Using a dedicated toolkit and the cloud to coordinate shared music representations

George Vlachakis, Nikolas Karadimitriou and Demosthenes Akoumianakis
Department of Informatics Engineering
Technological Education Institution of Crete (TEI-C)
Heraklion, Crete, Greece
gio.vlachakis@gmail.com ; karadimnik@gmail.com ; da@ie.teicrete.gr

***Abstract*— This paper describes the design and implementation of a toolkit for collaborative music making. The rationale for the toolkit stems from the commitment to considering music making as a representational practice that entails virtual work amongst distributed peers. Using a pictorial illustration of the toolkit, we discuss several application scenarios and highlight the innovative features of the approach. One key contribution of the toolkit is that it enables formation of online music ensembles capable of negotiating music making using computer-mediated boundary artifacts and generic cloud services.**

Keywords—collaborative music making; boundary artifacts; cloud services; virtual work

I. Introduction

Collaborative music performance typically engages musicians in the use of multiple media (auditory, visual, textual / spoken, physical / gestures, etc.) to attain a common agenda such as improvisation, rehearsal, stage performance, composition, etc. Technology has always played a key role, and in many cases, it becomes an indispensable aspect of the user experience. Nevertheless, and in spite of the plethora of efforts, most of available works tend to be '… focused on individual expressivity' while the potential for collaborative engagement has been under-utilized [1]. This paper revisits this claim and aims to provide insights into collaborative music making as virtual work.

During Network Music Performance (NMP), virtual teams use Internet connections to collaborate. Typically, team members meet on a 'digital space', which aims to simulate a recording studio (in the case of music making) or a stage (in the case of music performance). Prior to a NMP, teams may also collaborate on a broad range of issues that determine and catalyze the actual performance. Nonetheless, the conduct of these activities, the detailed treatment of their intrinsic properties and their importance for NMP are not sufficiently explored in the relevant scholarship.

The present work attempts to address this gap by anchoring collaborative music making as virtual work and then discussing a toolkit intended to facilitate non-trivial scenarios of collaborative music engagements prior or leading to NMP. The rest of the paper is structured as follows. Section II highlights our theoretical motivation and consolidates the research focus of the present work. Section III summarizes the design perspective and commitments that underline the proposed toolkit for collaborative music making. Then in Section IV, we present the architectural components of the toolkit and discuss representative scenarios. The paper concludes with a summary of key contributions and pointers to future work.

II. Theoretical Background & Motivation

The present work links with two primary scholarships. The first relates to virtual teams and genres of computer-supported collaborative work. The second stream anchors the domain of digital music in relation to some of the prominent representations used.

A. *Virtual teams and computer-supported collaboration*

Attempts to define virtual teams typically refer to "…a collection of individuals who are interdependent in their tasks, who share responsibility for outcomes, who see themselves and who are seen by others as an intact social entity embedded in one or more larger social systems and who manage their relationship across organizational boundaries" [14]. Such a definition, although valid and widely acknowledged, fails to grasp intrinsic properties of virtual work. In their recent analysis of the 'lure' of the virtual [2], Bailey and colleagues make this point explicitly and attempt to understand the virtual in relation to what is digitized versus what is virtualized as well as the types of operations that qualify virtuality. Thus, they point out regimes of virtual work and qualify them in terms of operations *with*, *on*, *through* and *within* representations.

Scholarship on remediation also acknowledges the importance of representations in distributed and collaborative settings. Specifically, in the new media literature, remediation refers to the process of representing one medium into another [18]. According to this view, paintings, pictures and text are remediated when embedded in digital media, such as electronic encyclopedias, which in addition to text and graphics provide facilities for sound, video, searching and linking capabilities. Another conception of remediation points to the changing human routines and practices, as a result of using digital technologies. In this line of thinking, remediation prompts considerations at the level of embedded agencies and agency distribution. Specifically, Lanzara [4] defines this sort of remediation as a migration of an assemblage of embedded agencies established in a certain medium to a new assemblage in a new or multiple media. For our purposes, the baseline of this perspective shifts the focus of attention towards a more detailed study of how 'digital' assemblages occur and what effects their operation may have on virtual practices as well as on the routine activities that constitute these practices. As briefly outlined in earlier works [3], this is particularly relevant for collaborative music making since in the process of remediation, new representations may be favored that bring into sight aspects hidden or not visible in conventional representations or older media.

