University of Wollongong
Research Online

Faculty of Creative Arts - Papers (Archive)

Faculty of Law, Humanities and the Arts

2009

Spatial Grains: Imbuing Granular Particles With Spatial-Domain Information

E. Deleflie
University of Wollongong, etienne@uow.edu.au

Greg Schiemer
University of Wollongong, schiemer@uow.edu.au

Publication Details

This conference paper was originally published as Deleflie, E and Schiemer, G, Spatial Grains: Imbuing Granular Particles With Spatial-Domain Information, in Proceedings of ACMC09, Improvise, The Australasian Computer Music Conference, Queensland University of Technology, 2-4 July 2009.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

SPATIAL-GRAINS: IMBUING GRANULAR PARTICLES WITH SPATIAL-DOMAIN INFORMATION

Etienne Deleflie
Sonic Arts Research Network
Faculty of Creative Arts
University of Wollongong
School of Creative Arts

Greg Schiemer
Sonic Arts Research Network
Faculty of Creative Arts
University of Wollongong
School of Creative Arts

ABSTRACT

Granular synthesis techniques have been appropriated for 3D sound spatialisation in a number of ways, such as the spatial encoding of individual grains. This paper proposes a new technique that aims to use the spatial information already encoded in ambisonic signals, the principle hypothesis being that this encoding is actually retained at the granular level. This opens up exciting new possibilities for spatial sound. The paper outlines some of these possibilities but focuses primarily on the synthesis of non-point sources of sound which forms the basis for a second hypothesis, involving functions that relocate spatially encoded grains in time.

1. INTRODUCTION

Granular synthesis techniques have been used by a number of composers working with different spatialisation techniques from diffusion to ambisonics.

Truax, composing for diffusion environments, has used granular synthesis for the decorrelation of sounds projected over multiple speakers, thereby imbuing a sense of aural volume [1]. Barrett, working in ambisonics, has explored allocating individual spatial positions to grains [2].

Roads summarises the spatialisation of sound particles into two broad approaches [3]:

- the spatial encoding of individual particles and,
- the use of processing techniques to use particles as spatialisers for other sounds.

This paper proposes a different approach that retains both the original spatial encoding and the original sound content of an ambisonic source file; this is the first hypothesis.

A natural extension of the technique proposed in this paper, in combination with specific ambisonic source files, involves the synthesis of non-point sources of sound; this forms the second hypothesis.

2. PROPOSED TECHNIQUE

A new technique is proposed that uses the spatial encoding already inherent in the component channels of an ambisonic signal as a palette of pre-spatialised grains. The technique involves no spatial encoding other than that which is inherent in the source file.

In the granulation of source files, a grain is defined by the following micro-control properties [3, 4]; position (in the source file), duration, envelope and pitch (or speed of playback). In the context of the technique proposed, spatial encoding is retained by using the same micro-control parameters for each component (or channel) of an ambisonic signal. The resultant set of grains are the component parts of what is referred to hereafter as the 'spatial-grain'.

For example, a first order ambisonic signal typically contains 4 components (or audio channels) named W, X, Y and Z (for periphonic B-format). Since it is the combined components that define the spatial encoding, maintaining the same grain micro-control over W, X, Y and Z should maintain the spatial encoding for each spatial-grain.

> *" The grain is an apt representation of musical sound because it captures two perceptual dimensions: time-domain information (starting time, duration, envelope shape) and frequency-domain information (the pitch of the waveform within the grain and the spectrum of the grain)"* [3]

As an extension to this description, the spatial-grain also contains spatial-domain information.

2.1. Importance of the source content

Since both the sound content and the spatial encoding of the source material is retained in a spatial-grain, it is therefore the contents of that source material that will define the nature of the spatial opportunities that we intend to explore.

Consider the example of an ambisonic recording of a car moving left to right. The recording is granulated, or broken up into tens of thousands of spatial-grains each lasting 10-50 msecs. Since each spatial-grain retains the original recording's spatial-domain information, it is now possible to mix spatial-grains of the car in the left position with spatial-grains of the car in the right. The result would be granulated car sound coming from both the left and the right.

