Panoramix: 3D mixing and post-production workstation

Thibaut Carpentier

Panoramix: 3D mixing and post-production workstation

Thibaut Carpentier
UMR 9912 STMS IRCAM – CNRS – UPMC
1, place Igor Stravinsky, 75004 Paris
`thibaut.carpentier@ircam.fr`

ABSTRACT

This paper presents panoramix, a post-production workstation for 3D-audio contents. This tool offers a comprehensive environment for mixing, reverberating, and spatializing sound materials from different microphone systems: surround microphone trees, spot microphones, ambient miking, Higher Order Ambisonics capture. Several 3D spatialization techniques (VBAP, HOA, binaural) can be combined and mixed simultaneously in different formats. Panoramix also provides conventional features of mixing engines (equalizer, compressor/expander, grouping parameters, routing of input/output signals, etc.), and it can be controlled entirely via the Open Sound Control protocol.

1. INTRODUCTION

Sound mixing is the art of combining multiple sonic elements in order to eventually produce a master tape that can be broadcast and archived. It is thus a crucial step in the workflow of audio content production. With the increasing use of spatialization technologies in multimedia creation and the emergence of 3D diffusion platforms (3D theaters, binaural radio-broadcast, etc.), new mixing and post-production tools become necessary.

In this regard, the post-production of an electroacoustic music concert represents an interesting case study as it involves various mixing techniques and raises many challenges. The mixing engineer usually has to deal with numerous and heterogeneous audio materials: main microphone recording, spot microphones, ambient miking, electronic tracks (spatialiazed or not), sound samples, impulse responses of the concert hall, etc. With all these elements at hand, the sound engineer has to reproduce (if not re-create) the spatial dimension of the piece. His/her objective is to faithfully render the original sound scene and to preserve the acoustical characteristics of the concert hall while offering a clear perspective on the musical form. Most often the mix is produced from the standpoint of the conductor as this position allows to apprehend the musical structure and provides an analytic point of view which conforms to the composer's idea.

Obviously, the sound recording made during the concert is of tremendous importance and it greatly influences the post-production work. Several miking approaches can be used

(spaced pair, surround miking, close microphones, etc.), and advantages and drawbacks of each technique are well known (see for instance [1–4]). For instance when mixing Pierre Boulez's *Répons*, Lyzwa emphasized that multiple miking techniques had to be combined in order to benefit from their complimentarity [5]: a main microphone tree (e.g. surround 5.0 array) captures the overall spatial scene and provides a realistic impression of envelopment as the different microphone signals are uncorrelated; such a system is well suited for distant sounds and depth perception. However the localization of sound sources lacks precision, and thus additional spot microphones have to be used, close to the instruments. During post-production, these spot microphones have to be re-spatialized using panning techniques. Electronic tracks, if independently available, have to be processed similarly. Finally the sound engineer can add artificial reverberation in the mix in order to fuse the different materials and to enhance depth impression.

In summary, the mixing engineer's task is to create a comprehensive sound scene through manipulation of the spatial attributes (localization, immersion, envelopment, depth, etc.) of the available audio materials. Tools used in the post-production workflow typically consist of: a mixing console (analog or digital), digital audio workstations (DAWs) and sound spatialization software environments.

The work presented in this article aims at enhancing existing tools especially in regard to 3D mixing wherein existing technologies are ill-suited. Mixing desks are usually limited to conventional panning techniques (time or intensity differences) and they do not support 3D processing such as binaural or Ambisonic rendering. They are most often dedicated to 2D surround setups (5.1 or 7.1) and they do not provide knob for elevation control. Similarly, digital audio workstations lack flexibility for multichannel streams: most of the DAWs only support "limited" multichannel tracks/busses (stereo, 5.1 or 7.1) and inserting spatialization plugins is difficult and/or tedious. On the other hand, many powerful sound spatialization engines are available. As shown in [6] and other surveys, a majority of these tools are integrated into realtime media-programming environments such as Max or PureData. Such frameworks appear inadequate to post-production and mixing as many crucial operations (e.g. group management or dynamic creation of new tracks) can hardly be implemented. Furthermore, spatialization libraries are generally dedicated to one given rendering technique (for instance VBAP [7] or Higher-Order Ambisonic [8]) and they are ill-suited to hybrid mix.

