Journal of Universal Computer Science, vol. 19, no. 12 (2013), 1718-1735
submitted: 9/8/11, accepted: 24/6/13, appeared: 28/6/13

An Investigation into the Relationship Between Perceived Quality-of-Experience and Virtual Acoustic Environments: the Case of 3D Audio Telephony

Khalil ur Rehman Laghari, Tiago H. Falk
(Institut National de la Recherche Scientifique (EMT-INRS)
Montreal, QC, Canada, laghari|falk@emt.inrs.ca)

Mansoor Hyder
(Information Technology Center, Sindh Agriculture University, Tandojam
Pakistan
&
Eberhard Karls Universität Tübingen, Germany
mansoor.hyder@sau.edu.pk)

Michael Haun, Christian Hoene
(Eberhard Karls Universität Tübingen, Germany
michael.haun|hoene@uni-tuebingen.de)

Noel Crespi
(Institut Telecom SudParis, Paris, France
noel.crespi@it-sudparis.eu)

Abstract: Quality of Experience (QoE) is a human centric quality evaluation method which provides the blue print of human needs, perceptions, feelings and experiences with respect to a multimedia service. In a communications ecosystem, human interaction takes place alongside technological, contextual, and business domains, thus producing a holistic view on QoE formation. In this paper, we investigate the relationship between human perceived QoE and "context" for burgeoning 3-dimensional (3D) audio teleconferencing services. 3D audio teleconferencing applications are customizable by generating different virtual acoustic environments (VAE), where parameters such as virtual room size and competing talker conditions can be adjusted for a particular application. The impact of different VAE characteristics on perceived QoE, however, is still unknown. In this study, four QoE factors were investigated across different VAE scenarios. It was found that a) medium-size virtual rooms produce optimal perceived QoE, b) competing talkers of mixed gender could be easily located in the virtual space, and c) competing speaker gender had no significant effect on perceived audio quality.

Key Words: Quality of Experience, 3D Audio, Communication Ecosystem, Teleconferencing, Virtual Acoustic Environment

Category: H.1, H.5, L.2.7, H.5.4

1 Introduction

The last decade has witnessed significant growth in multimedia services, products, applications, and devices. As a consequence, quality demands and user experience requirements have become important benchmarks for gauging the effectiveness of new multimedia services and applications. Commonly, the quality of a new multimedia service is evaluated based on so-called Quality-of-Service (QoS) parameters, such as packet loss rates, delay, jitter, and frame rate. However, technology-centric QoS based approaches do not incorporate the user's aesthetic and hedonic needs. In this competitive era, the main goal of multimedia service providers is to have sustained growth, which requires reliable customer engagement (i.e., attract new customers and retain existing ones) [Accenture 2011]. The key for successful customer engagement is providing them with a rich 'quality of experience', thus fulfilling their expectations and needs. Hence, Quality of Experience (QoE) has emerged as the holy grail of human-centric multimedia services and products. QoE is defined as "an assessment on human expectations, feelings, perceptions, cognition and acceptance with respect to a particular product, service or application" [Laghari et al. 2011]. QoE is a human centric paradigm formulated within a communications ecosystem comprised of the "systematic interaction of living (human) and non living (technology, and business) beings in a particular context" [Laghari et al. 2012]. In other words, human QoE requirements can be broadly influenced by three main domains: technology, business and context.

Traditionally, QoE studies have investigated the effects of technological (i.e., QoS) parameters on perceived QoE. In this paper, an alternate approach is taken and an investigation on the effects of context on QoE is carried out. For this purpose, burgeoning 3-dimensional (3D) telephony was selected as a use case study. 3D telephony is a audio supported telephone and a teleconferencing system that provides a customizable virtual acoustic environment (VAE). A virtual acoustic environment helps participants in a conference call to spatially separate each other, to locate concurrent talkers in space and to understand speech with greater clarity [Hyder et al. 2010]. Furthermore, a virtual acoustic environment provides teleconferencing participants a level of freedom to modify specifications of the virtual environment such as room size, table size and even to place participants at a specific distance and direction as per their own requirements and ease. 3D telephony attempts to solve problems related to classic teleconferencing such as low intelligibility, limited ability of the participants to discern (in particular) unfamiliar interlocutors, to separate different speakers and to communicate over a long time without fatigue. The impact of these factors on perceived QoE, however, has yet to be studied. This paper aims to fill this gap by utilizing subjective user feedback on the effects of varying VAE characteristics on perceived QoE.

The remainder of this paper is organized as follows. In Section 2 we present background work. In Section 3, we present an overview to our QoE model in communication ecosystem based on work of [Laghari et al. 2012]. In Section 4, we present the 3D telephony architecture, research model for QoE-Context, and the methodology for the user study. In Section 5, we present test results and discuss our findings. We present our conclusions and future work in Section 6.

2 Background

Traditionally, audio telephony services, such as voice over internet protocol (VoIP), are assessed based on Quality of Service (QoS) parameters only. QoS metrics, such as packet loss rate, jitter, delay, and frame rate are typically used to indicate the impact on the audio quality level from the technological point of view [Radhakrishnan and Larijani 2010] but these QoS parameters fail to capture human perceptions and feelings. As such, QoE approaches have been introduced to overcome the limitations of current QoS-aware multimedia networking schemes and to introduce human perception and subjective-related aspects [Falk and Chan 2008, Falk and Chan 2009, Takahashi et al. 2008].