In other words, the temporal dimension of the recording has been abstracted so that individual spatially-encoded grains may be selected and played in any order. In effect, each spatial-grain is accessed by specifying a time-position in the source ambisonic file. Essentially, time is used as an index to the library of spatial information contained within the source.

3. SYNTHESISING NON-POINT SOURCES OF SOUND

While much has been written about synthesising point sources of sound, relatively little has been published about synthesising non-point sources of sound [5]. Perhaps due to complexity, the methods described [5, 6, 7, 8] have not yet found widespread use amongst composers.

The technique we have proposed could be used in combination with specific source files to synthesise non-point sources of sound, by choosing a collection of spatial-grains whose locations model a line, plane, or any volumetric shape. The term "volumetric" is used in this instance to refer to the dimensions of the space and is not to be confused with the term "volume" which is commonly used to refer to sound level.

Reconsidering the above example of a car moving left to right, imagine now that all grains -- all tens of thousands of them -- are played back simultaneously. This would encapsulate every position in which the car was recorded and would effectively model (instantaneously) the sound's spatial trajectory from left to right independent of its temporal trajectory. The result would effectively be a line source of sound.

Ambisonic recordings often contain sounds that emanate from many directions other than the principle subject (i.e. the car). It will be useful, therefore, to create an ambisonic source file which contains a point-source of sound that is completely isolated from ambient sound. This can be done by ambisonically encoding a point source of sound such as a synthetically generated sine tone.

Now consider how such a synthetically created point source might zigzag throughout a cubic volumetric space. Because the source file is ambisonic, the recording would contain spatial-grains that essentially occupy every possible position within the space. Removing time domain information reduces an ambisonic recording to a granular cloud spread throughout the entire cube -- and played in a single instant.

Lastly, instead of spreading the granular cloud evenly over the entire recording, imagine using an algorithm such as a statistical function to select which spatial-grains to play. If that algorithm includes a time component, then it becomes a sequence. This gives the potential to create dynamic, moving, morphing volumetric shapes projected in sound.

The success of this second hypothesis depends on how the clouds of spatial-grains are perceived psycho-acoustically. It is one thing to model a non-point source of sound, it is another for it to be heard that way.

4. PROCESSING ENVIRONMENT

Many audio processing platforms contain some form of support for granular synthesis techniques. These include independent software applications such as Road's Cloud Generator [3] and plugins for platforms such as MaxMSP. The implementation of spatial-grains has some extra requirements that need to be satisfied.

4.1.0. Control

To retain spatial information, the same micro-control parameters must be applied to every component grain of a spatial-grain.

Some granulators implemented exclusively in stereo may not be appropriate if the random number generators are encapsulated within the granulator; and can not be overridden.

4.1.1. Non real-time processing.

The ability to record a CPU intensive process to file. There are two ways in which processing with spatial-grains can scale up CPU usage beyond the capabilities of current processing power; the first is in an increase in the number of spatial-grains; the second is to offer more accurate localisation by increasing the ambisonic order.

The first scaling factor involves the number of component grains. When extending spatial-grains to higher orders, CPU usage will increase dramatically. For a periphonic B-format ambisonic source, each spatial-grain will contain 4 component grains. When this technique is extended to periphonic second order ambisonic material which has 9 component channels, each spatial-grain will contain 9 component grains. Periphonic fourth order is likewise modelled using 16 channels, thereby requiring 16 component grains.

The second scaling factor involves the number of grains used per second – or the granular density. As an indication, Truax in 1988 achieved a granular density of approximately 2000 grains per second for real-time processing [9]. Computer power has come a long way since 1988, but considerably more will be required now to produce granular synthesis where each individual spatial-grain is ambisonically encoded. For example, in a second order periphonic ambisonic signal where there are 9 component channels, 2000 spatial-grains per second will equate to 18000 component grains per second.

