Ambisonics Spatialization Tools for Max/MSP

Jan C. Schacher, Philippe Kocher
ICST Institute for Computer Music and Sound Technology
Zurich School of Music, Drama and Dance
jan.schacher@hmt.edu
philippe.kocher@hmt.edu

Abstract

Ambisonics is an effective way of describing and projecting spatial sound. Based on the theoretical works of the late Michael Gerzon, it is more and more frequently being used not only to reproduce recordings made with a soundfield microphone and encoded in the corresponding B-format but also to place virtual sources in a periphonic space and decode them to an arbitrary number and configuration of speakers. The tools developed at the ICST implement Ambisonics in the form of Max/MSP externals and allow the encoding and decoding in three dimensions of up to third-order Ambisonics. In addition they include a graphical control module for real-time manipulation of source placement and modules for algorithmic control of source motion in three-dimensional space.

1   Introduction

The Institute for Computer Music and Sound Technology and its predecessor, The Swiss Center for Computer Music, have been involved since 2000 in developing and applying Ambisonics in real time for the concert hall. The need to optimize and standardize the numerous methods used has led to the development of specialized modules to process the many channels typically present in Ambisonics and add intuitive graphical control modules to allow musicians to work with this technique without having to deal with the underlying mathematical and physical theory. The first attempts in this direction were the creation of a group of VST plug-ins by Dave Malham of York University [1, 2] for the use in more traditional DAW-systems but also in real-time systems such as Max/MSP [3]. For various reasons this didn't lead to the desired results, mainly in terms of usability and cross-platform compatibility. A pair of orchestras for Csound [4] were developed in parallel for use in offline composition methods. Taking these implementations as a basis the first DSP-objects for Max/MSP were created in 2004, refined and tested in a multitude of contexts and are now being released under the GNU LGPL [5].

2   Ambisonics

Ambisonic spatialization starts from the premise that the sound waves any source emits in space can be modeled using spherical harmonics. In order to obtain a precise decomposition of a sound wave several passes with increasing numbers of spherical components in a Fourier series are used, corresponding to the increasing order of the harmonics. These components are obtained in the form of audio-streams which are grouped into a multi-channel format called the B-Format. It contains one channel of omni-directional signal (the monophonic signal or the zeroth order), three channels for the "figure-of-eight pressure responses" along the three Cartesian axes, encoding a full sphere, using only three directional components. Higher order virtual and physical microphone techniques superimpose increasing odd numbers of spatial components, thus narrowing the directional response pattern and refining the localization at each step. In the second order five components and in the third order seven components are added to the multi-channel stream to make for a total of 16 channels. Conventional Ambisonic soundfield recordings generate a periphonic spatial image through the use of a tetrahedral four-capsule microphone, the soundfield microphone as developed by the late Michael Gerzon [6, 7, 8]. Its output, the so-called A-format, is matrixed to obtain a first-order B-format stream.

To play back the encoded format a recomposition is made taking into account the exact location of each speaker. In theory the number of speakers and their position can be freely configured, but practical experience shows that symmetrical setups using at least as many speakers as there are components in the B-format are preferable. The great advantage of using Ambisonics is that it is not only a compact periphonic recording technique but also an intermediate carrier-format in which compositions can be stored and manipulated and then decoded to a specific speaker-array for each performance, for example in the concert hall. Furthermore the composition process can be monitored on a different and usually substantially smaller speaker-setup than that available at the performance-venue. So the finished piece will remain independent of any specific speaker set-up.

At the decoding stage the directional response pattern of the soundfield can be narrowed or widened according to the needs of the room acoustics or the position of the speakers. This is achieved by weighting the order components differently. For example mixing 100% zeroth order with 50% first order will give a very diffuse soundfield, as the directional information is lowered in

favor of the omni-directional component. Through the mixing of first, second and third order the directional response pattern gets narrowed and the side lobes of the response can be lowered considerably.

Based on the assumption of mathematically perfect microphones, speaker characteristics and room acoustics, Ambisonics has proven to be surprisingly resilient in maintaining a coherent spatial image under adverse conditions. This may be due to the fact that with each higher order an additional layer of spatial "sampling" occurs. This superposition together with the relatively large "sweet" area created by interference patterns allows covering spaces where the traditional panning techniques fail. Naturally there is a tradeoff between precision in localization and size of listening area. When decoding to asymmetrical and odd speaker layouts such as an ITU 5.1 setup the general symmetrical formulas still produce a coherent image which does appear to be biased in the direction of the more evenly spaced speakers.

