Networked Performances and Natural Interaction via LOLA: Low Latency High Quality A/V Streaming System*

Carlo Drioli[1,3], Claudio Allocchio[2], and Nicola Buso[1]

[1] Conservatorio di Musica G. Tartini, Trieste, Italy
[2] Consortium GARR, Rome, Italy
[3] University of Udine, Udine, Italy
carlo.drioli@uniud.it,
claudio.allocchio@garr.it,
nicola.buso@conts.it

Abstract. We present LOLA (LOw LAtency audio visual streaming system), a system for distributed performing arts interaction over advanced packet networks. It is intended to operate on high performance networking infrastructures, and is based on low latency audio/video acquisition hardware and on the integration and optimization of audio/video data acquisition, presentation and transmission. The extremely low round trip delay of the transmitted data makes the system suitable for remote musical education, real time distributed musical performance and performing arts activities, but in general also for any human-human interactive distributed activity in which timing and responsiveness are critical factors for the quality of the interaction. The experimentation conducted so far with professional music performers and skilled music students, on geographical distances up to 3500 Km, demonstrated its effectiveness and suitability for distance musical interaction, even when professional players are involved and very ”tempo sensitive” classical baroque music repertoire is concerned.

1 Introduction

Distributed collaborative environments have been recently the subject of considerable interest and investigations. In some specific cases, collaborative activities include tasks in which the speed of interaction plays a critical role, and the growing demands in terms of performance has highlighted the limits of the currently available multimedia data processing and transmission technology [1]. In particular, the field of interactive musical collaboration, referring to scenarios such as geographically distributed musical performance or distance music education, poses specific problems related to the management of high quality audio-video streaming, to the transmission speed and delays, and to the impact of transmission delays on musical performers accuracy [2,3,4,5].

Videoconferencing systems have always been the common answer to remote interaction among people in distant locations. However all of them were conceived for enabling people to hold a meeting where participants just talk and discuss. Until recent

* A project by Conservatorio di Musica G. Tartini and Consortium GARR.

P. Nesi and R. Santucci (Eds.): ECLAP 2013, LNCS 7990, pp. 240–250, 2013.

times, most of these systems did not even try to emulate the presence of the remote parties, and were designed with much higher delay (latency) tolerances then usually required for natual live interaction. Also the recent ”telepresence” immersive solutions (e.g., Polycom RealPresence or Cisco Telepresence System) only aim at a very fixed interaction scheme, such as a meeting where participants all sit together around a table, and talk.

Remote music education indeed tried to use and in some cases adapt existing videoconferecing tools, and adopted some of them as anyhow useful tools. Legacy H.323 systems usually offer poor qualiy audio codecs (80Hz-8KHz), put priority on video data over audio data in case of problems, use Automatic Echo Cancellation (AEC) mechanisms which are optimised for voice patterns, and cancel a wide range of audible frequencies. Some work has been done by some vendors [6] in collaboration with the music education community to obtain better results, but the codec latency remains still high (above 200ms), well beyond the possibility to play music together during the interaction. DVTS (Digital Video Transport System) [7] and ConferenceXP [8] have a much better audio codec quality and do not use AEC, but also have high latency (above 300 ms) which generates echo, and makes interaction complex and non-natural. Only ”walkie-talkie” style interaction was thus possible, wich required a highly structured human communication protocol, too.

The system we illustrate here was conceived for a completely different environment, where latency must be below what individuals can perceive, where people perform actions they normally do when they interact live (e.g., playing, singing, dancing), and where also remote sound fidelity must be the highest possible quality.

To specifically address this class of problems, an high-quality, low trasmission latency system aimed at distributed music performance over high-end packet networks, was designed and developed. Low transmission latency refers to a very specific feature of an Audio/Video communication system where the transmission delay among remote sites is very small and therefore negligible for the human eye and ear. Since its origin, the main goal of the LOLA project (the acronym stands for ”LOw LAtency”) was to create a system fulfilling this specific requirement, thus building an high quality tool for remote musical education and real time distance musical performance. To date, the system has been extensively tested and used in a wide number of demonstrative performances, and is ready to be used for production and musical education by institutions connected to academic networks.