For the assessment of multimedia audio and speech communication services, the International Telecommunication Union (ITU-T) [G.1080 2008] and the European Technical Committee for Speech, Transmission, Planning, and Quality of Service (ETSI) [ETSI 2009] have been working to devise methodologies for QoE based evaluations. The ITU has produced a subjective study guideline called Recommendation P.800 [P.800 1996] for speech quality evaluation. ITU-T has also proposed two processes for objective assessment of audio/speech services and applications, namely parametric or signal-based. Parametric models use network QoS parameters to estimate different impairment factors and to aggregate them onto a final quality score, thus serving as a transmission planning tool. The most popular of such models is the so-called E-model, first proposed by ETSI in the early 1990's and later standardized by the ITU-T as Recommendation G.107 [Bergstra and Middelburg 2003]. The E-model assumes that transmission impairments can be transformed into psychological impairment factors, which in turn, are additive in the psychoacoustic domain. Impairment factors related to speech transmission (e.g., quantization distortion), delay (echoes), and "effective equipments" (e.g., packet loss effects for different codec types) are quantified and mapped to either a 0-100 impairment rating scale or a 1-5 mean opinion score (MOS) scale.

Signal-based methods, as the name suggests, measures speech/audio quality based on analyzing the actual multimedia signals [Möller et al. 2011]. Signal-based methods can be further classified as double-ended (or intrusive) or single-ended (non-intrusive), depending on the need or not of a clean reference signal, respectively. The most widely used double-ended model for speech signals is the

ITU-T Recommendation P.862 (also known as PESQ, Perceptual Evaluation of Speech Quality [P.862 2001]). PESQ was recently succeeded by Recommendation P.863 (also known as POLQA, Perceptual Objective Listening Quality Assessment, [P.863 2011]) which allows for speech signals of higher bandwidth, such as those observed with emerging VoIP applications, to be handled. For audio/music signals, in turn, the ITU-R Recommendation BS.1387 (PEAQ, Perceptual Evaluation of Audio Quality, [BS.1387 2001]) has been used since 1998. Single-ended systems, on the other hand, are still in their infancy and in 2010 two models were standardized, one by ITU-T (Recommendation P.563 [P.563 2004]) and the other by the American National Standards Institute (ANIQUE+ [Ansi 2011]). Despite the wide usage of these objective models for speech/audio quality assessment, QoE-related factors such as context and human characteristics are not taken into consideration. As emphasized in the communication ecosystem paradigm, contextual and business aspects need to be modelled as they play an important role in influencing human behaviour and driving QoE perception.

Having this said, one important aspect that is often overlooked is that of "context," or the representation of the situation, environment and circumstances within the communication ecosystem. For example, a person participating in a teleconference call who is sitting in a small quiet office will have different QoE requirements than a person conducting a conference call from a large noisy room surrounded by multiple people. In order to minimize the negative impact of physical environment on a conference call and to provide more ease and comfort to end users, 3D audio teleconferencing systems have been developed. Such systems make use of binaural processing techniques to produce a virtual acoustic environment (VAE) [Begault 1994], either by placing loud speakers at different positions in the listening space, or by headphones [Pulkki 2001]. The basic principle is to control the sound field at the listener's ears so that the reproduced sound field coincides with what would be produced *in situ*.

When dealing with 3D or spatial audio applications, the majority of the published works have focused on sound localization [Blauert and Allen 1997], virtual 3D space generation [Kitashima et al. 2008, Kobayashi et al. 2010], recognition of unfamiliar voices with the help of spatialization and visual representation of voice location [Kilgore 2008], spatialized audio and video multi-way conferencing [Inkpen et al. 2010], and the cocktail party effect [Brungart et al. 2007]. It is widely known that spatial audio a) offers advantages in teleconferencing applications over stereo or mono systems [Best et al. 2006, Kilgore et al. 2008], b) reduces talker localization errors [Raake et al. 2007], c) improves "divided listening" performance [Best et al. 2006], and d) improves intelligibility in competing *noise* [Kitashima et al. 2008, Kobayashi et al. 2010] and competing multi-source *speech* (e.g., from male and female speakers) [Hawley et al. 1999]. To the best of our knowledge, however, none of these studies have looked at the effects of these parameters on the generated VAEs and, ultimately, on the user's perception of QoE. Such information will be invaluable for the creation of emerging technologies that meet user/customer QoE requirements.

3 QoE in the Communications Ecosystem

As mentioned previously, Quality of Experience (QoE) is influenced by various domains in a multimedia communication ecosystem such as business, technology, and context. Figure (1) depicts a high level overview of the different domains and their interaction in the communications ecosystem. To understand total quality of experience, it is pertinent to know the inter and intra-domain interactions in a communication ecosystem. The major interactions are between (i) the human and technological domains, (ii) the human and business domains, (iii) the human and contextual domains, (iv) the technological and business domains, and (v) the contextual, technological and business domains (see Fig.1).

Human-to-technology domain interaction in a communication ecosystem is normally affected by various technological characteristics such as service features, end-user device functionalities, and QoS parameters, thus degradation in these parameters can annoy users. Human-to-business interaction is also an important factor, as when a person wants to subscribe a service, s/he does care about various business characteristics such as price, brand image, customer care and promotions. In turn, business-to-technology domain interaction represents service providers' strategies and business models for their technological infrastructure, and how effectively they can utilize their resources to increase their profit by retaining customers as well as attracting new ones. Context represents situations and circumstances within the communications ecosystems. Context can be real, virtual, or social. During human-context domain interaction, contextual characteristics could also influence human mood and behaviour, thus directly affecting QoE.

