Requirements and Application Scenarios in the Context of Network Based Music Collaboration

Chrisoula Alexandraki
Department of Music Technology & Acoustics, Technological Educational Institute of Crete
chrisoula@stef.teicrete.gr

Ioannis Kalantzis
AKMI Vocational Training School
Athens, Greece
ikalantzis@hotmail.com

Abstract

Recent advances in computer network technology have greatly enhanced the feasibility of networks that allow remote collaboration in performing music. This paper presents a research study on the user and technical requirements for systems in this context. User requirements have been gathered through a questionnaire-based survey, whereas the reported technical ones are the result of a qualitative study on the relevant research projects and the existing technological tools in the area of live streaming of multimedia content. Furthermore, the paper attempts one step further, by classifying the effective application scenarios that can emerge for remote music collaboration, when the reported requirements have been met.

1. Introduction

The growing need for innovative network-collaboration environments for live music performance has been a challenging field for a number of academic and research institutions all over the world [1, 2, and 3]. An overview of the music and sound art projects involving the use of network infrastructures can be found at [4]. According to this article, the advent of computer network music dates back to the 1970s, when the commercialization of personal computers in the United States began.

Currently, the latest advancements in the field of broadband networking and of computer technology in general, have allowed for a variety of music collaboration scenarios to be considered feasible not only in research, but also in a commercial context. It is worth noticing for example, that live streaming of multimedia content is becoming so apparent that scenarios of network music collaboration are used by network providers to advertise the quality of the services they provide. Such scenarios, usually involving a popular Greek performer, have been used in a number of TV commercials in Greece.

In practice however, using computer networks for music collaboration is not trivial. The effectiveness of such attempts depends on various factors that range from the quality of service (QoS) provided by the underlying network, to a number of psychophysical, perceptual and artistic aspects [2]. Furthermore, the success of such experiments is strongly dependent on the means provided to the user in order to interface with the environment and communicate with other performers.

In this paper we attempt to enumerate the requirements of network based music collaboration environments and classify the application scenarios that emerge in this context.

2. Research context

The study reported in this paper has been carried out, in part requirement of a Greek national research project, which is currently in progress. The title of this project is "DIAMOUSES – distributed interactive communication environment for live music performance".

The main objective of the DIAMOUSES project is the development of an integrated platform, which will allow for remote collaboration throughout a distributed live music performance environment. Musicians-members of an orchestra, whilst geographically spread, will be able to simultaneously perform the same piece of music. At the same time, this 'network-performance' will be witnessed by an audience located elsewhere, breaking the barriers set by geographical distance, thus resulting in a new network collaborative community.

The system under development will support signal transition in heterogeneous computer networks, including IP networks as well as a pilot DVB-T network platform which operates in the island of Crete. The combination of these two types of networking allows for simultaneous support of various routing schemes such as broadcasting, multicasting and unicasting. Moreover, it enables application scenarios which involve a broad range of target users with diverse skills and preferences, such as digital TV subscribers for interactive and non-interactive television services.

3. Research methodology

In this section of the paper we present the methodology which was adopted for performing the study whose results are reported in the sections that follow. The objective of the study was to define a set of requirements that must be met in the context of network based music collaboration. These requirements concern the ones set forth by users and also the technical ones for performing music through networks effectively.

In respect with user requirements, we followed a quantitative approach, by performing a questionnaire-based survey. This survey was targeted towards two groups of potential users of our system. The first group was concerned with users that have a high level of involvement in music. The users of the first group were performers, composers, conductors, instructors, as well as recording engineers and professionals from the area of music technology. The second group of users took into account the general public, which can act as an audience of a distributed music performance, having a general interest in music.

Audience involvement in distributed music performance has been taken into account since the early experiments of network performance. However, to the authors' awareness, these experiments silently assumed that all members of the audience were to be situated at the same location and therefore occupy a single node in the network, where high quality video projections and an appropriate sound reproduction system were provided [2]. In our analysis, we additionally consider the situation in which not only the various musicians, but also the members of the audience can be distributed in different locations (e.g. in the area of coverage of a broadcasting network infrastructure, such as a digital TV network).

