The Projection of Sound in Three-Dimensional Space

Gerald Bennett, Peter Färber, Philippe Kocher, Johannes Schütt
Hochschule für Musik und Theater Winterthur Zürich

This text reports on four years of research on three-dimensional sound projection in electroacoustic music. The work was supported by and carried out at the Hochschule für Musik und Theater Winterthur Zürich. The goal of the research was to provide composers of electroacoustic music flexible and easy-to-use tools for creating the illusion of the movement of sound in three-dimensional space. This report has four parts. The first is a brief discussion of the theoretical background of the Zürich projects. The second part describes the development of the work since 1999. The third part discusses prospects for future work, and a fourth part reflects very briefly on aesthetic consequences of this work.

I. The Theoretical Background

Since its beginnings, electroacoustic music has seemed to offer composers the chance to place their music in three-dimensional space. Stockhausen considers his *Gesang der Jünglinge* (1955/56) the first piece of "Raum-Musik". Varèse's *Poème Électronique*, written for the World's Fair in Brussels in 1958, was played over a large number of loudspeakers affixed to the walls of the Philips Pavilion. By continuously routing the three tape tracks to different loudspeakers the sound could be made to seem to dance along the surface of the building. John Chowning took an important technical step forward in his composition*Turenas* for four-channel tape (1972). By very carefully panning sounds between the four speakers placed around the audience, by controlling the relation between direct and reverberated sound, and finally by using the Doppler effect, Chowning created very realistic illusions of movement in two-dimensional space. Beginning in 1984, another important advance was made by Gary Kendall and his collaborator William Martens. Using only two loudspeakers, they realized extraordinary illusions of position and movement in three-dimensional space by synthesizing not only the primary position of a sound and its diffuse reverberation, but also the first two or three echoes from the walls, ceiling and floor of a reverberant room.

Human hearing makes use of many perceptual cues to judge the position of a sound in space. Some of these cues are:

1. The difference of intensity between the two ears. A sound straight ahead sounds equally loud in both ears. As the sound moves to one side, the head masks the sound, creating up to a 20 dB difference in its intensity at the two ears. This difference in intensity is the principle cue for position of a sound in the horizontal plane around the listener.

2. The difference in arrival times of a sound at both ears. A sound coming from the right side reaches the right ear slightly earlier than the left ear. This time difference is very small, at most about 0.6 millisecond for a sound 90 degrees off

center. Time differences between the two ears of as little as one ten-thousandth second can be easily perceived with headphones, but for electroacoustic music in a concert hall the delays and reflections of the room would mask such fine differences. When composing we must exaggerate the temporal differences between the channels.

3. The overall intensity of a sound. Of course, near sound appears louder than distant sound. It is less clear how much louder a sound should be to seem nearer. We have had good results by using a decibel (i.e. logarithmic) scale per distance unit (e.g. reducing a sound's intensity by 3 dB per unit of distance).

4. The amount of high-frequency energy in a sound. Air absorption affects high frequencies more strongly than low frequencies. A distant sound is not only softer than the same sound nearby, but also less brilliant. This effect can be easily simulated electroacoustically with a low-pass filter.

5. The ratio between direct and reverberated sound. As Chowning showed in *Turenas*, this ratio is a very important perceptual cue for distance. The intensity of a sound's reverberation decreases with distance more slowly than the intensity of the direct sound. To simulate a sound moving away from the listener, one can decrease its intensity logarithmically (e.g. 3 dB per distance unit as above), while decreasing the intensity of its reverberation linearly.

6. The overall spectrum of a sound. All of us use a sound's spectrum to judge its horizontal and vertical angle. The torso, the head and particularly the outer ears, act as a filter depending on a sound's angle of incidence. This resulting spectrum change can be synthesized and simulates very well position and movement in three dimensions. Unfortunately, the spectral differences are so fine that this important technique only works with headphones and so is not of interest for composers who wish to play their music in public spaces.

On the basis of the first five of these perceptual cues, composers of electroacoustic music have traditionally derived five techniques to simulate position and movement of sound in space:

1. Adjusting the amplitude of the sound in the two channels to correspond to the horizontal angle of the sound;

2. Decorrelating the stereo signal temporally so that the signal appears earlier in the channel which has the greater amplitude;

3. Adjusting the overall amplitude of the sound to simulate distance;

4. Filtering high frequencies to simulate distance (the farther the sound from the listener, the less high frequency energy;

5. Adding reverberation and adjusting the ratio of direct to reverberated sound to improve the illusion of distance.

These techniques work very well in two dimensions, but because they are only for a stereo sound field (or, by extension, for several loudspeakers around an audience where each loudspeaker is in a “stereo” relationship to its neighbor as in classical four- or eight-channel configurations), we found them inadequate for our purposes. We also did not consider working with the classic surround formats (Dolby 5.1, 7.1, etc), because they have a fixed front/rear orientation and treat the rear channels differently from the front. Instead, we chose a technique developed in the 1970’s by Michael Gerzon in Great Britain called “Ambisonics.”

