A Few Notes on Stem-based Composition: A Case Study.

Constantin Popp
Novars
University of Manchester, United Kingdom
constantin.popp@postgrad.manchester.ac.uk

Abstract

Stem-based multichannel compositions promise greater flexibility in different listening situations and therefore improve the chances of a rewarding listening experience. The idea in stems is to separate materials or spatial layers into a collection of sound files which can be reproduced in various ways depending on the situation's needs. Also, splitting musical layers into separate sound files could result in higher spatial depth and improved off-centre listening. However, composing and performing stem-based compositions poses its own challenges which shall be laid out in the paper. The discussion will focus on the benefits, limitations and compositional methodology of spatial stems.

1   In: David Holland, Louise Rossiter (Eds.) Proceedings of Sound, Sight, Space and Play 2013
Postgraduate Symposium for the Creative Sonic Arts
De Montfort University Leicester, United Kingdom, 5-7 June 2013.

1. Introduction

1.1. What are stems?

In popular music or film sound the name "stems" refer to sub-mixes of similar materials within a multi-track production based on heterogeneous sounds. For example, a soundtrack of a movie usually consists of multiple tracks of dialog, environmental sounds, music and special effects. Those tracks can be grouped together according to their type/function and stored as separate sound files.[iii] During the final mixing the mixing engineer then does not need to worry about the element's mix within each stem, but rather focuses simply on the level balance between the stems themselves. The concept of stems facilitates therefore the mixing process and also the creation of alternate versions – like a movie in different languages – as only some of the stems have to be replaced.
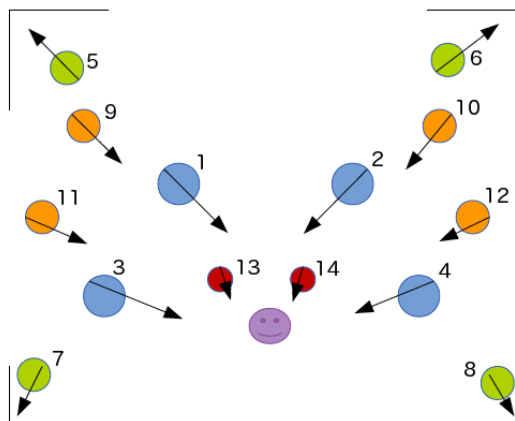

Illustration 1: Loudspeaker setup for "beeps". Each ring is colour coded to indicate which speakers correspond to which multichannel stem.

1.2. Stems in electroacoustic Music

The idea of stems could be applied to electroacoustic music in a variety of ways. Each way offers specific benefits in terms of the temporal evolution of a piece, the employed spatialisation formats or flexibility in loudspeaker

mappings. So far the author has found three approaches, the benefits of which shall be briefly discussed:

1. Stems grouping materials,
2. stems representing a hybrid format, implementing different spatialisation approaches, or
3. stems grouping spaces or spatial positions.

Stems based on different materials could improve temporal flexibility. For example, all gestures and textures of a piece could be stored as separate sound files. They could be then imported into, say, an Ableton Live session and triggered depending on the situation. The resulting structure and the timing of a piece could be easily adjusted during the performance within certain constraints. In a way, as Ben Ramsay puts it, the piece would be "unfixed", slightly attaching the idea of the open work to acousmatic pieces (Ramsay, 2011).[iii] It would also be possible to create compositions where the spatialisation of each material stem would be calculated in real-time according to the current available number of loudspeakers. This would enable the (spatialisation of the) piece to be scalable depending on the resources of the listening situation. Jonty Harrison and Scott Wilson reported having implemented this approach in their some of their compositions (Harrison, 2011).[iv]

Stems based on a variety of spatialisation formats make up a second category of stem approaches which I would like to call "hybrid formats". Here, a composition could make use of the benefits of different spatialisation formats. For example, a piece could be made up of both a stereo track meant for manual diffusion **and** a multichannel image with pre-composed spatialisation. The elements would be spread around the stems according to what kind of spatialisation would be appropriate. The flexibility and performative aspect of manual diffusion would be paired with the opportunity for elaborate automated multichannel spatialisation. David Berezan has been reported to have tried that approach (Harrison, 2010: 245).[v] Another possible idea would be to mix spatialisation based on ambisonics with channel-based amplitude panning (Harrison, 2010: 245).[vi] This, too, would mix the best of both worlds as the scalability and "naturalness" of ambisonics could be amended by sounds placed "within" a loudspeaker via amplitude panning.

