Public Viewing Service with High Presence by Using Binaural 3D Audio IP Phone

Yuki Yoshida, Yoshihiro Ito and Kaori Iwata
Department of Computer Science
Nagoya Institute of Technology
Nagoya, 466-8555
Japan
Email: {yuki@en, yoshi@, iwt@en}.nitech.ac.jp

***Abstract*—This paper proposes a new public viewing service with high presence by utilizing binaural 3D audio IP phone. The authors evaluate the effectiveness of the proposal by experiment. In the experiment, the authors utilize actual live concert movies and perform subjective and objective evaluation. The subjective assessment is done by the SD method with 9 evaluation terms. On the other hand, the objective evaluation is performed by Objective Difference Grade (ODG). From the experimental results, the authors confirm that a public viewing with 3D audio gives higher feelings of presence than that with monaural audio.**

I. Introduction

Recently, we can watch moving pictures in a public space; we often refer to the service as a public viewing (screening) service. Now we can enjoy various events by a public screening, for example, a sporting game, a live concert, and so on. Thus, as public viewing services are widely spread, they are expected to provide higher presence. In a public viewing service, audiences generally share not only video but also audio. Especially, since the audio is generated by fixed speakers, the artificial audio may degrade the feel of presence of audiences. Even if each audience has a pair of headphones and listen the audio with it, it does not always improve the feel of presence.

This paper proposes a novel public viewing service with binaural 3D audio IP phone service, which has been developed by the authors[1]. In our proposal, the audio of the public viewing is recorded as binaural sounds, which is one of 3D audio recordings, and transmitted over IP networks. The 3D audio can make the feel of presence of a public viewing high.

The paper is organized as follows. Section II shows experiments. Sections III presents the results. Finally we conclude our research in Sect. V.

II. Experiments

In our experiment, we utilize three movies of a rock music recorded at an actual live concert and assess the presence of a public viewing service by subjective assessment with subjects. We refer to the movies as Movie-1, Movie-2 and Movie-3. Since we target a public viewing, the video of each movie is projected onto a screen without transmitting over networks while its audio is encoded and transmitted over IP networks.

Figure 1 shows our experimental environment. Under this environment, the video player/audio encoder connects with audio the video player/audio encoder projects the video of the recored movie onto a screen with a projector; it also encodes the audio of the movie and transmit it to the audio decoder/player. The network emulator[2] connects between the video player/audio encoder and the audio decoder/encoder; it can degrade the quality of the IP network between them by causing packet losses. The packet loss rates are 0 % and 3 %. The subjects watch the movie projected onto the screen while listening the audio output by the audio decoder/player.

The audio is recorded by binaural recording. For comparison, we also treat monaural recording audio. We adopt ITU-T H.264[3] as the encoding method of the movie; we treat LPCM and ITU-T G.711[4] μ-law as the one of the audio. However, for monaural audio, we use only LPCM encoding.

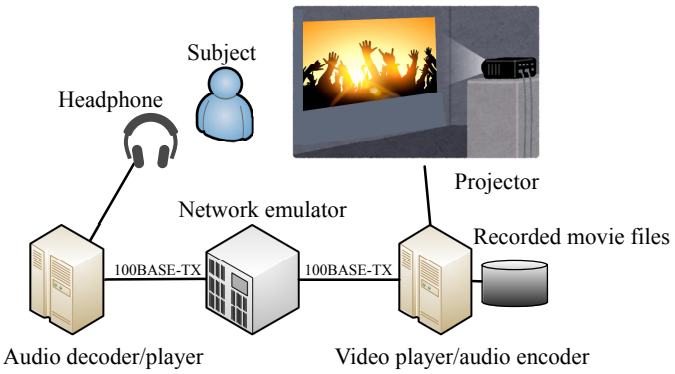

Fig. 1. Experimental envinronment.

In order to perform subjective assessment, we utilize SD method with 9 evaluation terms. The evaluation terms are the following pairs of adjectives.

- beautiful–dirty
- acoustic–not acoustic
- visible–invisible
- clear–unclear
- congruous–incongruous
- natural–artificial
- roundish–edgy
- cubic–flat

- high presence–low presence

We adopt a 5-grade rating-scale method[5] for the above evaluation terms as subjective assessment and treat the mean value of the scores voted by subjects as presence score.

In addition to the subjective assessment, we also perform objective evaluation. We utilize the objective difference grade (ODG) as the objective evaluation method. The ODG is calculated by perceptual assessment of the audio quality algorithm defined in ITU BS.1387-1(PEAQ)[6]. It corresponds to the subjective difference grade and ranges from 0 to -4.

III. Results

Figure 2, Fig. 3 and Fig. 4 display the presence score for Movie-1, that for Movie-2 and that for Movie-3, respectively. We also plot ODG values in these figures.