Each domain is classified into entity roles and attributes/characteristics levels. An entity is a real-world concept or item that exists on its own. In the proposed model, there are four entities: human, contextual, business, and technological. Each entity could have multiple roles; for example, a human entity could perform the role of a user or customer; similarly, a business entity could be a service provider or device manufacturer. Each entity has some attributes and characteristics, as shown in Table 1. The most important part of the communication ecosystem is the 'human QoE factors' which itself is part of the human entity. Human QoE factors can be further classified as subjective or objective, as shown in Table 2. Subjective QoE factors represent human perceptions, feelings and intentions. Primarily, subjective human factors are based on human psychological aspects. These factors are normally obtained through surveys, customer interviews, and ethnographic field studies [Cooper et al. 2007]. Objective

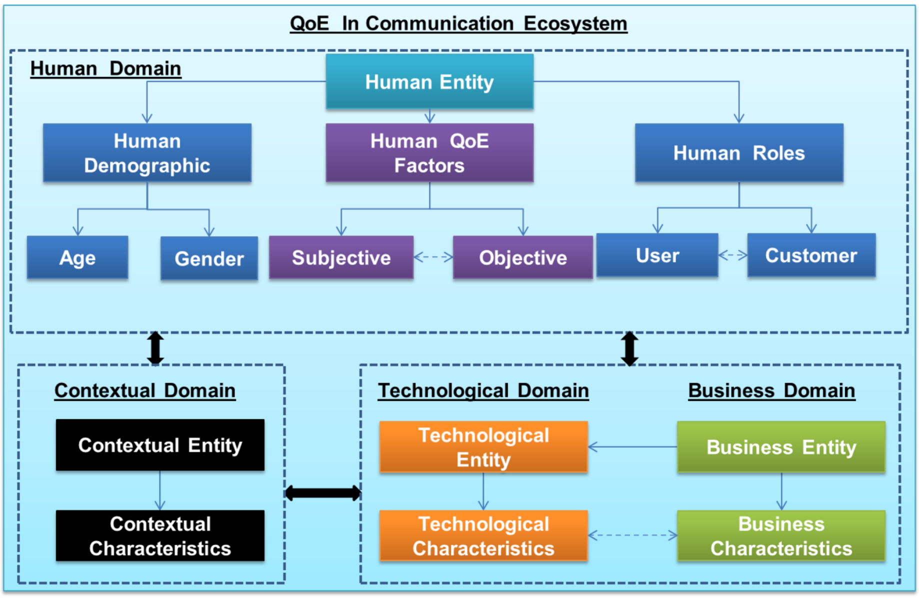

Figure 1: QoE Based Model for Communication Ecosystem

QoE factors in turn, are related to human physiology, psycho-physical, and cognitive aspects. Some examples of human objective factors are the human audio-visual system, event related potentials (ERP), heart rate, blood volume pressure, memory, attention, language, task performance and human reaction time. Human Objective QoE Factors can be obtained using various neuroimaging tools such as electroencephalography (EEG), near-infrared spectroscopy (NIRS), and magnetoencephalography (MEG), as well as biosensors that harness measures of heart/respiration rate, skin conductance/temperature, and muscle activity. The interested reader is referred to [Laghari et al. 2012] for more details on the consolidated QoE model. The mapping between different domain characteristics is needed to establish complex but important relationship between them. But the challenge is how to map diverse set of characteristics belonging to different domains. In psychology, causal relationship is commonly used to establish relationship between two variables. In a causal relationship, outcome is caused by some influencing factors. Following it, we consider QoE factors as outcome factors which are actually caused by different influencing factors related to technological, business and contextual domain. Influencing factors are independent factors and they are used to explain or predict changes in outcome factors (QoE). For detail on interdomain mapping, it is suggested to refer to [Laghari et al. 2012]. In next section, we will present results of a 3D Audio teleconferencing user study in order to assess the influence of contextual characteristics on QoE factors.

4 3D Telephony Based Virtual Acoustic Environments

In this section, we first briefly describe the architecture of the 3D telephony application used in our studies. We then present an applied QoE model for 3D telephony to show the relationship between QoE and virtual context and describe the methodology used in the user study.

4.1 3D Telephony Architecture

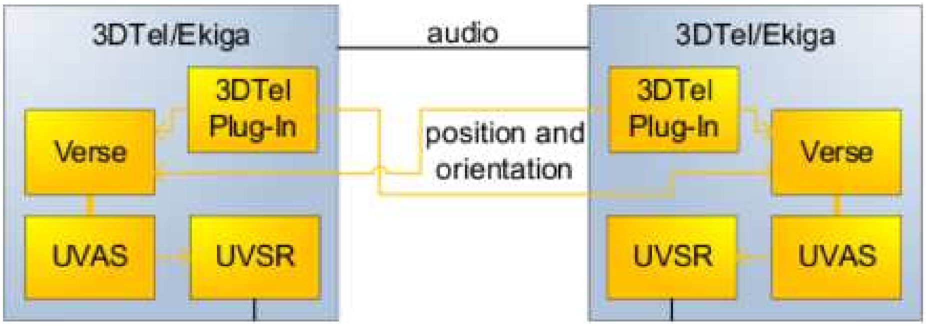

Figure 2: 3D Telephony Setup Architecture

A 3D telephony [Hyder et al. 2010, Hyder et al. 2010b] setup is based on a point-to-point architecture using a different virtual environment for each user, while each user maintains full control over the virtual environment placed at his or her end of the connection. All audio streams are only rendered locally and