Finally, stems could also represent different spatial positions or room acoustics, improving spatial depth as each stem could be mapped to dedicated rings of loudspeakers. For example, in my piece "beeps" (Popp, 2013)[vii], the spaces are grouped into three quadrophonic rings with increasing diameters, plus an added close stereo pair (Illustration 1). These stems correspond to the "abstract" positions of a British sound diffusion system like Birmingham Electroacoustic Sound Theatre (BEAST)[viii] or Manchester Theatre in Sound (MANTIS)[ix]. So one stem would be mapped to a ring of very distant loudspeakers, another to very close loudspeakers, and so on. This concept greatly improves spatial depth and detail as each loudspeaker ring reproduces its own, specially designed sound image.

This paper will now focus on the spatial stems approach, discussing the benefits, limitations and compositional methodology in greater detail.
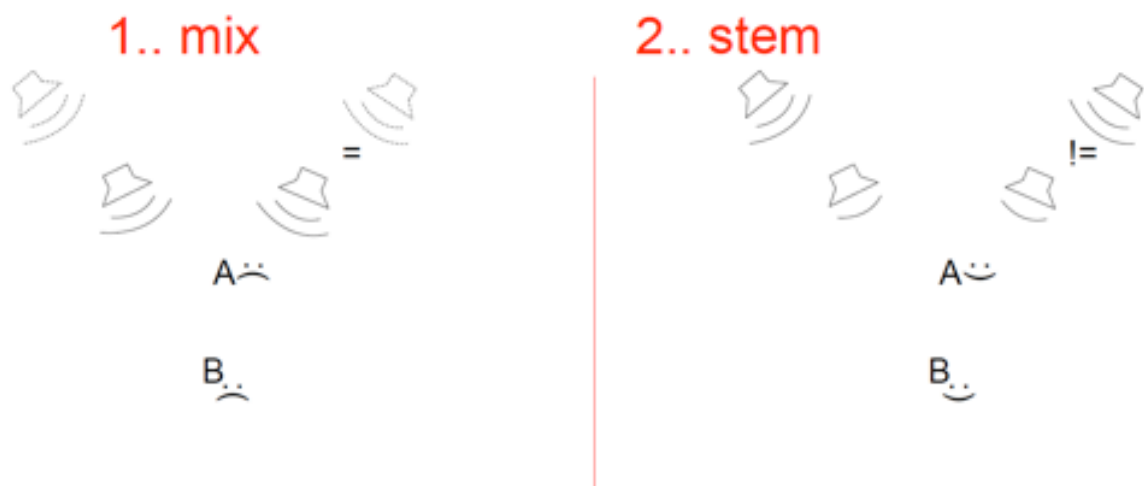

Illustration 2 : The illustration compares to what extend close and distant sounds are perceivable by listeners from various listening positions. The dashed lines indicate the approximate position of the simulated sound sources.

2. The case of spatial stems – benefits and limitations

2.1 Impact on spatialisation and space

As stated before, spatial stems improve spatial resolution. But what does this actually mean? Consider the following situation (Illustration 2). Two listeners sit in a reverberant venue, one sits relatively close to the frontal

momentarily supposed to appear either far away or close but doesn't work well when different musical layers mix distant and close sounds at the same time. Stems however would be beneficial here as those layers could be split to different stems and the stems with the distant sounds could be assigned to the distant loudspeakers while the close ones to the close loudspeakers. This means that each loudspeaker would reproduce its
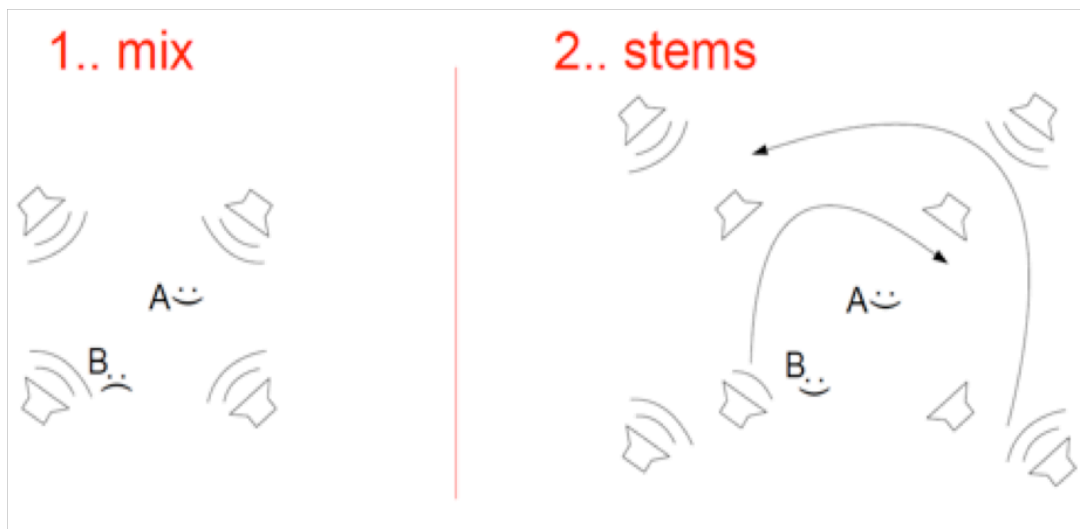

