SemanticHIFI

IST-507913

Public Report of WP4 "Rendering"

Covering period: December 2003 – October 2006

Report version: **1.0**

Report preparation date:

Writers: **Olivier Delerue, Samuel Goldszmidt**

Control: **Hugues Vinet, IRCAM**

Classification: **Public**

Contract start date: **December, 1st 2003**

Duration: **36 months**

Project co-ordinator: **Hugues Vinet, IRCAM**

Involved Partners: IRCAM-Room Acoustics (WP Coordinator), IRCAM Hypermedia Studio, BGU

Project funded by the European Community under the "Information Society Technology" Programme

Table of contents

1 WP Overview

1.1 Objectives

The goal of the rendering work package is to provide techniques, either in the form of signal processing modules or as graphical user interfaces, which allow new ways of listening or interacting with the music, based on associated metadata. As opposed to the features developed in the Browsing Project workpackage, which concern more specifically *inter-document browsing*, the ones developed here focus on *intra-document browsing* as an important and specific feature for novel hi-fi systems. Various complimentary approaches have been taken therefore:

- navigation within the music polyphony, through the use of different audio tracks and real time spatial audio simulation techniques. The key underlying technology for this work package is the IRCAM Spatialisateur that allows creating the sensation of sound source localization and room acoustic
- navigation through the temporal structure of music pieces, using automatic analyses developed in WP2
- navigation using hypermedia interfaces, showing internal structures of musical excerpts, as the result of a musical analysis

1.2 Partners' roles

IRCAM Room Acoustics team coordinates the rendering work package and provides the "Spatialisateur", a set of signal processing libraries allowing generating localization and distance effect to anechoic audio signals: these are the necessary signal processing modules and interfaces to handle the last processing stage of an Hifi System. At the rendering stage, an arbitrary number of audio streams (multi channel audio content) have to be mixed into a format compatible with the user's restitution system. Along with this technology, the Room Acoustic team provides applications and graphical user interfaces that allow controlling sound spatialization in a end-user – non expert user – context. IRCAM Hypermedia contributes as well to this task by designing and providing relevant interfaces for temporal navigation and hypermedia navigation. Finally the BGU partner contributes to this work group by providing de-mixing algorithms that enrich a given stereo signal in order to better exploit the mixing tools of the rendering phase

1.3 WP contribution to the project

The contribution of the WP4 to the project remains essentially in terms of spatialization prototypes and scientific research in the field of human – machine interface. Indeed the complete unavailability of the necessary source-separated audio content on the end-user market and the weakness of actual source separation algorithms led the integration partner of the project not to integrate spatialization tools in the main features of the Hifi System.

However some of the prototypes developed will be fully integrated to the final Hifi using and alternate architecture and sound server. This is also true for the temporal navigation and hypermedia interfaces.

1.4 Synthesis of main achievements

- Original prototypes in the context of the control & use of sound spatialization have been created. These projects demonstrate original means for interacting with a 3D sound scene, focusing on domestic use and needs. Some of these prototypes will be remain as fully working features in the final Hifi system.
- SPAT. The IRCAM Spatialisateur has been ported to the PC platform as the preferred environment for prototyping in the context of the SemanticHifi project. The Spatialisateur is a set of digital signal processing modules designed for the Max/MSP environment and dedicated to sound spatialization (sound source localization and room effect simulation).
- Integration process. A number of propositions have been made in the context of integration in order to reach the needs of the particular semanticHIFI device (dynamic creation of players, processing and rendering modules,…). These prototypes take advantage of the Spatialisateur and adapt themselves automatically to the rendering setup used by the listener (stereo loudspeakers, headphones, 5.1 setup,…).
- A set of 6 audio examples has been prepared in order to experiment and demonstrate with the tools developed in the rendering work-package. These examples are issued from a recording provided (for the project internal use only) by the Music Conservatory in Paris. The separated audio tracks were extracted from the original format, trimmed and converted into an interlaced audio format in order to be adapted to the rendering system playback modules.
- An interface for temporal navigation within the piece structure has been developed and is directly integrated in the Hi-fi system
- Several hypermedia interfaces providing analyses of excerpts of musical works have been developed and run on the Hi-fi system as specific applications.
- A de-mixing algorithm, that enables to separate, under certain conditions, several independent audio tracks from a mixture.
- Scientific publications and presentations.

2 WP Results and Achievements

2.1 Cocktail Party Browser

Responsible partner: IRCAM-RA

2.1.1 Functional description

The "Cocktail Party Browser" provides an original means for browsing in a database of music pieces based on streams separation. Different titles are simultaneously presented to the user as audio streams organized in a sound scene paradigm. The browsing process is translated into spatial navigation. The audio stream corresponding to the title under focus is presented in the frontal position, while secondary related streams are presented in the background. The challenge in this application lies in the creation of the desired "cocktail party effect" that allows putting the focus on a musical piece while remaining conscious of several other pieces played in the background.

2.1.2 Method description

The method used for this prototype consists in performing the spatialization simultaneously over three different music title. A number of "spatial positions" have been defined: front, front left, front right, back, back left and back right. When the user chooses either the front left or front right title to replace the front one, a smooth transition operates and the selected title comes to the front whereas the two other titles are soberly replaced by two new corresponding titles.

2.1.3 Position over state-of-the-art

In all existing traditional applications, browsing in a musical database is handled visually, by manipulating titles and categories. The key idea in this project is to provide a complete auditory user interface for navigating in the database.

Related work in auditory display and sonification has been carried for instance by Fernström & Brazil (see "Sonic Browsing: an auditory tool for multimedia asset management" in Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland) and shows that the efficiency in tasks such as "searching for an item in a multimedia database" can be improved by using simultaneous audio streams.

The Challenge in our work is obviously to maintain intelligibility between different streams that do not match in terms of musical dimensions (tonality, tempo, rhythm…) and avoid masking effects between the audio streams. A first proposition has been made by "aligning" the different streams in terms of loudness. Other perceptual aspects should be handled as well in order to completely the capacity of emergence of each stream. For instance, our prototype would probably benefit from the use of the audio summary feature developed in the WP3.

2.1.4 Implementation

The original prototype was created using the MAX/MSP environment for designing the DSP & rendering modules (see Figure 1).

The rendering engine is in charge of streaming the different titles and of performing the spatialisation of these streams. Additionally, a pre-processing module called "loudness-compensation" aligns the 3 streams in terms of perceptual level: the loudness of each stream is measured in real time and a gain factor is applied continuously to each of them in order to avoid that a secondary source take over the primary one.

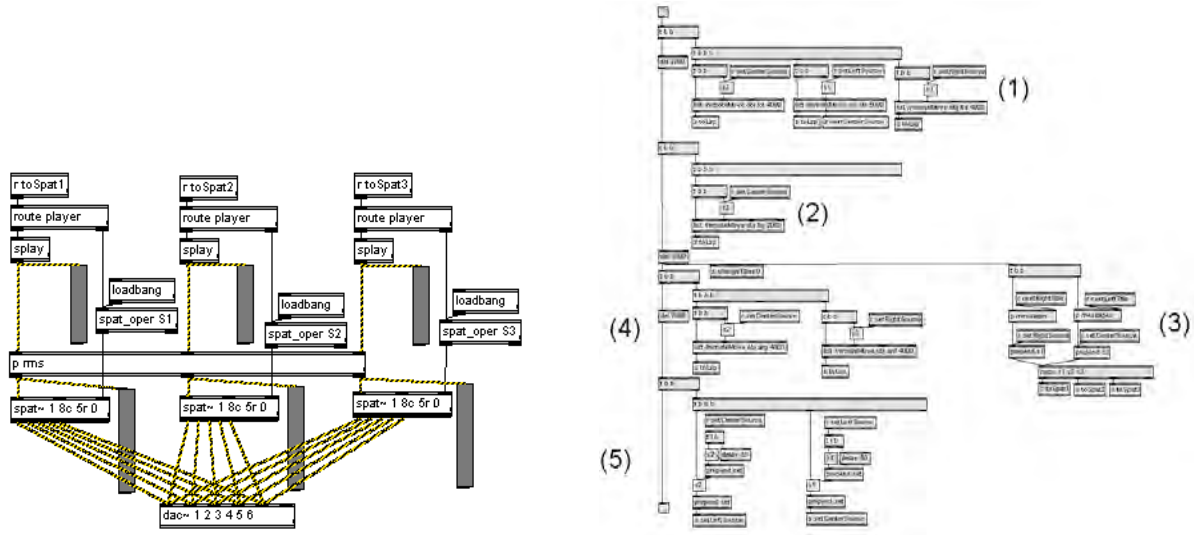


Figure 1 : view of the main window of the audio rendering engine (left) and the interaction component (right).