2.1 Formulas

Encoder and Decoder Formulas

The ICST implementation uses the semi-normalized form of the Furse-Malham set both to encode and to decode higher order Ambisonics in symmetrical and equidistant speaker setups. Angles are given in a polar form using radians. In the following table azimuth is designated by θ and elevation by δ. Note the standardized use of uppercase letters to name each component.

Order	Name	Coefficient
0th	W	$\sqrt{2}/2$
1st	X	$\cos(\theta)\cos(\delta)$
	Y	$\sin(\theta)\cos(\delta)$
	Z	$\sin(\delta)$
2nd	R	$1.5\sin^2(\delta) - 0.5$
	S	$\cos(\theta)\sin(2\delta)$
	T	$\sin(\theta)\sin(2\delta)$
	U	$\cos(2\theta)\cos^2(\delta)$
	V	$\sin(2\theta)\cos^2(\delta)$
3rd	K	$\sin(\delta)(5.0\sin^2(\delta) - 3)0.5$
	L	$\cos(\theta)\cos(\delta)(5\sin^2(\delta) - 1)$
	M	$\sin(\theta)\cos(\delta)(5\sin^2(\delta) - 1)$
	N	$\cos(2\theta)\sin(\delta)\cos^2(\delta)$
	O	$\sin(2\theta)\sin(\delta)\cos^2(\delta)$
	P	$\cos(3\theta)\cos^3(\delta)$
	Q	$\sin(3\theta)\cos^3(\delta)$

Table 1. The coefficients for encoding/decoding into third order Ambisonics

Distance Formulas

The Encoder incorporates a gain stage that corrects the incoming signal for the third dimension of the polar coordinate system. The amount of decrease in amplitude related to the increase in distance can be scaled with a factor which generates a more or less pronounced distance effect. By default this value is set to a 3 dB decrease in amplitude when doubling the distance. In the following table distance is designated by the letter d and dB per unit by the letter u.

Component, Distance	Gain Coefficient
Omni, d <= 1	$1 - (1 - \sqrt{2}/2)d^2$
Omni, d > 1	$\sqrt{2}/2\ \text{pow}(10, ((d - 1)u)/20.0))$
Higher order, d <= 1	d
Higher order, d > 1	$\text{pow}(10, ((d - 1)u)/20.0)$

Table 2. The coefficients for distance correction of the omni-directional and the higher order components

To compensate for the fact that higher orders will increase gain exponentially when calculated for positions within the unit circle and approaching the zero position, the omni-directional signal is raised while the higher orders are lowered all the way to zero, effectively transforming the sound to a monophonic signal at the origin of the coordinate system. Other aspects of distance encoding such as air-absorption are not applied at this stage; they have to be implemented upstream of the encoder.

3 Implementation

A collection of externals for Max/MSP has been developed to cover the audio processing as well as the control aspects of working with Ambisonics. They include all the basic functionality required and should reduce complexity as much as possible without compromising flexibility.

3.1 DSP modules

The Ambisonics encoder "ambiencode~" and decoder "ambidecode~" have been implemented as a pair of externals for Max/MSP. The encoder takes N channels of sound-input and generates the full multi-channel B-format up to the third order. The decoder takes the multi-channel B-format stream corresponding to the desired order and generates M speaker-feeds. The separation into two modules is maintained mainly for practical reasons. It gives the user the option to mix and match discrete audio-channels with B-format streams and leaves open the access to the B-format for further processing. B-format files with missing or disordered components can easily be rerouted and fed into the decoder.

Mixing B-format and conventional sound sources is done by simply mixing two encoded streams before sending them into the decoder. Global transformations on the soundfield such as rotations around the three axes (pitch, yaw, roll) can be inserted in the B-format stream. Recording the freshly encoded B-format into a multi-channel sound file can easily be added after the encoder. To control the encoder the position information (azimuth, elevation and

distance) is input for each source either on a per channel basis or using indexed lists.

The input channel gets placed at the specified position in the soundfield as if it were a virtual microphone. One additional variable to the encoder is the slope-scaling factor in dB per unit for the distance correction.

The decoder takes azimuth and elevation information for each speaker as a grouped message or as indexed lists and regenerates the signal corresponding to the exact position of the speaker in the soundfield. A gain-factor for each order is available to permit the weighting of the components.