Finally, high-spatial resolution microphones such as the EigenMike [1] are essentially used in research labs but they

---

[1] `http://www.mhacoustics.com`

remain under-exploited in actual production contexts, in spite of their great potential.

As a consequence, we have developed a new tool which provides a unified framework for the mixing, spatialization and reverberation of heterogeneous sound sources in a 3D context.

This paper is organized as follows: Section 2 presents the process of recording an electroacoustic piece for use in 3D post-production. This paradigmatic example is used to elaborate the specifications of the new mixing engine. Section 3 details the technical features of *panoramix*, the proposed workstation. Finally Section 4 outlines possible future improvements.

2. PARADIGMATIC EXAMPLE

2.1 Presentation

Composer Olga Neuwirth's 2015 piece *Le* Encantadas *o le avventure nel mare delle meraviglie*, for ensemble and electronics [2] serves as a useful case study in 3D audio production techniques. The piece had its French premiere on October 21st in the Salle des Concerts de la Philharmonie 2 (Paris), performed by the Ensemble intercontemporain with Matthias Pintscher conducting. As is often the case in Neuwirth's work, the piece proposed a quite elaborate spatial design, with the ensemble divided in six groups of four or five musicians. Group I was positioned on-stage, while groups II to VI were dispatched in the balcony, surrounding and overlooking the audience (cf. Figure 1). The electronic part combined pre-recorded sound samples and real-time effects, to be rendered over a 40-speaker 3D dome above the audience. Different spatialization approaches were employed, notably Higher-Order Ambisonic (HOA), VBAP, and spatial matrixing. Throughout the piece, several virtual sound spaces were generated by means of reverberators. In particular, high-resolution directional room impulse responses, measured with an EigenMike microphone in the San Lorenzo Church (Venice), were used in a 4th order HOA convolution engine in order to simulate the acoustics of the church – as a reference to Luigi Nono's *Prometeo*.
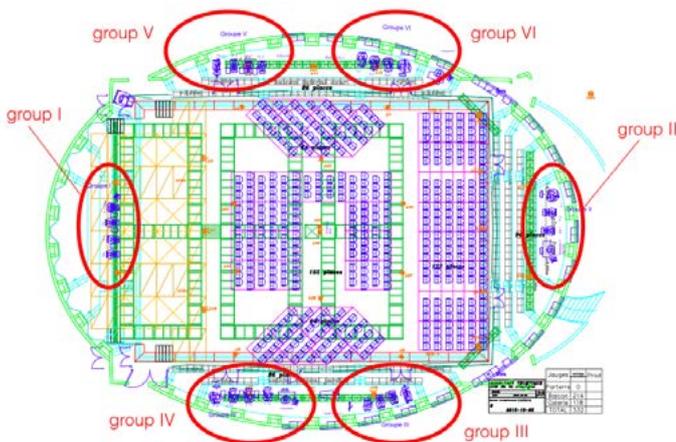

Figure 1. Location of the six instrumental groups in the Salle des Concerts – Philharmonie 2, Paris.

[2] Computer music design: Gilbert Nouno / Ircam

group	instruments	miking
I	saxophone, trumpet 1, bassoon, electric guitar	4 microphones: AT4050, AKG214, C535, AKG214
II	synthesizer 1, clarinet 1, trumpet 2, cello 1	5 microphones: KMS105, DPA4021, AKG214, KM140, AKG411
III	flute 1, oboe, french horn 1, trombone 1, percussion 1	11 microphones: DPA4066, KM150, C353, KM150, Beta57, SM58 (x2), SM57 (x2), C535, AKG411
IV	synthesizer 2, violin 3, violin 4, viola 1, cello 2	5 microphones: DPA4061 (x3), DPA2011, KM140
V	percussion 2, trombone 2, french horn 2, clarinet 2, flute 2	10 microphones: SM57 (x2), SM58 (x2), MD421, C535, Beta57, KMS105, AKG414 (x2), DPA4066
VI	synthesizer 3, violin 1, violin 2, viola 2, double bass	10 microphones: DPA4061 (x3), AKG414 (x4), KM140, C535 (x2), SM58 (x2)

Table 1. Spot microphones used for the recording.