LOLA was originally conceived and designed at the Tartini Music Conservatory of Trieste, and at present is being developed and tested with the collaboration of GARR (the Italian Education and Research Network organization). The Lola development team is composed of Massimo Parovel (conception and supervision), Paolo Pachini (general coordination), Nicola Buso (testing and musical advice), Carlo Drioli (system design and programming), Claudio Allocchio (testing and networking advice).

2 Architecture and Implementation

The LOLA A/V streaming system is conceptually an high quaility videoconferencing hardware/software system (see Fig. 1). However, the operating conditions that it was

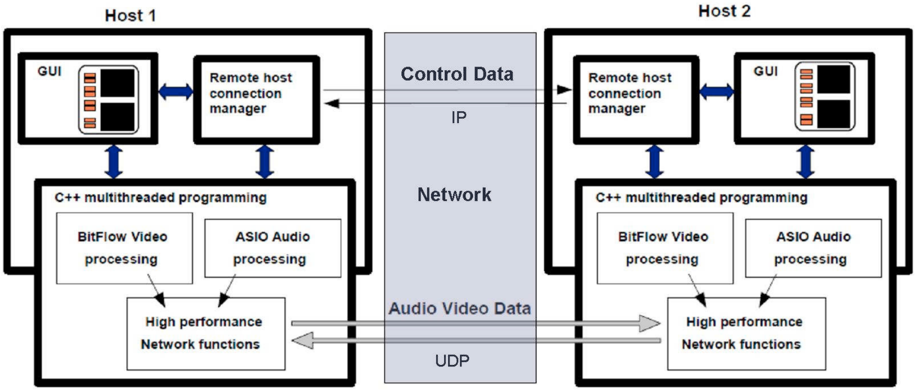
Host 1
GUI
Remote host connection manager
C++ multithreaded programming
BitFlow Video processing
ASIO Audio processing
High performance Network functions
Control Data
IP
Network
Audio Video Data
UDP
Host 2
Remote host connection manager
GUI
C++ multithreaded programming
BitFlow Video processing
ASIO Audio processing
High performance Network functions

Fig. 1. System architecture overview

designed for, and the hardware and software design solutions adopted, makes it different from any other videoconference system available to date. It was designed to fulfill a number of fundamental requirements: 1. to be suitable for musical performances relying on both audio and visual communication, with the goal to provide a natural and transparent end-user interaction; 2. to be low cost and portable; 3. to exploit bandwidth and robustness of dedicated high performance networks (e.g., LightNet Project, GARR, GÉANT, Internet2).

2.1 System Engineering

In order to achieve a low transmission and presentation latency, LOLA relies on software optimization and on high performance audio and video devices. Fast video acquisition and streaming relies on a family of industrial video grabbers by BitFlow Inc., which provides high hardware performances and a versatile programming API for low-level video processing control, and on industrial class progressive video cameras. Low latency audio performance is achieved by relying on robust hardware and driver equipment (RME Hammerfall and ASIO drivers). Both audio and video streaming are optimized for speed by relying on accurate audio acquisition, transmission, and rendering threads synchronization, and the system supports multiple indipendent audio channels. Figure 2 illustrates the threads organization implemented in the software design: accurate synchronization of acquisition and transmission threads is required to transmit audio and video data as fast as possible, and similarly accurate synchronization of receiving and rendering threads is essential to receive, decode and render audio and video data as fast as possible. On the networking side, a low level network packets handling system has been created and used, to avoid hidden queueing provided often by common network software. Finally, to mitigate the effects of jitter that might arise in public network due to presence of irregular network traffic, a network jitter compensation mechanism is provided for both audio and video, through read/write ring buffers.