Table 1: Highlevel QoE In CEC Taxonomy Table

Domain	Entity-Roles	Characteristics
Human	Customer, User	Demographic Attributes (Age, Gender), QoE Factors (Subjective, Objective)
Technological	Products, Services, Networks, Devices	Delay, Jitter, Packet loss, Frame rate, resolution
Business	Service Provider, Network Operator, Vendor, Manufacturer	Price, promotions, brand image, SLA, customer churn rate
Context	Physical, Virtual, Social	Location, Virtual room size, social pressure

Table 2: QoE Factors

QoE Factors	Examples	Evaluation Methods
Subjective QoE	Overall Quality, Naturalness, Ease, Comfort, Satisfaction	Surveys, User Studies, Focus Groups, and Interviews
Objective QoE	Reaction Time, Performance, Brain waves (P300, N200), Heart Rate, Blood flow (oxy/deoxy)	Quantitative methods (GOMS, User studies), Neuro-Physiological Methods (Electroencephalography EEG, Near Infra-red Spectroscopy, Bio-sensors)

played back directly to the headphones of the respective users. Multiple avatars, one for the local caller and one for each remote call party, are created. The incoming audio stream is then forwarded to the rendering engine and output to the headphones of the local caller. Head-tracking is enabled by connecting to all the hosts that supply local virtual environments and by modifying the positions of the local as well as of the remote avatars, (see Fig. 2).

The system implemented here is based on the open-source VoIP soft-phone Ekiga enhanced by a plug-in to control the virtual environment in order to support QoE requirements. As a rendering engine, the Uni-Verse acoustic simulation framework was utilized, which is an open-source software for developing 3D games [Kajastila et al. 2007]. In our research, we only use the features that are required for spatial audio rendering. The Asterisk telephony tool kit was employed as a conference bridge and enhanced by a dial-plan application that connects to the rendering front-end; Asterisk is an open-source telephony software framework [Spencer 2008]. The current prototype system can be installed on any desktop computer or laptop running an Ubuntu/Debian-based operating system. Further details about 3D telephony and associated information can be found in [Hyder et al. 2010b, Hyder et al. 2010].

4.2 QoE and the Virtual Acoustic Environment

To evaluate the implemented 3D audio teleconferencing tool and analyze the impact of different VAEs on perceived QoE, we present a research model for 3D telephony. From the high level QoE interaction model presented in Section 3, it was emphasized that QoE interacts with the technology, business and contextual domains. Here, focus is placed on the QoE-context domain interaction. More specifically, we investigate the effects of two important VAE factors, namely virtual room size and competing talker gender on perceived QoE.

When dealing with 3D telephony, the two most important QoE factors are localization of talkers in the virtual conference room and the perceived audio quality experience in the virtual acoustic environment. Figure (3) depicts the research model used in our study. As can be seen, three different room sizes

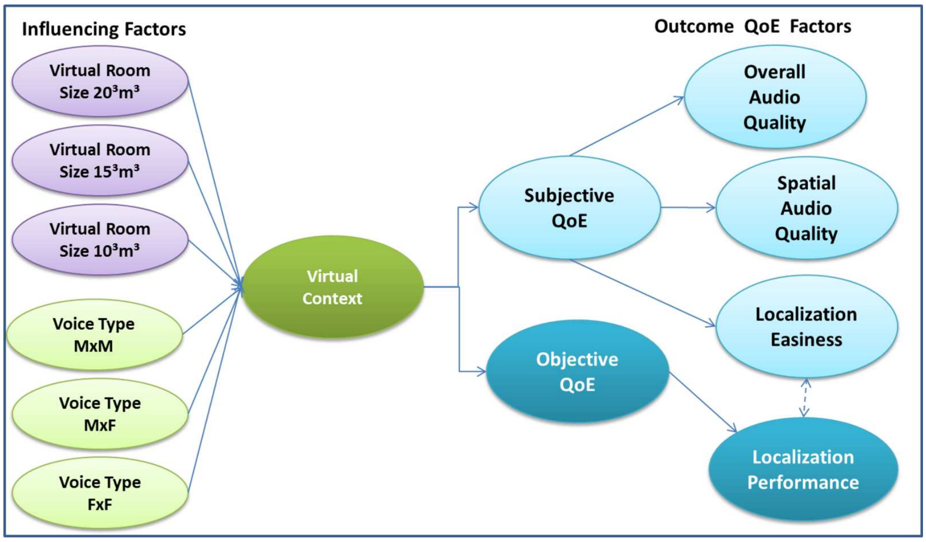

Figure 3: QoE Factors and Virtual Acoustic Environment(VAE) Relationship

(10^3m^3, 15^3m^3, and 20^3m^3) and three different competing speaker scenarios (two males, two females, one male and one female) are tested and their effects on four QoE factors are investigated, namely overall audio quality, spatial audio quality, localization easiness, and localization performance. In order to gauge the effects of varying VAE characteristics on perceived QoE, a user study was performed, as detailed below.

