5G-Enabled Augmented Audience Services for Live Events

Robert Hupke[1], Marcel Nophut[1], Stephan Preihs[1], Jürgen Peissig[1]

[1] *Leibniz Universität Hannover, Institut für Kommunikationstechnik, 30167 Hannover, Germany*

Email: hupke@ikt.uni-hannover.de.de

Abstract

PMSE-xG is a research project studying 4G+/5G technologies and their usage in wireless equipment for "Programme Making and Special Events" (PMSE) applications. The PMSE-xG project [1] has the vision of delivering important advancements in the creative industry ecosystem, focusing on audio and video applications as well as potential new audience services. This paper presents some of those services and proposes different use cases of innovative audio applications for listeners at live events that were developed within the project.

Using an augmented reality audio headset that is able to provide both environmental sounds and supplemental audio content, the services and their applications are aiming at improving or enhancing the listening experience without losing the auditory space of a live event. With the possibility of blending additional content into the existing sound experience of every single listener, the services provide individualized live content to audiences at concerts, sport or voice-based events.

Introduction

The PMSE-xG project considers the future-oriented use of mobile broadband radio technology and network infrastructure (4G+/5G) for PMSE applications. As presented in Figure 1 the project incorporates all technologies which are used within a live event production, like wireless microphones, in-ear monitors, video cameras and conferencing systems. In the depicted scenario all the PMSE equipment is connected to a local high quality network in which the recording, mixing and live editing can be conducted. The same network can deliver content to mobile devices of the audience which enables new possibilities of augmented audience services.

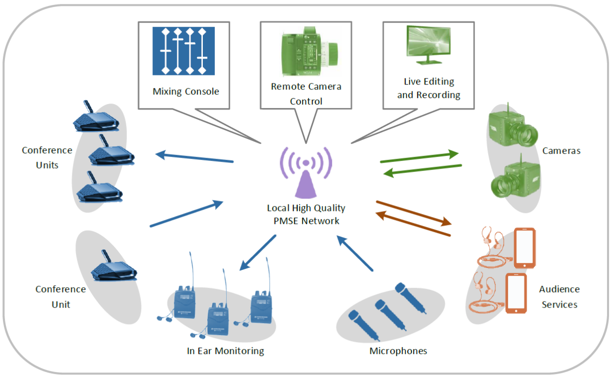

Figure 1: Summary of Use Cases [1].

As a new feature of PMSE applications, these new possibilities of using innovative audience services at live events, which are enabled by the future low-latency wireless technologies were investigated within the project PMSE-xG.

In this paper, the so called "Assistive Live Listening Service" and its possible applications at live events are presented. A first perceptual evaluation of the developed service was conducted to evaluate if it has the potential of improving speech intelligibility without the loss of live experience.

Assistive Live Listening Service

Nowadays, at a concert or a voice-based live event audio content is presented via a PA system. The novel "Assistive Live Listening Service" additionally provides individualized audio content to listeners at such a concert or voice based live event by the use of a low latency wireless audio link. The service can be used to achieve enhancements of the listening experience, e.g. to overcome poor acoustic conditions or to meet special preferences of the users regarding their listening habits.

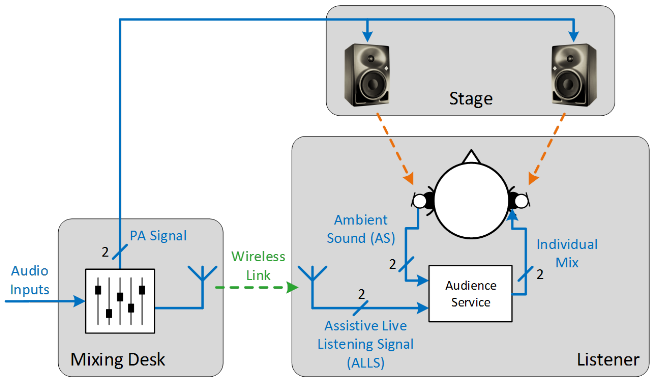

Figure 2: Assistive Live Listening Service setup.

The enhancement is achieved by mixing the Ambient Sound (AS) of the listener with supplementary audio content, the Assistive Live Listening Signal (ALLS), and presenting this individual live mix through in-ear headphones during the live event. The setup of the "Assistive Live Listening Service" for live concerts is presented in Figure 2. The listener stands in the sound field of the PA system. The AS is picked up by microphones placed at the outside of the headphones close to the listener's ear canal entrance and the ALLS, e.g. a dry stereo live mix or a dry speech signal, can be received through the low-latency wireless link from a central mixing unit.

Since the microphones of the headset capture the PA loudspeaker playback at the live event, a temporal offset between AS and ALLS can cause irritating echoes or comb filtering effects. Thus, the presented ALLS has to be time-aligned to the sound that reaches the ear from the PA system [2, 3].