The technical requirements were approached through a qualitative study which involved literature review, study of the relevant standardized technologies (e.g. RTP/RTSP protocols), and hands-on evaluation of the existing software tools that have been implemented in the area of live streaming of multimedia content.

4. User Requirements

This section presents the user requirements collected by sending questionnaires to potential users of the system under development. Two types of questionnaires were distributed: one for users actively involved in music and one for the general public which can be thought of as the audience of the distributed music performance.

Each question has a number of alternative responses. Users were asked to give a preference value to each response. So if for example a question had 3 alternative responses, then users would give a preference value 3 to the alternative response of their top preference, a preference value 2 to their second preference, and so on. According to the preference values, a normalised average was calculated for each alternative response-j, as follows:

$W_j = \frac{P_j}{\sum_j P_j}$, where P_j is the sum of the products of the preference values given, multiplied by the number of users that have assigned the particular value, for all preference values of alternative response-j. The analysis of the user requirements is based on the normalised average values W_j, which were calculated for each response. These values are given as a percentage in the diagrams that follow.

The analysis of the results takes into account aspects which are vertically related to the requirement in question. For example, questions regarding preferences in performing music are arranged according to music genre.

Each of the questionnaires was accompanied by a cover letter which was featuring the context of the research study and introducing the users to the concept of remote music collaboration.

4.1. Users actively involved in music

In this group of users a total of 58 replies was received. Requirements were classified according to the users' type of involvement in music and according to music genre. The form of the questionnaire was such that a user could have more than one type of involvement in music. However, if somebody was involved in more than one music genres, a separate questionnaire for each genre had to be completed. The following table shows the distribution of users among different music genres.

Table 2: Distribution of the music genres

Music Genres	No.	Perc.
Classical/Contemporary	22	38%
Jazz/Blues	8	14%
Pop/Rock	11	19%
Electronic/Electroacoustic	10	17%
Ethnic/Folk	7	12%

The rest of this section is structured as follows. Firstly, we provide an English translation of the questions in the questionnaire, as these were originally formulated in Greek. Following, is the diagram which depicts the average values (W_j) of the declared preferences for each alternative answer and for each music genre. Finally, some observations on the resulting diagram are provided.

Question1: Give your preference in musical instruments and musical interfaces when performing music.

a) Acoustic instruments

b) Electric instruments

c) Electronic instruments

d) Computer (interaction solely through mouse or mouse pad and keyboard)

e) MIDI controllers (keyboards, sliders, knobs, etc.)

f) Experimental sensors for gesture recognition

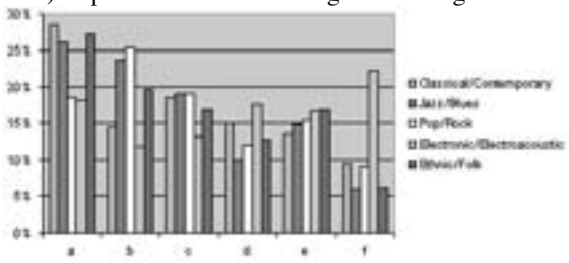

Figure 1: Musicians' preference in musical instruments and musical interfaces

As expected, the top preference is in acoustic instruments for all music genres, apart from musicians of pop/rock who prefer electric instruments, and musicians of electronic/electroacoustic music who prefer experimental sensors. It is interesting to notice that a) experimental sensors are top priority for musicians of electronic/electroacoustic music, and b) the use of MIDI controllers is almost equally preferred by all music genres.

