Project proposal for SFR Agorantic
ITACA: Information Technology and Artistic CreAtion

Coordinator: Eitan Altman

1 Participants

Equipes Projets INRIA:

- **EPI MAESTRO, Sophia-Antipolis:**
  Eitan Altman, Coordinator, Senior Researcher (Direciteur de Recherche) at INRIA Sophia-Antipolis. Has been working on economic modeling of the Internet as well as on the interplay between information technology, culture and society [1, 2]. Expert in networking, game theory and music composition. Graduated in the Composition Department of Tel-Aviv University, and studied two years in CIRM - Centre International de Recherche Musicale, à Nice).
- **EPI REVES, INRIA Sophia Antipolis**:
  George Drettakis. He is the scientific leader of the REVES (Rendering for Virtual Environments with Sound) research group. His research interests are in efficient 3D rendering for graphics and sound, including textures, lighting, shadows and perceptual considerations. He is also interested in immersive 3D interfaces.
- **EPI METISS, INRIA Rennes:**
  Emmanuel Vincent, Rémi Gribonval and Frédéric Bimbot, experts in signal processing and information retrieval techniques for speech and music, especially in audio scene analysis, source separation, speaker/speech recognition and music transcription.
- **Service Dream** Jean-Christophe Lombardo. Expert on acoustic signal processing, and on virtual reality.

Labs from the University of Avignon.

- **Laboratoire Identité Culturelle, Textes et Théâtralité (ICTT) of Univ of Avignon**: Liza Kharoubi.
  Liza Kharoubi, is lecturer in English, Drama and Theatre at the University of Avignon, France. She currently works on anglophone and francophone contemporary theatres, more particularly from Canada and New Zealand, involving new technologies in scenographic writing (cf. the Works of Emmanuel Shwartz and Jeremie Niel in Quebec for instance). She is also involved in the theoretical repercussions of digital interactional technologies on the audience in recent productions and experimentations, with a special interest for ethically related issues. Indeed, what does the intrusion of digital data into the dramatic texture entail for stage writing? What are the consequences of this "cyborg-writing" in its relation to the public?
- **Laboratoire Informatique d'Avignon (LIA) University of Avignon**
  Renato De-Mori and Georges Linares. Dr Linares (director of LIA) and Dr De-Mori specialise in signal processing for speech and audio signals. Since two years ago, Dr Linares has been applying these to music. De-Mori is the author of the book "spoken Dialogs with computers" Academic Press, 1998 and of "spoken language understanding" J whiley, forthcoming.

Other groups:

Artists, Composers and Art institutions

- **Centre National des Ecritures du Spectacle (CNES) de la Chartreuse de Villeneuve lès Avignon**
  Participants: Franck Bauchard, Artistic Director of the Chartreuse-CNES (see `http://cri.histart.umontreal.ca/cri/fr/cdoc/fiche_personne.asp?id=17722`)
  and Emmanuel Guez, Chargé de mission (http://writingmachines.org/vita)
- **Conservatoire National à Rayonnement Régional de Nice**
  Participiants: Michel Pascal and Sebastian Rivas, Composers, Professors in Electro Acoustic Music.
- Jean-Marie Adrien, composer, graduated from the Conservatoir Supérieur National de Musique de Paris, did his Master and PhD and later in IRCAM. Lead several creation projects involving researchers and composers.

2 Objective of the proposal and role of each group

The collaboration aims at initiating and coordinating research on the interplay between information technology and artistic creation, as well as a direct collaboration between researchers and artists. The essential goal of this project is to design and develop the **MediaThor:** A multimedia creation instrument, and to create with it.

Our focus will be on

- **Intelligent Interface:** real-time analysis of signals originating from one or more simultaneous stimulation sources and recorded from one or more sensors (music, monologs and dialogs and environmental noises, signals that come from mobility detectors of a dancer, or theater actor, or video signals).
- **Synthesis:** A tool that allows the artist to translate features extracted from the analyzed real time input signals into audio or video output signals.