Applications

The listener can take advantage of an improved sound experience by wearing headphones connected to his wireless Assistive Live Listening Device (e.g. future 4G+/5G enabled smart phones). The "Assistive Live Listening Service" enables a lot of applications by further audio processing steps to suit the needs of the listener. Several applications for enhancing the live listening experience are possible:

- Enhancement of speech intelligibility.
- Equalizing the ALLS and AS to suit the listener's needs or listening habits.
- Reducing the reverberant sound components while preserving the main characteristics of the venue sound.
- Highlighting of different instrument groups.
- Blending in augmented audio content, which is not present at stage or not presented at the PA system.

Setup and Test Design

A subjective listening experiment was carried out to evaluate if the ALLS may enhance the speech intelligibility without the loss of perceptual live experience.

For this, a 17 s long male speech mono signal from the EBU SQAM CD [4] was chosen as the ALLS. The same signal was filtered with a room impulse of a large church ($T_{60,mean} = 3.2\,\text{s}$)) and was presented over loudspeakers ($\pm 60°$) 2.4 m in front of a dummy head. The AS was recorded from in-ear microphones attached to a prototype headset (Figure 3), which was placed at the dummy heads ear canal.

Figure 3: Prototype headset with external microphones.

To prove the similarity between the loudspeaker playback and the recorded AS, the recording was conducted at the same position as the test subjects were sitting during the listening test. Before mixing the signals, the ALLS was time aligned to the AS signal. The aim of the listening experiment was to evaluate the relation between the enhancement of speech intelligibility for different mixing ratios to the effect of the dry ALLS on the live experience. Thus five mixing ratios of linear amplitude were chosen for the first perceptive evaluation:

– Ratio #1: 100% AS and 0% ALLS,
– Ratio #2: 75% AS and 25% ALLS,
– Ratio #3: 50% AS and 50% ALLS,
– Ratio #4: 25% AS and 75% ALLS,
– Ratio #5: 0% AS and 100% ALLS.

Speech Transmission Index (STI)

The STI values of all five mixing ratios were computed using their impulse responses and are presented in Figure 4. As expected, the STI value increases with a higher percentage of ALLS in the mixing ratio, while the STI is poor for the pure AS signal. Between the mixing ratios of 50/50 and 25/75 AS to ALLS, an excellent STI is achieved.

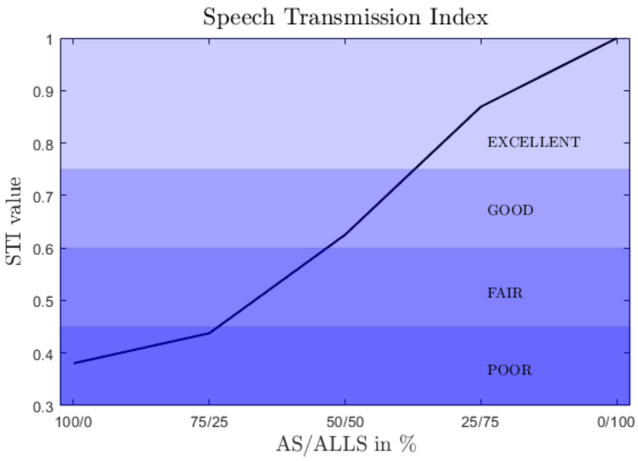

Figure 4: Calculated Speech Transmission Index (STI) of all five mixing ratios.

Perceptual Listening Test

The Spatial Audio Quality Inventory (SAQI) [5] test was used for a first evaluation of the service. Since the original SAQI test is rather time-consuming, a customized setup with a subset of six perceptual qualities was used. The six perceptual qualities were:

- Difference,
- Reverberation Time,
- Speech Intelligibility,
- Naturalness,
- Presence,
- Degree-of-Liking.

These qualities should give indication whether the service could improve speech intelligibility and still remain the feeling of being in a live performance. The listening experiment was implemented in Matlab using WHISPER [6], a package of scripts providing several listening test procedures as well as the SAQI. For every subject, the order of perceptual qualities as well as the conditions per quality were randomized. The subjects had to listen to each stimulus and reference for each auditory quality before being able to rate the quality, since the feedback sliders were disable beforehand. This led to a time period of approx. 45 minutes per subject. The subjects were a group of 11 normal hearing listeners. All subjects used open headphones (Sennheiser HD800) and for the presentation over loudspeaker, two Neumann KH120 were used. The listening test was conducted in the Immersive Media Lab [7] shown in Figure 5. The listening experiment was divided into two subtests:

- **Test 1:** The subjects were asked to rate the difference of six perceptual qualities for the presentation of the loudspeaker playback and the recorded AS signal through headphones.
- **Test 2:** The subjects were asked to rate six perceptual qualities of all five mixing ratios with respect to the reference (the recorded AS), both presented through headphones.