Question 2: Rate your preference in deciphering the flow of a musical piece while performing with others.

a) Through a musical score

b) Performing from memory

c) Prima vista or performing according to a score that is dynamically generated

d) Performing musical patterns based on your choice or on indications by others

e) Performing a score comprised of predefined graphical symbols

f) Improvising on a given musical theme

g) Free improvisation based on movement or eye contact

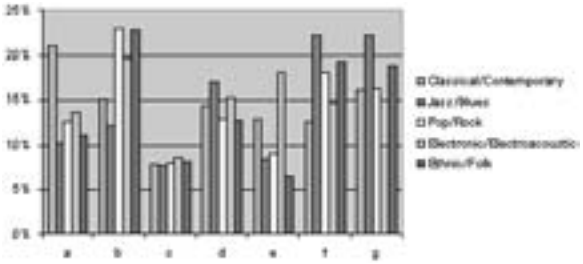

Figure 2: Preference in deciphering the flow of a musical piece

This question was included in the questionnaire in order to indicate requirements on the graphical user interface provided in circumstances of distributed music performance. It can be inferred from the diagram that musicians of classical and contemporary music have a strong preference in the presence of a score whereas jazz and folk musicians show a preference in improvisational music. It is interesting to notice that musicians of electronic/electroacoustic music would prefer to memorise the piece, rather than have to use any means for supporting them in following the flow of the music.

Question 3: Rate your preference in trying to synchronize with the other performers.

a) Conductor

b) Metronome

c) Visual metronome (usually a light, which flashes according to tempo and rhythm)

d) Score scrolling

e) Arithmetic visualization of tempo and rhythm (e.g. tempo: 120, bar: 27, rhythm: ¾, second quarter, would result in something like '120 27 ¾ 2')

f) No means of synchronization other than auditory and visual contact

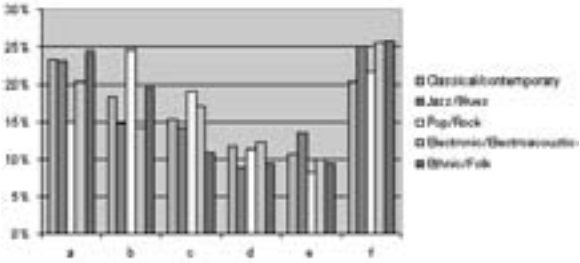

Figure 3: Preferences in synchronising with the other performers

All music genres show a very strong preference in visual contact with the other performers, which – in the perspective of a distributed performance – implies that

video communication should be provided among the musicians. Another interesting conclusion is that musicians of Pop/Rock prefer the metronome more than any other means of synchronization. This should be provided as a utility of the client software when performing pop music in a distributed environment.

Question 4: Rate your preference in the sound reproduction system for listening the other performers in the absence of visual contact.

a) Headphones

b) Loudspeakers

c) Multichannel audio

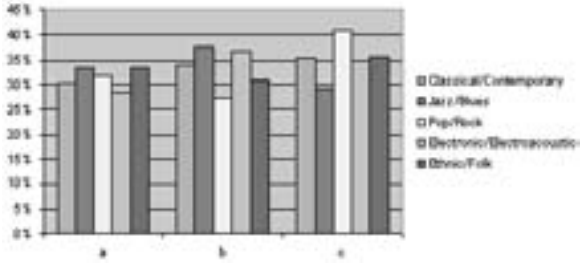

Figure 4: Preference in the sound reproduction system

It appears that there is a slight preference for multichannel audio. Although the majority of users questioned did not have an experience in distributed performance, it seems that musicians want to hear music reflected from the surrounding area, as it would do in a concert hall. There is strong evidence in prior experiments that sound reflections are desirable in this context [2].

Question 5: Rate your preference for special monitoring facilities in the absence of visual contact with the other performers

a) Monitor the dry mixed signal from participants

b) Monitor the mixed signal from participants after audio effects processing (e.g. reverberation)

c) Listen to one performer at a time with the possibility to choose another performer whenever needed (dry signal)

d) Listen to one performer at a time with the possibility to choose another performer whenever needed, after audio effects processing.