2.2 Sound recording

Given the spatial configuration of the piece, the recording session [3] involved a rather large set of elements:

- 45 close microphones for the six instrumental groups (see Table 1),
- distant microphones for capturing the overall image of the groups: spaced microphones pairs for groups I and II; omni-directional mics for the side groups,
- one EigenMike microphone (32 channels) in the middle of the hall, i.e. in the center of the HOA dome,
- one custom 6-channel surround tree (see [5]) also located in the center of the hall,
- 32 tracks for the electronics (30 speaker feeds plus 2 subwoofers),
- direct capture of the 3 (stereo) synthesizers as well as 3 click tracks.

In total, 132 tracks were recorded with two laptop computers (64 and 68 channels respectively) which were later re-synchronized by utilizing click tracks.

2.3 Specifications for the post-production workstation

In spite of its rather large scale, this example of recording session is representative of what is commonly used in the electroacoustic field, where each recorded element requires post-production treatment. As mentioned in the introduction, various tools can be used to handle these treatments, however there is yet no unified framework covering all the required operations.

Based on the example of *Encantadas* (and others not covered in this article), we can begin to define the specifications for a comprehensive mixing environment. The workstation should (at least) allow for:

- spatializing monophonic sound sources (spot microphones or electronic tracks) in 3D,
- adding artificial reverberation,
- encoding and decoding of Ambisonic sound-fields (B-format or higher orders),
- mixing already spatialized electronic parts recorded as speaker feeds,

[3] Sound recording: Ircam / Clément Cornuau, Mélina Avenati, Sylvain Cadars

- adjusting the levels and delays of each elements so as to align them,
- combining different spatialization approaches,
- rendering and exporting the final mix in several formats.

With these specifications in mind, we developed *panoramix*, a virtual mixing console which consists of an audio engine associated with a graphical user interface for controlling/editing the session.

3. PANORAMIX

Like a traditional mixing desk, the *panoramix* interface is designed as vertical strips depicted in Figure 3. Strips can be of different types, serving different purposes with the following common set of features:

- multichannel vu-meter for monitoring the input level(s),
- input trim,
- multichannel equalization module (where the EQ is applied uniformly on each channel). The equalizer comes as a 8-stage parametric filter (see ⑥ in Figure 3) with one high-pass, one low-pass (Butterworth design with adjustable slope), two shelving filters, and four second-order sections (with adjustable gain, Q and cutoff frequency),
- multichannel dynamic compressor/expander (Figure 2) with standard parameters (ratio, activation threshold, and attack/release settings),
- mute/solo buttons,
- multichannel vu-meter for output monitoring, with a gain fader.

In addition, a toolbar below the strip header (Figure 3), allows for the configuration of various options such as locking/unlocking the strip, adding textual annotations, and configuring the vu-meters (pre/post fader, peakhold), etc. Strips are organized in two main categories: input tracks and busses. The following sections describe the properties of each kind of strip.

3.1 Input tracks

Input tracks correspond to the audio streams used in the mixing session (which could be real-time or prerecorded). Each input track contains a delay parameter in order to re-synchronize audio recorded with different microphone systems. For example, spot microphones are recorded close to the instruments and so their signals arrive faster than microphones placed at greater distances. Miking a sound source with multiple microphones is also prone to tone coloration; adjusting the delay parameter helps reducing this coloration and can also be used to vary the sense of spatial envelopment. In practice, it can be effective to set the spot microphones to arrive slightly early, to take advantage of the precedence effect which stabilizes the perceived location of the combined sound.