Ambisonics was originally a microphone technique used to make recordings which captured spatial information. Using a special microphone called the “Soundfield Microphone”, recordings were made in four channels which effectively encoded the sound sources’ spatial characteristics. In order to listen to the recordings, this information had to be decoded by special circuits. An advantage of this technique, apart from the remarkable “spaciousness” of the result, was that the recording could be decoded for an arbitrary configuration of loudspeakers—two, three, four or more—with very little change its spatial quality. Michael Gerzon, who was a mathematician and not a sound engineer, was able to show that under certain conditions the decoded signal represented precisely the wave front of the original sound.

Figure 1 shows the basic ambisonic principle.

Insert Figure 1 about here.

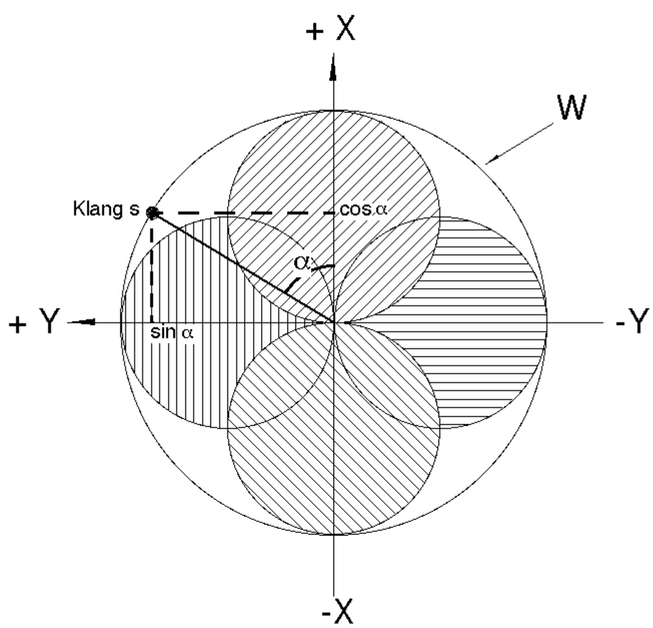

Figure 1. The Ambisonic Principle

Imagine two microphones with figure-of-eight characteristics at right angles to each other, one along the X-axis, one along the Y-axis. Now imagine sound *s* with amplitude 1.0 on a circle centered around the microphones (i.e. the radius of the circle is 1). Because of the directional characteristics of the microphones, sound *s* will be recorded at less than full amplitude by each one. In fact, its amplitude in the two microphones is equal to the cosine and the sine respectively of the angle □ times the amplitude of *s*. Finally, we also record the total amplitude of the sound with an omni-directional microphone (represented by the circle W in Figure 1). Hence the encoding of the spatial information in two dimensions for sound *s* is done by the following formulas:

$$w = s * 0.707$$
$$x = s * \cos \square$$
$$y = s * \sin \square$$

If we now imagine a third microphone pointing straight up and down (the Z-axis), we are able to represent the energy of the sound in three dimensions as follows (□ is the angle of elevation of the sound):

$$w = s * 0.707$$
$$x = s * \cos \square * \cos \square$$
$$y = s * \sin \square * \cos \square$$
$$z = s * \sin \square$$

The decoding of these four signals to derive the signals to be sent to the loudspeakers is essentially identical to the encoding, except that the angles refer to the position of each loudspeaker. This is the formula for the signal S_L sent to one loudspeaker of an array of arbitrarily many loudspeakers (here □ is the horizontal angle of the loudspeaker L, □ its elevation):

$$S_L = 0.707 * w + x * \cos \square \cos \square + y * \sin \square \cos \square + z * \sin \square$$

These equations correspond to the simplest (so-called zero-th and first order) equations for Spherical Harmonics and allow the listener to localize a sound within one quadrant (90 degrees). Higher order equations for Sperical Harmonics give greater precision of localization, but they require more channels of information. We have found second-order representation (nine channels of information) to be a good compromise between precision of localization and amount of information to be managed.