One of the challenges to be addressed will be to free as much as possible the performers from wearing a large number of sensors so as to make the technology less apparent. For instance, close-field microphones worn by an actor might be replaced by far-field microphones whose output could be processed in real-time by a source separation, dereverberation and characterization engine. This part requires the complementarity between the INRIA groups specializing in the analysis and in the synthesis of sound, the expertise in LIA needed at the intelligent interfaces for analysis of features in the real time voice input or in graphical input, along with the specialists in theater, music and dance.

MediaThor will be the main part of the project and will involve close collaboration between the participants with Engineering background and the artists. The role of the artists will be to test each module of the Mediathor, and to participate in the developement of the synthesiser part by proposing concepts in the way to exploit the input from the intelligent interfaces. In the testing part, the whole composition class of the Conservatoire Nationale de Region of Nice will participate together with the composers, in addition to the composers.

3 Scientific activity and collaborations

Information technology (IT) has been exploited to develop instruments for assisting the process of artistic creation (e.g. software for composition), and to creating new tools for artistic expression. For example, synthesizers or music software allow composers to create new musical instruments, or to directly create and shape the sound without the need of a musical instrument. I.T. has not only transformed the traditional art forms (music, photography, theater). It also (i) allowed artists to create new forms of art going beyond

those traditional ones; (ii) IT brought new types of relations and interactions between the creation and the audience. An example illustrating these two points is the Interactive Installation form, in which often the art piece interacts to audience location or movements. For example, a musical installation may involve precomposed elements. Various elements may be activated simultaneously in different places thus adding a whole new spatial dimension to the music. Each element may be triggered by the movements of the audience. The audience's movement may further serve to determine other parameters of the music, such as the pitch, the speed, the harmonic content.

The possibilities that this interactivity offers may be exploited in dance or in theater, where the dancer or actor may respond to sound or to light yet at the same time trigger musical events and control various musical and visual features.

We thus focus on man-machine interactions in which some information is retrieved in real time from a human (e.g. audience or actor or dancer), and is transformed as input to a unit that creates a sequence of visual (video) or audio signals (music). We decompose this into four units:

A1. The performer,

A2. A sensor network that conveys information from the performer,

A3. An interface that transforms the information into signals

A4. Reactive Synthesizer: A reactive audio or video unit. The way it reacts to the signals it receives from the interface unit may be part of a fixed system design. It could also involve some programable aspects which is left to the artist (e.g. composer).

We view the last three points as a sophisticated instrument to which the performer is connected, which we call a Interactive MediaThor.

4 Objectives

Some Mediathors already exist (see next section). Our contribution will be to design a novel intelligent interface based on advanced signal processing. This will allow the Mediathor to react (in part A4 i.e. the synthesizer) not only to signals directly obtained from the sensors (e.g. to the volume of speech, the position of a dancer or the intensity of light) but also to react to features extracted from this signal. Examples of such featues are

- the movements of a dancer, obtained by tracking its trajectory over time and classifying it into predetermined movement classes
- Pitches or volume of each of several music instrument that play.
- the intonation, the feelings and/or the keywords pronounced by a speaker, which may be obtained by analyzing the audio speech signal.

We will focus on designing the functionalities of this interface and providing a proof-of-concept of the use of advanced signal processing techniques such as source separation/dereverberation and speaker/speech recognition in the considered real-world applications.

We shall develop modules of the interface related to audio signal processing. Full development of the interface will be one of the goals of the envisaged ANR project proposal.

- To develop a reactive synthesizer for Mediathor. For examples, it would play randomly audio syllables where parameters such as intonation and pitch (or even the emotional content) will be a function of the detected features in the speech signal of a theater actor. The synthesizer can be based on prerecordered syllables.
- One interesting possibility will be the development of gesture-based interfaces in the immersive space at INRIA Sophia-Antipolis. The existing setup includes accurate finger tracking in front of a wall-size

stereoscopic display, or in the immersive iCube. Both systems include high-fidelity audio restitution either over speakers or using wireless headphones and binaural HRTF-based audio rendering.

- We plan also to investigate alternative architectures in which the intelligence of the MediaThor is mainly at the synthesizer, so that simpler inputs are mapped to synthesized output (which in contrast may involve controlled emotional content). This simpler architecture will be used as a starting point in a design of a Mediathor that reacts to video signals.
- Another advanced aspect would be to compare the spoken text to a written version, to detect deviations and to react to them.