Illustration 3 : The illustration compares the effect of local masking on two listeners with four close speakes or four close and distant loudspeakers

loudspeakers (listener "A"), one further away (listener "B"). If a sound is supposed to appear distant listener A might still perceive the sound to be physically close as he is sitting close to the loudspeakers. The opposite is true for listener B as he might hear almost everything as distant as he is sitting far away from the loudspeakers and the difference between the simulated distant and close might not be audible for him due to the venue's own acoustics. Sound diffusion could ease this problem with adding loudspeakers in distance to the listeners and routing the piece to those loudspeakers for moments when it should sound distant. This works well if the piece is

dedicated mix and elements of a composition could appear spatially **independently** of each other. As a result the spatialisation would be richer in depth.

Secondly, spatial stems help to reduce local masking in certain circumstances. Consider the following situation (Illustration 3): Two listeners sit in an area created by four loudspeakers placed as a square with one listener sitting in the sweet spot (listener "A") and the other one in proximity of one the rear loudspeakers (listener "B"). When for instance a soundscape is played back on all loudspeakers to define a space the loudspeaker near the listener B will mask most sounds

coming from other directions. When using stems with multiple spatial positions, however, such a soundscape would less likely mask other sounds if it is reproduced only on distant loudspeakers as it would appear further away from **all** the listeners. Also, close sounds could appear quieter because they would be less masked by distant sounds due to their spatial difference. Utilising motion helps to reduce local masking further as moving sounds hide other sounds only temporally.

2.2 Flexibility

Spatial stems introduce format flexibility to cater for various listening situations. Although my piece "beeps" asks for quite large sound diffusion systems for maximum effect, it is also possible to mix the stems depending on the available loudspeakers and their placement. For instance, a stereo-mix can be obtained through mixing all stems together while adding reverb to keep the spatial layers separated and collapsing the front and back channels. This stereo-mix still provides an adequate experience in personal home-listening situations as the listener(s) can be expected to sit in the sweet spot (where the room simulation cues can be perceived). This works similarly well for purely quadrophonicsetups with the difference of keeping the front and back relationship intact. Furthermore, in small sound diffusion systems all the distant stems could be assigned to a single ring of distant loudspeakers, whereas the close stems to one ring of close loudspeakers. Larger sound diffusion system offer more routing possibilities. Like the hybrid format, spatial stems enable a composition to be scaled to the available resources, too.

The spatial stem idea also caters for the idea to employ room simulation on a case by case basis. For instance, "beeps" exists as a dry and a reverberated version. Reverb had been used during production to simulate different room acoustics and to further expand the spatial distance between the stems. During mix-down into stems one mix had been created where the room simulation had beed printed into the stem and another one where the room simulation was turned off. In theory, the dry version could be chosen in rooms with prominent reverberation to avoid having the room decrease intelligibly of already pre-reverberated sounds. However the author has not used this option, yet, since any performance asked for it and therefore he's unsure if such an approach does make sense in the end.

2.3 Other benefits

In terms of abstract spatial positions, a similarity between the production setup and the performance setup can arise when spatial stems are used. If a composition implicitly expects the performance via a sound diffusion system then it could do so explicitly with creating stems fitting the abstract spatial positions of those systems. To be able to hear their effect during composition time a small scale sound diffusion system would have to be installed in the studio as well. In this train of thought, the production environment would be a bit closer to the performance environment through similarity in the loudspeaker setup.

Spatial stems may also reduce the need of rehearsal time and simplify performance. Because most of the spatialisation is done during production time, not much is left to do during performance time other than adjusting

the overall loudness, the subwoofer level or remapping stems for special effects. However, because the loudspeaker mapping can become quite complex, a lot of time is spend during rehearsals to find suitable settings, including moving loudspeakers to more appropriate positions and employ room equalisation and re-balancing the level of the stems. In case of only two stems and a massive sound diffusion system additional time for rehearsing manual diffusion-style spatialisation is required, too, although this is purely optional as the piece stays presentable without live manual input. So in terms of rehearsal time and performance complexity stems may actually not yield much time saving.

2.4 New problems

Since spatial stems ask for quite large sound diffusion systems for full effect, *ideal* playback opportunities can become limited. If no sound diffusion system is available the piece stays performable but loses its spatial richness for many listeners as all the stems will have to be collapsed. Also, large sound diffusion systems need a rather large performance space to offer enough space for both listeners and loudspeakers. Although many loudspeakers might be available for deployment the venue might limit loudspeaker placements. This in turn means that in most cases only two rings with different diameters might be possible plus some extra loudspeakers for effect. Similarly the size of the production room puts constraints on how many loudspeakers can be set-up to adequately monitor each stem and therefore limits the number of spatial stems the composition might utilise.