Current release supports audio at 44100 samples/sec, 16 bit, and 640x480 resolution video, at 60 or 30 fps, colour or black and white. Audio and video are non compressed,

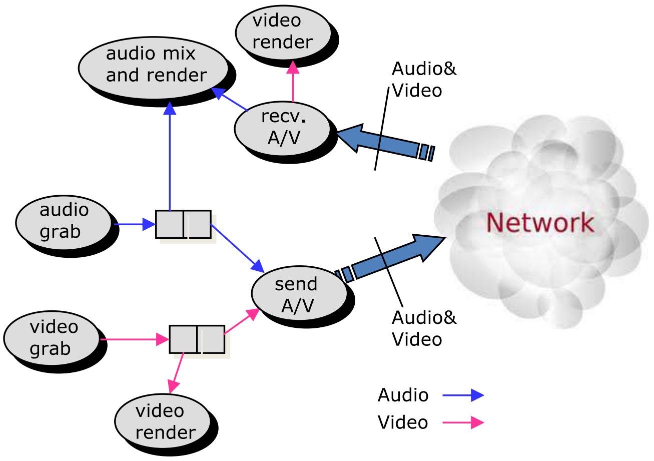

Fig. 2. Scheme of the threads involved in the audio and video streaming

to avoid introducing time delays in the encoding/decoding process. This allows to reach RTT delays (not considering network delay) as low as 5 msec for audio, and as low as less than 20 msec (estimated perceptually) for video. Network latency to be considered in the total RTT delay estimates is <1 ms on LANs, ∼1 ms per 100 Km on WANs. Jitter might become sensibly high when operating on network branches with public traffic, thus a buffering mechanism is provided to prevent data loss do to network delay oscillations. In terms of badwidth usage, LOLA requires at least 100 Mbps in minimal configuration (standard definition, b/w, 30 frames per second) up to 500 Mbps in full configuration (standard definition, color, 60 frames per second), and generates a very high Packet per Second (PPS) rate, as it uses 1 Kb data packets. Thus the minimal end-to-end connectivity must be at least 1 Gbps[1].

3 Distributed Performances

The LOLA project was conceived in 2005, and after a set of preliminary studies, the core system was developed between 2008 and 2010. LOLA first public showcase was in November 2010, with a live performance during the Network Performing Arts Production Workshop, by the piano duo Zaccaria-Trevisan (both professional performers and music teachers at the Tartini conservatory). The distributed piano concert was performed between the Music Conservatory ”G. Tartini” in Trieste and IRCAM in Paris, over a distance of 1300 Km, and featured some movements of Bach’s Brandenburg Concertos (transcribed for piano 4 hands by Max Reger). During next year’s Network Performing Arts Production Workshop, held in June 2011, the B. Bartok duets for violin were performed between the Conservatory of Trieste and the Gran Teatre del Liceu of Barcelona, at a distance of 2700 kilometers. The violin duo was made up of two brilliant

[1] The software and documentation is freely available for non commercial use at: `http://www.conservatorio.trieste.it/artistica/lola-project`