Typically one will add air-absorption and Doppler-effect before the encoder and distance correction delay lines after the decoder to obtain more control of the spatial imaging. Since these are standard routines they are not included in the Ambisonics-externals.
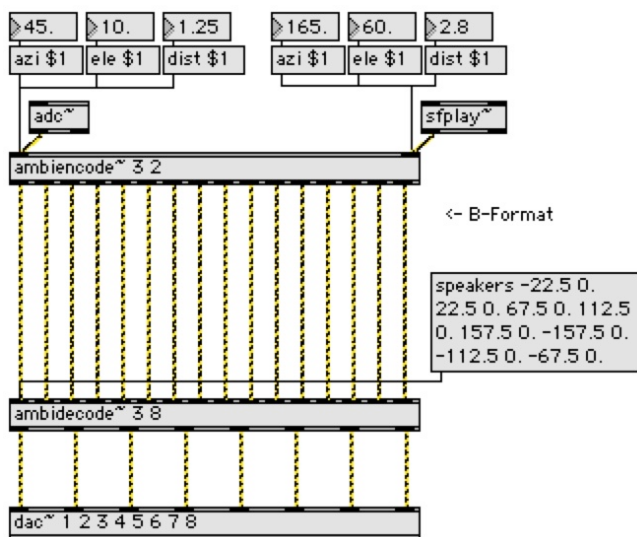

Figure 1. A basic application using two sources, eight speakers and transcoding in third order B-format

3.2 Graphical User Interface

To give the user an intuitive access to positions of the source-sounds the GUI-object "ambimonitor" was developed. It integrates seamlessly with the encoder, transmitting the correctly formatted indexed lists containing the position information for each point.

The graphic display is used to visualize positions on a two-dimensional surface or with half- or full-sphere three-dimensional display using two views (top and front). Sources are displayed as dots and are labeled either with symbolic names or their indices. If desired, the coordinates for each point can be displayed in cartesian or polar form.

Furthermore, points can be selected and moved with the mouse and a variety of key commands gives easy access to the points, allowing their creation and erasure and the control of a few basic motion types. The data from groups

of points can be stored in "snapshots", recalled, removed, interpolated, and finally exported in an XML-formatted text file.

A companion-class called "ambicontrol" connects to the data-structure containing the source-points and performs motion behaviors and transformations such as rotation, translation and random walks within boundaries. User-defined trajectories [9] given in a breakpoint format (time, coordinates) are executed as well. These trajectory descriptions can be imported and exported using XML-formatted text files. More complex interdependent behaviors are achieved by building logical chains with several of the "ambicontrol" modules.
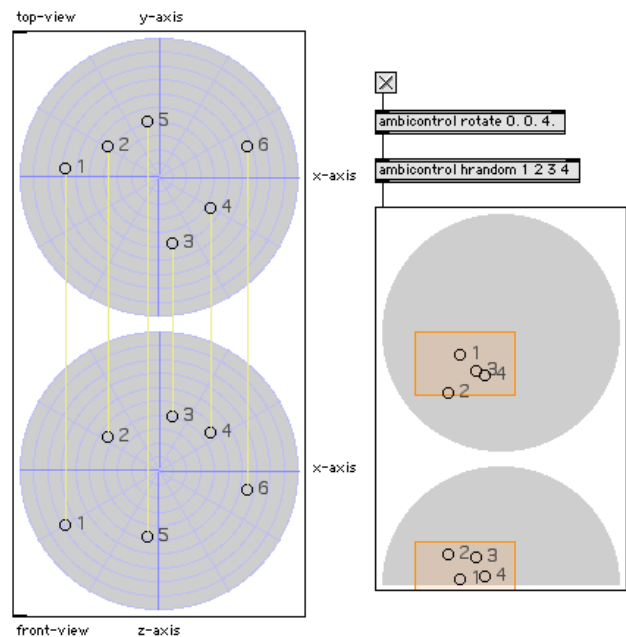

Figure 2. ambimonitor's full-sphere display and two instances of ambicontrol generating a rotating constrained random walk

4 Applications

A typical application of this technique in Max/MSP consists of some source-modules such as microphone-inputs, file- or buffer-based playback modules followed by any kind of modification and sound design processes. The resulting audio streams get fed to the encoder/decoder pair before reaching the output-stage where the final corrections of the signal take place. In the control domain the traditional pair-wise panning controls are replaced by a more complex spatial control system that generates the coordinates for source placement in the soundfield. Elements such as reverberation, filtering or Doppler-effect form a normal part of the signal-chain and in the case of reverb can go to separate spatial diffusion points in the soundfield to emulate the difference in incidence between early and late reflections.