Leaving room size and loudspeaker constraints aside, each additional stem also increases complexity. Consequently, the effort to compose and perform the piece in the best-most fashion increases, too. More stems mean that more precise spatial decisions have to be made during production time, slowing down the compositional process as more and more (spatial) parameters have to be defined and set in the software (DAW). During rehearsals the loudspeaker mapping becomes equally more complicated[x] and very detailed tech notes are needed to describe mapping strategies and recommendations which in turn make it more unlikely that performers other than the composer him-/herself will perform the piece. Similarly, the composer might not be happy to have to define space into such detail during composition time, preferring to rely on spatial interpretation during performance.

2.5 Compromise

To facilitate the production of spatial stems I would suggest basing each stem on quadrophonic rings. This suggestion results from a compromise between technical limitations, production hassles and desired spatial resolution. A quadrophonic image is produced relatively easy through multichannel recording or spatialisation, needs few channels or loudspeakers to be fully represented, can be reduced to stereo if needed with fewer problems than eight-channel images[xi] and the consumption of cpu and disc space only doubles compared to a stereo image.

Composing with multiple stems within constrained production environments can be improved through the use of room simulation. Each stem can be

made to sound distinct when real-time reverberation is added for monitoring purposes only. This works well if the simulated rooms sound distinctly different in colour and room size. However, it is to be noted that using room simulation to differentiate between stems can lead to false impressions if the stems can't be monitored separately at least during the final mixing period as the level balance between the stems might be skewed in separate playback.

Also, as a rule of thumb two stems with an additional stem for effects seems to be sufficient for most cases. Sounds can be placed either far away or close or on special effect loudspeakers. This keeps the requirements on the diffusion systems to a minimum and reduces production and performance effort (see discussion above).

It may be possible to incorporate n-channel spatialisation methods like VBAP or ambisonics to escape the limitations of channel based spatialisation. But more research needs to be done to estimate if the benefits of such an approach outweigh the increased technical hurdles it introduces.

material acquisition	studio / field recordings
conform to 4+ channels	SuperCollider / Reaktor
transformation / editing	
spatialisation	Reaper
mixing / tracking	
mastering	
diffusion	Max/MSP

Illustration 2: Compositional steps with regard to used software.

3. How to compose with spatial stems?

Illustration 4 outlines the author's approach in creating compositions based on spatial stems, utilising a mix of DAW and sound programming environments. A DAW (Reaper) was selected to facilitate mixing and editing, while sound programming languages (SuperCollider, Reaktor, Max/MSP) were mainly used to generate, transform or spatialise sounds. These particular computer programs were selected due to their flexibility in the handling and transformation of multichannel sound files.

Sound transformation or generation is mainly done in SuperCollider. The author created a modular environment for this purpose which handles the number of input and output channels of transformation processes in flexible way (Illustration 5) in order to avoid having to constantly split and combine sound file channels. This is very, very important as each splitting and combining slows down the compositional process even further. The output of a transformation process is then recorded in real-time into a sound file (green arrows).