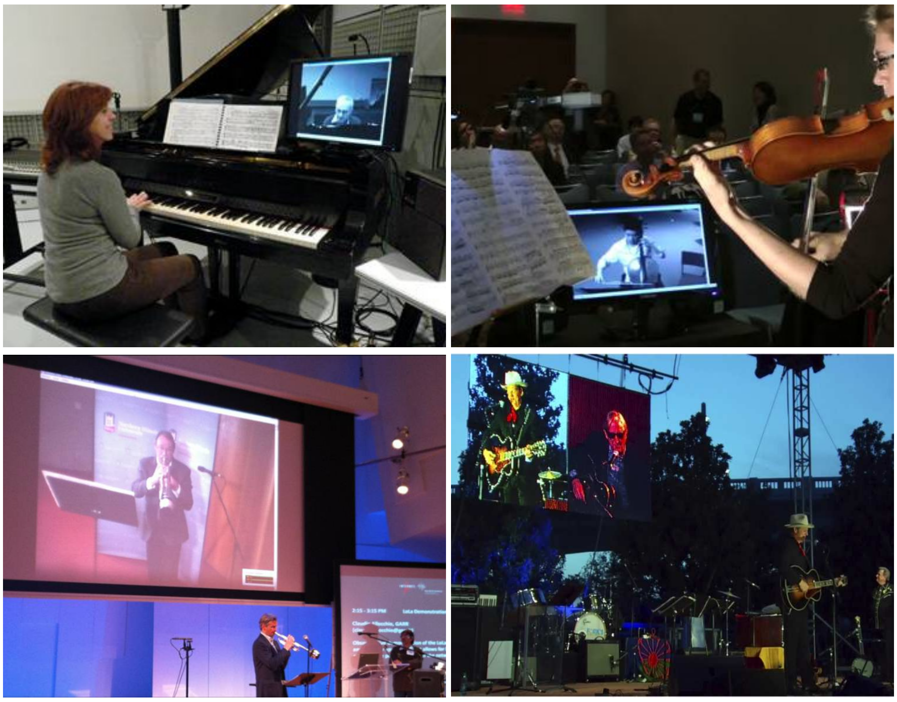

Fig. 3. Public networked live performances at four different moments of the LOLA system development. Top-Left: piano duo playing classical music repertoire (Bach, Reger) during 2010 Network Performing Arts Production Workshop at a distance of approx. 1300 Km (Trieste, Tartini Conservatory - Paris, IRCAM) ; Top-Right: violin-cello duo playing Haendel during 2011 Internet2 Fall Members Meeting at a distance of 1850 Km (Chicago, NIU - Raleigh, Congress Center); Bottom-Left: trumpet duo playing Bozza during 2012 Performing Arts Production Workshop, at a distance of approx. 2400 Km (Chicago, NIU - Miami, NWS); Bottom-Right: guitar-voice duo playng country tunes at a distance of 3500 Km (Chattanooga, TN - Los Angeles, CA) during 2012 Roots Riverside Country Music Festival.

students, Laura Agostinelli and Sebastiano Frattini, which have previously played together but still in their way to reach a complete technical maturity, proving that LOLA is also suitable to be used in the learning process and does not requires necessarily the experience of a mature concertist. A decisive test was then made in October 2011, during the ”Internet2 Fall Members Meeting”, between the NIU (Chicago, IL) and the Congress Center (Raleigh, NC), at a distance of 1200 Miles (1850 Km), featuring the execution of the Passacaglia for violin and cello by Handel by Marjorie Bagley (violin) and Cheng-Hou Lee (cello). In this case the two musicians had never played together before and had never met in life: after a one day rehearsal using LOLA (without a specific training), the day after they were able to successfully give a concert as if no distance nor technology was there. Since the first experimentations with LOLA, a number of successful performances have taken place during several other public events. In all these occasions, the system proved to be an effective transparent, non-invasive tool

Table 1. Some of the LOLA performances during public events in years 2010-2013

Event	Locations	Distance	Repertoire
Network Performing Arts Production Workshop, 2010	Tartini (Trieste)-IRCAM (Paris)	1300 Km	Piano Duo: Bach, Reger
Network Performing Arts Production Workshop, 2011	Tartini (Trieste) Gran Teatre del Liceu (Barcelona)	2700 Km	Violin Duo: Bartok
Internet2 Fall Members Meeting, 2011	NIU (Chicago, IL) Congress Center (Raleigh, NC)	1850 Km	Violin Cello: Haendel
Network Performing Arts Production Workshop, 2012	NIU (Chicago, IL) NWS (Miami, FL)	2400 Km	Trumpet Duo: Copland
JANET Performing Arts Networkshop, 2012	Royal College of Music (London) Napier Univ. (Edinburgh)	700 Km	Clarinet Piano: Mozart, Jazz
AEC Annual Meeting for International Coordinators, 2012	Tartini (Trieste) SS. Marcellino & Festo (Naples)	1200 Km	Trumpet Piano: Baroque, Romantic, Morricone, Dalla
Trieste Next International Festival, 2012	Tartini (Trieste) Ca Foscari (Venice) Academy of Music (Ljubljana)	750 Km (on GARR shared network) + 100 Km (on a 1Gbps Lambda)	Cello Quartett, Cello Duo Liute: Mozart, Vivaldi
Internet2 Fall Members Meeting, 2012	NIU (Chicago, IL) Sheraton Hall (Philadelphia, PA)	1700 km	Violin Cello: Haendel Halvorsen
Roots Riverside Country Music Festival, 2012	River Park (Chattanooga, TN) USC (Los Angeles, CA)	3500 Km	Guitars, Voice: Country
La Musica Viaggia Veloce, 2013	Tartini (Trieste) - Conservatorio (Frosinone)	1000 Km	Sax, Drums, Piano, Bass: Jazz Quartet

to remotely teach, rehearse and perform together. Table 1 gives an overview of the most relevant demonstrations to date.