The ListenSpace application was used in order to handle the geometrical description of the scene: position of frontal primary source, frontal secondary sources and background source position (see Figure 2, left). Finally, a very simple dedicated graphical user interface (see Figure 2, right) composed of two push buttons allows the user moving the focus to either the left or the right secondary source.

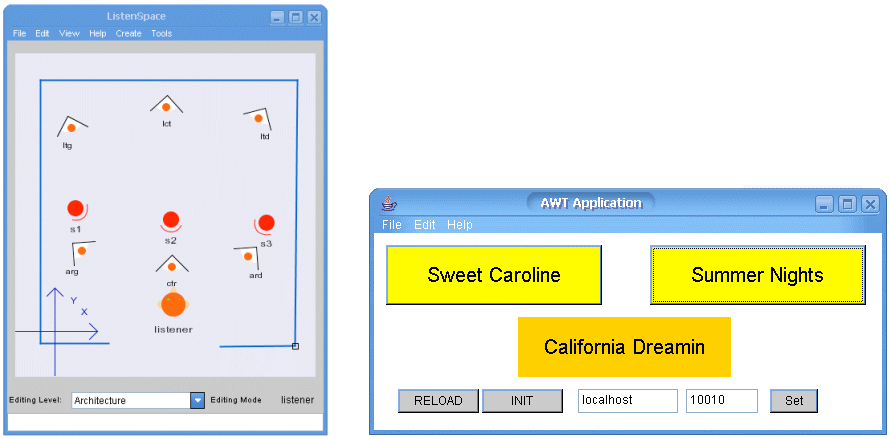


Figure 2 : view of the ListenSpace application with the corresponding audio scene (left) and the graphical user interface (right)

The Figure 3 shows the overall original architecture as specified in deliverable D1.2.1. In our implementation, the "blur processor" (referred as module WP4-M3) has been replaced by a module named "RMS compensation" in order to control the overall amplification of the secondary sources according to the RMS of the main stream.

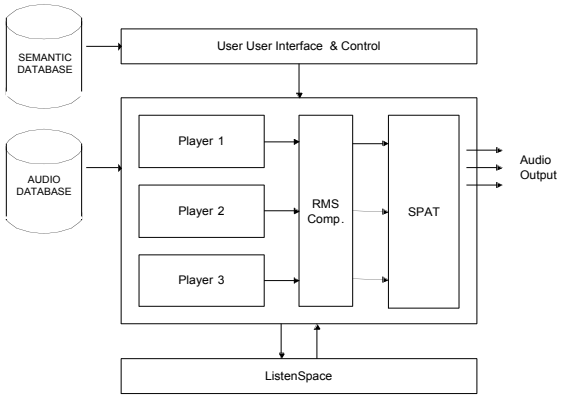


Figure 3: System architecture implemented for the Cocktail Party Browser demonstration.

The communication between the several software components is done via networked communications using the UDP protocol and the OpenSoundControl encoding convention

2.2 L-Scanner

Responsible partner: IRCAM-Room Acoustics

2.2.1 Functional description

This research relates to the domain of visualization and man – machine interaction in order to imagine meaningful ways of controlling the auditory sound scene. Indeed, as specified in use case WP4.3 (man/machine interface section) it is imperative that the sound scene (in the example of "on-the-fly mixing" application for instance) be controlled from a remote position matching preferably the listening position of the user.

The "L-Scanner" prototype is composed of a portable device that allows scanning the physical environment and representing a three dimensional model of the environment as well as the sound sources according to their spatial position in the manner of augmented reality systems.

2.2.2 Method description

The system displays a 3D model of the world that has been prepared beforehand. On top of this scene are represented the several sound scene whose positions are transmitted by the ListenSpace environment. Finally a tracking system attached to the control interface allow determining the position of the listener and its orientation in order to adapt the graphical representation.

2.2.3 Position over state-of-the-art

This project proposes a new way of interacting with a sound scene. Most of the existing works in this field propose a symbolic graphical representation in egocentric coordinates (top view with listener centered).

In our case, we take advantage of the natural "augmented-reality" aspect of audio diffusion to propose a concrete application of see-through interfaces. Thus, this paradigm is well known in the field of interaction in visual augmented environments (see for instance MacIntyre & Mynatt, "Augmenting Intelligent Environments: Augmented Reality as an Interface to

Intelligent Environments." In AAAI 1998 Spring Symposium Series, Intelligent Environments Symposium, March 23-25, 1998, Stanford University, CA.) but, to our knowledge, has never been applied to the control of sound spatialization.

Finally this project put to evidence a particular need in the field of user interface for the control of sound spatialization. Indeed, recent techniques for spatialization and new rendering setup tend to transform the original "punctual" sweetspot (usually defined as the edge of an equilateral triangle that the listener makes with the two loudspeakers) to a wide area in which the listener can walk through and sense new relief effect in the audio rendering. Thus, the man-machine interfaces for designing such sound scenes or controlling the spatialization in such rendering environments will require of being portable and able to accompany the user within the rendering space.

2.2.4 Benchmarks

The prototype we demonstrated proved to be reactive enough when running on a device with a processing power comparable to the power of the pocket-PC used as a remote device. However scaling the tracking system (using infrared cameras) to the restriction of an end-user device is difficult at this time. Different methods or technologies (video, ultrasonic…) need to be investigated.

2.2.5 Implementation

The Figure 4 shows a user manipulating the L-Scanner. The tablet-PC provides a "see-through" view interface to which is superimposed the sound sources represented by yellow spheres. In this experiment the tablet is tracked in space in real time using a sophisticated infrared camera based tracking system.

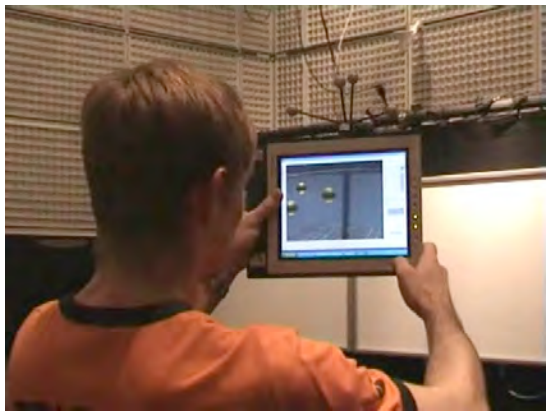


Figure 4: utilization of the L-Scanner interface for controlling spatialization other a wave field synthesis restitution system.

The user can grab a sound source by pressing the "grab" button, move it around in space by pointing the device in the desired direction and finally release the hold on that source by pressing again the grab button.

The device demonstrated at IRCAM during the Resonance festival makes use of an expensive tracking system in order to track the position and orientation of the Tablet PC. This prototype should be simplified in order to reach realistic requirements of a HiFi system).

A video recording of the demonstration of this prototype has been made and is available for project partners.

2.2.6 Dissemination materials

2.2.6.1 Scientific publications

Delerue, Olivier, Warusfel, Olivier, "Mixage Mobile", proceedings of the IHM06 (Interaction Homme – Machine) conference in Montreal.(April 18th – 21st 2006).

2.3 Visualization tool for the control of Spatialization

Responsible partner: IRCAM-Room Acoustics

2.3.1 Functional description

The visualization tool aims at providing useful information to the user while monitoring the spatialization of sources in a sound scene. The key idea is to exploit the background area of interactive graphical user interfaces to superimpose valuable information for the user. This information describes the dependency of a given perceptual criterion with regard to the parameter currently controlled by the user and is represented in the form of a grey scale. Thus, the user can visualize and anticipate what would be the effect of a given action such as the movement of a sound source in the scene according to this criterion.

The system requires that a source is defined as "selected" and represents for each location of the sound scene what would the value of the criterion if the target source was moved to this location.