Spatialisation is done both in SuperCollider and Reaktor. Reaktor has the advantage to run as a plugin within Reaper in real-time and includes full automation support. This means that any of its parameters can be adjusted without having to re-render a sound file. On the other hand, if a sound is to be moved quickly and randomly, SuperCollider's spatialisation objects feel easier to use as no automation curves have to be drawn and the need to change spatialisation later is negligible.
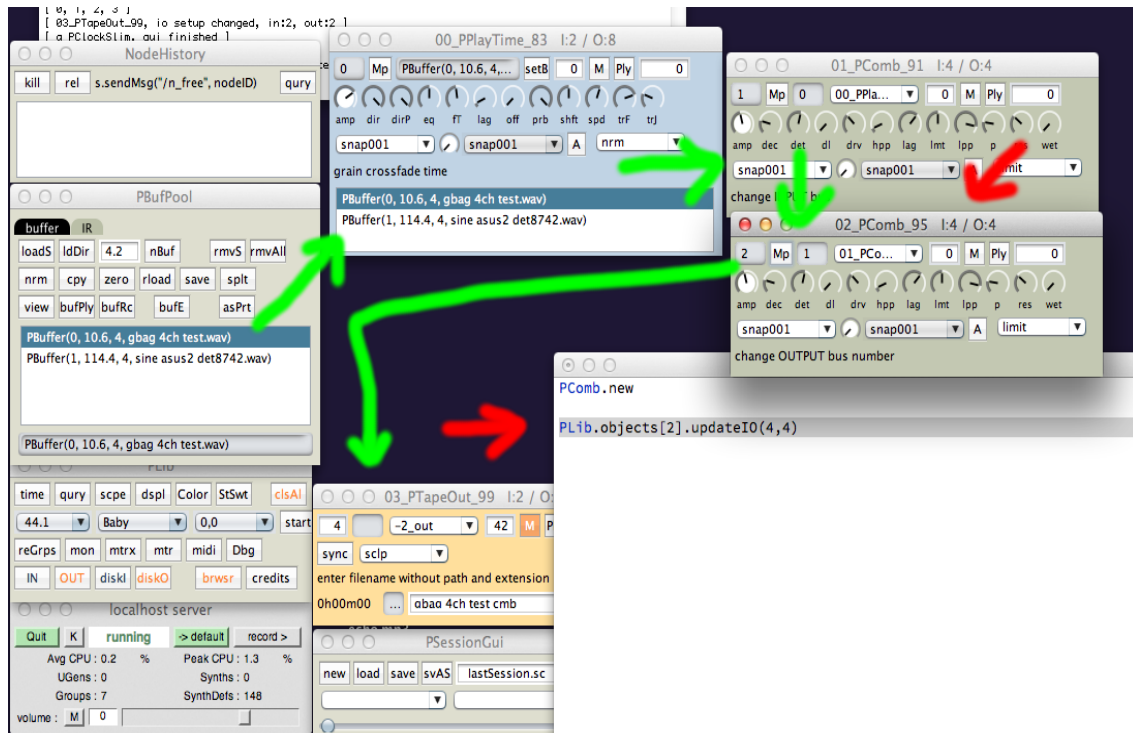
Illustration 5: Screenshot of the PLib. It handles various numbers of inputs and outputs in a flexible way (red arrows). The signal flow is indicated with the green arrows.

For editing, mixing and mastering, the DAW Reaper has proved to be well suited due to its flexible design. It not only allows merging of multichannel files with stereo files within the same tracks, but also incorporates a very flexible routing system which can be used to implement the spatial stem idea. Here, the stems are created via the send and receive functionality of tracks (Illustration 6). Each receive track represents a stem and corresponds to a dedicated spatial position or space (B). Other tracks will send sound to some extent to those receive tracks – the difference in level among the send amount and direct send to the main mix will determine the sound's final spatial position (A). In a way, the send and receive functionality is used to mimic sound diffusion style spatialisation within Reaper for each (!) sound file. Then, the receive tracks are routed to dedicated channels of the master track. The channels of the master track are mapped to the loudspeakers / sound card outputs. If not enough sound card outputs or loudspeakers are available to monitor the stems the stems can be mixed together purely for monitoring purposes (C); the rending of the master track stays unaffected from the sound card output routing.

For mixing a few VST-plugins proved to be essential in dealing with multichannel tracks. This is important as those plugins can be applied across all channels in a simple way without having to create multiple instances of the same plugin, routing the audio to each instance and linking their parameters manually. I use multichannel plugins from Voxengo (2013)[xii] and purpose built patches running in Native Instrument's Reaktor (2013)[xiii] in plugin mode. Other plugins
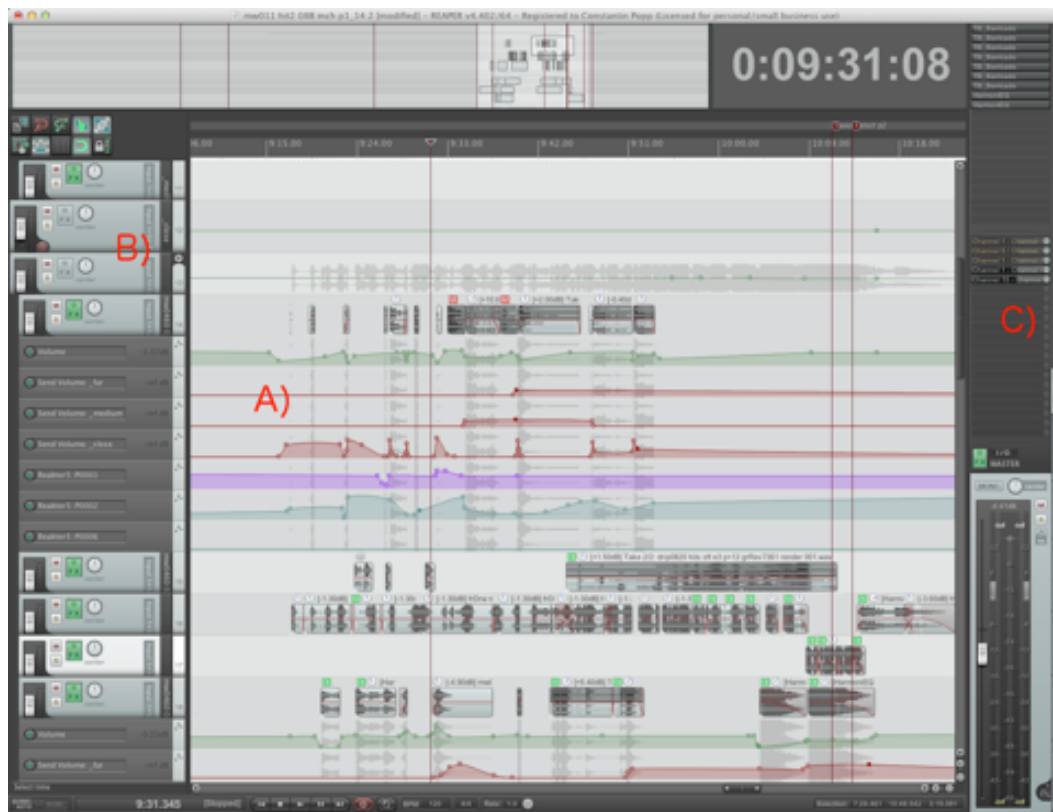
Illustration 6: Example of a mixing session of "beeps".