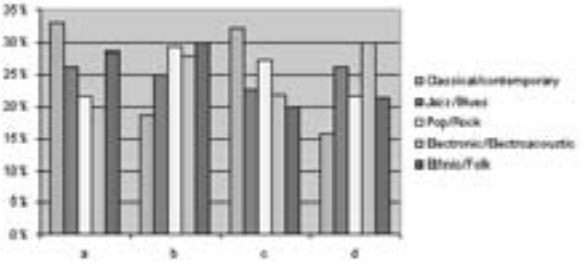

Figure 5: Preference in sound monitoring

In this diagram a preference for listening to all participants at the same time (mixed signal) is apparent for all music genres apart from the electronic and electroacoustic music. Furthermore, it appears that musicians of this genre find the presence of audio effects necessary, in contrast to musicians of classical/contemporary music who prefer to hear the dry signal.

Question 6: Suppose that you are remotely located from the other performers and that you are able to have visual contact with them through digital video. Rate your preference in the video communication provided.

a) One-way visual communication with the conductor

b) Bilateral visual communication with the conductor

c) One-way visual communication with one of the other performers, with the possibility to view another performer whenever needed

d) Bilateral visual communication with one of the other performers, with the possibility to view another performer whenever needed

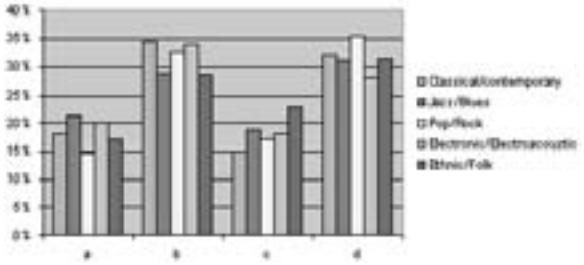

Figure 6: Preferences in visual communication

There is an obvious preference for bilateral visual communication for all music genres. The musicians of classical/contemporary and electronic/electroacoustic music prefer to have visual communication with the conductor than with the other musicians, which is not the case for the other music genres.

4.2. Members of the audience

Although users of this group were asked to rate their preference in different music genres, the analysis of their requirements is not arranged according to genres. The reason for this is that the audience have a more passive role than musicians who affect the outcome of a distributed performance scenario. This section will concentrate on the results of the survey, without getting in detail in formulation of questions or statistical data.

A total of 35 completed questionnaires was received, which were arranged according to users' education level and the kind of music of their top preference. Users were more or less evenly distributed

among the different music genres. The provided questionnaire form allowed them to declare their favorite music genre if this was not included in the list provided. The answers in this field were the genres of Heavy Metal, Soul, Disco and Byzantine-Hymnology. The educational level of the users ranged from school graduates to PhD holders, with the majority of users holding a university degree.

Users were introduced to the concept of remote distributed music performance and they were asked about their preference in the following aspects: facilities for watching a performance, sound reproduction system, video information content, metadata provided, provision of video on demand services and provision of event rating services. Finally users were prompted to comment on the concept of distributed music performance and give their own suggestions.

Regarding facilities for watching a distributed performance, users exhibited equal preference in the alternatives provided, which were a computer terminal, a home television or a centralized screen projection. The preferred sound reproduction system appeared to be the multi-channel system instead of conventional stereo sound reproduction systems, with a higher preference in surround speaker systems (of type 5.1 or 7.1), although polyphony (e.g. 8-speaker system) was provided as a separate option. In respect with the content of the video information, users seemed to be interested in having the possibility to choose when to view each of the distributed performers alone and when to view all of them on separate frame portions of the same display.

The interest in metadata information about the performance and the music performed was rated as shown in figure 7. It can be seen that users are more interested in having information about the music performed, rather than having information about the performance itself or the performers.

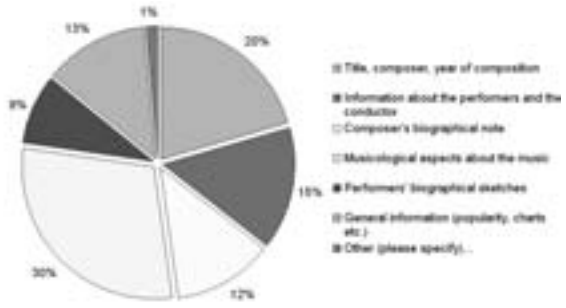

Figure 7: Audience preferences in the information content of the provided metadata

Users were also asked about their interest in a video on demand service which was related to the performance and could be offered to them, and a majority of 51% were highly interested. Finally, there was a higher interest in having the possibility to rate the performance in relation to its artistic aspects, rather than in relation to its technical coverage and the underlying technology.