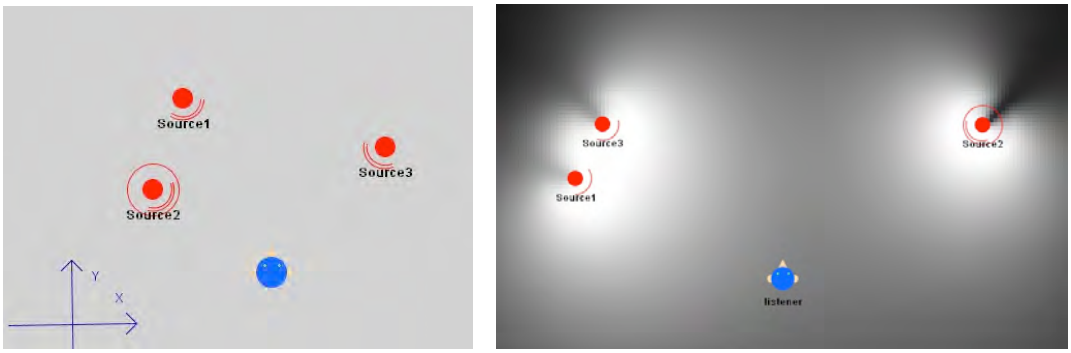


Figure 5: 2D representation & interaction space for a sound scene (left) and superimposition of the radiation pattern of sound sources (right)

Figure 5 gives an example of background representation: it consists in superimposing the radiation pattern of sound sources to the actual sound scene. This indicates to the user the positions of the listening reference point where the sources will be hearable and the positions where they can not be heard.

This system can be applied to a number of perceptual criteria. The representation of these criteria will guide the user in order to help reaching a consistent auditory result. Each of these criteria can be based on metadata either calculated in advance or estimated in real time. Example of such metadata is the loudness value of each sound source, it's global attenuation (resulting from spatialization parameters) or general contextual parameters such as the rendering setup description and the rendering algorithm used.

Several criteria have been implemented and demonstrated. These criteria include for instance the "spatial uniformity" (depending on the rendering setup and the panning law used) that claims that "a good mix implies a homogeneous spatial distribution of energy at the ears of the listener" and uses the loudness value of each sound source calculated in real time as metadata information.

2.3.2 Method description

This prototype does not define a graphical user interface in itself but adds a method for drawing the background of an eagle-view like user interface representing a sound scene. The Figure 6 gives an example of such representation for a sound scene composed of three sources. The criterion represented is the "spatial uniformity" in the context of a stereo (amplitude panning) rendering. The "Source2" is the target source: its location in a mid grey area indicates that the auditory result is slightly unbalanced and the system shows with a bright color the areas where this source could be moved in order to best satisfy the criterion.

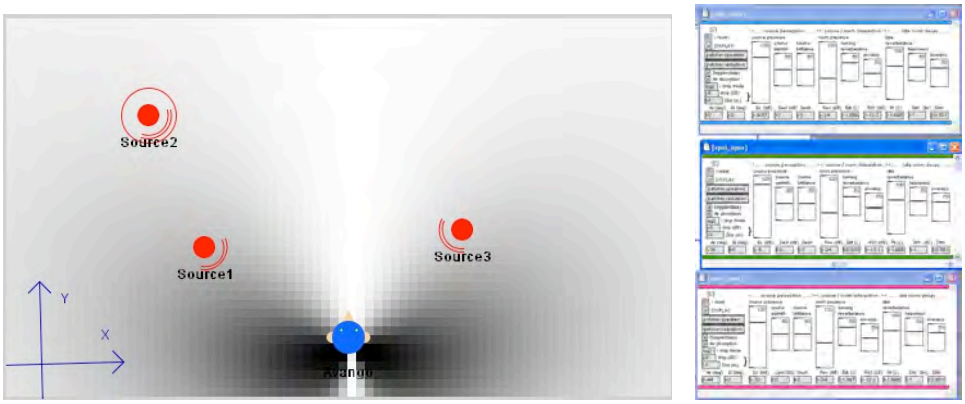


Figure 6: spatial uniformity criterion.

2.3.3 List of Criteria studied

2.3.3.1 Criterion #1: Spatial homogeneity

Problem: a good mix implies a homogeneous spatial distribution of energy at the ears of the listener

What is represented: for a given selected sound source, we represent in each location of the space the quality of the resulting mix in terms of "spatial unity" as if this sound source was moved to this location.

The notion of spatial unity is dependant on the restitution system. For a stereo setup for instance it can be defined as the difference in the resulting loudness coming from each loudspeaker.

Metadata: loudness value for each sound source, panning method, location of sound sources.

Method: for a given number of sound sources, a target point and a target source, we consider the loudness contribution of all sound sources at each loudspeaker if the target source was moved to the target point. We then estimate that the resulting mix would be highly unbalanced for that action if the resulting loudness difference between each loudspeaker exceeds 10 dB.

In our implementation, we use the Spatialisateur in a stereo rendering setup and implemented both a XY coincident microphones simulation panning law as well as a simple cosinus

panning law. We used the "Es" parameter of the Spat_Oper (the Spatialisateur perceptive control interface) as an estimation of the signal attenuation for a given reference distance and apply a generic $1/r^2$ attenuation law to evaluate the loudness of the target source at different positions.

Example:

In the following example, we use 3 sound sources (Source1, Source2, and Source3), and the question is "where should Source2 (the target source) be put in order to optimize the Spatial uniformity criterion. As described previously a light background describes an area where the given criterion would be maximized.

Since the sound scene is already balanced with Source1 and Source 3 that have the same loudness energy, the answer is that Source2 should remain ideally in the central axis of the scene.

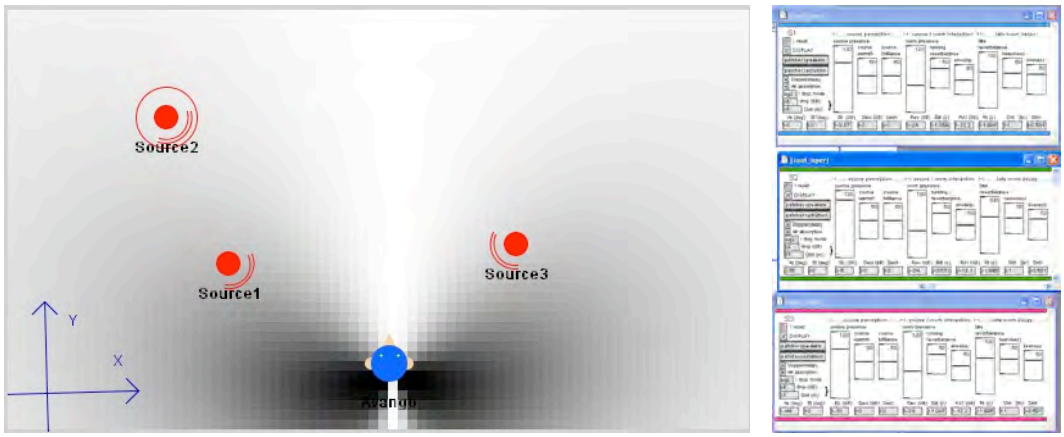


Figure 7: spatial uniformity criterion #1. Moving Source2 in a scene which is already balanced

If we now lower the energy delivered by Source3 (by adjusting is "presence" parameter in the Spatialisateur), we see that the sound scene is unbalanced and that Source2 should ideally be located in a circular-like area centered on the right side of the listening reference position

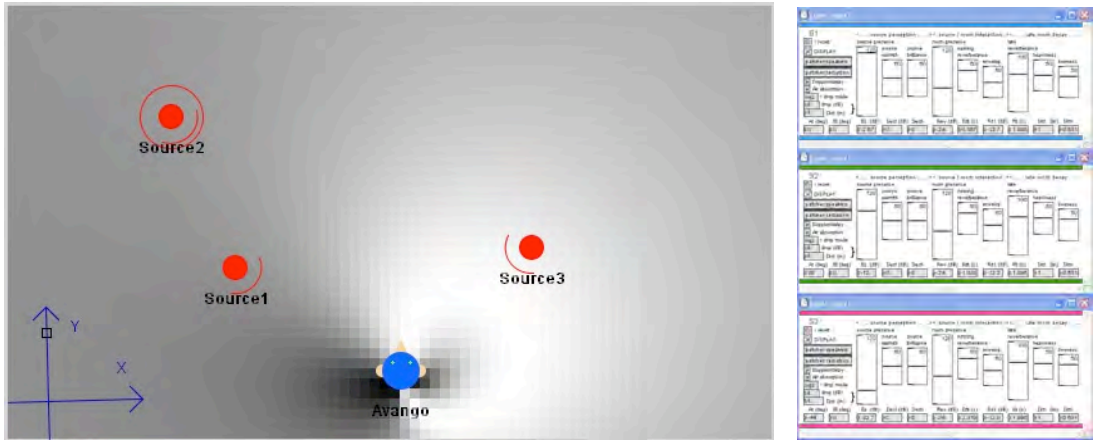


Figure 8: Spatial Uniformity Criterion #2. Moving Source2 in a scene with a high loudness difference between Source1 and Source3

Finally, in the last example we see that Source2 is located in a bright area: this signifies that the Spatial Uniformity Criterion is satisfied.