4.1.2. Score-like environment.

An ability to score the orchestration of spatial-grains in such a way that this orchestration can be repeated in non real-time.

Tests were conducted on a Core2Duo Macbook Pro. To leverage the power of both CPUs, it was decided that Jackdmp [10] (a multi-processor version of Jack [11]) be used such that the granulation could be done on one CPU, and the ambisonic decoding could be done on the other CPU. Jackdmp enables routing audio signals between applications running on different CPUs. Ambisonic decoding was done with AmbDec [12], a recommended ambisonic decoder [13].

The first implementation was done with PureData.

4.2. Exploration in PureData

Whilst PureData (Pd) does not, strictly speaking, implement a non-real-time mode it can execute expensive processes at 100% CPU and still successfully write audio signals to file. Pd also does not have a purpose built score feature but can easily communicate with other score-like applications (such as Csound).

Initial explorations in Pd exposed an important distinction between control data and signal data. When control messages were used to organize the macro structure of grains (which involves triggering grains in time) audio artifacts were created. These artifacts were caused because control data is not processed as often as signal data. In a vanilla Pd install, control data are only processed at the start of a block of 64 samples of a signal. If a control message does not land squarely -- in time -- at the start of a block of samples then it is processed in the next block. The software architecture of separated signal data and control data is common to much audio software.

Since granular synthesis involves triggering grains of sound which are typically 10-50 msec in length (grains whose size is not aurally distinguishable as an event), many grain triggers will not be time-accurate, since many triggers will not fall squarely at the start of a block size.

There are a number of solutions, within Pd, that can be used to work around this issue. One is to use a sub-patched defined block size of 1 (using [block~]), such that control messages are processed at every sample (restricted to a chosen sub-patch). Another is to use a set of sample-accurate trigger externals (known as [t3~]) bundled in the 3rd party lib IEMLIB [14]. Yet another would be to write C externals that behaved in the desired manner.

However, all of these methods would introduce a performance degradation which can be circumvented by working within Pd's control data rate. Whilst this limits the control one has over certain granular parameters, it allows for very efficient processing. Triggering 20 msec grains at 4 msec intervals, within a sampling rate of 48 kHz allows perfect synchronization of control messages with a block size of 64 samples. Using this method 20,000 1st order spatial-grains per second could be generated. This is equivalent to 80,000 grains per second.

The Pd patch shown in Figure 1 was used to confirm the first hypothesis, that spatial information can be retained in spatial-grains.

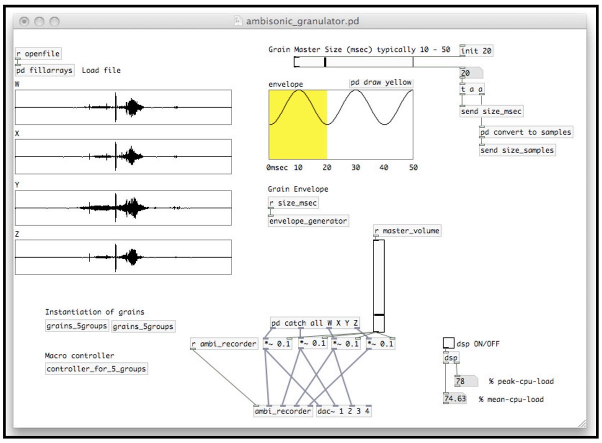

Figure 1. Screenshot of Pd patch shows the 4 channel B-format source file and some micro-control parameters. The patch also abstracts objects used for macro-control.

5. INITIAL RESULTS

A recording of a steam train moving right to left was initially used. When incrementally spreading the grain cloud from a time width of 4 msec to a time width of 20 sec (covering the entirety of the train's movement), a spatial gesture somewhat like a controlled audio 'explosion' (exploding right to left) was created. At the end of the 'explosion' granulated steam train sounds could be identified from the right to the left of the sound field.

This is not to say anything about the apparent width of the resultant sound.