B. *Digital music*

Although there have been plethora of research studies on digital music [5][6][7][10] and music making [9][12], the actual conduct of virtual work with certain representations remains challenging. This may be due to the complexity involved as the quality of a musical piece is not only affected by music principles and the knowledge of the composer (i.e. music theory, harmony , etc.) but also from factors such as mood, psychology, music preferences and other subjective factors. Another influential factor appears to be the social context of making music and the interaction of the composer with colleagues and third parties. As this interaction is computer-mediated, it relies as much on digital representations as on conventional media. Moreover, it may be argued that due to the digital medium and the computational capacity offered by computers, such remediation enhances creativity and communication, while it also simplifies procedures of music composition or learning.

Regarding digital music representations, there are four basic types [6] of interest to the present work; the symbolic musical notation, textual representations such as MIDI, the XML representation of music data, and the audio (acoustic) representation. In digital settings these representations may be exploited synergistically or even interchangeably (with certain limitations). For instance, the symbolic music notation data can be rendered as a digitized score or a MIDI roll. Such a representation is suitable for dictating notes, and/or filling lyrics and chords, while it also affords interaction with the end user through several I/O devices (i.e. MIDI keyboards, computer keyboard, mouse, etc.). Similarly, XML representation of the symbolic music notation affords coding in a structured textual format, which in turn, can accommodate additional information such as lyrics, comments related to notes, history of revisions and other similar functionalities that may be coded as metadata. Finally, the audio representation can be generated by rendering MIDI files or other formats that depict the recorded performance of co-located or remote music ensembles. In summary, Table 1 consolidates the four basic representations of digital music and highlights their intended use as well as the respective file formats.

TABLE I. TABLE 1 MUSIC REPRESENTATIONS & THEIR USAGE

Representation	***Usage***	***File Format***
Music notation	Note dictation	Score, MIDI roll
XML	Synchronization	musicXML
MIDI	Quick playback	MIDI file
Audio	Playback	mp3, wav

Progress on digital music representations catalyzed the development of new applications for different purposes. For example, for music composition there have been proposals such as PSO's [8], Daisyphone [9], Reactable [10] and FMOL [11]. Nonetheless, it is the advent of high-speed IP networks which in combination with digitized representations of music create novel virtualities, with the most prominent being the real-time NMP. NMP systems range in scope and include new musical instruments [10], laptop ensembles [12], but also the enacted capacity for virtual teaming between artists, composers and third parties.

C. *Research focus*

In spite of recent progress and advancements, collaborative music making is still pending for further attention. Most efforts reported in the relevant literature dismiss key questions such as what is virtualized, what sort of linguistic vocabularies are appropriate for collaboration and what is actually enacted through co-engagement in music practices? The present research attempts to address the above by considering music making as a representational practice that entails virtual work across sites and amongst distributed peers. Thus, the focus is on the 'activity-action-operation' system (in the activity-theoretic sense) that dictates co-engagements, the linguistic patterns that emerge and the tools needed to support them.

III. DESIGN TARGETS AND APPROACH

Collaborative engagement of peers with music can be anchored in various ways. One possible classification can be in relation to the typical phases such as composition, preparation where participants review and negotiate musical materials (asynchronously and synchronously), group rehearsals, improvisation, as well as the actual performance of the musical work. Another classification may be derived by focusing on the distribution of collaborative activities across multiple (i.e., online and offline) contexts [19]. It is also possible to envisage and qualify virtual work by examining the type of operations on the digital representations entailed. In all cases, the toolkit for conducting virtual work presents a prominent design challenge that requires detailed consideration.