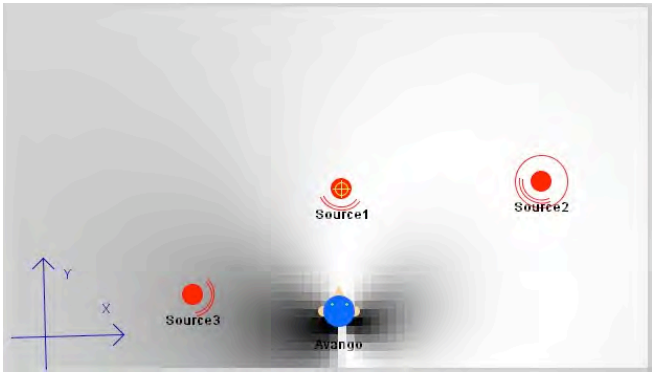


Figure 9: representation of a scene where the quality criterion is optimized

To compare the effect of the panning law used (which depends on the chosen spatialization technique and the rendering setup) we implemented a different panning law for the same Spatial Homogeneity Criterion.

Importance of the panning law:

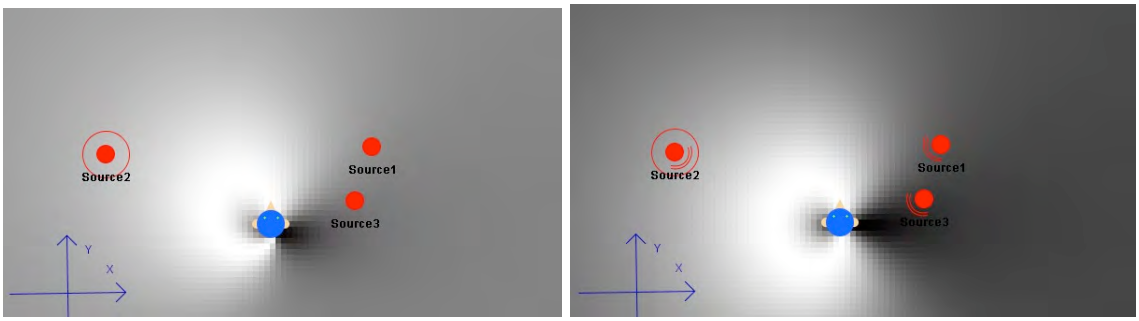


Figure 10: comparison of the effect of the panning technique over the spatial homogenity criterion
(XY panning law on the left, and cosinus panning law on the right).

The resulting area where the criterion is optimized is more symmetrical in the front / back positions and more lateralized in the cosinus law method than in the XY couple simulation. These results would differ with 5.1 rendering techniques (either ambisonics or pair wise panning).

2.3.3.2 Criterion #2.1: Spatial Masking Effect

Problem: the spatial coincidence of sound sources may induce a masking effect and / or a loss of intelligibility.

What is represented: in a shade of grey, for a given number of sound sources, and a given "target" source, we represent the risk for that target source to be masked by others.

Method: We consider the contribution in loudness of each sound source in the direction corresponding to the target point as described in Figure 11.

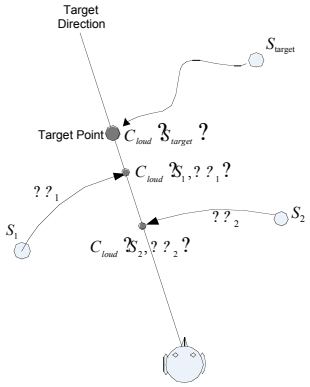


Figure 11: estimation of the loudness contribution of the sound sources at the target point

We then estimate that the target source will be masked if the difference between its loudness at the target point and the overall loudness from all other sources exceeds -10dB. This loudness difference is converted into a grey shade using a Gaussian function.

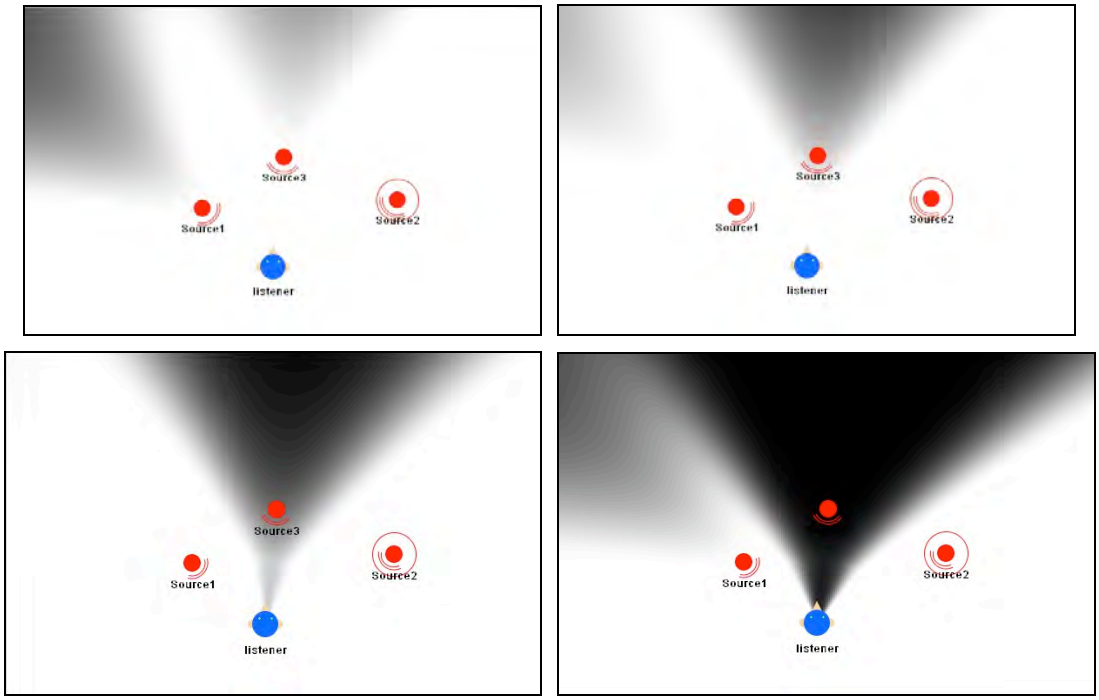


Figure 12: representation of the masking zones according to the loudness of the sound sources
(a) (upper left) E_{S1} = -10dB, E_{S2}=-10 dB, E_{S3}=-10dB, (b) (upper right) E_{S1} = -15dB, E_{S2}=-5 dB, E_{S3}=0dB
(c) (lower left) E_{S1} = -20dB, E_{S2}=-10 dB, E_{S3}=0dB, (d) (lower right) E_{S1} = -20dB, E_{S2}=-20 dB, E_{S3}=-0dB

The Figure 12 represents the "risk of being masked" criterion for Source2 according to the loudness value and the position of each sound source.

2.3.3.3 Criterion #2.2 Spatial Masking effect based on Zureck's Model

The criterion #2.1 has been refined in order to better correspond to the human auditory perception. The main improvement this model brings is the fact that it takes spectral aspects of the signal into account.