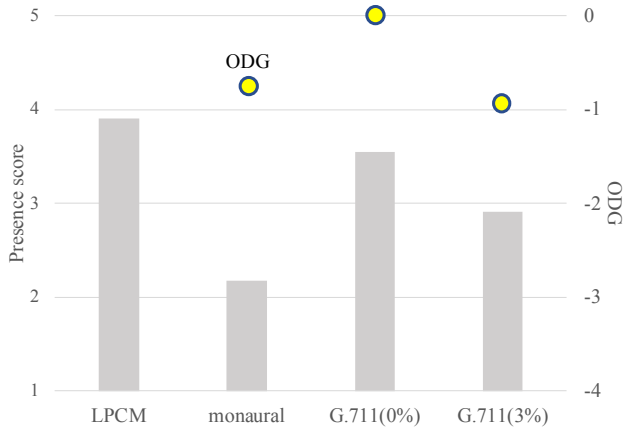

Fig. 2. Presence score and ODG for Movie-1.

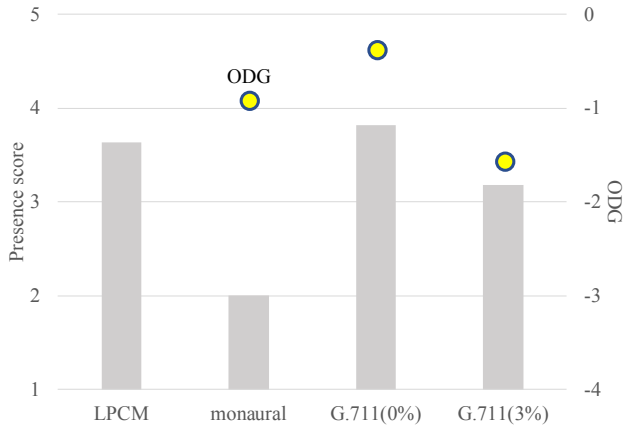

Fig. 3. Presence score and ODG for Movie-2.

From Fig. 2–4, we find that the presence score with 3D audio (LPCM, G.711) are higher than that with monaural for all of the three movies. Especially, although the objective score (ODG) of monaural is better than that of 3D audio with 3% packet losses, the presence score of the former is higher than that of the latter. Here, we should note that utilization of 3D audio can improve the feeling of presence of audiences more than the improvement of the audio quality. That is, even if we

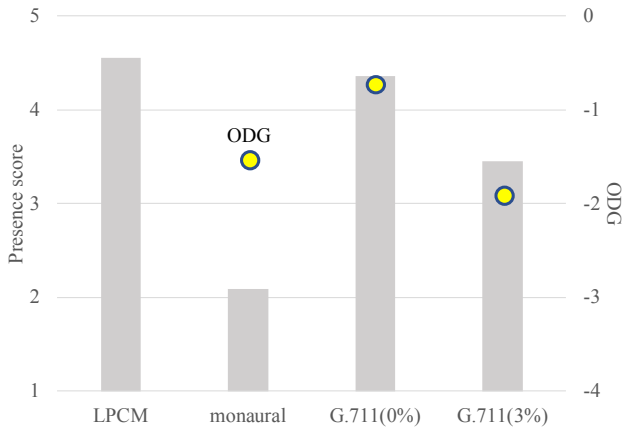

Fig. 4. Presence score and ODG for Movie-3.

cannot keep enough communication quality, high feelings of presence for a public viewing can be realized by adopting 3D audio IP phone service.

IV. Conclusions

This paper discusses a novel public viewing service with binaural 3D audio IP phone service. In the experiment, we targeted three movies recorded at an actual live concert and performed both of the subjective assessment and the objective evaluation. From the experimental results, we showed that the feelings of presence of audiences with 3D audio is higher than that of monaural. Therefore, the effectiveness of the proposal was confirmed.

We have some issues for our future work. First, we will treat other contents except for live concerts. Second, it is necessary to evaluate the effectiveness of the proposal for actual public viewing services.

Acknowledgment

This work was supported by MEXT KAKENHI Grant Number JP 16K00122.

The authors thank Mr. Takuma Yoshida for his assistance in our research.

References

[1] M. Tsubota and Y. Ito, “Assessment of Sense of Presence for 3D IP Phone Service Under Actual Environments,” Proc. GCCE 2015, pp.3–4, Oct. 2015.

[2] “The dummynet project,” http://info.iet.unipi.it/ luigi/dummynet/.

[3] ITU-T,“H.264: Advanced video coding for generic audiovisual services,” Apr. 2017.

[4] ITU-T, “G.711 : Pulse Code Modulation(PCM) of Voice Frequencies,” Nov. 1988.

[5] J.P.Guilford, “Psychometric methods,” 1954.

[6] ITU-R, “BS.1387-1 : Methods for objective measurement of perceived audio quality,” Nov. 2001.