Telepresence-Enhanced Network Music Performance

Introduction

Network Music Performance (NMP) can greatly benefit from 5G, as it requires 30-40 ms of latency and large amounts of bandwidth. Volumetric video can really transform NMP to an XR experience, if 5G can support its bandwidth needs. The TENeMP project carried out experiments in the Berlin 5G SA testbed provided by SPIRIT and the Athens 5G NonSA testbed at the MMLab, to see if 5G can make XR-enabled NMP a reality.

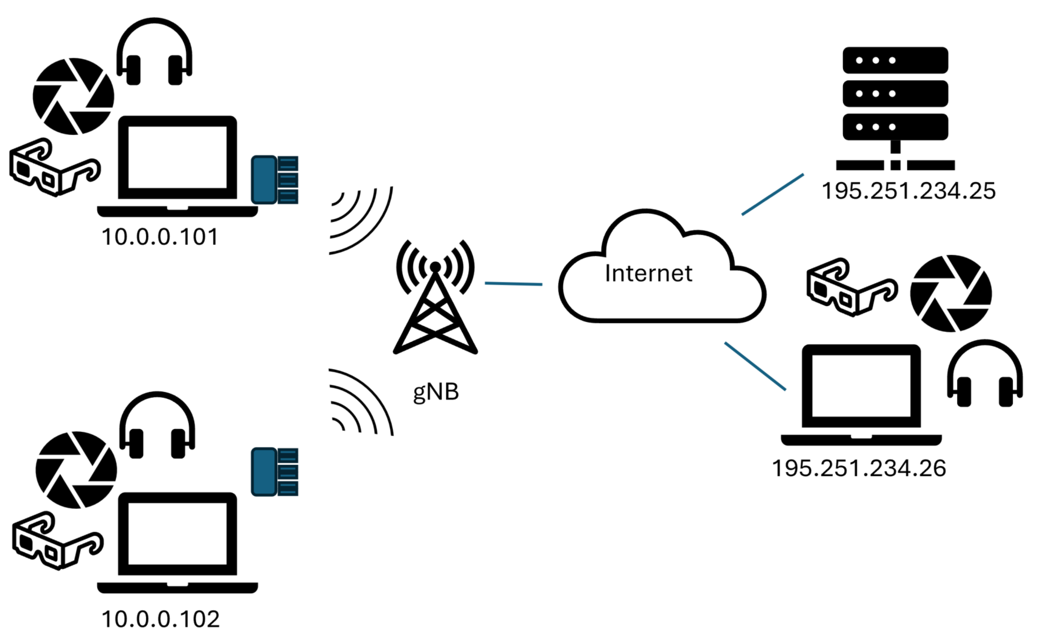

Fig. 1: Athens testbed

Testbed Setup

We tested two basic communication scenarios: Peer-to-Peer (P2P), where each musician directly communicates with all others, and Client-SFU (CSFU) scenario where musicians communicate indirectly, via an SFU, ideally located close to the endpoints at a MEC server.

In addition to the SPIRIT 5G testbed in Berlin, we created our own 5G testbed in Athens for development and testing (Fig.1). The two testbeds had three major differences:

1. In Athens we only had 5G-NonSA connectivity (COSMOTE/TELEKOM); Berlin had 5G-SA connectivity.
2. In Berlin we had MEC servers in the 5G cell, while in Athens they were deployed in our LAN. This made a real difference when using an SFU.
3. Berlin was isolated from the Internet, using private IPs. Athens used a public 5G network, which used NAT; we resorted to UDP hole punching to traverse it.

Measurement procedures

For audio, we measured the Mouth to Ear (M2E) latency, which is the time between a user producing a sound and the sound reaching the ears of another user. We used the reflected sound method to measure latency (Fig. 2).

For 2D video, the Glass to Glass (G2G) latency is to the time it takes for a frame to be captured by a camera and the corresponding frame to be presented at a screen. We used the flashing LED method to measure it, with an Arduino controlling the LED and detector (Fig. 3).

For volumetric video, we measured latency by analyzing the logs of the consumer / producer applications.