3.1.1 Mono Track

A Mono Track is used to process and spatialize a monophonic signal, typically from a spot microphone or an electronic track. The strip provides controls over the localization attributes (azimuth, elevation, distance), spatial effects (Doppler, air absorption filtering) and reverberation. The artificial reverberation module is derived from the Spat architecture [9] wherein the generated room effect is composed
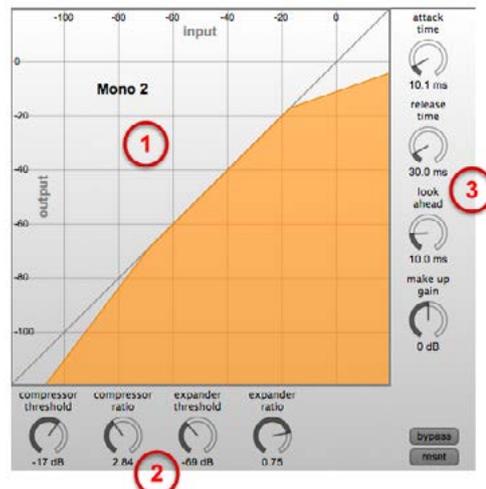

Figure 2. Compressor/expander module. ① Dynamic compression curve. ② Ratios and thresholds. ③ Temporal characteristics.

of four temporal sections: direct sound, early reflections, late/diffuse reflections and reverberation tail. By default the Spat perceptual model is applied, using the source distance to calculate the gain, delay, and filter coefficients for each of the four temporal sections. Alternatively, the perceptual model can be disabled (see slave buttons ③ in Figure 4) and the levels manually adjusted. Each temporal section may also be muted independently. In the signal processing chain, the extended direct sound (i.e. direct sound plus early reflections) is generated inside the mono track (Figure 7), while the late/diffuse sections are synthesized in a reverb bus (described in 3.2.2) which is shared among several tracks in order to minimize the CPU cost. Finally, a drop-down menu ("bus send") allows one to select the destination bus (see 3.2.1) of the track.

Moreover all mono tracks are visualized (and can be manipulated) in a 2D geometrical interface (⑦ in Figure 3).

3.1.2 Multi Track

A Multi Track is essentially a coordinated collection of mono tracks, where all processing settings (filters, reverberation, etc.) are applied similarly on each monophonic channel. The positions of each of the mono elements are fixed (i.e. they are set once –via the "Channels..." menu– for the lifetime of the session). Such Multi Track is typically used to process a multichannel stream of speaker feeds signals (see paragraph 2.3).

Similar results could be obtained by grouping (see 3.5) multiple "Mono" tracks, however "Multi" tracks make the configuration and management of the session much more simple, rapid and intuitive.

3.1.3 EigenMike Track

As its name suggests, an "EigenMike" Track is employed to process recordings made with spherical microphone arrays such as the EigenMike. Correspondingly, the track has 32 input channels and it encodes spherical microphone signals in the HOA format. Encoding can be performed up to 4th order, and several normalization flavors (N3D, SN3D, FuMa, etc.) are available.