5 State of the Art

Interactive mediathors have already been developed in the last century. Jean Claude Risset has developed an instrument allowing live interaction between a pianist and a computer playing the same acoustic piano and which reacts to the pianist. The piano is connected to the Max music programming environment running on a Macintosh II computer, a program developed by the composer. It allows the user to define how the computer should react to the pianist. The program was developed in the eighties (last century). The composition [3] based on this instrument was composed and performed on 1989.

Interactive installations can also be viewed as mediathors. The synthesizer part requires typically requires interactive gesture control of sound patterns [4], which may require coding of features in order to control sound as a function of the code [5]. Composers participating in our project have created various Ineractive installations.

We next provide some background related to the expertise involved in the intelligent interfaces of MediaThor.

Most of speech processing applications aims to extract, from a speech signal, some high level descriptors related to linguistic or semantic content, the language, the speaker identity or his emotional state, etc. This issue may be viewed as an inversion of the speech production process, that transforms thinking or communication-intents into understandable acoustic realizations. This process involves numerous factors related to context, to the linguistic and phonetic systems, to the speaker anatomy and his feelings... Therefore, speech processing consists in modeling relationships between concepts, intents, the linguistic and acoustic levels [9]. We shall use these features when the input for MediaThor is speech signals (e.g. a theater show).

Even if the role of concepts or intents clearly differ in speech and music, many analogies have been recently investigated, both on linguistics and expressivity of speech and music [8, 7, 6]. In this project, we propose to study how these analogies may be used as tools for music creation and concretely integrated into a MediaThor for handling music input signals or for synthesizing music or sound.

Speaker/speech recognition performance degrades when using far-field microphones or when several speakers and/or environmental noises are active simultaneouly. These situations often occur in a performance context when placing microphones on the stage. Advanced source separation/dereverberation techniques have been developed in the last few years [10]. Their application in a real-world performance context raises interesting challenges due to speaker movements and real-time constraints.

6 Budget

The project is already operational. INRIA has financed a stage of Julien Gaillard which was partly on this topic, supervised by Eitan Altman. Two other stages on Mediathor are financed by INRIA (under the supervision of Eitan Altman and Julien Gaillard).

We seek for finance of the following items:

- Two meetings of the whole group at INRIA (demos at the virtual reality lab to which Mediathor is being connected by Julien Gaillard and the two stages by Adam and John) as well as two mutual visits of one weeks each between groups participating in the project. Expected cost: 3000 euros.

- We seek for financing 3 months per year of the PhD thesis of Julien Gaillard. This would be around 7 Keuros per year.

References

[1] E Altman, S Wong and J Rojas-Mora, "P2P business and legal models for increasing accessibility to popular culture", The International Conference on Digital Business London, UK, 17 - 19, June 2009

[2] S Wong, E Altman, M Ibrahim,"P2P Networks: The interplay between legislation and information technology", Sulan Wong , Eitan Altman and Julio Rojas-Mora, "Internet Access: Where Law, Economy, Culture and Technology Meet", Computer Networks, special issue on Wireless Future Internet, 55(2): 470-479 (2011)

[3] Jean-Claude Risset, Duet for One Pianist Eight Sketches for MIDI Piano and Computer, 1989

[4] Alistair RIddell, Towards Interactive Gesture Control of Sound Patterns Using the Wiimote, In Proceedings of the ACMC07 Conference. ANU. Canberra. June 19th. 2007

[5] Alistair Riddell, Gesture and Musical Expression Entailment in a Live Coding Context In the proceedings of the ACMC'09 conference. QUT Brisbane. July 2-4. 2009

[6] J.T. Hogan, "A parallel between music and speech : Tonality and tone", Linguistica atlantica, v. 20, pp. 73-84, ISSN 1188-9322, Memorial University of Newfoundland, Linguistics Department, St. John's, NF, CANADA.

[7] Ray Jackendoff,"Parallels and Nonparallels between Language and Music",Music Perception, volume 26, n 3, pp 195-204, 2009.