supporting multichannel operation are available from Ircam / Flux-Tools (2013)[xiv], Melda Production (2013)[xv] and VirSyn (2013)[xvi].

The mastering and preparation of the various distribution formats is done in a dedicated Reaper session (Illustration 7 on next page). Since I tend to run out of cpu power quickly, each piece is split into sections, each of which has its own session and is mixed down to a set of stems that are combined during mastering. The mastering session then combines all those stems and sub-mixes (A). Because the stems appear as a separate channels of the mix the mastering session contains also tracks which are fed from the main mix to sum stems into various other formats if necessary (B).

The performance of a stem-based composition can then be done via the MANTIS (Popp, 2013)[xvii] diffusion software of the Novars Research Centre, Manchester (Novars, 2013).[xviii] It is built in Max/MSP and allows the flexible routing of various channels of a sound file and sound card inputs to the available loudspeakers (illustration 8).

4. Conclusion

So are spatial stems worth the effort? Yes, definitely! The increase in spatial depth, control, resolution and rendition, as well as flexibility towards distribution of a piece in various formats, spatial stems outweighs the extra effort needed during production and performance. Further research needs to be done how spatial stems behave during object-based spatialisation methods and to what extend other approaches to stem based composition can be incorporated into spatial stems.
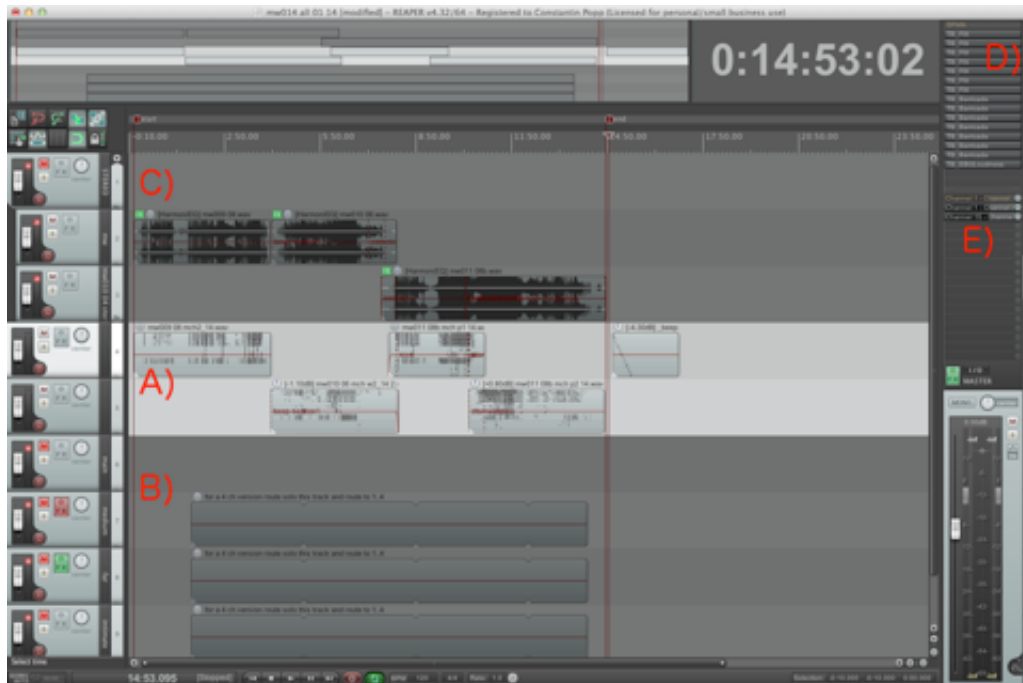
Illustration 7: Example of my mastering sessions of my pieces. The section's mixes are mixed (A) and mastered (D), as well sub-mixed via send and receives (B or C) if needed.
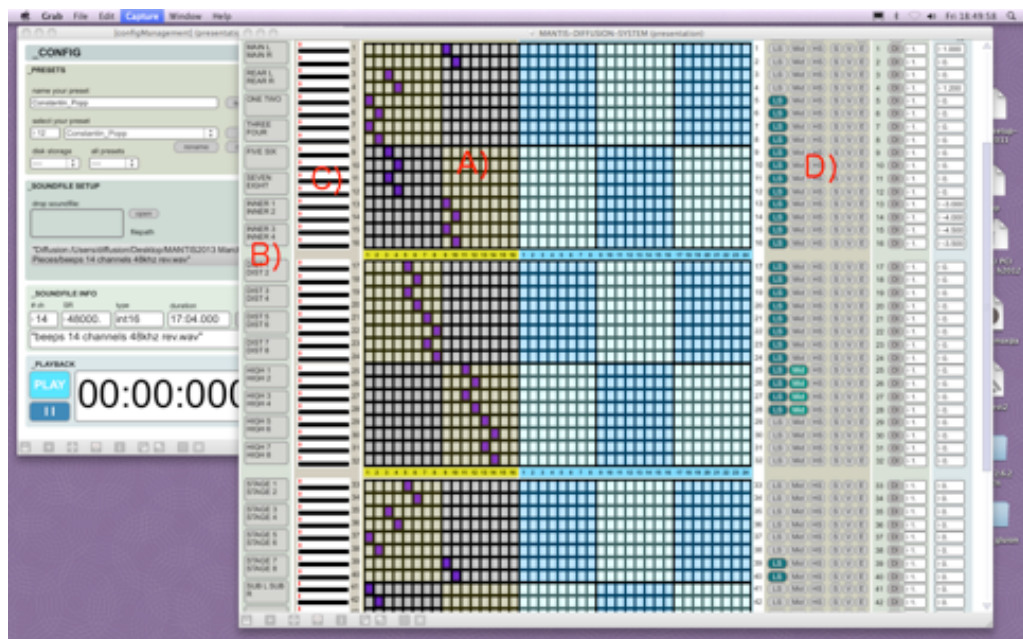