Modal-domain operators can later be applied to spatially transform the encoded sound-field, for example rotating the
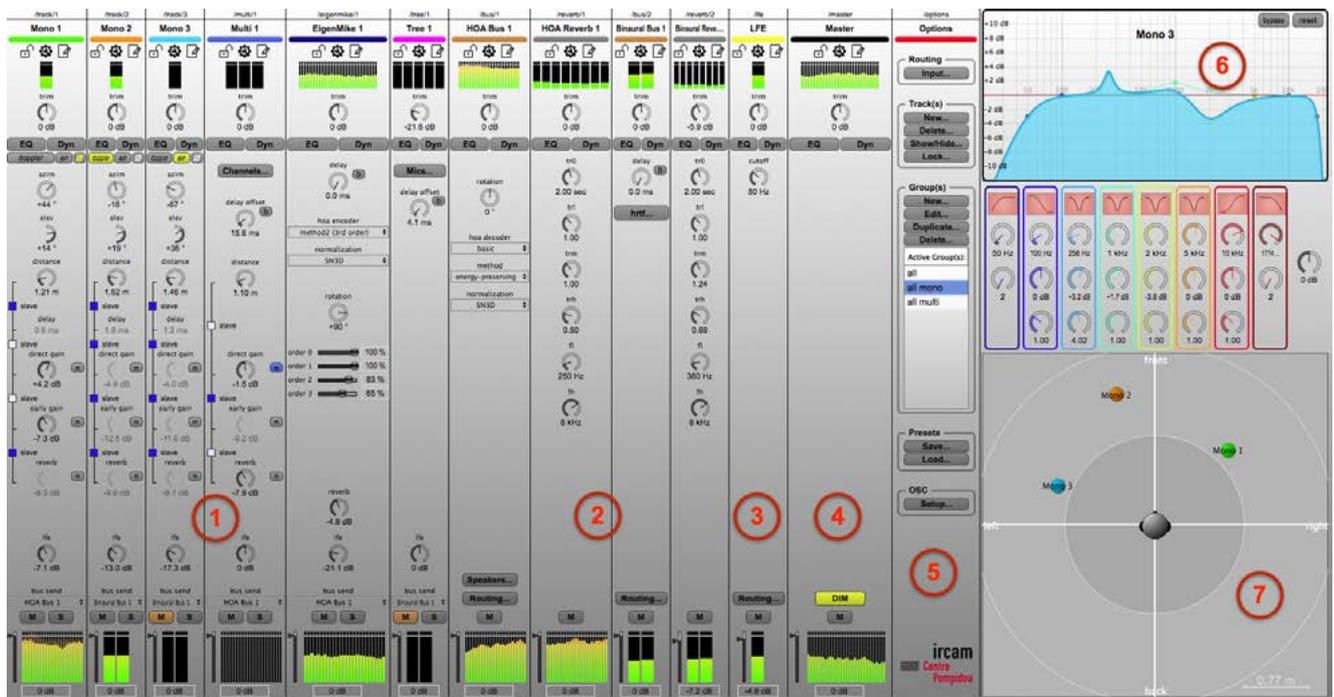
Figure 3. Main interface of the *panoramix* workstation. ① Input strips. ② Panning and reverb busses. ③ LFE bus. ④ Master track. ⑤ Session options. ⑥ Insert modules (equalizer, compressor, etc.). ⑦ Geometrical interface for positioning.

whole sound scene, or weighting the spherical harmonics components (see ④ in Figure 4).

Signals emanating from an EigenMike recording are already spatialized and they convey the reverberation of the concert hall, however a reverb send parameter is provided in the track, which can be useful for adding subtle artificial reverberation, coherent with the other tracks, to homogenize the mix. The reverb send is derived from the omni component (W-channel) of the HOA stream.

3.1.4 Tree Track

A "Tree" track is used to accommodate the signals of a microphone tree such as the 6-channel tree installed for the recording of *Encantadas* (section 2.2). The "Mics..." button (cf. Track "Tree 1" in Figure 3) pops up a window for setting the positions of the microphones in the tree. It is further possible to align the delay and level of each cell of the microphone array.

As microphone trees entirely capture the sound scene, the "Tree" track does not apply any specific treatment to the signals.

3.2 Busses

Three types of bus are provided: panning busses, reverb busses, and one LFE ("low frequency enhancement") bus.

3.2.1 Panning/Decoding bus

The role of panning busses is threefold: 1) they act as summing busses for the track output streams; 2) they control the spatialization technique in use (three algorithms are currently supported: VBAP, HOA and binaural); 3) panning busses are used to control various parameters related to the encoding/decoding of the signals. For speaker-based rendering (VBAP or HOA), the "Speakers..." button allows for the configuration of the speakers layout (Figure 6); in

case of binaural reproduction, the "hrtf..." button provides means to select the desired HRTF set. Finally, HOA panning busses decode the Ambisonic streams, and several decoding parameters can be adjusted (see "HOA Bus 1" in Figure 3).

The selection of rendering techniques (VBAP, HOA, binaural) was motivated by their ability to spatialize sounds in full 3D, and their perceptual complementarity. Other panning algorithms may also be added in future versions of *panoramix*.