In the present work, the focus is primarily on facilitating collaborative engagements in music composition and the preparatory stages leading to group rehearsals, improvisation and / or performances. To this effect, three sets of design issues stand out very promptly. The first entails the anchoring of collaborative music making as operations *with*, *on*, *through* and *within* digital representations, thus qualifying regimes of collaborative engagements as either, remote control, simulations or virtual teams [2]. The second set of issues accounts for devising an appropriate mix of digital representations for music and ascribing certain affordances such as activity awareness, social translucence, interoperability across platforms, synchronization of remote users, etc., which are important for virtual group work [3]. Finally, we are equally concerned with the tracking and management of digital trace data 0 that provide insights to dynamic aspects of collaboration and the sorts of negotiations that take place between participants. This is particularly useful for untangling the historical contexts that gave rise to a designated piece, while it may also follow the actual performance of the piece as added-value metadata.

A. Managing multiple representations for music

The use of multiple digital representations of music in collaborative settings is compelling for several reasons. Firstly, it is widely acknowledged that humans in certain contexts (i.e., composing vs. improvising) and certain roles (i.e., composer versus performer) prefer different representations that suit better their work. Secondly, different digital representations exhibit variable transformative capacity. For instance, it is possible to transform score into MIDI and vice versa but is quite challenging to (automatically) compile score or MIDI from physical (i.e., non-MIDI) instruments. Thirdly and as a result of what is digitized, it may be appropriate to recruit more than one digital representations (i.e., MIDI roll and visual score) to express certain human intentionalities (i.e., changes in pitch, addition of lyrics, etc.).

The above make compelling the need for our toolkit to exhibit the capability of handling multiple different representations of music so as to empower users to convey their intended goal while co-engaging in digital settings. In response to such a need, the research community has provided various partial solutions as indicated by the systems briefly reviewed in the previous section (see II.B). Nonetheless, these solutions address specific tasks (ignoring others), are platform-specific, and more importantly, they are not compatible, thus not easily combined in a unified manner. As a result, it was deemed appropriate to devise a three-layered design strategy in order to (a) augment interaction facilities supported by certain toolkits so as to improve the handling of established interaction components such as staves and MIDI rolls; (b) expand libraries to provide programmatic support for new and customized interaction objects, such as new top-level containers and object hierarchies; and (c) integrate different libraries and platforms through appropriate Application Programming Interfaces (API) that promote interoperability, thus appropriation of features such as cloud-based file sharing, communication and reuse.

B. Scenario-based analysis of regimes of virtual work

Another set of design-oriented challenges entails a deeper analysis of the interaction patterns proliferating in virtual settings when certain technologies mediate the conduct of virtual work. To attain these challenges, we rely on semiotic engineering and the classification of virtual group work as operations *with*, *on*, *through* and *within* representations [2]. To establish a baseline, scenario-based techniques are recruited as a context for detailed design [16] combined with appropriate retooling methods [17]. Specifically, scenarios may be retooled in a variety of ways to derive alternative concrete configurations of socio-material relations. One retooling strategy is purely technological and entails choices of specific tools, technologies and/or services for each regime of virtual work. This will highlight and confirm options about digital representations of music (e.g., audio, video, symbolic), contexts of use (desktop, mobile, laptop ensembles) and range of interaction facilities (e.g., music toolkits, HTML 5.0, cloud services, platforms, etc.). Another retooling strategy is to rely on musicological lenses to bring into sight alternatives based on music genres (jazz, classical music, folklore, rock & roll, electro-acoustic) and the corresponding prevalent practices. A final retooling strategy entails accounting of user types (experts versus novices) and their preferences (e.g., reliance on score for orchestras) or scope of activities (e.g., composing, rehearsing, learning, performing, etc.). So far, and for the purposes of the present work, we have explored the first and the third retooling strategy aiming for a mature prototypical implementation. Clearly, however, retooling scenarios from a musicological perspective is of critical importance and high priority in the plan for future work.