[8] "Analyse et modèle génératif de l'expressivité. Application à la parole et à linterprétation musicale", PhD Thesis, Université de Paris 6 - IRCAM, Unité mixte UMR IRCAM-STMS, 2009.

[9] F. Bimbot, J. Bonastre, C. Fredouille et al, "A tutorial on text-independent speaker verification", EURASIP Journal on Applied Signal Processing, 2004(4):430-451, 2004.

[10] E. Vincent, M.G. Jafari, S.A. Abdallah, M.D. Plumbley, and M.E. Davies, "Probabilistic modeling paradigms for audio source separation", In Machine Audition: Principles, Algorithms and Systems. IGI Global, chapter 7, 2010.

7 Requested budget

Most of the budget would be used for paying Master stages or internships for students who will be working on MediaThor. Some will be used for inviting artists, for traveling for collaboration purposes, and for organising meetings once every 3 months that will enable to do brainstorming and to collaborate. The requested budget per year is thus:

- 8000 Euros for travel and local expenses for collaboration and for the meetings of the ARC.
- 3000 euros for hosting visitors
- 12000 euros for stages and internships,
- 2000 euros for participating in conferences related to the issues of this ARC.
- 1000 euros participating in the organization of the international symposium "Travail et création artistique en régime numérique : Images et Sons" that will take place in 24 - 27 may 2011 at university of Avignon.

We thus request 26000 euros for the first year and 25000 euros for the second. In addition we reqeuest a financing of a postdoc for one year.

8 Postdoc

We plan to hire a postdoc with background in signal processing of audio or of speech. Further background in musical signal processing is welcome. The postdoc will mainly work on MediaThor. two third of his time will be spent on the intelligent interface and one third on the synthesizer. The postdoc will be located in Avignon, and will be supervised by Georges Linares (Avignon University) and Eitan Altman (affiliated with INRIA Sophia Antipolis, but physically located at Avignon University under a collaborative contract between INRIA and the university). The postdoc will be expected to spend two months a year in mobility to other groups.

of Support.pdf

Letter of Support

to INRIA ARC Proposal ITACA

Coordinated by Eitan Altman

The research activity of the Acoustic and Cognitive Spaces team of IRCAM is dedicated to the analysis, the reproduction and the synthesis of 3D sound environments. The goal is to provide models and tools, which will enable composers to integrate the spatial organization of sound into their work from its conception to the concert situation. As such, the objective is to promote spatialization as a key parameter in the musical composition process.

In the recent past, we observed a growing interest of composers for interactive sound installations involving the participation of the listener (e.g. navigation within a musically organised space or gestural interaction). Under these conditions, the listening experience is augmented since the auditory sensorial modality is now solicited in combination with other sensorial modalities such as vision and proprioception. This motivated us to develop a research axis on auditory spatial cognition, notably via the study of multi-sensorial integration. We can also notice that the software and hardware instrumentation used for sound installations show significant convergence with those developed for the virtual reality domain. This motivated our participation to the European project CROSSMOD coordinated by George Drettakis from the REVES team and dedicated to crossmodal perceptual interaction and rendering.

We support the ITACA project, coordinated by Eitan Altman. We would be specifically interested in collaborating to the project around the second research axis Mediathor, as we can see strong links with the developments needed for sound installations that often involve 3D audio rendering in combination with navigation and gesture tracking. An interesting application would be for instance to exploit the research and technological environment developed at REVES to allow for interactive sound installations sketching and planning. Indeed, sound installations may be complex to develop and tune. Moreover they may be designed for public sites (a public garden, an exhibition venue) that do not allow for a long rehearsal and fine tuning period. It would be very interesting to be able to sketch and assess such sound installation in a virtual environment providing an accurate rendering of the acoustic, visual and interaction conditions.

ITACA will be of help to our team who intends to collaborate to the project.

Name : Olivier Warusfel
Title : PhD, Head of Acoustic and Cognitive Spaces Team
Address : IRCAM, 1 place Stravinsky 75004 PARIS
Email : warusfel@ircam.fr
Tel: +33 (0)1 44 78 48 85
Date : October 12, 2010