Initial explorations involving the adjustment of granulation parameters and temporally changing spatial-grain cloud spread confirmed that rich and complex spatial textures could be created.

It became quickly apparent that the 'cleanliness' of the source file has a strong impact on the result. In this context, 'cleanliness' is used to refer to the extent of isolation of source sounds from ambient sounds. When engaged in open-ended exploration of spatial sound textures, a 'dirty' source file may have more interest. When attempting to create a designed spatial result, 'clean' source files may offer better results.

Lastly, it was evident that an enormous number of spatial-grains are required to granulate the entirety of a source file (depending on the length of that file).

6. A MORE APPROPRIATE AUDIO PROCESSING PLATFORM

Once the initial results were confirmed, the Pd patch described above quickly revealed its limitations. The support of a strong score and non-real-time processing environment became a priority. Sample-accurate triggering of grains became necessary.

Research returned to audio processing platforms and granulators. Searching for sample-accurate grain triggering revealed other solutions such as GMEM's bufGranul~, an external for Max/MSP [15].

A quick look at SuperCollider, however, revealed that it was better suited to the task.

6.1. Exploration in SuperCollider

SuperCollider (SC) caters well for non real-time processing. It provides an OSC message based score syntax which can be executed in non-real-time, and the results saved to disk.

SC includes a range of granular synthesis unit generators (uGens) which provide sample-accurate triggering.

SC's architecture also supports a powerful multichannel expansion feature. Passing an array of signals to the input of a single uGen causes that uGen to automatically copy itself for each signal in that array. This made short work of converting GrainBuf (a granulation uGen) for processing multi-channel source files. The following code triggers 4 sample-accurate grains (all with identical micro-control parameters) from 4 input buffers:

```
GrainBuf.ar(
        // number of channels
        1,
        // trigger (sample accurate)
        Impulse.ar( triggerFrequency ),
        // duration of grain in seconds
        duration,
        // the audio buffer
        [w, x, y, z],
        // playback rate
        1,
        // position in file 0 to 1
        pos,
        // pitch shifting interpolation
        // 1 is none, 2 linear, 4 cubic
        1,
        // pan left or right (none)
        0,
        // window envelope
        windowbuf
)
```

Converting GrainBuf for the creation of spatial-grains -- where each component grain used identical micro-control parameters -- only required passing an array of buffers (instead of a single buffer) into its audio buffer argument.

7. SYNTHESIS OF NON-POINT SOURCES OF SOUND

Potard [16] identifies the following approaches for synthesizing apparent source size:

1. Boosting the zeroth order (known as W for B-format) in ambisonic signals.
2. Turning an ambisonic signal inside out, to define a point source radiating outwards. This is known as O-format [6].
3. VBAP spread. Spatialising multiple point sources around the main source. A technique proposed for use with a simple method of spatial encoding called vector based amplitude panning (VBAP).

Potard also proposes and perceptually evaluates a fourth approach involving the decorrelation of multiple point sources.

Much research has been conducted into the psycho-acoustic perception of apparent source size [7]. The principle factors indentified as affecting the perception of apparent source size are [16]:

- The inter-aural cross correlation coefficient,
- sound loudness,
- pitch,
- signal duration.

It is interesting to note that no sophisticated spatial modeling is required to deliver three of the above factors: sound loudness, pitch and signal duration. This explains, perhaps, how composers using sound diffusion techniques (where typically stereo music is diffused over multiple speakers placed amongst the audience [1]) have been able to intuitively create apparent size in sound [17].

The second hypothesis proposes a technique for generating non-point sources of sound. The success of this technique will not be perceptually evaluated in this paper.

Generating non-point sources of sound essentially involves the design of the content in the source ambisonic file. It should be noted that this approach may not be limited to ambisonic source files. A similar approach could perhaps be applied to the channels of a 5.1 file, or to the speaker feeds derived from an ambisonic signal.

7.1. Generating a line of sound

A point source of sound, spatialised ambisonically, is recorded moving from A to B. This recording will contain spatial information describing the sound at every position from A to B.