5. Technical requirements

According to a number of scientific articles ([3] & [5]), real-time audio streaming is one of the most intensive applications in networking. The technological innovation of applications for network based music performance has been somewhat discredited due to the broad proliferation of teleconferencing technologies. However in music, accuracy in time and quality of the information delivered is a lot more crucial than in teleconferencing applications.

In respect with the network infrastructure, in order to accomplish network-based music collaboration a high level of QoS must be ensured, which requires cooperation at all network layers so as to minimize delay and quality variation of the information delivered. There have been a few scientific publications which enumerate the technical requirements in network based music collaboration. In this paper we will concentrate on latency sensitivity, bandwidth demand, synchronisation and error susceptibility.

5.1. Latency sensitivity

There are a number of factors causing latency in delivering live data streams in distributed music collaboration scenarios. These are due to the hardware equipment, the software applications involved, the operating system and the network infrastructure. If we concentrate on transmitting raw PCM audio streams and simplify the process of signal transmission between two participants, then we can identify causes of latency in the entire lifecycle of a data packet. Specifically, in a one-to-one transmission the lifecycle of this packet will involve the following steps: data capturing (analogue-to-digital conversion included), data packetisation, network transmission, data depacketisation and finally data playback (including digital-to-analogue conversion). What is more, an additional delay is caused by the process of loading the data buffer, which should be of adequate size in order to follow the above procedure and get reproduced at the receiver's playback equipment without producing additional distortion.

It appears to be a good analogy and has been suggested in a number of publications in this area, that the target of maximum tolerable round-trip delay ought to be comparable with the amount of the acoustic latency produced due to physical separation. Estimating latency according to the speed of sound in dry air (344m/sec) and assigning the spatial separation of musicians a value of the order of 10m result in a tolerable delay of approximately 30 milliseconds. According to prior evaluations and psychoacoustic experiments, this value is highly dependent on the music performed and the performing schema. A 20 to 30 millisecond delay is tolerable for traditional ensemble performance although this value will vary depending on the tempo of the music performed ([2] & [6]), as well as the acoustic properties and in particular the timbre of the musical instruments involved [2].

5.2. Bandwidth demand

Bandwidth demand is directly related to the information content of the transmitted data. Network music collaboration, may require apart from audio, also video transmission and possibly other types of information content (MIDI data, or gesture data, etc.)

In the case of audio information, transmission of CD-quality audio requires a data rate of 1.4Mbps. When employing multi-channel or better quality of audio (e.g. sampled at 48, 96 or 192 kHz, or providing 24-bit resolution), bandwidth demand is further increased. Therefore, it seems reasonable to find ways to minimise data overload for live audio streaming. In this direction, two main approaches are being discussed: audio compression and alternative encodings for representing sound and music.

It has to be taken into account, that lowering the bandwidth of sound information has major drawbacks, either in the quality of the reproduced sound or in the overall latency. For instance, sound compression algorithms that achieve sufficient compression ratios with decent audio quality result in a significant delay overhead, especially during the encoding process [7]. At the other end of the spectrum, a number of possibilities appear for low-bitrate representation of sound information, such as the conventional MIDI streams or the more recent OpenSound Control protocol, the standard for MPEG-4 Structured Audio and the IEEE standard for Symbolic Music Representation in MPEG [8]. The disadvantage in these approaches is that they cannot reproduce expressiveness in performing music, and that they are not appropriate for all types of music. Vocal music can be considered as an example.