If compared to inherent latencies introduced by standard videoconferencing systems (e.g. DVTS, Conference XP, Skype), usually not below 0.5 sec for both audio and video, the inherent low latency provided by LOLA was received very favourably by musicians who had former experiences related to distributed musical performances. In all cases in which the RTT was kept below the 75 msec threshold, performers reported to be able to play comfortably and to feel the system as transparent after a short while (this operating situation was met in all public events, see e.g. Fig. 3).

3.1 The Impact on Distance Music Education

Standard videoconferencing systems as the ones mentioned, also limit the effectiveness of distance teaching, since the teachers is not able to play along with students, because of the high transmission latencies, and the teaching action goes necessarily through verbal communication in deferred time: first the student plays, and then the teacher comments on his performance. The ability to play together guaranteed by LOLA allows to reintroduce a decisive factor in the praxis of distance music education, i.e. the non-verbal communication: a peculiar trait of music lessons, where the teacher accompanies the student by marking the rhythm, the phrasing articulation, and checking and suggesting the performing gestures. With this respect, other systems for distance learning, characterized by higher latencies, do not allow the purely musical interaction between the teacher and the student, forcing to resort to the mediation of speech for the most part.

The high quality audio rendering also allows to accurately deal with timbre aspects. The ability to intervene in real time on the timbric nuances of touch, on the gestures and with gestures, significantly expands the horizon of the distance musical education.

Last but not least, the significant reduction of the traveling and accommodation costs has a relevant impact on the opportunities for increasing cultural and technical skills, both from the teaching and from the professional point of view.

3.2 Considerations on Delay Impact and on Remote Acoustic Scene Rendering

During laboratory test sessions, in which latencies were artificially rised or lowered on request, a number of observations were collected concerning the different factors participating to the perception of interaction delay and on their impact on performance quality: the musical repertoire, the timbre and dynamic characteristics of the musical instruments, the reverberation and remote instrument rendering, among others.

To further discuss the importance of audiovisual communication delay in distributed interaction, let us refer to the following scenario: two musicians in different locations, connected through the network, must play together a sequence of four notes, lasting one second each, so that the notes played from the first musician are synchronous to those performed by the second musician. If the transmission channel and the signal encoding/decoding system introduce a noticeable delay, the second musician has to wait for the note from the first musician to arrive. When the first note of the musician arrives and is rendered, the second player can finally start playing his note. The execution of the second musician, in turn, is sent through the network to the first musician, who will receive it with the inherent delay of the transmission and reproduction system, plus the delay introduced by the second musicians. Only then the first musician will be allowd to play his second note, and so on, note after note. The delays due to technology sums up to the delays introduced by musician's due to their reaction time, and the concatenation of waiting and transmission delays gives rise to a rallentando, eventually leading to the loss of musical coordination bewtween the two. The overall delay is determined by the human-machine-human interaction. To be able to play together, the latency of the system should not be perceptible, i.e. should not exceed the order of magnitude of the thresholds of perception for temporal segregation (approx. 30 ms). To give an approximate indication, the round trip delay introduced should be no more than 75 ms (which is however an indicative value, which is subject to many variables: the class of musical instruments involved, the music repertorie and, last but not least, the musical skill of musicians). If the delay is kept below the threshold of perception, the human factor exits the dynamics of the growing delay loop, and the round trip delay reaches a stable value determined for the most part by the signal encoding/decoding, transmission timings, and network performance.