Illustration 8: shows an example of the loudspeaker mapping of "beeps" for it's performance during the MANTIS festival in March 2013. The stems are mapped (A) to loudspeakers (B), the level balance is adjusted in (C) and the stems's equalisation had been corrected to fit the room acoustics (D)

References

Flux::. (2013). *Flux:: sound and picture development*. [online]. Available from: http://www.fluxhome.com/ [Accessed September 28, 2013].

Harrison, J. (2011). The Final Frontier? Spatial Strategies in Acousmatic Composition and Performance. In *Toronto Electroacoustic Symposium*. [online]. Available from: http://cec.sonus.ca/econtact/14_4/harrison_spatialstrategies.html.

Harrison, J. and Wilson, S. (2010). Rethinking the BEAST: Recent developments in multichannel composition at Birmingham ElectroAcoustic Sound Theatre. *Organised Sound*, 15(3), pp.239–250.

MeldaProduction. (2013). *MeldaProduction, professional audio processing software*. [online]. Available from: http://www.meldaproduction.com/ [Accessed September 28, 2013].

Native Instruments. (2013). *Komplete : Synths & Samplers* : Reaktor 5 | Products. [online]. Available from: http://www.native-instruments.com/en/products/komplete/synths-samplers/reaktor-5/ [Accessed September 28, 2013].

Novars Research Centre. (2011). *MANTIS (Overview The University of Manchester)*. [online]. Available from: http://www.novars.manchester.ac.uk/mantis/overview/index.html [Accessed September 28, 2013].

Novars Research Centre. (2013). *NOVARS Research Centre (NOVARS - SAHC, The University of Manchester)*. [online]. Available from: http://www.novars.manchester.ac.uk/ [Accessed September 28, 2013].

Popp, C. (2013). MantisDiffusionSystem. *GitHub*. [online]. Available from: https://github.com/constantinpopp/MantisDiffusionSystem [Accessed September 28, 2013].

Ramsay, B. (2011). T*ools, Techniques and Composition: Bridging Acousmatic and IDM* by Ben Ramsay. [online]. Available from: http://cec.sonus.ca/econtact/14_4/ramsay_acousmatic-idm.html [Accessed June 30, 2013].

VirSyn Software Synthesizers. *BARK*. [online]. Available from: http://www.virsyn.de/de/Products/BARK/bark.html [Accessed September 28, 2013].