4.3 Methodology

In order to evaluate the relationship between QoE and 3D virtual acoustic environments, formal listening-only tests were conducted with 31 paid subjects (13 female and 18 male) in accordance to ITU-T P.800 Recommendations [P.800 1996]. All tests were conducted in a quiet listening room on a computer using a specially designed user interface on a 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 by their spoken text. Each participant was asked a series of questions to be answered for each talker within each sample. Four QoE-related metrics were used, two related to perceived quality and two related to localization. For quality perception, participants were asked to rate the spatial audio quality (SAQ) and the overall audio quality (OAQ) using a five-point scale: [1 (bad), 2 (poor), 3 (fair), 4 (good), and 5 (excellent)]. The former assessed listener perception of spatial talker separation and 3D audio

QoE Factors	Scenario		Cronbach Alpha
Localization Performance (LP) Localization Easiness (LE) Spatial Audio Quality (SAQ) Overall Audio Quality (OAQ)	Virtual Room Size	$10m^3$	.867
		$15m^3$	.844
		$20m^3$	.813
		Overall	.893
	Voice Type	M*M	.813
		F*F	.839
		M*F	.814
		Overall	.903

Table 3: Verification of Reliability and Internal Consistency of QoE Factors through Cronbach Alpha

quality, whereas the latter investigated their overall acceptance of the generated 3D sound effects. This five-point scale was also used to assess the listener's "localization easiness" (LE) in separating the competing speakers based on their locations. Lastly, localization performance (LP) was measured by presenting listeners with a map with four possible talker locations distributed around a table. LP is computed as the percentage of correctly identified speaker locations in the virtual teleconferencing room. In this experiment, six anechoic speech samples (three male, three female American English speakers) from the ITU-T Rec. P.50 Appendix 1 library were used. Speech files were processed by the open-source 3D audio rendering engine Uni-Verse [Kajastila et al. 2007] at a sampling rate of 16 kHz. More details about data generation can be found in [Hyder et al. 2010].

5 Results and Discussions

5.1 Reliability and validity verification

Prior to performing data analysis, the reliability and validity of the obtained QoE constructs (LP, LE, SAQ, and OAQ) were investigated via the Cronbach Alpha [DeVellis 1991] and the Confidence Interval (CI) tests. Results of the two tests are presented in Tables 3 and 4, respectively. As can be seen from the tables, all QoE factors obtained a Cronbach alpha parameter greater than 0.6 and 95% CIs of approximately 0.1, thus are considered to have a high level of reliability and validity.

Virtual Acoustic Environement		⇒	Localization Performance	⇒	MOS LE (with 95% CI)	MOS SAQ (with 95% CI)	MOS OAQ (with 95% CI)
Room Size	20 m³	⇒	63,44%	⇒	3,68 ± 0,11	3,84 ± 0,10	3,86 ± 0,10
	15 m³	⇒	71,77%	⇒	3,77 ± 0,10	3,79 ± 0,10	3,74 ± 0,10
	10 m³	⇒	72,31%	⇒	3,62 ± 0,11	3,58 ± 0,11	3,53 ± 0,11
Voice Type	Male Talkers	⇒	63,44%	⇒	3,68 ± 0,11	3,84 ± 0,10	3,86 ± 0,10
	Female Talkers	⇒	48,66%	⇒	3,70 ± 0,13	3,81 ± 0,11	3,81 ± 0,11
	Mixed Talkers	⇒	76,61%	⇒	3,83 ± 0,12	3,97 ± 0,10	3,87 ± 0,11
Parameters		⇒	Performance Comparison	⇒	MOS score Compariosn		

Table 4: Results of Localization Performance and MOS scores for Virtual Acoustic Environment

5.2 Relationship Between QoE Factors and Virtual Room Size

In this experiment, the QoE factors are analyzed based on changes in the size of a virtual teleconferencing room. The results presented in Fig. 4 and Table 4 suggest that there is only a subtle decrease in localization performance when we switch from a small room ($10^3 m^3$) to a medium-size room ($15^3 m^3$). However, when we switched the room size to that of a large room ($20^3 m^3$), a substantial decrease in localization performance score could be observed. The Pearson correlation coefficient was calculated between the size of virtual rooms and the LP scores and it was found to have a strong negative correlation of *(-0.89)*.

A paired-samples t-test was conducted to compare the Localization Performance scores between different virtual room sizes. The difference in mean LP scores were not found to be statistically significant between the small and medium room sizes *(p=0.85)*. However, the differences in LP scores between the small and large-size *(p=0.007)*, and medium and large-size rooms were found to be statistically significant *(p=0.009)*. These results suggest that as the size of virtual conferencing room increases, LP tends to decrease.

In Addition to LP, we also gathered subjective scores on localization easiness. The subjective scores of LE are presented in Table 4 and in Figure (5. As can be seen, LE has the highest score in medium size room and the lowest in small size room. Using the non-parametric test Wilcox Rank sum test (LE data did not pass the Shapiro-Wilk normality test, thus the t-test could not be used), it was found that LE scores were significantly different between small and medium-size rooms *(p=0.01)*. Taken together with the localization performance figures mentioned above, this suggests that while the speaker localization performance is

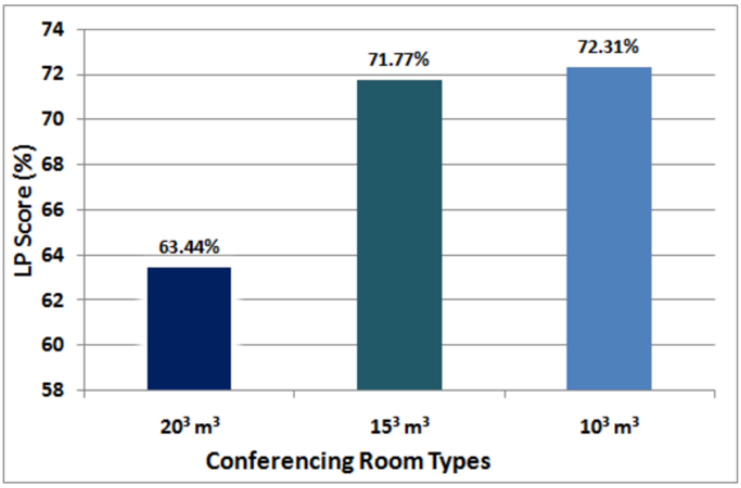

Figure 4: Comparison of localization performance for different virtual acoustic room sizes

similar in small and medium-size rooms, listeners found that locating concurrent talker in medium-size rooms was easier. On the other hand, the difference in LE scores between medium and large-size rooms were not statistically significant *(p=0.2)*. In summary, it is found that in terms of localization, the medium-size room is better suited for audio conferencing applications, as both LP and LE scores converged towards optimal values.