The microphones and sound diffusion setup may vary consistently, depending on the instruments involved and the acoustic characteristics of the rooms where the performers and the public are located. In principle, the use of small diaphragm condenser microphones, with cardioid polar figure and located fairly close to the acoustic source, is a good choice for a punctual recovery of the acoustic source nature and to minimize the electroacoustic feedback. To render the sound of the remote acoustic sources, a cluster of loudspeakers is used, directed radially with respect to the position of the virtual

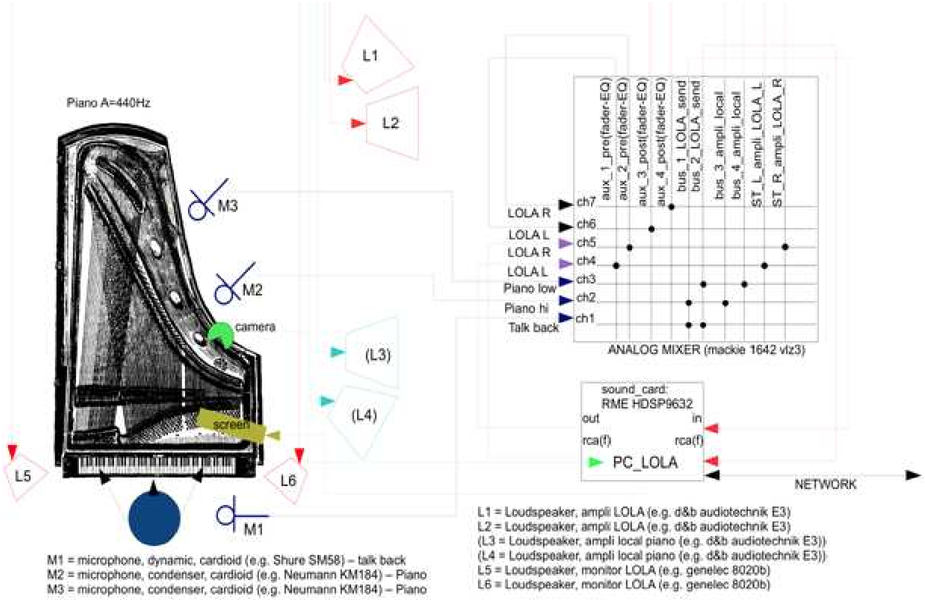
L1
L2
Piano A=440Hz
M3
M2
camera
(L3)
(L4)
screen
L5
L6
M1
aux_1_pre(fader-EQ)
aux_2_pre(fader-EQ)
aux_3_post(fader-EQ)
aux_4_post(fader-EQ)
bus_1_LOLA_send
bus_2_LOLA_send
bus_3_ampli_local
bus_4_ampli_local
ST_L_ampli_LOLA_L
ST_R_ampli_LOLA_R
LOLA R ch7
LOLA L ch6
LOLA R ch5
LOLA L ch4
Piano low ch3
Piano hi ch2
Talk back ch1
ANALOG MIXER (mackie 1642 vlz3)
sound_card:
RME HDSP9632
out
in
rca(f)
rca(f)
PC_LOLA
NETWORK
M1 = microphone, dynamic, cardioid (e.g. Shure SM58) – talk back
M2 = microphone, condenser, cardioid (e.g. Neumann KM184) – Piano
M3 = microphone, condenser, cardioid (e.g. Neumann KM184) – Piano
L1 = Loudspeaker, ampli LOLA (e.g. d&b audiotechnik E3)
L2 = Loudspeaker, ampli LOLA (e.g. d&b audiotechnik E3)
(L3 = Loudspeaker, ampli local piano (e.g. d&b audiotechnik E3))
(L4 = Loudspeaker, ampli local piano (e.g. d&b audiotechnik E3))
L5 = Loudspeaker, monitor LOLA (e.g. genelec 8020b)
L6 = Loudspeaker, monitor LOLA (e.g. genelec 8020b)

Fig. 4. The LOLA setup used for the distributed piano due performaces

remote instrument to simulate its sound radiation. Figure 4 illustrates the typical setup used so far for distributed performances involving two grand pianos.