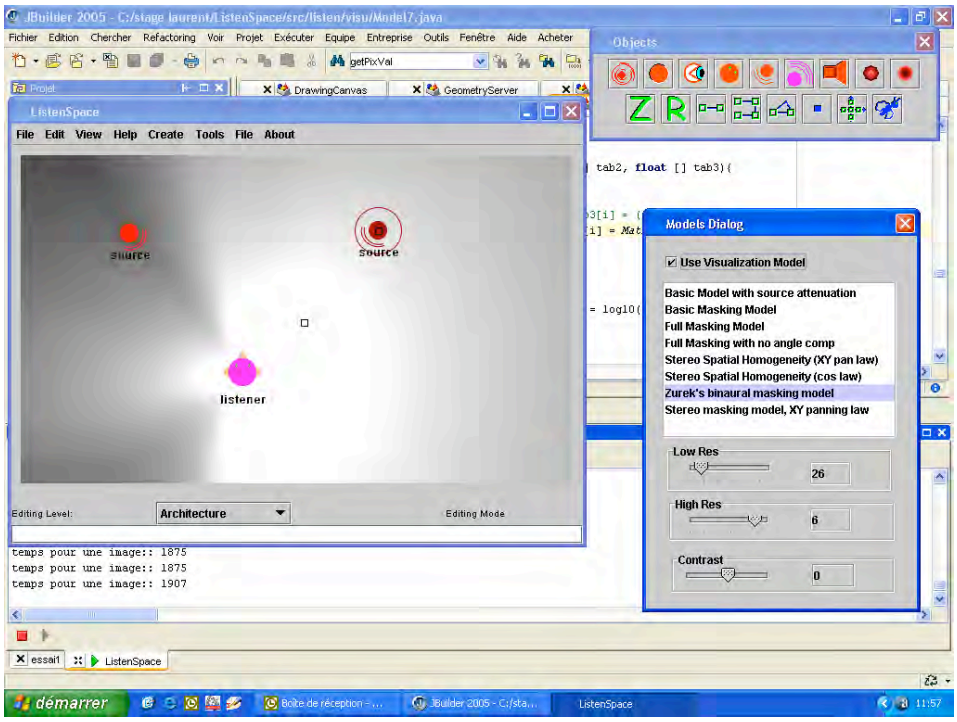


Figure 13 : view of the spatial masking criterion in ListenSpace, based on the Zureck's perceptual masking model.

These spectral properties of the signal need to be analyzed beforehand and stored as metadata that will be exploited on the fly during playback.

The two main references concerning these techniques are the following:

- Patrick M. Zurek. Acoustical factors affecting hearing aid performance, chapter Binaural advantages and directional effects in speech intelligibility. Allyn and Bacon, Boston, ii edition, 1993.
- Patrick M. Zurek, Richard L. Freyman, and Uma Balakrishnan. Auditory target detection in reverberation. J. Acoust. Soc. Am, 115(4) :1609-1620, April 2003.

2.3.3.4 Criterion #3: Crosstalk and source separation artifacts

Problem: multi-sources audio content coming from either an acoustical recording or from a source separation processing algorithm have crosstalk or separation artifacts properties. The correlated channels should remain spatialized in the same region of space in order to avoid either a loss of precision in the perception of source localization (in the case of crosstalk) or the apparition of audible source separation artifacts

What is represented: for a given target source, a light background represents the acceptable zones for interaction. Dark areas represent areas that might be avoided in order to present crosstalk or source separation artifacts from becoming hearable.

Some research have been started for this criterion but are still in progress.

2.3.4 Position over state-of-the-art

To our knowledge, no other application (at least in the field of control of the spatialization) allows showing to the user what would be the result of one of his actions. This type of representation is particularly suited to real-time applications where it is not possible to experiment "offline" and each action of the user has an immediate consequence. Thus, this project should find further application in the field of computer programming for live performances.

2.3.5 Benchmarks

The main issue for benchmarking is the reactivity of the system. Indeed some of the criterion can require time and CPU for calculation. A number of improvements have been brought to the system in order to guarantee the fluidity of the user interaction as well as the consistency of the information displayed. These improvements include mainly switching between two resolution modes according to the user's activity in order to enforce reactivity when the user is interacting with the system and improve the precision when the user is not interacting.

2.3.6 Implementation

So far, the model is implemented within the ListenSpace application. Final demonstration of the prototype will require that this background visualization can be streamed through the network and decoded within the Flash environment in order to be displayable on the remote Pocket-PC within the flash spatialization control interface.

2.3.7 Dissemination materials

2.3.7.1 Scientific publications

- Olivier Delerue, "*Visualization of perceptual parameters in interactive user interfaces: Application to the control of sound spatialization*", presented at the AES 120th Audio Engineering Society Convention, May 20th – 23rd 2006, Paris

2.3.7.2 Related scholar works

- Laurent Simon, "Dispositif d'Aide à la Spatialisation basé sur des critères perceptifs", Master's degree internship report.

2.4 Constraint based control of spatialization

Responsible partner: IRCAM-RA

2.4.1 Functional description

The MusicSpace application is a control interface for sound spatialization that describes a sound scene (sound sources and the listener's avatar) from a top view. The listener can at any time move the sound sources or its avatar (in order for example to get closer to an instrument he/she likes) and perceive the changes in the auditory scene in real time.

Moreover, a constraint solver algorithm and authoring tool is included to the system in order to specify relations between sound sources and / or the avatar, so that a naïve user (e.g. music amateur that has no particular knowledge in audio engineering) can perform actions on the sound scene while being guaranteed that the resulting 3D audio mix will remain consistent.

This project was originally a midi-based application (e.g. could only play midifiles and use very simple panning techniques to perform spatialization). A more advanced version allowed audio sources but was restricted to the use of the Microsoft DirectX mixing capabilities.

The context of the SemanticHIFI project was the opportunity to create a bridge between MusicSpace and the IRCAM Spatialisateur in order to take advantage of both advanced control possibilities using MusicSpace's constraints and high quality spatial audio rendering brought by the IRCAM Spatialisateur.

2.4.2 Method description

This prototype integrates three different software components:

1. A graphical user interface giving an overall representation of the sound scene and allowing user interaction.
2. A spatialization system performing real time signal processing operations on the different audio streams
3. A constraint solver that maintain and verifies in real time information on the sound scene expressed in the form of constraints.

A given set of constraints is provided in the form of metadata for each set of audio streams and ensures that the auditory results remain consistent, when a non expert user interact with the system.

2.4.3 Position over state-of-the-art

Applying constraint programming to control the spatialisation of sound sources is a concept that had never been exploited before. The challenge in this application is first to imagine constraints that allows the sound engineer to express conveniently the relations between the sound sources that should be verified. Another important aspect of this research consists in designing an efficient dedicated constraint solver that handles the particular context of non linear and non functional constraints, with possibly cycles in the constraint graph, and that remains reactive to reach the needs of real-time interaction.

2.4.4 Benchmarks

As mentioned above, the main difficulty lies in the reactivity of the system, and especially of the constraint solver. Our solver, adapted from existing local propagation algorithm proved to be reactive enough for the kind of application we address and for a reasonable number of constraints.

2.4.5 Implementation

Figure 14 describes both the graphical user interface MusicSpace (left) used in this example and the rendering engine (right) composed of a Max/MSP patch taking advantage of the Spatialisateur. The scene is composed of 3 sound sources (piano / bass / drums) and a number of constraints represented as colored spheres linked to the constrained sources.