Moreover, Table 4 and Fig. 5 depict the effects of virtual room size on perceived spatial and overall quality. As can be seen, both SAQ and OAQ scores gradually improve with an increase in the size of the virtual room. This is the opposite of what was observed with the LP parameter and high Pearson correlation coefficients were obtained with room size: SAQ (0.94) and OAQ (0.98). One possible explanation for this result could be due to echoes and reverberations, since they are stretched in larger rooms. As reported in [Zahorik 2002], reverberation in acoustic environments is considered to be a reliable cue in identifying sound source distance, but it also modestly degrades sound source directional perception [Santarelli 2008] and speech intelligibility. In addition, it has been reported that reverberation enhances the distance perception but degrades localization performance [Shinn and Ihlefeld 2008]. In summary, small-size rooms provide better localization performance, but at the cost of lower localization easiness, spatial and overall audio quality. Large-size rooms, in turn, provide poor localization performance, but high spatial and overall audio quality. In between lies the medium sized room, which strikes a balance between acceptable localization performance and perceived quality.

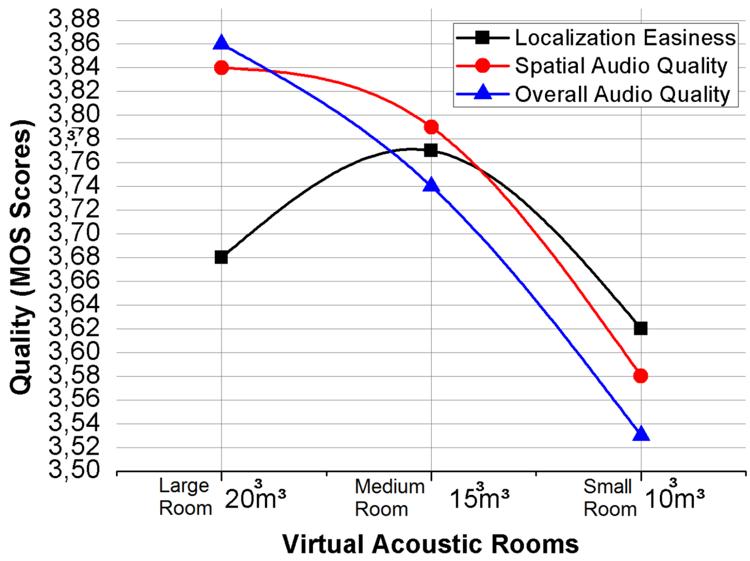

Figure 5: Comparison of quality and localization easiness scores for different virtual acoustic room sizes

5.3 Relationship Between QoE Factors and Competing Voice Type

In this experiment, we selected the large-size room and varied the voice types of simultaneous talkers in order to verify if there are any changes in QoE values and ratings. The results for the localization performance metric are shown in Table 4 and in Fig. 6. As can be seen, when both competing talkers were female, localization performance decreased substantially to less than 50%. When both competing talkers were male, LP remained at approximately 63%. Performance then increased to 76% when competing speakers of mixed gender were tested, thus suggesting that pitch differences may assist in the speaker localization task. Paired t-tests between all three voice type conditions and LP scores were found to be statistically significant *($p \leq 0.05$)*. Similar patterns were obtained with the localization easiness metric (see Fig. 7), which showed the mixed-gender competing speakers to be more easily localized. A paired Wilcox test was used to evaluate statistical significance of the LE data, and significant differences were found only with mixed-gender competing speakers. Lastly, the plots in Fig. 7 show that improved perceived spatial audio quality was obtained with competing speakers of mixed gender. A paired Wilcox test suggested a significant difference only be-

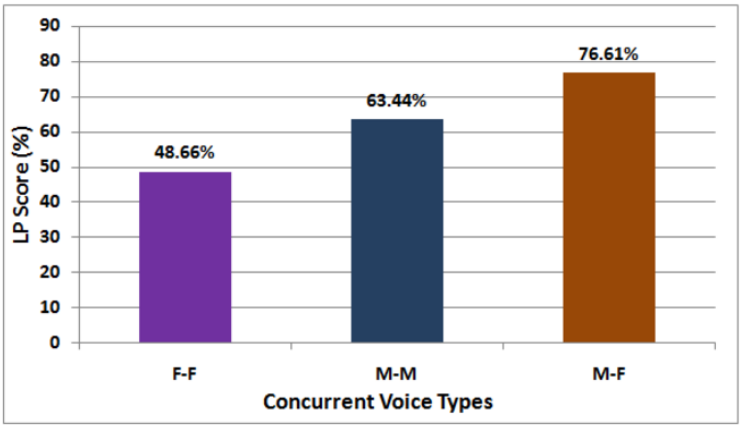

Figure 6: Comparison of localization performance for different competing voice types

tween the female-only condition and the mixed-gender condition *($p \leq 0.03$)*. For overall audio quality, all voice type conditions did not prove statistically significant as per a Wilcox paired test, thus suggesting that competing talker gender played little effect on overall quality.