In addition to sound information and according to the user requirements presented in this paper, it appears that video information is also necessary for remote music collaboration. Video information has two major advantages in this context. The first is related to the fact that video information can be recognisable, even when it has very low quality. For example, the Simple Profile of MPEG-4 Video supports bitrates, which are as low as 64kbps. The second advantage in employing video data is concerned with the directness of visual information in communication. The need for visual communication is evident in the user requirements section of this paper. Furthermore, the example of large orchestras, where performers synchronise by watching the conductor should be considered as a proof of the directness of visual communication. In this case, the delay of the visual information from the conductor to each of the performers is practically zero. In the context of the DIAMOUSES project, we are adopting an approach, in which musicians will receive low-fidelity video, for communicating with each other and the audience will receive high quality video. Sending high quality video to the audience is made feasible due to the fact that communication with the audience does not have to be synchronous.

5.3. Synchronisation

In respect with network based music collaboration, synchronisation refers to the time adjustments which need to be made when multiplexing multiple streams of audio or video data. There are two preconditions for achieving this type of synchronization. The first is that the clocks of the participants must agree with great accuracy and the second is that timing information must be sent along with the network stream.

The suggested approaches for synchronizing the clocks of multiple participants in a network music performance are to synchronise either by using the Network Time Protocol (NTP) [3], or via GPS signals [2]. The first solution offers an accuracy of 200μsec under optimal conditions in a LAN and a few milliseconds in WANs. The GPS solution offers an accuracy of approximately 10μsec or better. However, even if synchronizing the connected participants through the network, one must take into account clock inaccuracies caused by the operating system itself. This is in fact the main reason why some operating systems are considered inappropriate for network music performance.

Timing information sent along with the data packet can be ensured by the network protocols that operate at the application layer of the computer network. For

example, protocols that are normally used in multimedia streaming (e.g. RTP/RTSP) ensure the delivery of NTP timestamps included in the header of the network packet, as a built-in functionality.

When the above conditions are met, synchronising multiple streams is only a matter of calculation.

5.4. Error Susceptibility

Sound information is particularly sensitive to errors. The major cause of transmission errors is packet loss over the network. Errors due to lost packets are inevitable while at the same the strict requirements in minimizing all sorts of latencies renders the task of data correction even more complicated.

Most applications that involve network based music collaboration facilitate the UDP protocol. Although a fast protocol, UDP offers no guarantee for the reliability of the data delivered, as packets may arrive out of order, appear duplicated, or go missing without notice. However the RTP protocol, which operates at a network layer above UDP, offers mechanisms for detecting packet loss. Such a mechanism is the provision of the 'RTP sequence number' (i.e. the index of the packet), which is included in the network packet and is increased by one for every new RTP packet.

In cases of excessive packet loss, there has to be a mechanism, which will compensate for this loss. As presented in article [9], data correction algorithms can be classified in two main categories: Automatic Repeat Request (ARQ), which requires retransmission of the lost packet, and Forward Error Correction (FEC), which is based on transmitting redundant information along with the original information. Obviously, ARQ mechanisms are not acceptable for live audio applications over the network, as they dramatically increase the end to end latency. However, FEC data correction algorithms have been used in network music performance before, as they offer data reliability, without causing significant overhead on the overall latency and the required bandwidth ([2] & [3]).

6. Application Scenarios

Different application scenarios, or different variants of application scenarios put forward different requirements, both from the perspective of the user and the one of the technological infrastructure needed to support the specific scenario. For instance, a piano master class distributed within a Campus Area Network (CAN), will have different requirements from a piano master class distributed among different continents (WAN), both from the perspective of instructor-to-student communication and the one of the underlying network infrastructure.

In this context, an application scenario may be formed by assigning attribute values to a number of parameters. These parameters will be referred to as 'interaction parameters' hereafter, due to the fact that they affect the type of interaction in an application scenario for remote music collaboration. This section follows by attempting to provide an overview of all the interaction parameters that comprise an application scenario for network-based music collaboration and which can have a direct impact on the requirements which need to be satisfied.

Figure 8: The interaction parameters that comprise an application scenario for network-based music collaboration

Obviously, one of the most determinant parameters is the operational intent of the scenario, namely the purpose of the event. Different requirements are raised in the context of a live concert, than in the context of a master class. As for recording in a remote studio for example, strict requirements are posed in terms of bandwidth and tolerable data loss. Although user roles are related to the operational intent, they are included in the above figure as a separate node because different user roles raise different requirements in the interaction environment. It was apparent from the user requirement analysis preceded, that user roles, similarly to music genres, significantly affect the requirements of the application scenario.