C. Managing digital trace data

The third set of design-oriented commitment relevant to the present work addresses the challenge of tracing and making sense of dynamic aspects of collaboration. The rationale of this commitment is grounded on the belief that there are intrinsic details behind NMP that can only be untangled when considering the historical context of activities leading to the actual NMP. Such historical data can be traced, made explicit and if needed, follow the final 'cultural' artifact of the NMP as metadata. In addition, digital trace data may also be useful as anchors of the micro-level negotiations taking place between remote peers in certain scenarios. For instance, in music theory lessons it is useful to disentangle the aggregate outcome by responding to questions such as: Who makes errors repeatedly and in which parts of the music piece? Who receives most recommendations about his work? Who is recurrently fast or slow when performing certain actions? In light of the above, being able to retain digital traces of the co-engaging peer practices and transform these into meaningful information is relevant, appropriate and valuable for any sort of collaborative setting, including music making. Having said this it is important to highlight that addressing questions such as the above is not possible through conventional provisions of NMP systems.

Finally, it should also be mentioned that commitment to digital trace data serves an additional target which is that of facilitating virtual ethnographic analysis of online music ensembles in their undertakings of virtual work. This form of analysis can provide substantial insights into the questions raised earlier by recruiting exploratory visualizations to make explicit and accountable micro-negotiations. Thus, embedding functionality that enables tracing actions of users is a highly relevant design target as it allows re-construction of the past through its digital traces and augmentation of NMP with its traceable history.

IV. Architectural Components

Having outlined the design commitments relevant to the present work, this section will briefly describe the architectural underpinnings of a toolkit intended to provide support for collaborative music making. At the highest level of abstraction, the toolkit comprises one module intended to facilitate local and situated articulation of collaborative music making and one boundary artifact that resides in the cloud and synchronizes all users co-engaging in the practice of collaborative music making. The practice-specific component makes provisions for creating, manipulating and negotiating music materials (based on their current state as retrieved from the cloud), inviting and communicating with users, as well as managing digital trace data. A synchronization module (which is part of the client software) undertakes to interoperate with the boundary artifact through API calls and to translate data received into the local context of each user.

A. *Client side components*

The practice toolkit implements a range of functionalities, including articulation of digital representations of music, transcoding music representations, annotation mechanisms aligned in time and event logging. One primary function of this toolkit entails the handling of instances of the MusicSheet object that allows users to work on different representations of musical information. Figure 1 depicts a representative example of a user working in the music composition mode. Each representation exhibits different affordances. Thus, notes can be added to and modified in the score while some of their properties such as pitch or dynamic, may be determined by dragging notes Up/Down and/or Right/Left in the MIDI roll.

Once the user creates a piece of work, peers can be invited to co-engage and refine it. Figure 2 depicts a situation where two users have been invited, but only one is online (i.e., logged in with their accounts) reviewing the comments attached to the highlighted note. As shown, evidence of social interaction is presented by the counter just above the specific note which sums the comments related to this note. The dialog offers additional details such as who raised the comment and who has responded (or is responding). Another scenario, useful for highlighting some of the capabilities offered by the toolkit to remote users is depicted in Figure 3. This time, the screen highlights relative performance of two users co-engaged in the joint session. This form of artifact awareness using highlighters anchors errors or misalignments, thus prompting users to consider appropriate corrective action. It is also worth mentioning that the set of highlighters can be invoked in non-collaborative settings with the remote user's activity being simulated.

In the current version of the toolkit, the implementation of the music representations have been effected by augmenting and expanding features of the basic 'stave' object of the JMusic library [13]. Augmentation aimed at injecting new capabilities to the Stave object. Expansion led to new objects such as the MusicSheet, the Sheet, the Controller and the Handler. The MusicSheet object is the application's main container and represents the digitized score. The Sheet object is an intermediate container. By default, three sheets are placed on a MusicSheet making it three pages long. The Controller object is responsible for the MusicSheet view properties. The Stave object represents the music and can be Treble Stave or MIDI roll. User actions such as inserting, editing or removing notes and chords are implemented as operations on a Stave object. Furthermore, Staves invoke functionalities such as commenting (see Figure 2) or adding lyrics in a manner similar to handling comments. The Stave Handler object is responsible for Stave object control (i.e., commenting and note changes) as well as for synchronizing staves. It also undertakes to highlight scores and indicate ownership, depending on the current users' access rights.