6 Conclusions

This paper has investigated the effects of contextual factors on perceived QoE within the framework of a 3D audio telephony application. More specifically, the effects of two virtual acoustic environment parameters (virtual room size and competing speaker gender) were explored across four important QoE-related parameters: spatial audio quality, overall audio quality, localization easiness, and localization performance. It was observed that virtual room size plays a critical role not only on localization performance but also on overall audio quality. A medium-size conferencing room ($15^3 m^3$) was found to strike a balance between localization performance and perceived quality, thus is suggested for an optimal quality of experience. When comparing the effects of competing talker gender, it was observed that the mixed-gender condition resulted in the best localization scores little influence over perceived quality. Our ongoing studies aim at conducting objective assessments using neuro-physiological tools to better understand the intriguing relationships between the various domain characteristics within the communications ecosystem.

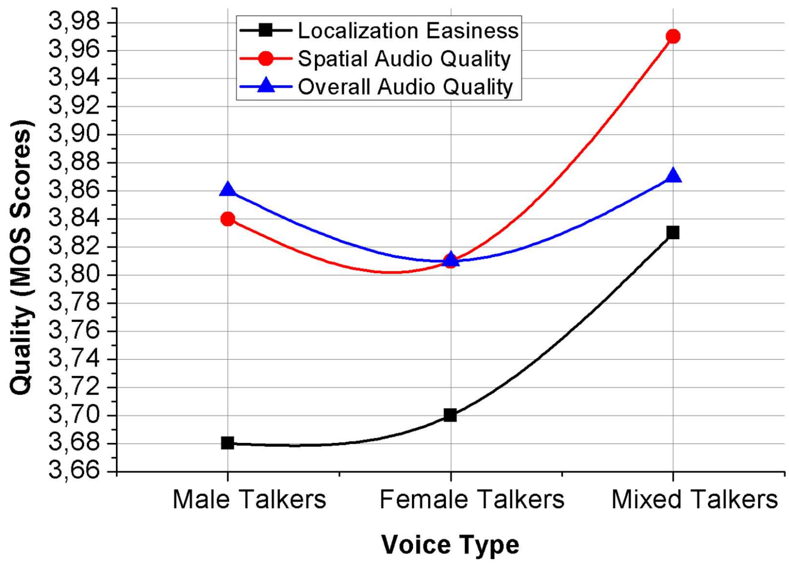

Figure 7: Comparison of quality and localization easiness scores for different competing voice types

References

[Accenture 2011] Accenture, "Accenture 2011 Global Consumer Study, The New realities of dating in Digital Age,", Consumer Survey report (2011).

[Ahrens et al. 2010] Ahrens, J., Geier, M., Raake, A., Schlegel, C., "Listening and Conversational Quality of Spatial Audio Conferencing", in Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space, (2010)

[Al-Qeisi 2009] Al-Qeisi, K.I.,"Analyzing the use of UTAUT model in explaining an online behaviour: Internet banking adoption", Brunel University, Brunel Business School, PhD Theses (2009).

[Ansi 2011] ANSI ATIS0100005-2006, Auditory Non-Intrusive Quality Estimation Plus (ANIQUE+): Perceptual Model for Non-Intrusive Estimation of Narrowband Speech Quality, American National Standards Institute, (2006)

[Begault 1994] Begault, D. R., "3-D sound for virtual reality and multimedia". San Diego, CA, USA: Academic Press Professional, Inc., (1994).

[Bergstra and Middelburg 2003] Bergstra, J. and Middelburg, C., "ITU-T Rec. G. 107: The e-model, a computational model for use in transmission planning", (2003).

[Best et al. 2006] Best, V. and Gallun, F.J. and Ihlefeld, A. and Shinn-Cunningham, B.G., "The influence of spatial separation on divided listening", The Journal of the Acoustical Society of America (2006), pp. 1506.

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

[Brungart et al. 2007] Brungart, D., 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).

[BS.1387 2001] ITU-R Recommendation BS.1387: Method for objective measurements of perceived audio quality -PEAQ (2001).

[Cooper et al. 2007] Cooper, A. and Reimann, R. and Cronin, D., "About face 3: the essentials of interaction design", Wiley-India (2007).

[Davis 1986] Davis, F.D., "A technology acceptance model for empirically testing new end-user information systems: theory and results", Sloan School of Management, Massachusetts Institute of Technology (1986).

[DeVellis 1991] DeVellis, R.F., Scale Development Theory and Applications, Applied Social Research Methods Series, Vol 26. (1991).

[Dey 2011] Dey, A. "Understanding and using context", Personal and ubiquitous computing, vol. 5, no. 1, pp.4-7, (2001).

[Dictionary 2011] Dictionary, H.M, "The American Heritage Science Dictionary, Houghton Mifflin Company", (2011). `http://dictionary.reference.com/`.

[ETSI 2009] ETSI, STQ, European Technical Committee for Speech, Transmission, Planning, and Quality of Service (2009). `http://www.etsi.org/WebSite/homepage.aspx`.

[Falk and Chan 2008] T. Falk and W.-Y. Chan, Hybrid Signal-and-Link-Parametric Quality Measurement for VoIP Communications, IEEE Trans. Audio Speech Lang. Process., Vol.16, No.8, pp.1579-1589, (2008).

[Falk and Chan 2009] T. Falk and W.-Y. Chan, Performance Study of Objective Speech Quality Measurement for Modern Wireless-VoIP Communications, J. Audio Speech Music Process., Vol. 2009, 11 pages, (2009).

[G.1080 2008] G. 1080:" Quality of experience requirements for IPTV services", ITU-T, Rec. (2008).

[Hawley et al. 1999] Hawley, M.L., Litovsky, R.Y., and Colburn, H.S., "Speech intelligibility and localization in a multi-source environment", The Journal of the Acoustical Society of America (1999), pp. 3436.