As mentioned before, different types of information content results in different requirements on the available network bandwidth. In the above figure, the term 'control data' is used to refer to the various alternative representations for sound and music that were mentioned at the section related to bandwidth demand. The interaction parameter 'networking' is included as a separate interaction parameter, because it is directly related to the type of services that can be supported in a certain scenario. Additionally, networking affects the scalability of application

scenarios, not only in terms of their geographical spread (e.g. LAN or WAN) but also in terms of the number of participants that may be supported by the infrastructure without causing network congestion (e.g. DVB vs. WiFi).

7. Conclusions and future work

In this article, we presented an overview of the requirements for environments that enable network-based music collaboration. Although requirements in this context have been previously reported for specific research efforts, we targeted towards a more generalised approach that takes into account the majority of the variations that exist in distributed music performance scenarios.

The requirements study, as well as the unraveling of the possible variations of an application scenario for remotely performing music, is a part of a larger research project. In this project, DIAMOUSES, three of the possible scenarios have been selected for evaluating the system under development. We expect that user and expert evaluation of the selected scenarios will enlighten valuable findings in the area of network-based music collaboration.

8. Acknowledgements

The DIAMOUSES project is being implemented in the context of the Regional Operational Programme of Crete 2000 – 2006 and it is co-funded by the European Regional Development Fund (ERDF) and the Crete Region, coordinated by the General Secretariat for Research and Technology, of the Ministry of Development of Greece. The partners of the DIAMOUSES consortium are: Department of Music Technology and Acoustics, Technological Educational Institute (TEI) of Crete – Project Coordinator; Department of Applied Informatics and Multimedia, TEI of Crete; Department of Electronics, TEI of Crete; Department of Computer Engineers and Informatics, University of Patras; FORTHnet S.A.; AKMI, School of Vocational Training.

9. References

[1] J. Lazzaro and J. Wawrzynek, “A case for Network Musical Performance”, *in Proceedings of ACM NOSSDAV ’01*, Port Jefferson, NY, June 2001, pp. 157–66.

[2] A.A Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos and C. Kyriakakis, “From Remote Media Immersion to Distributed Immersive Performance”, *in Proceedings of the ACM SIGMM 2003 Workshop on Experiential Telepresence,* November 7, 2003, Berkeley, California, USA

[3] X. Gu, M. Dick, Z. Kurtisi, U. Noyer, and L. Wolf, "Network-centric Music Performance: Practice and Experiments", *IEEE Communications Magazine*, June 2005, pp.86-93.

[4] A. Barbosa, “Displaced Soundscapes: A Survey of Network Systems for Music and Sonic Art Creation”, *Leonardo Music Journal* - Volume 13, MIT Press, 2003, pp. 53-59.

[5] W. T. C. Kramer, “SCinet: Testbed for High-Performance Networked Applications,” *IEEE Comp. Mag*., vol. 35, no. 6, June 2002, pp. 47–55.

[6] R. Rowe, “Real Time and Unreal Time: Expression in Distributed Performance”, *Journal of New Music Research*, vol. 34, Routledge, 2005, pp. 87-95.

[7] K. Sooyeon, L. JeongKeun, W. Y. Tae, K. Kyoungae, and C. Yanghee, “Hat: A High-quality Audio Conferencing Tool using mp3 Codec”, *INET 2002*, Washington, DC, USA, June 2002

[8] P. Bellini, P. Nesi, G. Zoia, "Symbolic Music Representation in MPEG", *IEEE MultiMedia*, Oct-Dec, 2005, pp. 42-49.

[9] J.-C. Bolot, H. Crepin and A. Vega Garcia: “Analysis of audio packet loss in the Internet”, *Proceedings of the NOSSDAV’95*, 1995, pp. 154 – 166.