The flexibility offered by Ambisonics for the compensation of site-specific constraints without compromising the integrity of the periphonic sound field is a real advantage.

The efficiency of this technique in terms of computing power makes the use of complex spatialization algorithms on one portable computer an attractive option, be it in a studio as a modeling tool or on a stage in a portable setup. Contrary to other higher level spatialization tools, in Ambisonics a large number of sources can be treated in real-time while leaving a considerable amount of headroom for other types of signal-processing.

The Max/MSP externals have already been used extensively in multi-channel concerts and multimedia installations. One primary application is in the playback of B-format recordings and tape-pieces stored in B-format. A more complex use is in pieces combining pre-edited B-format tracks with live-electronics and instrumental sound in real-time. On the other end of that spectrum are the performances where all sounds are generated completely live and encoded/decoded in real-time. The application of this technique in installations is very interesting, notably in a large scale interactive audio-visual piece in the French pavilion at the 2005 Expo in Aichi, Japan [10].

The ICST Externals are being incorporated into the Jamoma Project, an Open Source Initiative for standardized MaxMSP Modules [11].

B-format decoded to a 5.1 speaker layout is used for multichannel internet streaming of Ambisonics performances [12].

5   Outlook

A version of the DSP external is under development at ICST which incorporates the complete encoding/decoding cycle into one module and will be extended for higher orders of Ambisonics, where it becomes cumbersome to deal with the multitude of B-format channels in two separate modules. Tests in a variety of mixed encodings in two and three dimensions with changing orders will be more practical to handle inside one unified external.

Efforts are currently being undertaken to build a family of Pluggo-based plug-ins [13], compatible with all major DAW-hosts to facilitate studio editing processes in B-format.

Preliminary tests also indicate that the externals can be used as building blocks for Ambisonics to 5.1 decoding. In addition a binaural rendering of the B-format for headphones using HRTF-conversion is planned.

Finally the cross-compilation of the DSP-externals for Pure Data [14] will give a wider public the opportunity to experiment with this spatialization technique.

7   References

[1]   Malham, Dave G. *Ambisonics - A Technique for Low Cost, High Precision Three-dimensional Sound Diffusion* ICMC Glasgow 1990, pp. 118-120

[2]   Malham, Dave G. *Higher order Ambisonic systems for the spatialisation of sound* Proceedings, ICMC99, Beijing, October 1999

[3]   Zicarelli, David *An Extensible Real-Time Signal Processing Environment for Max*. In Proceedings of the International Computer Music Conference 1998. International Computer Music Association, Ann Arbor, 1998, pp. 463-466

[4]   Boulanger, Richard *The Csound Book*, The MIT Press, Cambridge MA, 2000

[5]   The ICST Max/MSP externals are available from http://www.icst.net URI valid in July 2006

[6]   Gerzon, Michael A. *Periphony: With-height sound reproduction*, JAES Jan/Feb. 1973, Vol. 21, No:1 pp. 2-10

[7]   Gerzon, Michael A. *The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound*, preprint of the 50th Audio Engineering Society Convention, London (March 1975)

[8]   Gerzon, Michael A. *Practical periphony: The reproduction of full-sphere sound*, AES Preprint 1571, London 1980

[9]   Chowning, John M. *The simulation of moving sound sources*. Preprint of the Audio Engineering Society 38th Convention, New York, 4-7 May 1970

[10] http://www.thierryfournier.net/gb/ce-qui-nous-regarde.html URI valid in July 2006

[11] Place, Tim; Lossius, Trond *Jamoma: A Modular Standard For Structuring Patches In Max*. Proceedings, International Computer Music Conference, New Orleans 2006: International Computer Music Association.

[12] Internet Streaming using the ogg/vorbis codec: www.spikamedia.ch URI valid in July 2006

[13] Pluggo: http://www.cycling74.com/products/pluggo URI valid in July 2006

[14] Puckette, Miller *Pure Data*. Proceedings, International Computer Music Conference. San Francisco 1996: International Computer Music Association, pp. 269-272.