II. A Summary of the Work at the HMT Since 1999

The first project in ambisonics, begun in October 1999, consisted primarily of realizing the first-order ambisonic formulas in software. As programming language we chose the well-known program for sound synthesis Csound. Besides implementing the formulas, we wrote programs to position sound in three-dimensional space and to describe simple movements. Working with the Csound programs gave very good results almost immediately but was quite cumbersome. Not only did the position or movement of every sound in a composition need to be defined by the composer, but the four-channel encoded sound file (called the B-Format file in ambisonics jargon) had to be decoded into as many individual monophonic files as there were to be loudspeakers. These files were then put into a mixing program like Pro Tools and played from the computer onto the eight-track tape used for the concert performance. The greatest disadvantage was that the composer could not listen to the ambisonic realization in the studio until the tape was finished. When we first heard ambisonics in a large space, our reactions were amazement and delight at the marvelous quality of the sound. After our first euphoria subsided however, it was very clear that we needed to offer composers a way of working interactively with ambisonics.

In Spring 2000, we began using the interactive program Max/MSP for ambisonics. One of the first results was a combination encoder/decoder which allowed the composer to control the position of a sound with the mouse. Figure 2 illustrates this program.

Insert Figure 2 about here.

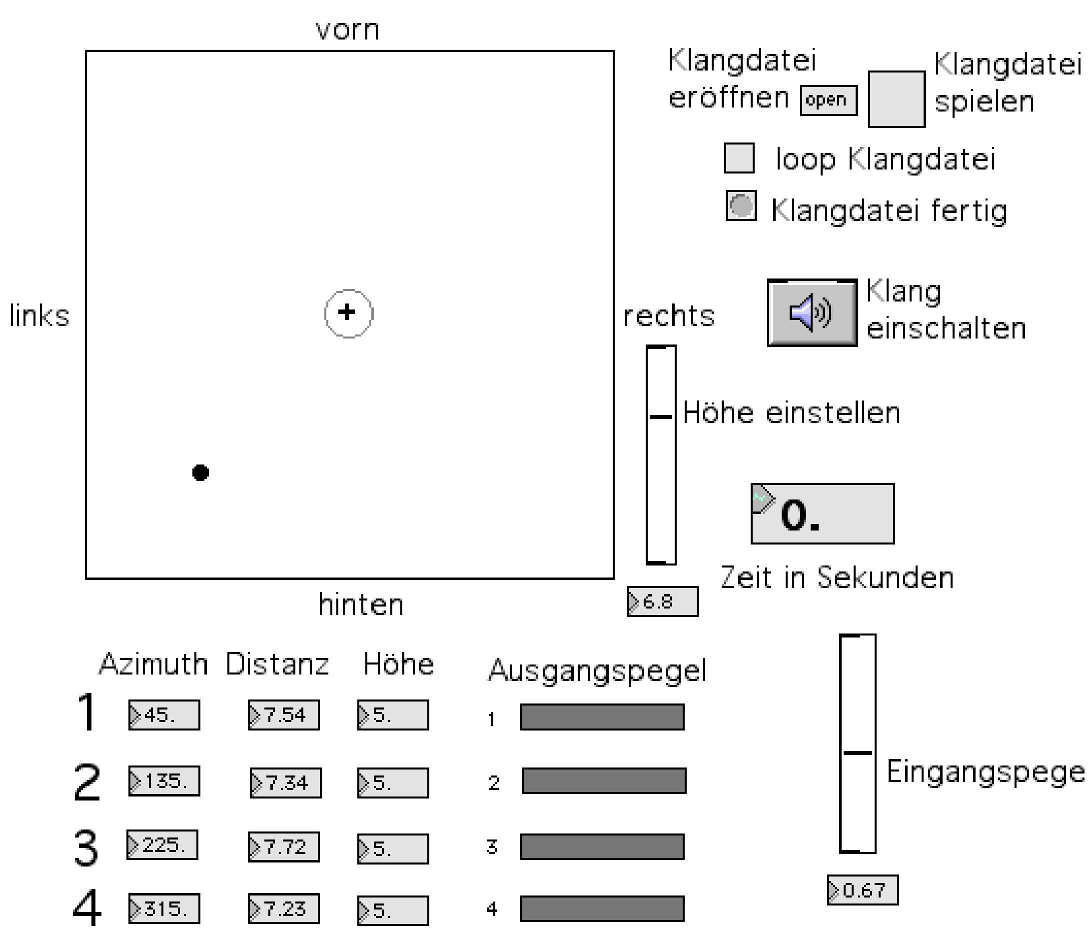

Figure 2. A screen shot of a combination encoder/decoder written in Max/MSP (Spring 2000)

The square at the left represents a space seen from above. The black dot shows the position of a sound in the horizontal plane and can be moved with the mouse. A slider to the right of the square regulates the sound's height. The fields below the square allow one to enter the position of each loudspeaker (here four); the program calculates the optimal decoding and also compensates for the loudspeakers' being at unequal distances from the center if necessary.