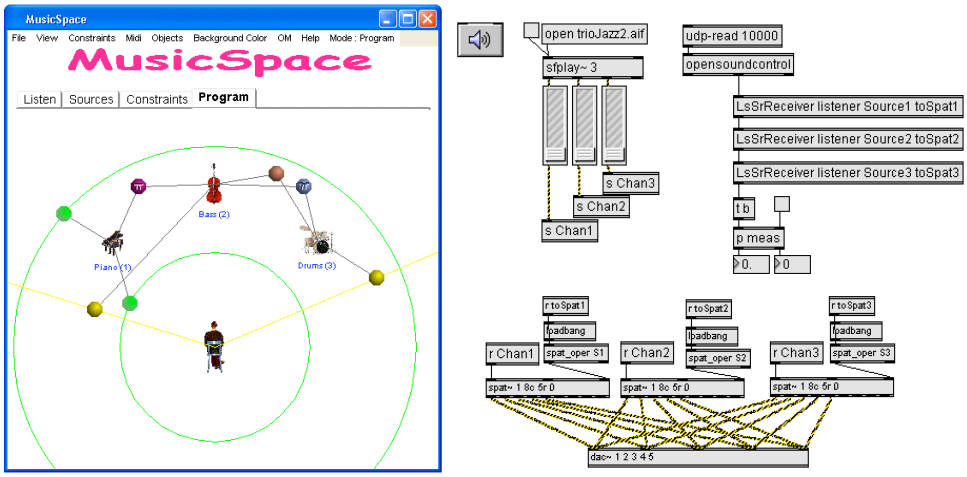


Figure 14: overview of the MusicSpace graphical user interface (left) and the corresponding implementation of the rendering engine in the Max/MSP environment (right)

2.4.6 Dissemination materials

2.4.6.1 Related PhDs

- Olivier Delerue. Spatialisation et programmation par contraintes : le système MusicSpace. PhD thesis, Université Paris 6, 2004.

2.5 OpenAL SoundServer & Flash remote application

Responsible partner: IRCAM-RA

2.5.1 Functional description

The goal of this part of the work is to port the spatialization prototypes to the final HIFI system for demonstration purpose. This work requires a special task since the final architecture of the systems is based on an environment that does not allow spatialization.

Thus, a dedicated sound server based on the OpenAL API has been built. The OpenAl library is compatible with the linux distribution chosen for the HIFI system. The server is written in Java using the JAOL binding for OpenAL. This server is capable of playing simultaneously several audio streams (e.g. multitrack audio content) and of performing spatialization on these streams. The quality of the spatialization is dependent on the OpenAL implementation used and on possible hardware acceleration provided by the sound card.

So far, the basic OpenAL implementation is used, but the audio rendering may be improve by replacing this library by one that would implement the IRCAM-Spat features.

The audio server is a command line tool with no user interface. However, a flash interface has been created in order to control the position of the sound sources in the audio rendering. It is This interface represents a basis for implementing the various prototypes (such as constraint based spatialization or visualization tool for controlling the spatizalisation as described in the previous sections of this document).

2.5.2 Position over state-of-the-art

This part of the work correspond to the porting of other research or prototypes to the final HIFI system platform. Therefore there is no positioning over the state of the art.

2.5.3 Benchmarks

The sound server has been tested successfully with multi-track audio content containing up to 12 channels without any problems.

2.5.4 Implementation

The audio server is implemented in java. It makes use of the Java binding for OpenAL as well as of the OpenAL library. The user interface is implemented in the Flash langage. These different software components communicate using network connection that can be typically wireless connections.

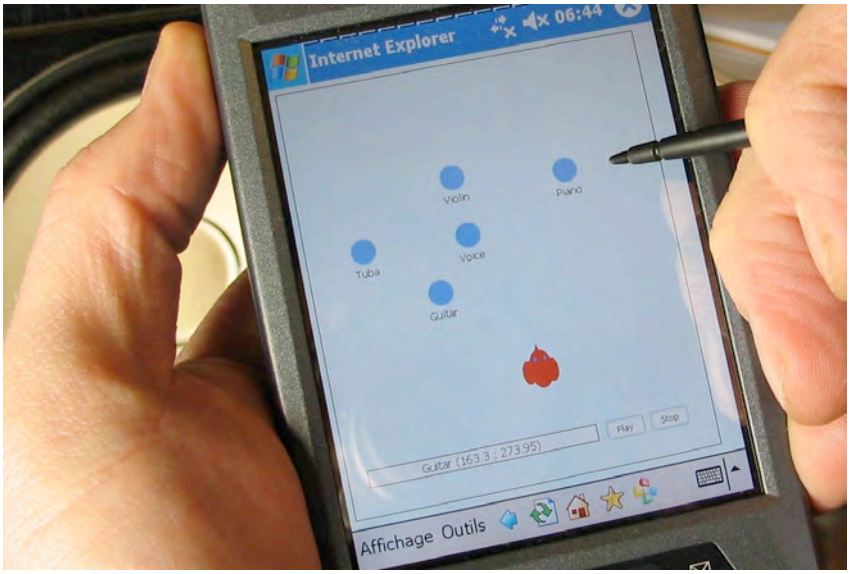


Figure 15: view of the flash interface for controlling the localization of sound sources.

2.5.5 Dissemination materials

2.5.5.1 Professional & Public communications

This work has not been presented yet outside of the project but will be specifically dedicated to public and professional presentation.

2.6 "Cera-Player"

Responsible partner: IRCAM Studio Hypermedia & IRCAM-RA.

2.6.1 Functional description

The application proposed consists in a "step sequencer" playing several sounds extracted (using filtering method and time segmentation) from a given original audio sample and reconstructed synchronously along the original sample. This example was integrated to the demonstrating system by transforming the main module into the form of a specific "player" module. A corresponding application of the Hifi System was created by associating to this

player a number of spatialization modules and choosing MusicSpace to control the spatialization of each instrument stream.

2.6.2 Method description

The originality in this prototype lies in the fact that the music, when delivered to the listener, is not "pre-recorded" as in actual audio CDs. In the case of the Cera Player, part of the music is predefined and part of it comes as a set of samples that will be arranged and played in real-time according to the interaction with the listener. This concept of synthesizing the music 'on the fly' pushes further the idea of "active listening" already brought up in the MusicSpace project.

Andrea Cera is an Italian composer. Having worked on a regular basis in commercial music, he went on to write various compositions, which can be situated in the field of contemporary music and sound installation. It is his belief, that dance music puts a particularly strong emphasis on rhythmical complexion. When faced with a rhythmical singularity, Andrea Cera seeks to obtain an analytical understanding to enable him to go beyond an unexplored fascination (and possibly to integrate this rhythmical singularity into his own work as a composer). To understand such rhythmical complexity, Andrea transcribes his listening onto a medium, allowing a more detailed investigation of the loop. Using the transcription as the basis for a more detailed study reveals some remarkable phenomena, via the addition of different layers. Suddenly dactyls or 'grooves' become apparent, which were impossible to hear a priori and which only become visible when laid out on the page or screen. However, these analytical sketches make the composer aware of them, thus allowing him to recompose his listening, provoking a less compulsive and more open fascination.

This listening practice has been analysed in order to propose the integration of an application into the Semantic HiFi system. In this context, we worked with Andrea Cera on the demix of 2 different loops: "Krupa-Appolo4-40" and "Machine Gun – Kelly".

2.6.3 Position over state-of-the-art

In the current state of our knowledge, there has been no other attempt of developing an hypermedia analyses the internal structures of techno music.

2.6.4 Implementation

The application comes as a Max/MSP patch connected to the MusicSpace software. The user can cross the transcription of one loop with the songs extracted from the other loop. He can also modify the transcription using the step sequencer.

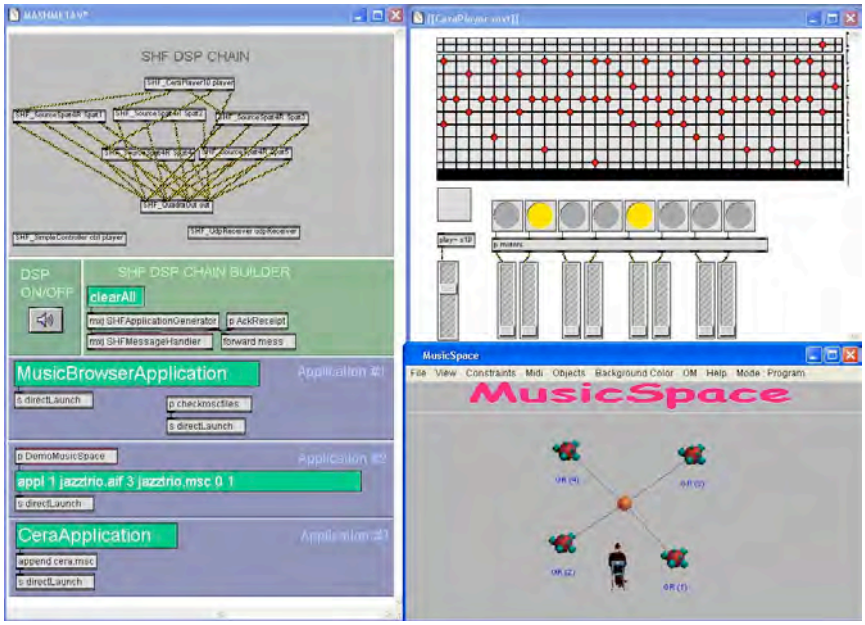


Figure 16: view of the integrated prototype for the "Kupra Appolo 4-40" example.