If the entire recoding is turned into grains, all of which are played back simultaneously, then a line source of sound will be modeled.

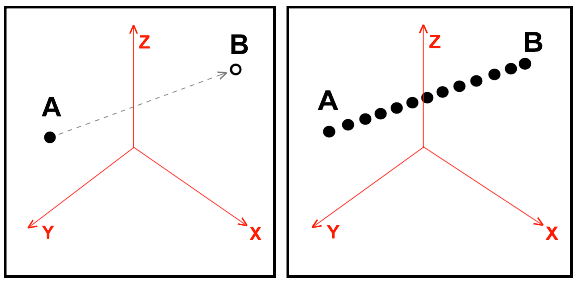

Figure 2. shows a spatial trajectory (on the left) which, when recorded, essentially contains the spatial encoding of that sound object at every point between A and B (illustrated by dots on the right). The dots can be viewed as a crude representation of spatial-grains. If all spatial-

grains are played back simultaneously, a line is modeled. In reality each spatial-grain can be as small as 10 msec in length.

7.2. Generating planar and volumetric sources of sound

Generating planar and volumetric sources of sound is similar to lines. This time, the point source of sound is moved on a spatial trajectory that covers either a plane or a volume.

A surface can not be covered in its entirety by a point source in the same way that a line is covered by a moving point. Generating a plane is ideally done from a moving line, but a well positioned moving point should achieve a similar result.

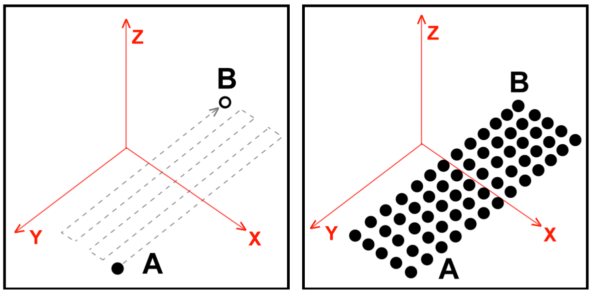

Figure 3. shows an example of a path (on the left) used to cover a planar area, and the resultant planar shape (on the right) when spatio-granularised.

By extension, creating a volumetric shape involves only defining a different more complex spatial trajectory.

7.3. Application of algorithms for variation

Algorithms can be implemented to alter the parameters of certain spatial-grains (like loudness), thereby creating non-uniform linear/planar/volumetric shapes.

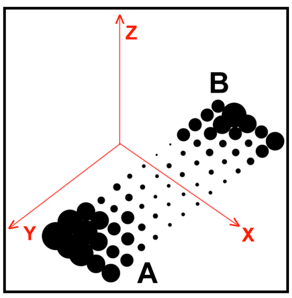

Figure 4. shows the potential effect of applying an algorithm which varies the loudness of spatial-grains. The size of each dot represents the loudness of a spatial-grain.

Consider a long spatial trajectory where a point source is moved to effectively cover every inch of space for a 100 x 100 x 100 m volumetric space. This would effectively create a spatial palette of every possible position within that space. Once that palette is created, algorithms can be designed to choose which of the palette's spatial-grains will be played. These algorithms could model any shape or form possible within that space, and affect any of the micro-control parameters used to create the spatial-grains.

7.4. Introducing time component to the algorithms

If the algorithms include a time component then the shapes created can begin to change in time. This could be used to create spatial gestures with movement or morphing of volumetric shapes.

7.5. Advantages of this approach

Granular synthesis has the secondary advantage of being an effective technique used to perceptually 'blend' – or decorrelate -- multiple sources of sounds into one larger source.

> *"The volume, or perceived magnitude, of a sound depends on its spectral richness, duration, and the presence of unsynchronized temporal components ... Electroacoustic techniques expand the range of methods by which the volume of a sound may be shaped. Granular time-stretching is perhaps the single most effective approach, as it contributes to all three of the variables just described… It should be noted that delays of only a few milliseconds are sufficient to decorrelate the different grains streams and thus increase their sense of volume."* [1]

Another advantage of the proposed technique is that there is no processing overhead involved in spatial encoding, since all spatial encoding is already present in the source file. Further, many (tens of thousands) of separate point sources can be used, with relatively little impact on CPU.