Simple tools like this enabled us to gain a great deal of experience with ambisonics. Being able to listen while making the movement was a great help for the imagination, for, much to our surprise, imagining sounds in three-dimensional space turned out to be more difficult than we had thought. But a simple mouse-driven instrument with no memory which only works with one sound at a time is obviously not a serious tool for electroacoustic composition. Therefore composition continued using the Csound programs, which were soon complemented by second-order versions. It became usual to calculate nine-channel B-Format files with Csound and to decode them in real time with a Max/MSP decoder. At the same time, three of the authors (Färber, Kocher and Schütt) worked intensively to expand the library of Max/MSP programs. Their interest was in using ambisonics interactively in concerts in combination with programs for the treatment and the synthesis of sound. Between Spring 2000 and the present, several concerts have been given using programs that were hand-crafted for each composition and hall. With input of up to 24 channels of sound to be transformed, treated ambisonically and output to as many as 24 loudspeakers, the demands made on the computers were considerable. There were never fewer than three fast Macintosh computers running in parallel at these concerts.

The problem of working polyphonically in the studio for tape pieces was still unsolved. In the Spring of 2001 we had the idea of writing a so-called plug-in for a commercial mixing program which would allow the composer to build up complex textures and hear the ambisonic result while working. In Autumn 2001 the HMT sponsored a research project, in collaboration with Dave Malham and Ambrose Field of York University (Great Britain), which resulted in a family of VST plug-ins for both Macintosh and Windows platforms. These plug-ins have been available on the Internet as freeware since April 2002.

In the academic year 2002/03 Färber, Kocher and Schütt worked to design an environment for interactive composition and performance with ambisonics. They were able to use the considerable experience they had acquired over the previous three years to design a program combining great flexibility with ease of use. Figure 3 shows part of the graphical user interface.

Insert Figure 3 about here.

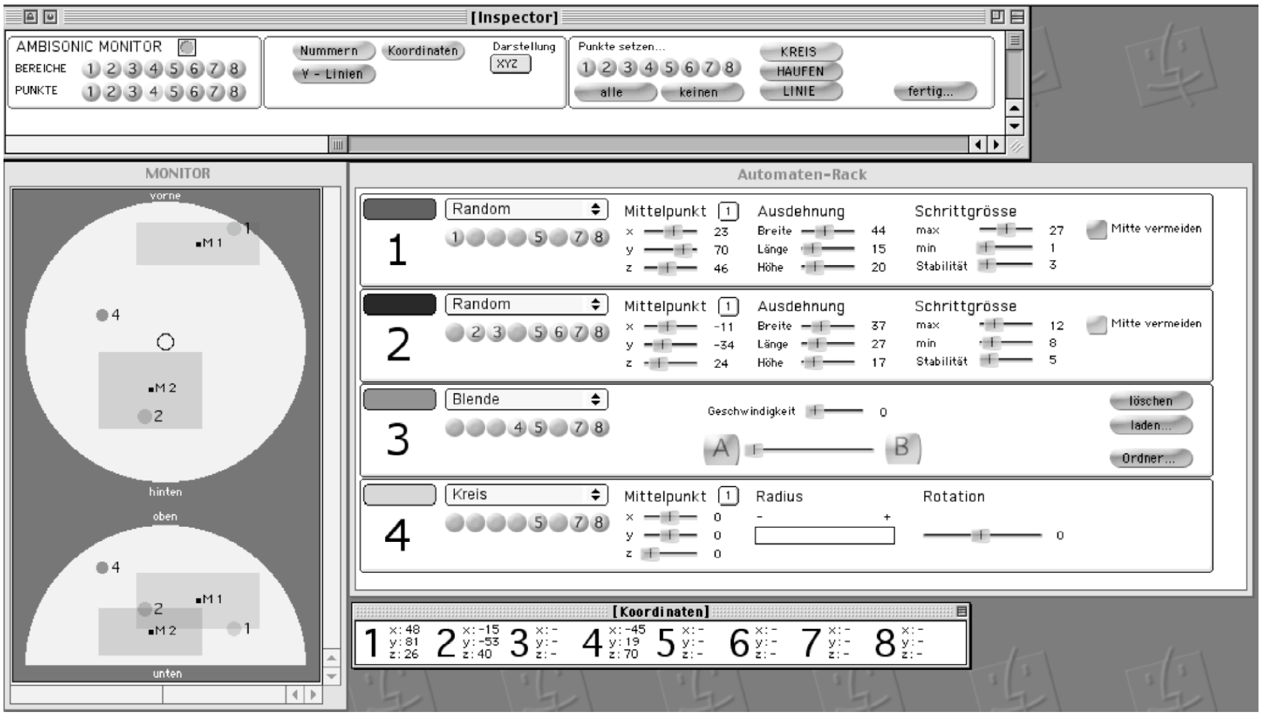

Figure 3. A screen shot of some elements of a graphical environment for interactive composition and performance with ambisonics.