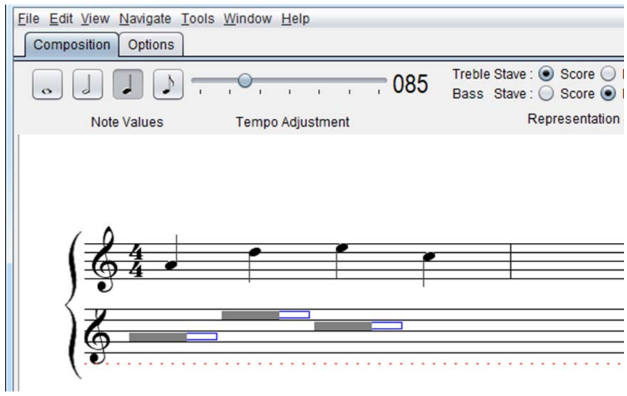

Figure 1: Work on multiple music representations

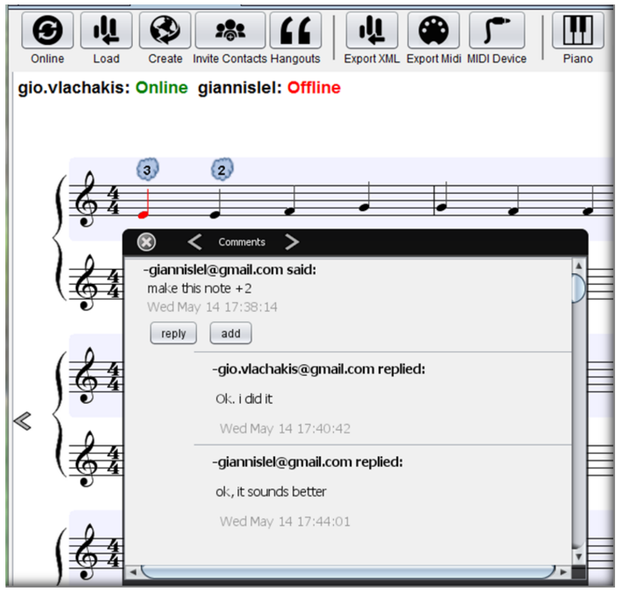

Figure 2: Collaborative engagements

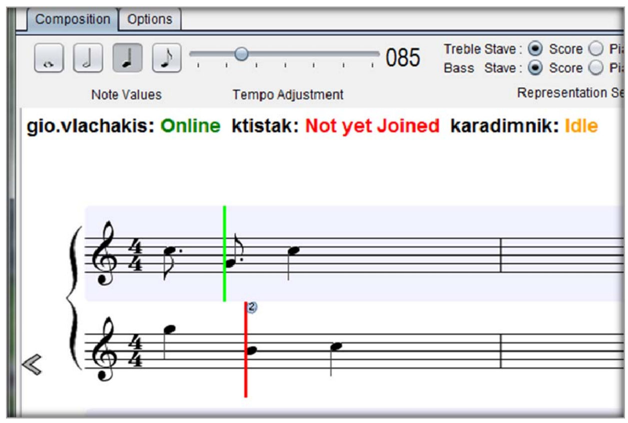

Figure 3: Coordinating and simulating peers

B. *The boundary artifact in the cloud*

One defining characteristic of the underlying architectural pattern is that the toolkit is completely relieved form the need to access local data stores. This is achieved by transforming events to a shared XML representation retained in the cloud and constituting the boundary artifact (see Figure 4). The Synchronizer maintains exclusive access to the boundary artifact and is responsible for coordinating different clients and the handling of the transactions with the boundary artifact. Our current implementation utilizes Google Cloud, Google Drive and Google Docs libraries for implementing profile management, data stage and synchronization of instances of the application. A side effect of this architectural choice is that collaborative work implicates users who are virtually present with their Google accounts and possessing access to the shared Drive designated to hold the object of collaboration. On the other hand, this arrangement allows transformation of every action triggered locally to XML, retained as boundary artifact in the Drive. Thus, every time a peer modifies the state of the shared object locally, the synchronizer commits this change to the boundary artifact and propagates its effect to all registered clients for local processing.