Output signals from the panning busses are sent to the Master strip. Each panning bus provides a routing matrix so as to assign the signals to the desired destination channel (② in Figure 5).

3.2.2 Reverberation bus

Reverberation busses function to synthesize the late/diffuse sections of the artificial reverberation processing chain. A reverb bus is uniquely and permanently attached to one or more panning buses, where the reverberation effect is applied to each track routed to this bus.

Panoramix builds on the reverberation engine of Spat which consists of a feedback delay network with an variable decay profile, adjustable in three frequency bands. The main parameters of the algorithm are exposed in the reverb strip (see ⑤ in Figure 4).

3.2.3 LFE Bus

Each track has a LFE knob to tune the amount of signals sent to the LFE bus which handles the low-frequency signals sent to the subwoofer(s) of the reproduction setup. The bus applies a low-pass filter with adjustable cutoff frequency.
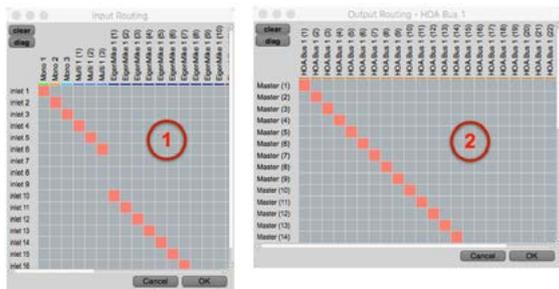
3.3 Master

The "Master" strip collects the output signals of all the busses and forwards them to the *panoramix* physical outputs.

Although the workstation only has one Master strip, it is possible to simultaneously render mixes in various formats. For instance, if the session has 26 physical output channels, one can assign channels 1–24 to an Ambisonic mix and channels 25–26 to a binaural rendering.
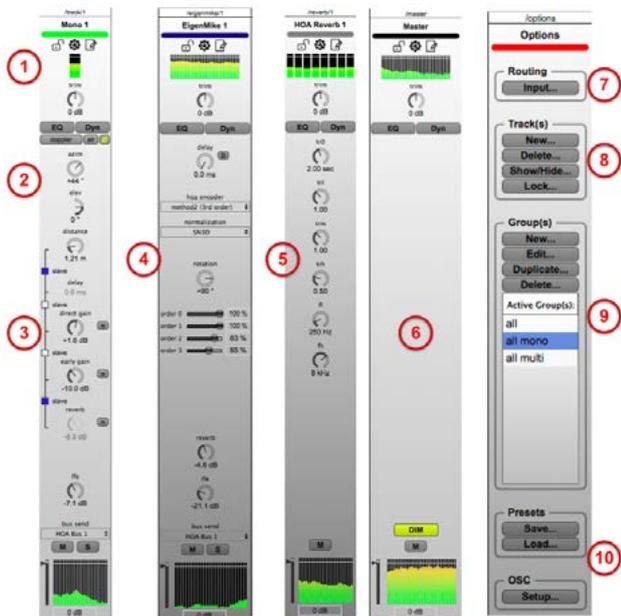

Figure 4. View of multiple strips; from left to right: mono track, Eigen-Mike track, HOA reverberation bus, master track, session options. ① Strip header: name of the strip, color, lock/unlock, options, annotations, input vu-meter, input trim, equalizer and compressor. ② Localization parameters (position, Doppler effect, air absorption). ③ Room effect settings (direct sound, early reflections, send to late reverb). ④ HOA encoding and sound-field transformations parameters. ⑤ Late reverb settings (reverberation time, crossover frequencies, etc.). ⑥ Master track. ⑦ Input matrix. ⑧ Track management (create, delete, etc.). ⑨ Groups management. ⑩ Import/export of presets and OSC configuration.

Figure 5. ① Input routing. Physical inputs (rows of the matrix) can be assigned to the available tracks (columns). ② Panning bus routing "HOA 1". The output of the bus (columns) can be routed to the Master channels (rows), i.e. towards the physical outputs.
Each channel can have multiple connections (e.g. one physical input can be routed to several tracks).