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

[Hyder et al. 2010b] Hyder, M., Haun, M., Weidmann, O., and Hoene, C., "Assessing virtual teleconferencing rooms", in 129th Audio Engineering Society (AES) Convention, San Francisco, CA, USA, Nov. (2010).

[Inkpen et al. 2010] Inkpen, K., Hegde, R., Czerwinski, M., and Zhang, Z., "Exploring spatialized audio and video for distributed conversations", in Proceedings of the 2010 ACM conference on Computer supported cooperative work. ACM, (2010), pp. 95 - 98.

[ITU-T 2007] ITU-T, Definition of Quality of Experience (QoE) (2007).

[Kajastila et al. 2007] Kajastila,R., Siltanen, S., Lunden, P., Lokki, T., and Savioja, L., "A distributed real-time virtual acoustic rendering system for dynamic geometries", in 122nd Convention of the Audio Engineering Society (AES), Vienna, Austria, May (2007).

[Kilgore et al. 2008] Kilgore, R., Chignell, M., and Smith, P., "Spatialized Audioconferencing: what are the benets¿', in Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research. IBM Press, (2003), p. 144.

[Kilgore 2008] Kilgore, R., "Simple Displays of Talker Location Improve Voice Identication Performance in Multitalker, Spatialized Audio Environments", Human Factors, vol. 51, no. 2, p. 224, (2009).

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

[Kitashima et al. 2008] Kitashima, Y. and Kondo, K. and Terada, H. and Chiba, T. and Nakagawa, K., “Intelligibility of read Japanese words with competing noise in virtual acoustic space“, Acoustical science and technology (2008), pp. 74 - 81.

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

[Laghari et al. 2012] Laghari, K.U.R. and Connelly, K.,“Toward total quality of experience: A QoE model in a communication ecosystemt“, Communications Magazine, IEEE, April 2012, Vol=50,No=4, pp.58-65.

[Laghari et al. 2011] Laghari, K., Molina, B., Crespi, N. and Palau, C., “QoE aware Service Delivery in Distributed Environment“, Advanced Information Networking and Applications Workshops, March 22 - 25, Biopolis, Singapore, (2011).

[Laghari et al. 2010a] Laghari, K.R., Yahia, B.I.G., and Crespi, N., “Analysis of Telecommunication Management Technologies“, International Journal of Computer Science (2010).

[Laghari et al. 2010b] Leghari, K.U.R., Yahya, I.G.B. and Crespi, N., “Towards a service delivery based on customer eXperience ontology: shift from service to eXperience“, in Proceedings of the 5th IEEE International Conference on Modeling autonomic communication environments. Springer Verlag, (2010). pp. 51–61.

[LiaisonStatemen 2007] Liaison Statement,“Definition of quality of experience (QoE)“, ITU-T, TD 109rev2 (PLEN/12), Jan. (2007).

[Möller et al. 2011] Mller, S. Chan, W.-Y. N. Cte, T. Falk, A. Raake, and M. Wltermann, Speech Quality Estimation: Models and Trends, IEEE Signal Process. Mag., Vol. 8, No. 6, pp. 18-28, Nov (2011).

[P.563 2004] ITU-T P.563, Single ended method for objective speech quality assessment in narrow-band telephony applications,” Intl. Telecom. Union, (2004).

[P.800 1996] P. 800: Methods for subjective determination of transmission quality, ITU-T, Rec. (1996).

[P.861 1998] P.861, “Objective quality measurement of telephone-band (300-3400 hz) speech codecs“, ITU.T, Rec. (1998).

[P.862 2001] P.862, “Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs“, ITU-T. Rec. (2001)

[P.863 2011] ITU-T Recommendation P.863: Perceptual objective listening quality assessment, Geneva, (2011).

[Pulkki 2001] Pulkki, V., “Spatial sound generation and perception by amplitude panning techniques“, Doctoral thesis, Helsinki University of Technology, Espoo Finland (2001).

[Raake et al. 2007] Raake, A., Spors, S., Ahrens, J., and Ajmera, J., “Concept and evaluation of a downward compatible system for spatial teleconferencing using automatic speaker clustering“, in 8th Annual Conference of the International Speech Communication Association, Aug.(2007), pp.1693-1696.

[Radhakrishnan and Larijani 2010] Radhakrishnan, K. and Larijani, H., “A study on QoS of VoIP networks: a random neural network (RNN) approach“, in Proceedings of the2010 Spring Simulation Multiconference. ACM, (2010), p. 114.

[Santarelli 2008] Santarelli, S.G.,“Auditory localization of nearby sources in anechoic and reverberant environments“, Ph.D. dissertation, Boston University (2001).

[Sheppard et al. 1988] Sheppard, B.H., Hartwick, J. and Warshaw, P.R., “The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research“, Journal of Consumer Research (1988), pp. 325–343.

[Spencer 2008] Spencer, M. Asterisk PBX, (2010). `http://www.asterisk.org/`

[Shinn and Ihlefeld 2008] Shinn, C., B. and Ihlefeld, A., “Selective and divided attention: Extracting information from simultaneous sound sources“, in Proc. ICAD, vol. 4, (2004).

[Takahashi et al. 2008] Takahashi, A. and Hands, D. and Barriac, V., "Standardization activities in the ITU for a QoE assessment of IPTV", Communications Magazine, IEEE (2008), pp. 78–84.

[Zahorik 2002] Zahorik, P., "Assessing auditory distance perception using virtual acoustics", The Journal of the Acoustical Society of America (2002), pp. 1832.