In terms of content, the boundary artifact comprises (a) music score related data including revisions and comments; (b) user data with indication of accounts, status and contributions (c) session data (i.e., sessionID) and summative details such as number of times the team has gathered online to work on a particular musical piece, etc.

C. Digital trace data

Digital trace data management is the responsibility of the Handler object which undertakes to store a number of changes on scores. Such changes may entail modification of notes, changes in pitch, duration, comments and annotations. These are traceable events which are made persistent together with other metadata indicating who triggered the event, the event's timestamp, etc. Stored events can be visualized either as a timeline (Figure 5, left hand side component of the toolkit) or as 2D exploratory visualization (Figure 5, middle part). In the timeline, different types of events are represented with different nodes, thus allowing users to review the history of collaboration at the note level. Back and forth navigation through the history of events is possible by selecting different nodes in the timeline. In this manner, the score can be navigated by following the history of revisions. The 2D visualization makes use of prefuse (http://prefuse.org/) to compile an isomorphic representation of the chain of events for post-session ethnographic analysis. This time, vertical rectangles stand for different staves, while horizontal rectangles represent notes within a stave. The use of color in notes denotes the pitch information (lower pitches are colored in darker colors). The size of a note depends on its duration (the higher the duration the bigger the x-size of the horizontal rectangle in the visualization). Labeled circles indicate the number of revisions for a specific note. For instance, in our example in Figure 5, the first note has been the result of 8 revisions with the 8th revision depicting the discussion between the owner of the stave and a peer. The turn taking between the users engaged is illustrated by distinct symbols that anchor contributions in the discussion thread. Finally, it is worth mentioning that the visualization is fully interactive. Thus, notes are selectable for reviewing their content and history. Users may also experiment with previous instances in a specific part of the musical piece by replacing notes with earlier revisions.

V. DISCUSSION, IMPLICATIONS & FUTURE WORK

The present research is a step in the direction of supporting collaborative music making as a representational practice. In this vein, several key design targets have been achieved. Firstly, the handling of multiple digital representations of music, concurrently and in synchronous collaborative settings, is important and challenging with the current generation of graphical toolkits. Secondly, the management of digital trace data to unfold the history of co-engagements makes compelling the need for appropriate visualizations. Thirdly, the use of cloud service to facilitate boundary spanning is a promising tactic that relieves from the burden of handling add-ons such as dedicated collaboration servers. It is also evident that our research supports several of the regimes for virtual work suggested in [2]. Specifically, collaborators can co-engage by operating *with* their voices (Google Hangouts) or their Google accounts. Once online, they can also operate *on* or *through* designated digital representations. The more demanding case of operating *within* representations was anchored by simulating user performance. However, the ultimate test is the conduct of real-time NMP where peers operating *within* representations create a new reality that lacks a single physical referent.