The subjects were instructed to assess auditory differences related to a reference stimulus. As reference stimulus in Test 1 the loudspeaker playback which was the ALLS convolved with the impulse response of a large church was chosen. The subjects were instructed to imagine that they are sitting in a large church, where a speech on the other side of the nave is given. The playback over loudspeakers was supposed to simulate this situation. The aim of this experiment was to evaluate the similarities between the original sound field of the loudspeakers and the AS recorded over the external microphones of the headset.

As reference stimulus for Test 2 the recorded AS was chosen, which corresponds to the signal the listener would hear at a live event without the use of the ALLS. The subjects were introduced to the same situation as in Test 1.

Figure 5: Listening Test Setup in the Immersive Media Lab.

Results Test 1

The subjects had to judge 6 perceptual qualities by comparison of the loudspeaker playback and the recorded AS stimulus (Ratio #1) over headphones. The perceived differences of all six qualities are very small as shown in Figure 6. Thus, the AS signal presented over headphones is very similar to the original playback over loudspeakers at the listing position. Some non-experienced listeners even thought the playback over headphones did not work because it sounded the same as the loudspeaker playback.

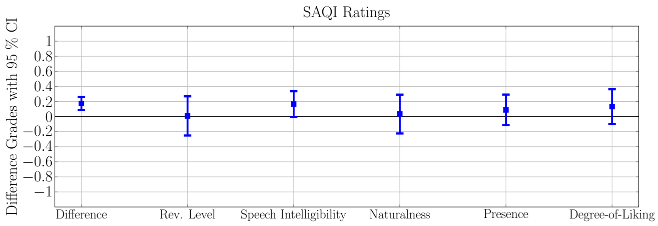

Figure 6: Result of Test 1: Average difference grades of headphone vs. loudspeaker presentation of six perceptual qualities with the 95% confidence interval.

Results Test 2

As Test 1 showed that there were no or only small audible differences between the perception of the AS signal and the real sound field produced by the loudspeaker at the listening position, in Test 2 the subjects had to judge the different mixing ratios presented exclusively over headphones.

The results of the perceptual quality "Reverberation Level" and "Speech Intelligibility" are presented in Figure 7. As expected, with a higher percentage of the ALLS in the mixing ratio, the perceived "Reverberation Level" decreases while the "Speech Intelligibility" increases. This demonstrates that the subjects were able to rate the mixing-ratios in a differentiated way. The perceived "Speech Intelligibility" corresponds to the objective expectations resulting from the calculated STI of the test ratios in Figure 4.

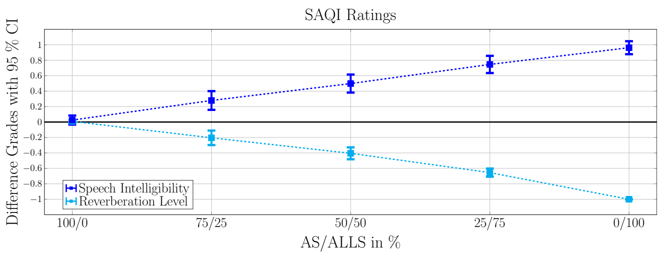

Figure 7: Result of Test 2: Average ratings with 95% confidence intervals. Qualities: "Reverberation Time" and "Speech Intelligibility".

In Figure 8 the results of the perceptual qualities "Naturalness" and "Presence" are shown. As both are room acoustical qualities and have nearly the same trend, they are summarized under the term of "Live Experience". Up to the mixing ratio of 25/75 AS to ALLS, the "Live Experience" is nearly the same as for the pure AS. For the mixing ratio of 0/100 AS to ALLS, which is the dry speech signal, the results show very little or no "Naturalness" and "Presence". Therefore it can be stated that no "Live Experience" exists. However, the result of the 75/25 ratio even gives a better "Live Experience" than the AS signal alone.

Thus, the "Live Experience" is preserved up to the mixing ratio of 25/75. The ALLS may potentially enhance the "Live Experience" in case of the mixing ratio close to 75/25.

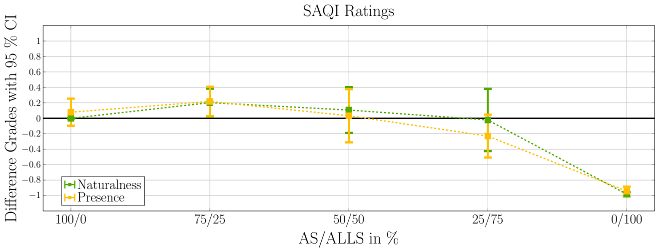

Figure 8: Result of Test 2: Average ratings with 95% confidence intervals. Qualities: "Naturalness" and "Presence".