2.7 Hypermedia Analysis of Webern Op.5 N°11

Responsible partner: IRCAM- Studio Hypermedia

2.7.1 Functional description

The hypermedia analysis of Webern Op.5 n°II is a flash prototype available integrated on the touch screen of the semantic Hi-fi or standalone on a PDA. It presents two score-based listening guides of a Webern piece based on the score written by Nicolas Donin, musicologist and researcher at Ircam.

The user can play the music track and follow it on the score, where some annotations guide its listening. Due to the small size of the touch screen on the Semantic Hi-fi system or the screen of a PDA, the analysis is entirely based on the usage of shapes and colors for the different annotations. The text is only present as a title for these two analyses.

The song and the score are linked by a cursor, but at any time, the user can:

- Navigate through the score with drag and drop,
- Reach other time code in the song with a click on the timeline.

While listening to the piece, the user can compare the two analyses, switching from one to another without the sound stops or the position of the score changes.

There are two types of annotation. Ones that are always represented on the score and others that appear and disappear at a time code, not necessarily when the cursor reaches it. This allows the author to help the user anticipate some song structure, important phrases, breaks …

2.7.2 Method description

The hypermedia analysis has been developed with the help of Nicolas Donin using Flash Macromedia Professional 8.

The first step was the multimedia synchronization of the score and the music track. After, annotations were added to the score for each of these two analysis.

All the material is embedded in a single Flash swf file.

2.7.3 Position over state-of-the-art

The APM team at Ircam has already experiment hypermedia score annotations. One of them is available online:

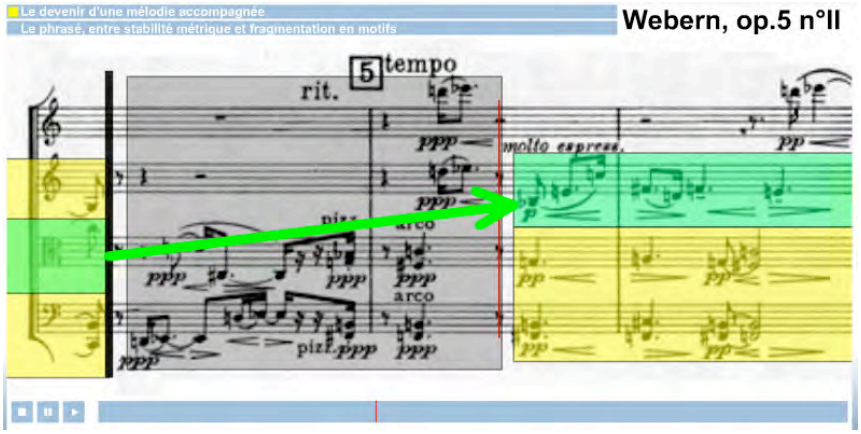
« Samson François jouant *Noctuelles* : notes de lecture », *DEMeter. Revue électronique du Centre d'Etude des Arts Contemporains – Université de Lille-3*, août 2005, accès *via* http://www.univ-lille3.fr/revues/demeter/

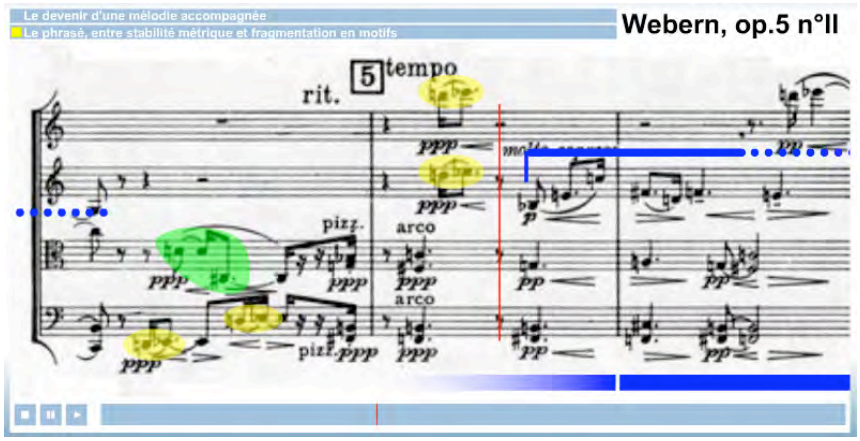
Also, INA-GRM has developed a sonogram annotation tool. Sonogram annotation is different than a score annotation, because of a horizontal temporal aspect of a sonogram.

Some examples are available online:
http://www.ina.fr/grm/acousmaline/polychromes/index.fr.html

Our interface is designed to be published on small screen size contrary to the 2 examples above.

2.7.4 Implementation





These two screenshots represent the two analyzes of the Webern piece, at the same time code in the song. On the top, the user can select the analyze he wants to visualize. In the middle of the interface, there is the score synchronized with the audio file. At the bottom, the control bar allows the user to play, pause, stop or reach any time code of the song.

Figure 17: Photo of the PDA with the Webern op 5 n°II analyze

2.7.5 Dissemination materials

2.7.5.1 Press articles and interviews

Nicolas Donin, « Vers l'annotation multimédia d'informations musicales », *L'inouï, revue de l'Ircam*, n° 2, 2006, p. 160-163.

2.8 Virtual Mixer

Responsible partner: BGU

2.8.1 Functional description

The objective of demixing modules is to enrich existing audio content by trying to reconstitute the original separated sound sources. This audio content can then be remixed at the rendering stage and, to some extent, adapted to the listener's preferences using applications such as MusicSpace.

The work done in Virtual mixing was mainly scientific work about the theory of the harmonic filter. This filtering problem has been formulated as a linear constraint minimum problem from which was derived the filter equation.

2.8.2 Position over state-of-the-art

The source separation problem is well known and constitutes the objective of many research projects. Due to its complexity, this problem is far from being solved. However, for remixing application, source separation does not need to be perfect and the extracted content can be remixed under certain conditions without making the source separation artifacts audible.

Despite the quality of the source separation itself, one of the main challenges in this application lies in its combining with a remixing tool such as MusicSpace and the Spatialisateur.

Thus, at the development of the source separation itself, our concern is to express (in the form of medatata along with the audio signal) to which extent the sources can be re-spatialized without the apparition of source separation artifacts (see section 2.3).

2.8.3 Implementation

This project comes in the form of a toolbox for signal processing that allows creating and adapting audio content specifically for the SHF Hifi System.

2.8.4 Dissemination materials

2.8.4.1 Scientific publications

- O. Hadar, D. Bykhovsky, G. Goldwasser, and E. Fisher, "A musical source separation system with lyrics alignment" in *WSEAS Transactions on Systems*, Issue 10, Vol. 5, pp. 2464-2467 October 2006

2.9 Additional Work

The last section of this document describes general achievements that were necessary for most of the prototypes presented previously. This work concerns mainly the development of general signal processing libraries for spatialization, the creation of suitable audio content and collaborative work concerning the general architecture of the system.

2.9.1 Audio Content Creation

Relevant audio examples have been created in order to experiment and demonstrate the functional prototypes in the rendering context.

- a Jazz trio (piano – bass and drums) example with three different constraint sets and a Tango example (three guitars and a singer) with two different constraint sets and karaoke capabilities were created to illustrate the constraint based spatialization prototype.

- a set of 6 acoustical examples containing typical crosstalk has been prepared in order to present the work about visualization. These examples are issued from a recording provided (for the project internal use only) by the Music Conservatory in Paris. The separated audio tracks were extracted from the original format, trimmed and converted into an interlaced audio format in order to be adapted to the rendering system playback modules.