Figure 6. Configuration of the speaker layout for a panning bus. Speakers coordinates can be edited in Cartesian ① or spherical ② coordinates. The reproduction setup can be aligned in time ③ and level ④; delays and gains are automatically computed or manually entered.
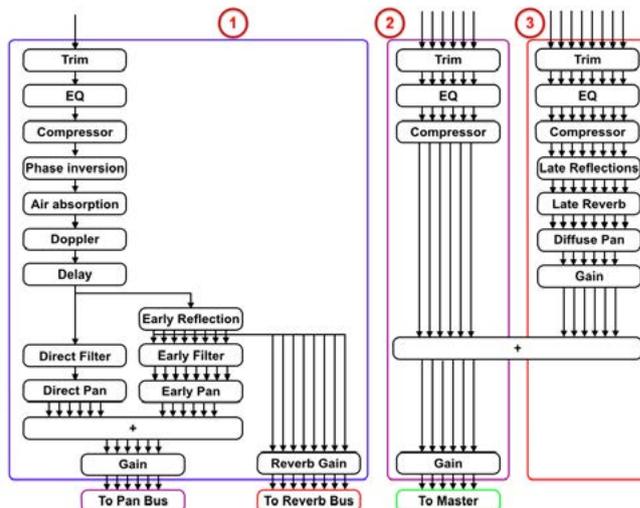

Figure 7. Audio architecture (simplified representation). ① Mono track. ② Panning/decoding bus. ③ Reverb bus.

3.4 Session options

The "Options" strip is used for the management of the mixing session. This includes routing of the physical inputs (see ⑦ in Figure 4 and ① in Figure 5), creation and edition of the tracks and busses (⑧ in Figure 4) as well as import/export of preset files (⑩ in Figure 4).

3.5 Group management

In a mixing context, it is frequently useful to group (or link) several parameters to maintain a coherent relationship while manipulating them. To achieve this, *Panoramix* offers a grouping mechanism where all modifications to one track parameter will also offset that parameter in every linked track. The "Options" strip provides a means to create, edit, duplicate or delete groups (see ⑨ in Figure 4 and Figure 8), and the ability to select the active group(s). Grouping effects all track parameters by default, however it is also possible to exclude some parameters from the group (e.g. mute, solo, send; see ③ in Figure 8).
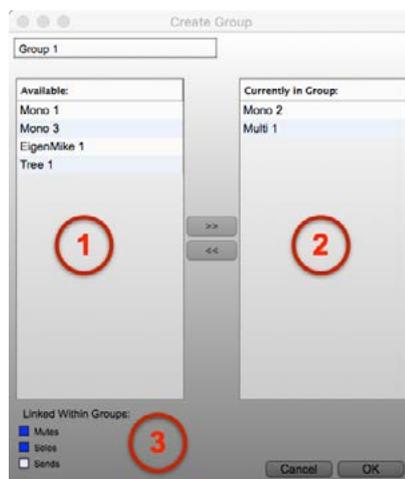
Figure 8. Creation/edition of a group. ① Available tracks. ② Tracks currently in group. ③ Group options.

3.6 OSC communication

All parameters of the *panoramix* application can be remotely accessed via the Open Sound Control protocol (OSC [10]). Typically, a digital audio workstation is used for edition and playback of the audio tracks while *panoramix* handles the spatial rendering and mixing (see Figure 9). Automation data is stored in the DAW and sent to *panoramix* through OSC via a plugin such as ToscA [11].
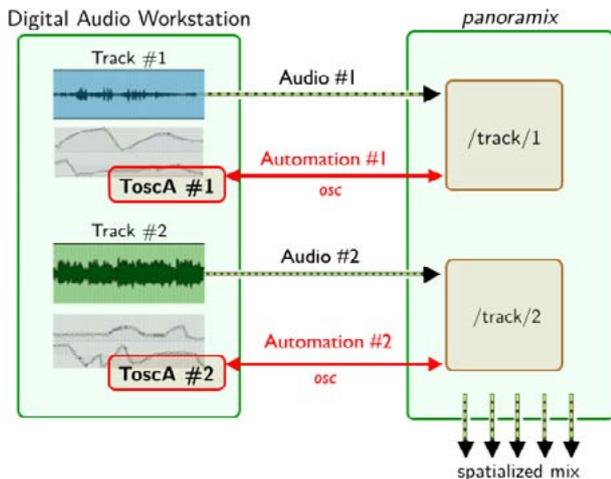
Figure 9. Workflow with *panoramix* and a digital audio workstation communicating through the OSC protocol and the ToscA plugin.