Finally, another important element of the performance is the eye contact between the actors, in order to strengthen the understanding, as well as a comfortable environment in terms of reverberation, sound immersion and concentration.

All these factors will be sistematically investigated in future research concerning the musical applications of LOLA.

4 Considerations on Network Traffic

Since the beginning of LOLA design we assumed that the network was able to deliver in a timely and reliable way the audio and video data. This has proven to be true on all the long distance academic networks used (GARR, GANT, RENATER, Red.Es, JANET, Internet2, etc.) (see Fig. 5), but issues were discovered in some Local Area Network (LAN) setups: indeed, the quite high Packet Per Second rate generated by LOLA can put a non negligible stress on some LAN switches, when other application and services also compete with similar requirements (like Voice over IP services), and in a few cases the equipment is unable to perform correctly. However, this has always been solved bypassing the troubled equipment to access directly the network backbone, and most modern devices do not show the problem.

In order to ensure LOLA data traffic to be delivered, we also assumed that ”overprovisioning” the network capacity is the best possible strategy: in fact we also successfully tested bandwidth resevation techniques, in situations where the situation could be critical (see Fig. 6, upper panel). But network latency also depends on network equipment

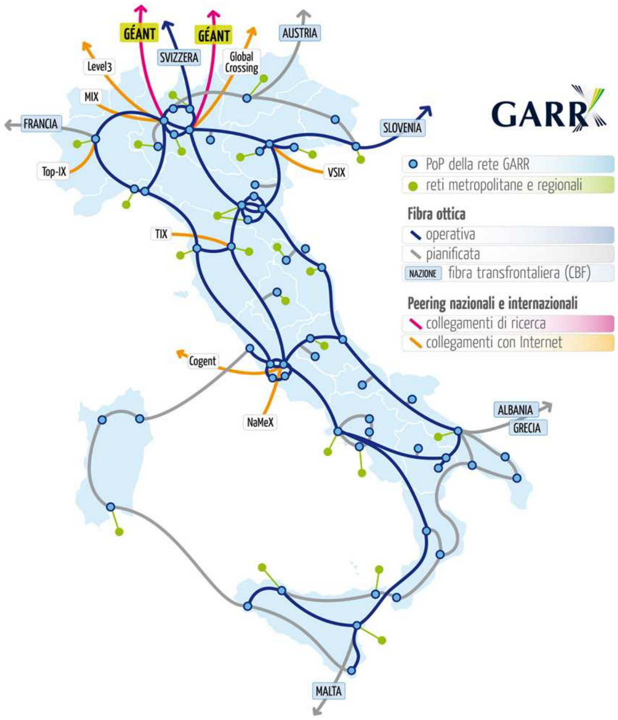

Fig. 5. The GARR-X dark fiber optical backbone

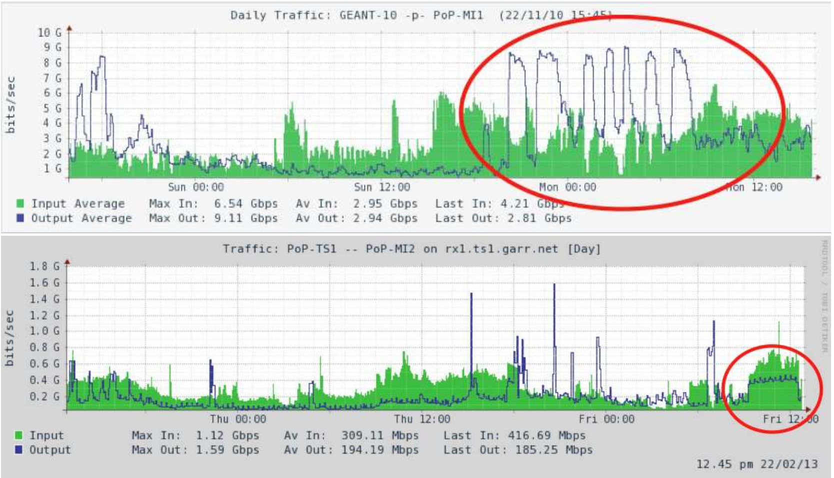

Fig. 6. Upper plot: network traffic close to backbone capacity (10 Gigabit at that time) close to the first public LOLA showcase in November 2010, when bandiwidt control techniques were used to ensure the minimal capacity for LOLA. Lower plot: LOLA traffic (in the red circle) on one of the GARR-X 10 Gigabit backbone generated by a single LOLA session.