The detail of audio extracts is as follow:

1: Jazz Trio (saxophone, bass, drums), 9 channels, 5’

2: Raga (sitar & double bass), 4 channels, 16’40’’

3: Contemporary saxophone duo, 3 channels, 10’

4: Jazz piano solo, 3 channels, 3’50’’

5: Jazz Quintet (saxophone, trumpet, piano, bass, drums), 12 channels, 13’40’’

6: Jazz Quartet (2 guitars, bass, drums), 9 channels, 3’45’’

2.9.2 Architecture description, Implementation & related developpements

This section provides details on the system architecture and the developments that were necessary in order to complete the implementation of the functional prototypes.

2.9.2.1 System Architecture:

The first demonstration prototypes were conceived as "standalone" applications and did not render of the possible transitions between specific application of the HIFI system (browsing, playing, performing,…). Moreover, it was not possible to switch dynamically from one to the other as would have to do a HIFI system. In this third step, we focused our research on a unified architecture description and proposed an implementation in the Max/MSP environment. Together with these works on general architecture and DSP components, research has been continued in the field of user interfaces in order to create convenient means for controlling the spatialization of sound sources.

A number of internal meetings have been held at IRCAM concerning integration aspects and the diagram presented on Figure 18 has been submitted as a proposition of the overall system architecture in the case of a user listening to a multi-sources audio file and using a microphone input for karaoke-like performing applications. This diagram allows identifying the "play back / rendering engine" as well as the various interactions that this component has with the rest of the system. Generally, the playback & rendering engine will consist of a number of "player" modules (e.g. modules that output audio streams coming from various sources such as files from hard disk, or removable storage), a number of "Audio-Fx" modules, typically for each audio stream coming from the player section, and finally a dedicated spatialization module for each audio stream coming from the "audio-fx" section.

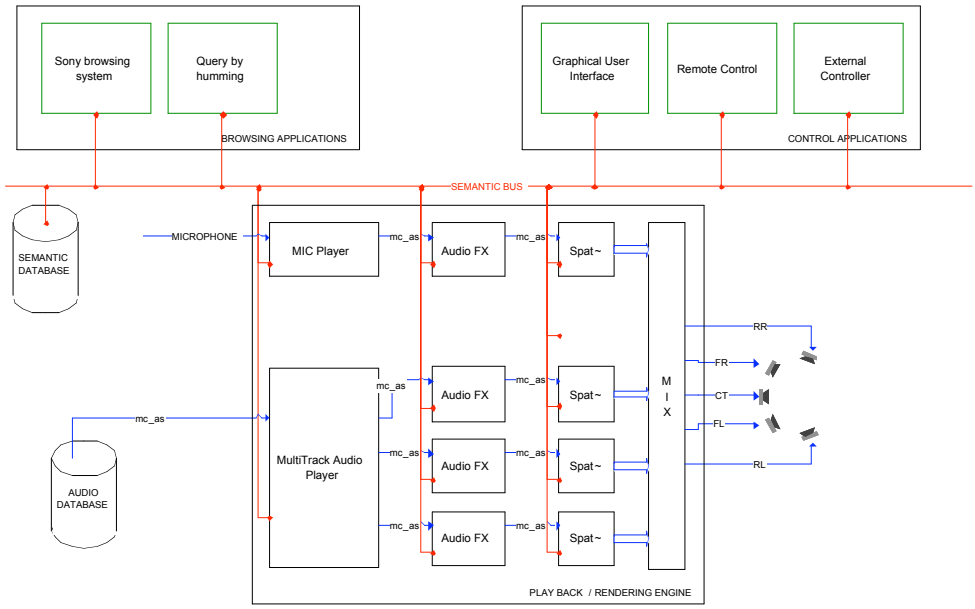


Figure 18: typical overall architecture example

2.9.2.2 Integration implementation:

An integration framework has been proposed and implemented in the Max/MSP environment. This work consisted in the realization of a Max/MSP "meta-patch", in charge of constructing dynamically the playback / rendering engine as well as the graphical user interfaces elements according to the audio content and the associated metadata. This patch, as represented on

Figure 19 is composed of two modules linked together: a SHFApplicationGenerator and a SHFMessageHandler. Both of these modules were constructed on the basis of Max / MSP Java external objects.

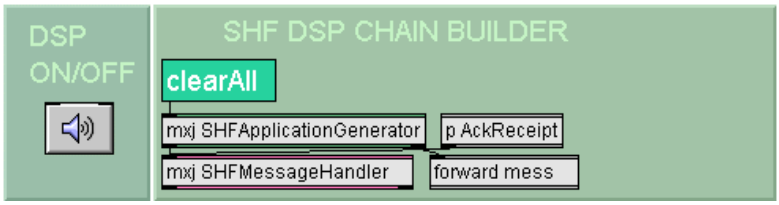


Figure 19: the SHF Meta application generator

The **SHFMessageHandler** is in charge of handling low level commands concerning the dynamic creation, deletion, and connection of signal processing or control interface modules. Thus, it ensures that all modules have been successfully and entirely instantiated before connecting them together and starting the DSP chain. Typically, if a connection command is sent before the corresponding modules are instantiated, this command will be buffered until it can be successfully handled.

The **SHFApplicationGenerator** takes as input the result of a browsing operation specifying what audio content the user wants to listen to, and how this content should be played and decides the list of signal processing modules to be used and their interconnection scheme. An example of hi-level command that the "SHFApplicationGenerator" can handle is illustrated in Figure 20.

Figure 20: example of a typical hi-level command resulting from a browsing operation.

In this typical example, the command line "appl 1 tango.aif 4 tangokaraoke.msc 1 1" should be the result of the "browsing" operation and signifies that the user wants to listen to the "tango" song, using the MusicSpace application and singing along as in a karaoke situation. It specifies as well that the loudspeaker setup is in a quadraphonic configuration.

The resulting "DSP chain", created on the fly, is represented in Figure 21: it is composed of two different "players", a "MultiTrack player" that reads an AIFF multi-track audio file and generates the corresponding audio streams and a "Karaoke player" that uses the analog audio input of the computer (e.g. the microphone) to create a corresponding audio stream. Each output of these players are connected to a dedicated spatialization module, and finally each spatialization module is connected to a general "output module" adapted accordingly to the rendering setup that has been chosen (quadraphonic speakers in this example).

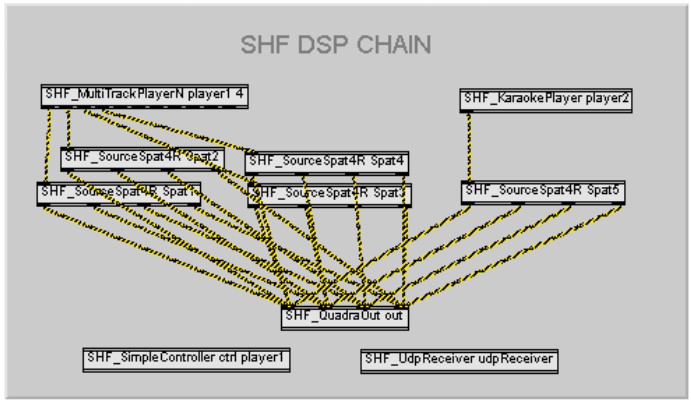


Figure 21: view of the corresponding playback / rendering engine constructed

As for the user control interface, the MusicSpace application is launched with the "tangokaraoke.msc" constraint set file (see Figure 22). It represents the different sound sources of the audio scene as well as a microphone icon corresponding to the microphone input of the HiFi system: constraints have been set between the original singer of the recording and the microphone input in order to create an exclusive relation between these two streams.

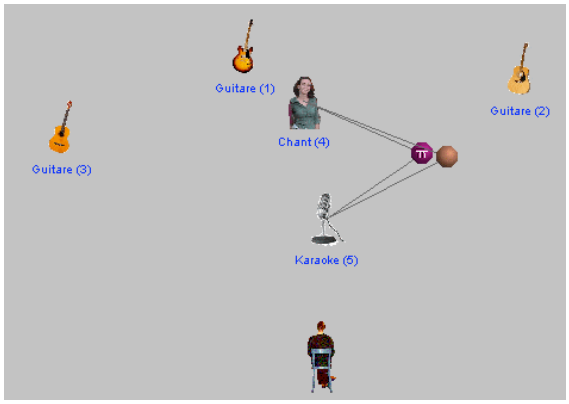


Figure 22: view of the corresponding scene in MusicSpace.