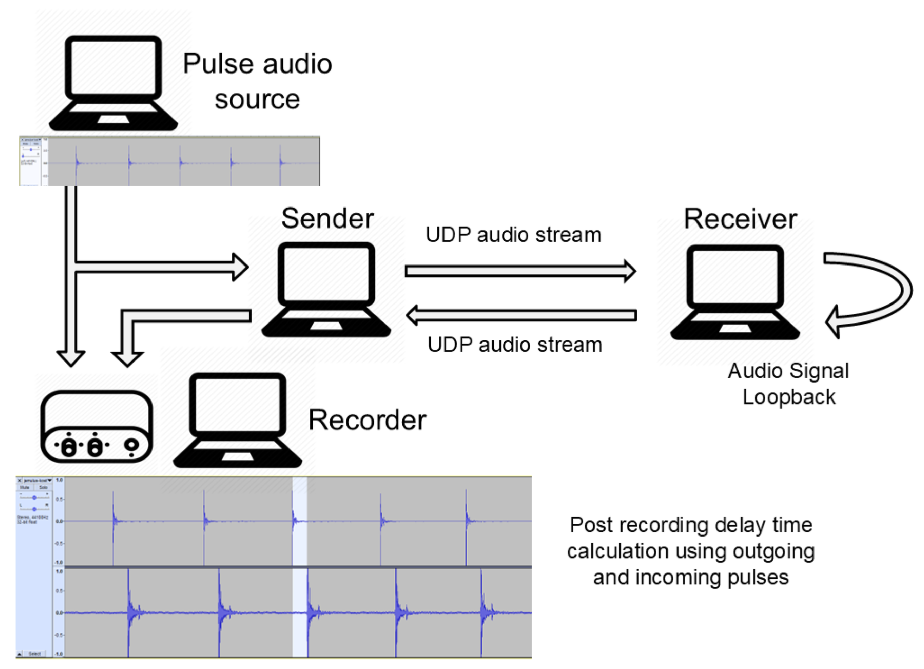

Fig. 2: Audio measurement setup

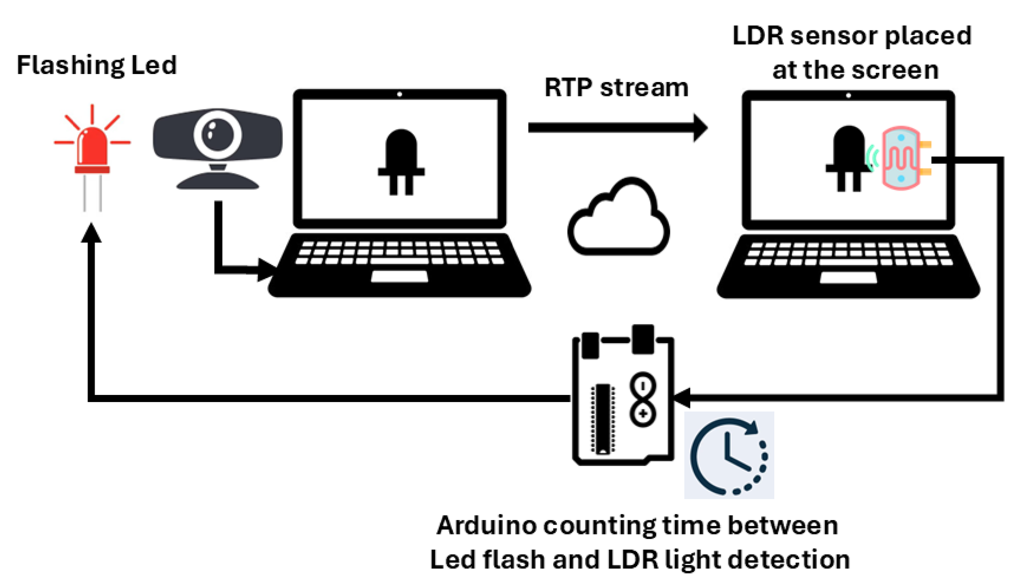

Fig. 3: Video measurement setup

Results

For audio, we used PulseAudio with Gstreamer to capture an audio channel at 44.1 KHz @ 16 bps. The buffer size was set to 132 samples. In CSFU mode, we used another Gstreamer pipeline, which simply relayed incoming UDP packets from one client to the other. Figure 4 shows boxplots for the M2E audio delay.