To the left of the interface is the visual representation of the position of up to eight sounds within a spherical space. The circle above shows their position in the horizontal plane looking from above, the half-circle below their height looking along the axis vorne/hinten. The points are individual sounds, the rectangles shows groups of sounds which will move together. Each of the numbered boxes to the right is a virtual “device” which can be programmed to carry out a specific kind of movement for up to eight different sounds over time (there can be a total of eight of these “devices” active simultaneously). Three types of movement are shown here: Random (the one to eight selected soundstreams change position randomly), Blende (the selcted soundstreams move gradually from one defined position to a second position) and Kreis (the selected soundstreams move in a circle together). Each “device” offers numerous control parameters for the basic movement. The panel at the top of the figure allows basic configuration of the sounds in space. The lowest panel shows the coordinates of the soundstreams at each moment. At the moment of writing, this interface is being completed and sent to several composers for beta-testing. After beta-testing it will be available on the Internet.

III. Prospects for Future Work

Two smaller projects are planned for the academic year 2003/04. The first is another plug-in which will complement and extend the current group of plug-ins. All our ambisonic work until now has assumed that the composer places his or her material in non-reverberant space, so to speak in open air. The plug-in to be written in the Winter Semester 2003/04 will allow a composer first to design a (closed) three-dimensional space by defining shape, dimensions and characteristics of the walls and then to describe the position or movement of a sound within this space. The program will calculate not only the B-format representation of the direct sound but also that of the first three echoes from the walls. These echoes are perceptually of great importance for our ability to localize sound in space, and we expect that localization will improve greatly thanks to this treatment. A second project, to be realized in the Summer Semester 2004, is to write an independent program (i.e. not a plug-in) for ambisonic treatment. Part of the project will be to define a new file format, containing both a monophonic sound and the information about its position and movement in three-dimensional space. The program will then calculate the ambisonic signal without having to store the many channels of the B-format representation. The composer can change the sound's movement interactively and save (or not) the new patterns of movement with the sound.

We hope to organize a larger-scale project in collaboration with York University for the year 2004/05, whose goal would be to create a rich compositional environment in ambisonics. There have been remarkable technical advances in sound spatialization in the world during the last year, including a report of 15^{th}-order ambisonics and a paper introducing a general theory for calculating perfect three-dimensional representation of sound on the basis of simple and straightforward recordings. Our theoretical background needs to be brought up to date. In addition, we need urgently to know more about the perception of ambisonic sound. How good is the localization? How great is the frequency dependency of one's perception? What are optimal loudspeaker configurations for the different orders of ambisonics? Perhaps most importantly, how does surround sound differ aesthetically from sound presented frontally? This psychoacoustic knowledge needs to be incorporated into the next generation of compositional tools. Such a project would take two or three years of time and would involve psychologists, engineers, physicists, programmers and of course musicians.

IV. Aesthetic Considerations

Finally, it seems appropriate to consider briefly the aesthetic consequences of composing with surround sound. When we say something important, we do not stand behind the addressee and speak softly, we stand in front and speak clearly. Surround sound is more closely related to ambient sound than to speaking clearly in front of someone. We tend to disregard ambient sound in everyday life, monitoring it in the background for signs of danger. On the other hand, "electronic" sounds whose source we can neither see nor identify, tend to elicit greater alertness in the listener. The elements of surround sound—seemingly real three-dimensional spaces, invisible sounds, motion in space—have traditionally been used by the perception to warn of danger. We have very little experience with their aesthetic heightening, nor do we know what connotations the manipulation of these primary perceptual elements will awaken in listeners. Grafting complex intellectual interpretive mechanisms onto reactions of the subconscious nervous system as old as mankind itself will definitely enrich music. It is too early to say just how.

By dissolving the traditional frontal orientation of musical discourse, surround sound, and in particular ambisonics, will certainly accelerate the development of new modes of listening. But the dissolution will also turn inward, changing the languages of music and the ways in which music speaks to us. In the music of the past, space has been imaginary, using for example an abrupt modulation or a change of instrumentation as metaphors for distance. In surround sound, space and movement become real for the perception. We have yet to discover the emotional realities for which they will become metaphors.