Voxengo. (2013). *VST plugins, AU plugins, Professional audio plugins - Voxengo*. [online]. Available from: http://www.voxengo.com/ [Accessed September 28, 2013].

Wikipedia. (2013a). Birmingham ElectroAcoustic Sound Theatre. *Wikipedia, the free encyclopedia*. [online]. Available from: http://en.wikipedia.org/w/index.php?title=Birmingham_ElectroAcoustic_Sound_Theatre&oldid=567369743 [Accessed September 28, 2013].

Wikipedia. (2013b). Stem (audio). *Wikipedia, the free encyclopedia*. [online]. Available from: http://en.wikipedia.org/w/index.php?title=Stem_(audio)&oldid=523635492 [Accessed September 28, 2013].

Recordings

Constantin Popp, *"beeps"*, 2013, 14 channels (stems), 14:45, premiered at Novars, 2nd March 2013. See: https://soundcloud.com/constantinpopp/beeps-2012-extract [accessed 26/05/2013].

Notes

1 See: Fluid Mastering. (2011). Fluid Mastering: Mastering from Stems: what it means, and how to do it. [online]. Available from: http://www.fluidmastering.com/mastering-from-stems.htm [Accessed September 28, 2013]. See also: Wikipedia. (2013). Stem (audio). *Wikipedia, the free encyclopedia*. [online]. Available from: http://en.wikipedia.org/w/index.php?title=Stem_(audio)&oldid=523635492 [Accessed September 28, 2013].

2 Ramsay, B. (2011). Tools, Techniques and Composition: Bridging Acousmatic and IDM by Ben Ramsay. [online]. Available from: http://cec.sonus.ca/econtact/14_4/ramsay_acousmatic-idm.html [Accessed June 30, 2013].

3 Harrison, J. (2011). The Final Frontier? Spatial Strategies in Acousmatic Composition and Performance. In Toronto Electroacoustic Symposium. [online]. Available from: http://cec.sonus.ca/econtact/14_4/harrison_spatialstrategies.html [Accessed June 30,2013]. See also: Harrison, J. and Wilson, S. (2010). Rethinking the BEAST: Recent developments in multichannel composition composition at Birmingham ElectroAcoustic Sound Theatre. Organised Sound, 15(3), pp. 248.

4 Harrison, J. (2010), p 245.

5 Harrison, J. (2010), p 245.

6 Constantin Popp, "beeps", 2013, 14 channels (stems), 14:45, premiered at Novars, 2nd March 2013. See: https://soundcloud.com/constantinpopp/beeps-2012-extract [accessed 26/05/2013].

12 In: David Holland, Louise Rossiter (Eds.) Proceedings of Sound, Sight, Space and Play 2013
Postgraduate Symposium for the Creative Sonic Arts
De Montfort University Leicester, United Kingdom, 5-7 June 2013.

7 Wikipedia. (2013). Birmingham ElectroAcoustic Sound Theatre. *Wikipedia, the free encyclopedia*. [online]. Available from: http://en.wikipedia.org/w/index.php?title=Birmingham_ElectroAcoustic_Sound_Theatre&oldid=567369743 [Accessed September 28, 2013].

8 Novars Research Centre. (2011). MANTIS (Overview The University of Manchester). [online]. Available from: http://www.novars.manchester.ac.uk/mantis/overview/index.html [Accessed September 28, 2013].

9 See illustration 8 as an extra example.

10 Due to the higher channel count of eight-channel images the loss of spatial information in a stereo reduction would be more dramatic than reducing quadrophonic images to stereo.

11 Voxengo. (2013). VST plugins, AU plugins, Professional audio plugins - Voxengo. [online]. Available from: http://www.voxengo.com/ [Accessed September 28, 2013].

12 Native Instruments. (2013). Komplete : Synths & Samplers : Reaktor 5 | Products. [online]. Available from: http://www.native-instruments.com/en/products/komplete/synths-samplers/reaktor-5/ [Accessed September 28, 2013].

13 Flux::. (2013). Flux:: sound and picture development. [online]. Available from: http://www.fluxhome.com/ [Accessed September 28, 2013].

14 MeldaProduction. (2013). MeldaProduction, professional audio processing software. [online]. Available from: http://www.meldaproduction.com/ [Accessed September 28, 2013].

15 VirSyn Software Synthesizers. BARK. [online]. Available from: http://www.virsyn.de/de/Products/BARK/bark.html [Accessed September 28, 2013].

16 Popp, C. (2013). MantisDiffusionSystem. GitHub. [online]. Available from: https://github.com/constantinpopp/MantisDiffusionSystem [Accessed September 28, 2013].

17 Novars Research Centre. (2013). NOVARS Research Centre (NOVARS -SAHC, The University of Manchester). [online]. Available from: http://www.novars.manchester.ac.uk/ [Accessed September 28, 2013].