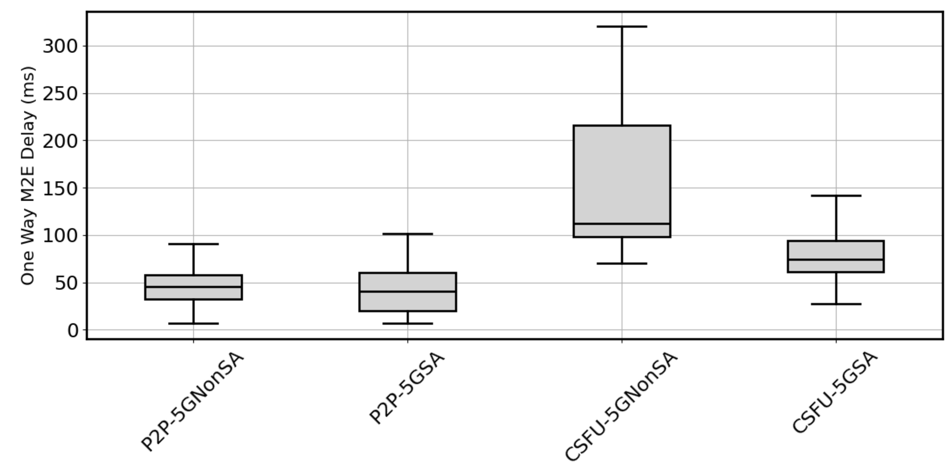

Fig. 4: Audio latency

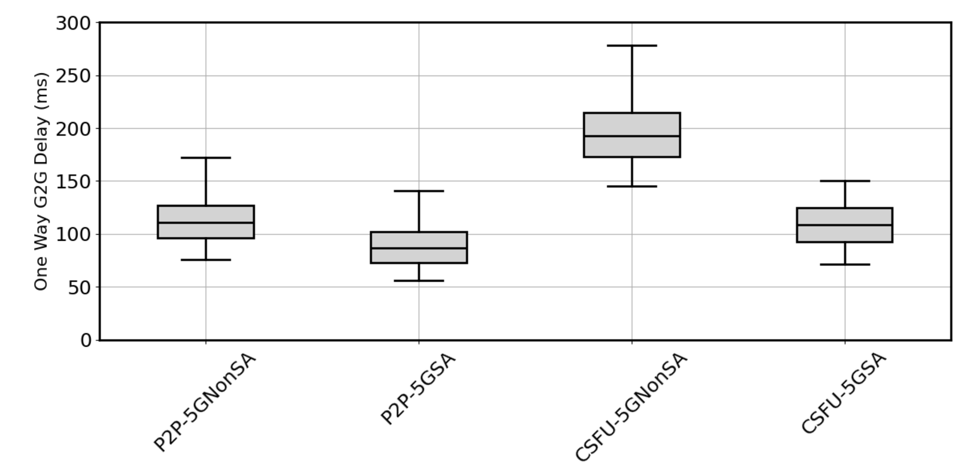

Fig. 5: Video latency

For 2D video we used Video4Linux with Gstreamer to capture video and H.264 to compress it. In Berlin we used 1400-byte UDP/RTP packets. For the CSFU topology, we again used Gstreamer. In Athens, we reduced the packet size to 300 bytes, for better quality. Figure 5 shows boxplots for the G2G video delay.

For volumetric video, we developed our own Point Cloud (PC) streaming tool, using the Draco encoder. Figure 7 shows the end-to-end latency of 1000 PCs when using 2 compression threads with 50% dropped points; this supported 30 FPS, with 97.2% of frames received correctly.

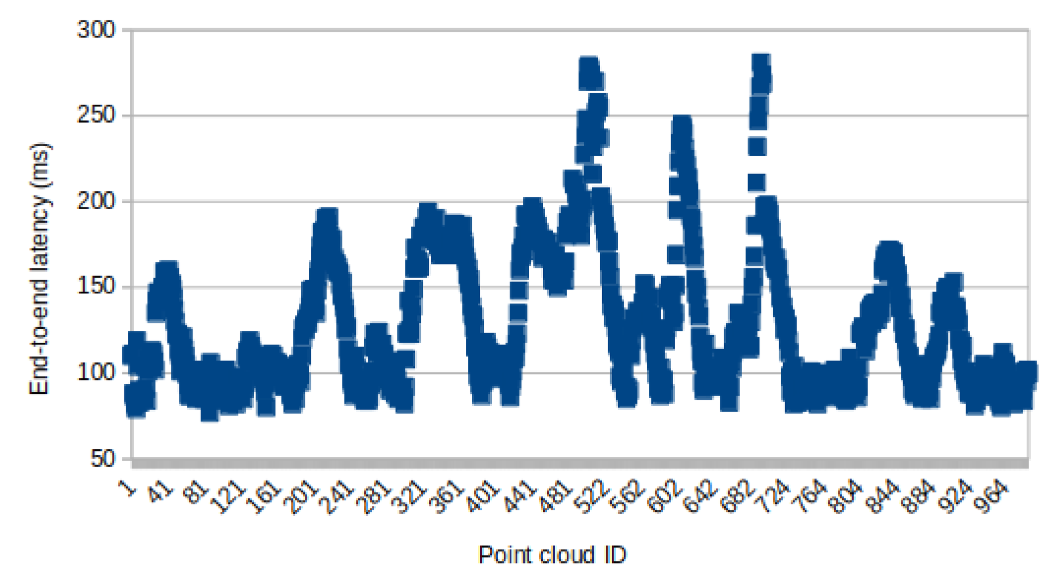

Fig. 6: Volumetric latency

Contact

Professor George Xylomenos
Mobile Multimedia Laboratory, AUEB
eMail: xgeorge@aueb.gr
LinkedIn: george-xylomenos-966808

References (Selection)

Bachhuber, C., and Steinbach, E. (2016) *A system for high precision glass-to-glass delay measurements in video communication*. IEEE International Conference on Image Processing (ICIP), pp. 2132–2136.

Tsioutas, K., Thomas, Y., Bistas, F., Barous, I., Xylomenos, G., and Polyzos, G.C. (2025) *Network Music Performance Beyond 4G*. IEEE International Wireless Communications & Mobile Computing Conference (IWCMC).

Thomas, Y., and Xylomenos, G. (2025) *Ultra-low Latency Point Cloud Streaming in 5G*. EuroXR International Conference.

Tsioutas, K., and Xylomenos, G. (2025) *Network Music Performance in 5G Networks*. IEEE International Symposium on the Internet of Sounds (IS2) (submitted).

Co-funded by the European Union
UK Research and Innovation

www.spirit-project.eu