7.6. Limitations and possible workarounds

7.6.0. Sounds limited to contents of source file.

The sounds generated, and the spatial encoding generated, are both sourced from the ambisonic source file. Resultant spatial sound designs are therefore limited (in both audio content and spatial character) to what is present in the source file.

A potential technique to circumvent this limitation might involve convolution to substitute the source audio content (e.g. a 440Hz sine tone, or an impulse) with a more temporally complex sound.

7.6.1. Restricted spectral character.

When dealing with spatial-grains which have a time dimension of 10-50 msec, there is a limit to the sound's spectromorphological [18] evolution (once converted to a non-point source). In effect, the temporal character of a source sound is limited, since the differences between the source sound at time t and time t_1 are (essentially) used to distinguish two separate spatial positions, and not two separate spectral profiles. In other words, when the source sound is moved from A to B, it will typically remain the same sound. Of course, one could explore using sound sources that change in time thereby associating specific and different spectral characters for each spatial position.

Again, using convolution to replace a known source sound may offer a way to include rich spectromorphological evolution.

7.6.2. Reverberation in the source file.

Time based effects such as reverberations (both early and diffuse) included in the source file could easily be lost. Reverberations may occur in a window of time larger than the spatial-grain size.

One technique to retain reverberation information would be to make sure that the spatial-grains containing the reverberations are included in the cloud. Another simpler workaround would be to increase the grain size to include reverberation effects. This may require extending the grain size to greater than 50 msec – where a grain may start to be recognized as a single event – which may require greater care in the granulation so as to avoid a pointillistic sound character.

7.6.3. Doppler shift in the source file.

Any Doppler shift applied to the movement of the point source (in the original ambisonic file) will not help localise the resultant line/plane/volumetric shape. Rather, it may obfuscate the cohesion of the resultant shape due to the variation in pitch which is no longer relevant to the re-mixed content.

A designed or synthetically created sourced file can easily omit simulation of the Doppler shift.

It should be possible, however, to simulate the Doppler effect in a cloud of spatial-grains. Since 'playback speed' is one of the granular micro-control parameters, it should be possible to define a per-spatial-grain playback speed, that mimics a Doppler shift, simply by knowing where the spatial-grain is, and in what direction it is moving.

7.6.4. Two step process.

The contents of the source file may be designed (synthesized) or recorded. In either case, this file must be generated ahead of time, before being processed into spatial-grains.

A source file could be generated in real time, with the granulation following closely behind it. If an entirely new spatial trajectory is to be modeled, its complete granulation can only occur after the movement has been completed.

8. FUTURE WORK

Future work is needed in the application of spatial-grains using time as an index. This involves some fine tuning in two areas:

8.1.0. Micro-control parameters

Experimentation is required to understand how micro-control parameters affect the spatial image of the spatial-grain cloud. This will involve tweaking parameters that control grain size, grain envelope and speed of playback.

In addition to the modification of the source sound as it accessed using time as an index we expect that increasing the spatial-grain size will produce interesting spatial sound design possibilities given that time based spatial cues may be embedded in the source file.

8.1.1. Statistical functions

Statistical functions will be useful for varying the volumetric shape and character of spatial-grain clouds affording new opportunities for the exploration of the spatial composition. Statistical functions involving a time parameter can introduce movement.

8.1.2. Convolution

Convolution may allow a method for substituting the character of the sound in the source ambisonic file, with a more spectro-morphologically interesting one. This would free the sound designer from being restricted to the sound content of the original source file, whilst still being able to explore its wealth of spatial information.

Initial explorations will focus on simple spatial gestures characterized by a 'dualism' [19] between two spatial states.

All three areas will be explored further using various ambisonic speaker configurations possible in the newly configured speaker array called CHESS [20].