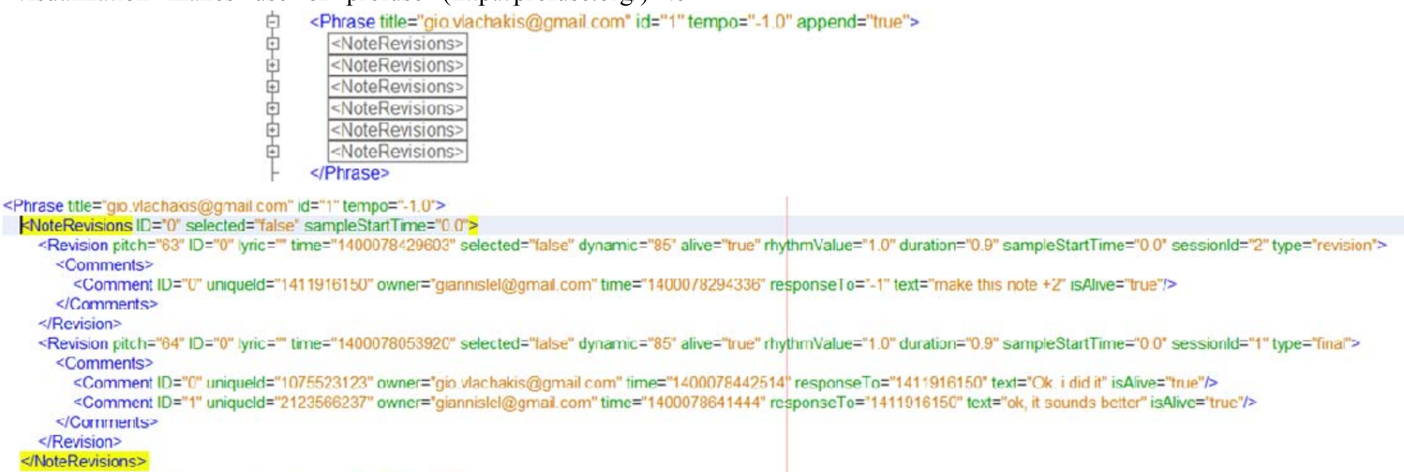

Figure 4: The boundary artifact

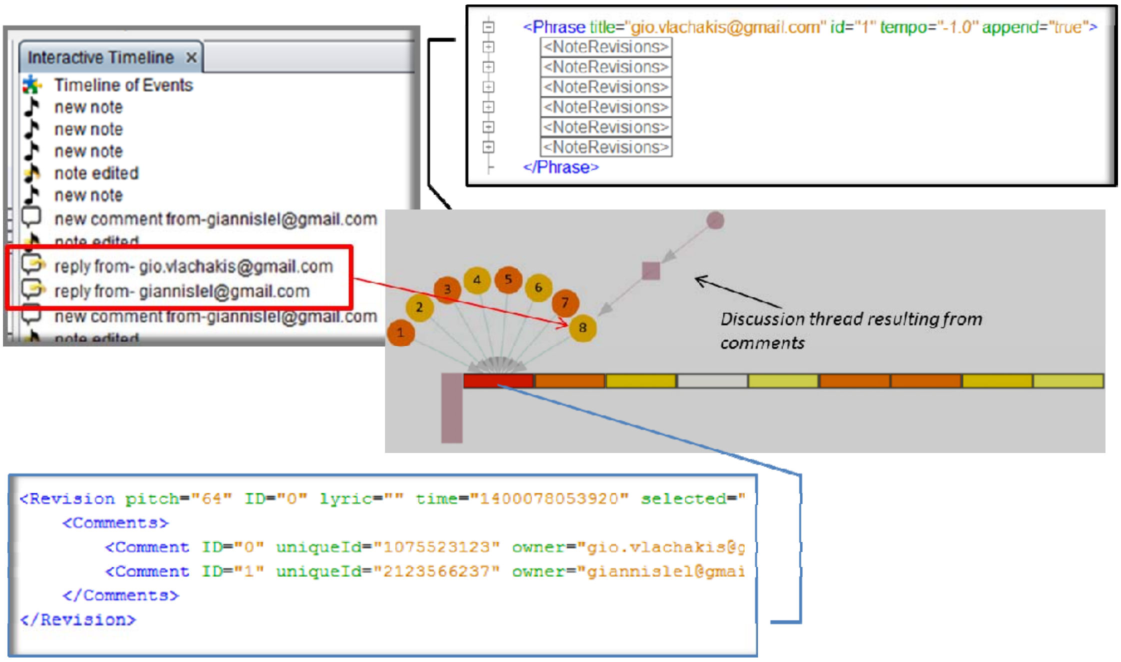

Figure 5: Digital trace data as timeline (left) and as 2D visualizations (right)

Future work concentrates on some of the challenges outlined above which are currently being detailed using appropriate scenarios, including real-time NMP. We are also exploring additional Web 2.0 services such as EchoNest (http://echonest.com/) and its API to capitalize upon machine listening techniques that simulate how people perceive music. In this respect, we can take advantage of our traced events and make a musicological analysis among several phases of collaboration. Finally, we are considering multimodal dialog patterns to augment affordances of digital music either prior or during real-time network music performance.