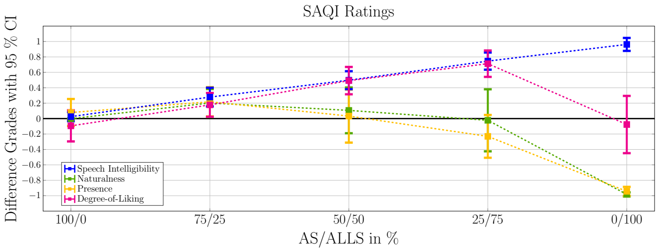

Figure 9: Result of Test 2: Average ratings with 95% confidence intervals. Qualities: "Speech Intelligibility", "Naturalness", "Presence" and "Degree-of-Liking".

In Figure 9 the perceived results of the qualities "Live Experience", "Speech Intelligibility" and "Degree-of-Liking" are presented. With a higher percentage of ALLS the "Degree-of-Liking" increases up to the dry ALLS signal which was rated nearly the same as the AS signal. Although the increase of "Degree-of-Liking" with a higher percentage of ALLS signal in the AS to ALLS ratio is an inverse effect to the qualities "Naturalness" and "Presence", it shows that the "Degree-of-Liking" seems to be directly related to the "Speech Intelligibility" until the dry signal dominates the mix and the "Live Experience" does not exist any more.

Thus, not only "Speech Intelligibility" can be improved but also the "Assistive Live Listening Services" can potentially enhance the "Live Experience" for the listener in terms of the qualities "Naturalness" and "Presence" as well as the overall "Degree-of-Liking". The results have shown that a mixing ratio between 50/50 and 25/75 might be a good compromise between an excellent speech intelligibility without the loss of "Live Experience".

Conclusions

In this paper a novel augmented audience service was presented. The so called "Assistive Live Listening Service" and its possible applications for live events are described. For a first evaluation of our service in terms of improvement of "Speech Intelligibility" vs. "Live Experience", a listening test was conducted and the results of the test are presented.

It has been shown that by the use of the "Assistive Live Listening Service", an enhancement in "Speech Intelligibility" is possible by mixing the dry ALLS to the AS. The listening experiment showed that neither the single AS nor the ALLS seem to be satisfying signals for a live event. A mix of both achieves a much higher "Degree-of-Liking", which was highest with a very high percentage of ALLS in the AS to ALLS mixing ratio (25/75 AS to ALLS ratio). Thus, a mixing ratio between 50/50 and 25/75 AS to ALLS ratio seems to be good a compromise between an excellent "Speech Intelligibility" without the loss of "Live Experience" and a much higher "Degree-of-Liking" than the single AS.

Further investigations have to be done to prove the enhancement of "Live Experience" by the service, while this first evaluation shows that it seems to be possible not only to improve "Speech Intelligibility" without the loss of "Live Experience" but also to enhance the complete "Live Experience" in terms of "Presence" and "Naturalness". For this, further investigations have to be conducted, e.g. in a live event situation.

Acknowledgements

The project PMSE-xG was co-funded by the German Federal Ministry of Transport and Digital Infrastructure (reference number: DG12/8331.3-4 PMSE-xG).

References

[1] "PMSE and 5G", white paper, PMSE-XG Project, 2017, `http://pmse-xg.research-project.de/`

[2] Nophut, M., Hupke, R., Preihs S., Peissig J., "Comparison of Methods for Estimating the Propagation Delay of Acoustic Signals in an Audience Service for Live Events," in *144th Convention of the Audio Engineering Society (accepted for publication)*, 2018.

[3] Nophut, M., Hupke, R., Preihs S., Peissig J., "Real-Time Estimation of Propagation Delays for Temporal Alignment of Audio Signals in Augmented Reality Applications," in *Fortschritte der Akustik: DAGA 2018, München: 9.-22. März 2018: 44. Jahrestagung für Akustik*, 2018.

[4] EBU – TECH 3253, "Sound Quality Assessment Material recordings for subjective tests," Geneva, 2008.

[5] Lindau A., Erbes V., Lepa S., Maempel H-J., Brinkmann F., Weinzierl S., "A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI)," in *Acta Acustica united with Acustica*, 100(5), 984-994, 2014.

[6] Ciba S., Wlodarski A., and Maempel H.-J., "WhisPER – A New Tool for Performing Listening Tests," in *126th Convention of the Audio Engineering Society*, 2009.

[7] Hupke R., Nophut M., Li S., Schlieper R., Preihs S., Peissig J., "The Immersive Media Laboratory: Installation of a Novel Multichannel Audio Laboratory for Immersive Media Applications," Convention e-Brief, in *144th Convention of the Audio Engineering Society (accepted for publication)*, 2018.