4. CONCLUSION AND PERSPECTIVES

This paper considered the design and implementation of a 3D mixing and post-production workstation. The developed application is versatile and offers a unified framework for mixing, spatializing and reverberating sound materials from different microphone systems. It overcomes the limitations of other existing tools and has been proved useful in practical mixing situations.

Nonetheless, the application can be further improved and many new features are considered for future versions. This includes (but is not limited to):

- support of other encoding/decoding strategies, notably for M-S and B-Format microphones,
- extension of the reverberation engine to convolution or hybrid processors [12],
- import and/or export of the tracks' settings in an object-oriented format such as ADM [13],
- implementation of monitoring or automatic down-mixing tools, based for instance on crosstalk cancellation techniques as proposed in [14],
- insert of audio plugins (VST, AU, etc.) in the strips,
- integration of automation data directly into the *panoramix* workstation,
- synchronization of the session to a LTC time-code.

Acknowledgments

5. REFERENCES

[1] D. M. Huber and R. E. Runstein, *Modern Recording Techniques (8th Edition)*. Focal Press, 2014.

[2] F. Rumsey and T. McCormick, *Sound and Recording (6th edition)*. Elsevier, 2009.

[3] B. Bartlett, "Choosing the Right Microphone by Understanding Design Tradeoffs," *Journal of the Audio Engineering Society*, vol. 35, no. 11, pp. 924 – 943, Nov 1987.

[4] R. Knoppow, "A Bibliography of the Relevant Literature on the Subject of Microphones," *Journal of the Audio Engineering Society*, vol. 33, no. 7/8, pp. 557 – 561, July/August 1985.

[5] J.-M. Lyzwa, "Prise de son et restitution multicanal en 5.1. Problématique d'une œuvre spatialisée : Répons, Pierre Boulez," Conservatoire National Supérieur de Musique et de Danse de Paris, Tech. Rep., May 2005.

[6] N. Peters, G. Marentakis, and S. McAdams, "Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis," *Computer Music Journal*, vol. 35, no. 1, pp. 10 – 27, 2011.

[7] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456 – 466, June 1997.

[8] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université de Paris VI, 2001.

[9] T. Carpentier, M. Noisternig, and O. Warusfel, "Twenty Years of Ircam Spat: Looking Back, Looking Forward," in *Proc. of the 41st International Computer Music Conference*, Denton, TX, USA, Sept. 2015, pp. 270 – 277.

[10] M. Wright, "Open Sound Control: an enabling technology for musical networking," *Organised Sound*, vol. 10, no. 3, pp. 193 – 200, Dec 2005.

[11] T. Carpentier, "ToscA: An OSC Communication Plugin for Object-Oriented Spatialization Authoring," in *Proc. of the 41st International Computer Music Conference*, Denton, TX, USA, Sept. 2015, pp. 368 – 371.

[12] T. Carpentier, M. Noisternig, and O. Warusfel, "Hybrid Reverberation Processor with Perceptual Control," in *Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, Sept. 2014.

[13] M. Parmentier, "Audio Definition (Metadata) Model – EBU Tech 3364," European Broadcasting Union, Tech. Rep., 2015. [Online]. Available: https://tech.ebu.ch/docs/tech/tech3364.pdf

[14] A. Baskind, T. Carpentier, J.-M. Lyzwa, and O. Warusfel, "Surround and 3D-Audio Production on Two-Channel and 2D-Multichannel Loudspeaker Setups," in *3rd International Conference on Spatial Audio (ICSA)*, Graz, Austria, Sept. 2015.