ACKNOWLEDGMENT

This work is part of the first author's thesis submitted to the Department of Informatics Engineering of TEI of Crete in partial fulfillment of the requirements for the MSc "Informatics & Multimedia'. The research has been co-financed by the European Social Fund (ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALIS–MusiNet.

REFERENCES

[1] I. Hattwick, and M. Wanderley, "A Dimension Space for Evaluating Collaborative Musical Performance Systems," NIME'12, May 21 – 23, University of Michigan, Ann Arbor, 2012.

[2] D. E. Bailey, P. M. Leonardi, and S. R. Barley, "The Lure of the Virtual,"Organization Science 23 (5), pp. 1485–1504, 2012.

[3] D. Akoumianakis, G. Ktistakis, G. Vlachakis, P. Zervas, and C. Alexandraki, "Collaborative music making as 'remediated' practice: Blending boundary objects and cloud services," in 18th International Conference on Digital Signal Processing (DSP), IEEE, 2013.

[4] G. Lanzara, "Remediation of practices: How new media change the ways we see and do things in practical domains," First Monday 15 (6-7), June 2010.

[5] N. Bryan-Kinns and F. Hamilton, "Identifying mutual engagement," Behavior & Information Technology 31(2), pp. 101–125, 2012.

[6] G. Wiggins, E. Miranda, A. Smaill, and M. Harris, "A Framework for the Evaluation of Music Representation Systems," Computer Music Journal 17 (3), pp. 31-42, 1993.

[7] A. Barbosa, "Displaced SoundScapes A Survey of Network Systems for Music and Sonic Art Creation," Leonardo Music Journal, vol. 13, 2003.

[8] A. Barbosa, "Public Sound Objects: a shared environment for networked music practice on the Web," Organised Sound 10(3), pp. 233–242, 2005.

[9] N. Bryan-Kinns, "Daisyphone: The Design and Impact of a Novel Environment for Remote Group Music Improvisation," in Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, NY, USA, 2004, pp. 135–144.

[10] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner, "The reacTable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces, in Proceedings of the 1st international conference on Tangible and embedded interaction (TEI'07), NY, USA, 2007, pp. 139 – 146.

[11] S. Jordà, "FMOL: Toward User-Friendly, Sophisticated New Musical Instruments," Computer Music Journal 26 (3), pp. 23–39, 2002.

[12] J. Freeman and A. V. Troyer, "Collaborative Textual Improvisation in a Laptop Ensemble,"Computer Music Journal 35 (2), pp. 8–21, Jun. 2011.

[13] A. R. Brown and A. C. Sorensen, "Introducing jMusic," in Creative Industries Faculty, Queensland Univ. of Technology, Brisbane, 2000.

[14] S. G. Cohen and D. E. Bailey, "What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite," Journal of Management 23 (3), pp. 239–290, 1997.

[15] J. Heer, S. K. Card, and J. A. Landay, "Prefuse: A Toolkit for Interactive Information Visualization," in Proc. of SIGCHI Conference on Human Factors in Computing Systems, NY, USA, 2005, pp. 421–430..

[16] J. M. Carroll, "Making use: Scenario-based design of human-computer interactions," MIT Press, Cambridge, Massachusetts, 2000.

[17] L. Erskine, D. Carter-Tod, J. Burton, J., "Dialogical techniques for the design of web sites," International Journal of Human Computer Studies, 47, pp. 169-195, 1997.

[18] R. A. Grusin and J. D. Bolter, "Remediation," Configurations 4(3), pp. 311–358, 1996.

[19] D. Akoumianakis, and Alexandraki C. "Collective Practices in Common Information Spaces: Insight from two case studies," Human–Computer Interaction 27 (4), pp. 311-351, 2012.

J. Howison, A. Wiggins, and K. Crowston. "Validity Issues in the Use of Social Network Analysis with Digital Trace Data," Journal of the Association of Information Systems 12 (12), pp. 767-797, 2011.