend work to other domains of the generalized QoE model and also to conduct more user studies to better map user/customer QoE requirements

VII. ACKNOWLEDGMENT

Authors of this article are very thankful to Dr. Ing. Christian Hoene and Michael Haun for their valuable feedback and suggestions. We are also thankful to the administration of Sindh Agriculture University for their support

VIII. REFERENCES

[1] Kilkki, K., Quality of experience in communications ecosystem, Journal of universal computer science (2008), pp. 615—624

[2] Khalil Ur Rehman Laghari, B. Molina and C. Palau,"QoE aware Service Delivery in Distributed Environment," in Advanced Information Networking and Applications Workshops, March 22 - 25, Biopolis, Singapore,2011.

[3] D. P. Peter James, "Virtual meetings and Climate Innovation in the 21st Century," 2009.

[4] M. Hyder, M. Haun, and C. Hoene, "Placing the participants of a spatial audio conference call," in IEEE Consumer Communications and Networking Conference - Multimedia Communication and Services (CCNC2010), (Las Vegas, USA), Jan. 2010.

[5] Y. Kobayashi, K. Kondo, and K. Nakagawa, "Intelligibility of HE-AAC Coded Japanese Words with Various Stereo Coding Modes in Virtual 3D Audio Space," Auditory Display, pp. 219–238, 2010.

[6] Yankelovich, N., Jonathan K., Joe P., Wessler,M., and Joan, M. D., "Improving audio conferencing: are two ears better than one?", ACM, pp. 333—342, 2006.

[7] Y. Bai and M. Ito, "A study for providing better quality of service to VoIP users," 2006.

[8] K. Radhakrishnan and H. Larijani, "A study on QoS of VoIP networks: a random neural network (RNN) approach," in Proceedings of the 2010 Spring Simulation Multiconference, p. 114, ACM, 2010

[9] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment," in Communications (ICC), 2010 , IEEE International Conference, pp. 1–5.

[10] S. Chatterjee, "Modeling, Debugging, and Tuning QoE Issues in Live Stream-Based Applications-A Case Study with VoIP," in 2010 Seventh International Conference on Information Technology, pp. 1044–1050, IEEE, 2010.

[11] J. Fajardo, F. Liberal, and N. Bilbao, "Study of the impact of UMTS Best Effort parameters on QoE of VoIP services," in Autonomic and Autonomous Systems, 2009. ICAS'09. Fifth International Conference on, pp. 142–147, IEEE, 2009.

[12] J. Blauert and J. S. Allen, Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press,1997.

[13] Y. Kitashima, K. Kondo, H. Terada, T. Chiba, and K. Nakagawa, "Intelligibility of read Japanese words with competing noise in virtual acoustic space," Acoustical science and technology, vol. 29, no. 1, pp. 74–81, 2008.

[14] Y. Kobayashi, K. Kondo, and K. Nakagawa, "Intelligibility of HE-AAC Coded Japanese Words with Various Stereo Coding Modes in Virtual 3D Audio Space," Auditory Display, pp. 219–238, 2010.

[15] R. Kilgore and M. Chignell, "Listening to Unfamiliar Voices in Spatial Audio: Does Visualization of Spatial Position Enhance Voice Identification," Human Factors in Telecommunication, 2006.

[16] R. Kilgore, "Simple Displays of Talker Location Improve Voice Identification Performance in Multitalker, Spatialized Audio Environments," Human Factors, vol. 51, no. 2, p. 224, 2009.

[17] . Inkpen, R. Hegde, M. Czerwinski, and Z. Zhang, "Exploring spatialized audio & video for distributed conversations," in Proceedings of the 2010 ACM conference on Computer supported cooperative work,pp. 95–98, ACM, 2010.

[18] D. Brungart, B. Simpson, C. Bundesen, S. Kyllingsbaek, A. Burton, and A. Megreya, "Cocktail party listening in a dynamic multi-talker environment," Perception and Psychophysics, vol. 69, no. 1, p. 79, 2007.

[19] Y. Kanada, "SIP/SIMPLE-based Conference Room Management Method for the Voice Communication Medium, voiscape ," 2008.

[20] M. Hawley, R. Litovsky, and H. Colburn, "Speech intelligibility and localization in a multi-source environment," The Journal of the Acoustical Society of America, vol. 105, p. 3436, 1999.

[21] Uni-Verse consortium, "Uni-verse webpage." http://www.uni-erse.org/, Mar. 2007.