9. CONCLUSION

The proposed technique involves granulating ambisonically encoded audio files in a way that makes use of the spatial encoding embedded in the file.

Two audio processing environments have been investigated. An initial proof-of-concept implementation was done with PureData. SuperCollider has shown promise as an appropriate platform for multi-channel processing.

Spatio-granulation of a B-format recording confirmed that spatial information could be retained, confirming the first hypothesis. Some simple manipulations of spatial-grain clouds also confirmed that this technique could offer exciting possibilities for spatial sound design.

A second hypothesis, in which non-point sources of sound can be synthesized, has been outlined and discussed. The exploration and implementation of the second hypothesis is ongoing.

10. REFERENCES

[1] Truax, B. "Composition and diffusion: space in sound in space". in *Organised Sound*, 1998, pp. 141-146.

[2] Barrett, N. "Spatio-musical composition strategies". in *Organised Sound*, 2003, pp. 313-323.

[3] Roads, C. *Microsound*. Cambridge, Mass. London: MIT Press, 2001.

[4] Opie, T. "Sound in a Nutshell: Granular Synthesis". *Music*. Melbourne: La Trobe, 1999.

[5] Menzies, D. "W-panning and O-format, tools for object spatialisation". *AES 22nd International Conference*. Espoo, Finland, 2002.

[6] Malham, D. "Spherical Harmonic Coding of Sound Objects - the Ambisonic 'O' Format". *AES 19th International Conference*. Schloss Elmau, Germany, 2001, pp. 54-57.

[7] Potard, G. and Burnett, I. "Decorrelation Techniques for the Rendering of Apparent Sound Source Width in 3D Audio Displays". *Conference on Digital Audio Effects (DAFx'04)*. Naples, Italy, 2004.

[8] Menzies, D. "Ambisonic Synthesis of Complex Sources". in *Audio Engineering Society*, 2007, pp. 864-876.

[9] Truax, B. "Real-time granular synthesis with a digital signal processor". in *Computer Music Journal*, 1988, pp. 14-26.

[10] Letz, S. *Jackdmp. A multiprocessor version of Jack.* 1.9.0 Edition., [cited 29 April 2009]. Available from: http://www.grame.fr/~letz/jackdmp.html

[11] *Jack. An audio connection kit.* Edition., [cited 29 April 2009] Available from: http://jackaudio.org/

[12] Adriaensen, F. *AmbDec*. 0.2.0 Edition., [cited 29 April 2009]. Available from: http://www.kokkinizita.net/linuxaudio/

[13] Heller, A. and Benjamin. "Is My Decoder Ambisonic?". *AES 125th International Conference*. San Francisco, CA, USA, 2008, pp. 7553.

[14] Musil, T. *iemlib for PureData*. 1.17 Edition. Graz [cited 29 April 2009]. Available from: http://pd.iem.at/iemlib/

[15] Bascou, B. "A flexible granular synthesis environment in Max/MSP". *Sound and Music Computing*. Salerno, Italy, 2005, pp.

[16] Potard, G. *3D-audio object oriented coding [manuscript] / by Guillaume Potard*, 2006.

[17] Austin, L. "Sound diffusion in composition and performance: an interview with Denis Smalley". in *Computer Music Journal*, 2000, pp. 10-12.

[18] Smalley, D. "Spectromorphology: explaining sound-shapes". in *Organised Sound*, 1997, pp. 107-126.

[19] Doornbusch, P. and McIlwain, P. Integrating spatial parameters in composition practice. In: Vickery L, (ed) *Converging Technologies: Aust Computer Music Assoc Conf 2003*. Edith Cowan Uni, Perth, Australia: Australasian Computer Music Association, 2003.

[20] C. Ritz, G. Schiemer, et al "An Anechoic Configurable Hemispheric Environment for Spatialised Sound," in Proc. *Australasian Computer Music Conference (ACMC 2008)*, Sydney, Australia, 10-12 July 2008.