processing time, and the more processing is added to packets, the more network latency is increased, thus overprovisiong shall be the preferred method, as it also ensures a much lower network jitter, and limits the number of buffers needed to counteract possible jitter effects on audio/video data delivery. On the basis of laboratory measures, we established that a single jitter audio buffer introduces approximatey 0.75 msec of additional delay (one-way). Whenever a shared network circuit was used, without any specific protection for the LOLA traffic and in regular traffic conditions, we observed that avarage jitter is normally around 1.5 msec, and 3 to 5 audio buffers are necessary to ensure a perfect audio communication. When protected circuits (including virtual ones) are used, for which jitter is still present but usually below 1 msec, 2 to 3 audio buffers are usually required to avoid audio drop-outs. When jitter is below the observable threshold, e.g. when using lambda or ad-hoc circuits, it is possible to avoid the use of buffering.

Last, but not the least, LOLA can generate, just for a single node running in standard definition, a non neglegible data traffic, which can be still very well visible also on high speed backbones (fig. 6, lower panel) where shared traffic runs. This thus requires a careful network planning, to avoid transforming LOLA into a network ”killer application”.

5 Conclusions

A new audio visual streaming system, characterized by high quality audio/video rendering and by very low data acquisition, transmission and presentation delays, has been illustrated. It was designed to accomplish time-critical distributed interactive tasks such as distributed music performance. It was assessed with respect to this specific aim by involving professional music performers playing classical repertoire over wide geographical distances, and it proved suitable and effective for distance musical rehersal, performance and production.

The remote musical interaction, either aimed at distributed public performances or at remote teaching purposes, is only one of the many possible applications of the LOLA system, which not only lends itself to experimentation in related fields such as dance theater, performing arts and teaching methods, but also in the more general range of disciplines concerning distributed interaction in the presence of hard time constraints, e.g immersive virtual reality and medical applications, for which there is an high interest in exploring the new fronteers of remote interaction offered by low delay audiovisual streaming technologies and high performance networks.

References

1. Gergle, D., Kraut, R.E., Fussell, S.R.: The impact of delayed visual feedback on collaborative performance. In: Proceedings of CHI 2006, pp. 1303–1312. ACM Press (2006)
2. Konstantas, D., Orlarey, Y., Carbonel, O., Gibbs, S.: The distributed musical rehearsal environment. IEEE MultiMedia 6, 54–64 (1999)
3. Sawchuk, A.A., Chew, E., Zimmermann, R., Papadopoulos, C., Kyriakakis, C.: From remote media immersion to distributed immersive performance. In: Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence, ETP 2003, pp. 110–120. ACM, New York (2003)

4. Chafe, C., Gurevich, M., Leslie, G., Tyan, S.: Effect of time delay on ensemble accuracy. In: Proceedings of the International Symposium on Musical Acoustics (2004)
5. Zimmermann, R., Chew, E., Ay, S.A., Pawar, M.: Distributed musical performances: Architecture and stream management. ACM Trans. Multimedia Comput. Commun. Appl. 4, 14:1–14:23 (2008)
6. Orto, C., Karapetkov, S.: Music performance and instruction over high-speed networks. Polycom White Paper (2008)
7. Ogawa, A., Kobayashi, K., Sugiura, K., Nakamura, O., Murai, J.: Design and implementation of DV based video over RTP. In: Proceedings of the Packet Video 2000 Workshop, Forte Village Resort, Sardinia, Italy (2000)
8. Needham, T.: ConferenceXP and advanced collaborative scenarios. In: Proceedings of International Symposium on Collaborative Technologies and Systems, Las Vegas, Nevada, USA (2006)