A Poor Man's Techniques of Sound Diffusion

Gerald Bennett

In April 1994, I invited what was then called the GMEB to Zurich for a concert with the Gmebaphone. I had heard various incarnations of the Gmebaphone over the years at the festival in Bourges, and I liked both the idea and the realization of an instrument for interpreting electroacoustic music, even if I was often daunted by the difficulty of actually making the various versions work (I remember the beautiful copper control desk with no indication whatsoever of what something might mean). I had also heard the then newest Gmebaphone twice outside Bourges, once in Madrid in a round, highly reverberant space, where I was impressed at how well it presented music in that difficult hall, and once in a very dry theater in Geneva, where I thought the results were all right but not spectacular.

The concert in Zurich was in the Musikhochschule in a traditional rectangular hall with a quite high ceiling. The hall is used for concerts ranging from solo to chamber orchestra combinations. It is fairly reverberant. The Gmebaphone performance there was perhaps the most beautiful concert of any kind – electroacoustic or otherwise – that I have ever heard. The sound itself had a marvelous golden quality while remaining absolutely transparent and clear, and the plasticity of the sound – its three-dimensional quality – was perfectly remarkable, without being in the least dramatic or anecdotal. This concert convinced me that I should be paying much more attention to how my music is conceived in terms of space.

On the other hand, neither I personally nor the Swiss Center for Computer Music will ever have the financial means to own a Gmebaphone or any similarly conceived instrument. On those few occasions when my electroacoustic music is played, it is frequently on systems I do not know and over which I have no control. Usually there is a stereo signal being fed to four loudspeakers placed in the four corners of the hall. In any case, I am interested in having as much spatial information as possible on my tape in as robust a form as possible. There is a big drawback to this point of view: the more detailed the information on the tape about the projection of the sound in space, the less freedom of interpretation there can be in the concert. I have decided to accept this disadvantage in view of the advantages of the careful realization of the original tape. Furthermore, if I know my music will be played more or less as I have heard it, I will have more incentive to include the placing of the music in a virtual space in my compositional ideas.

I would like to speak very briefly about three models for spatial representation of sound. In all three cases, I am thinking of techniques to incorporate into the composition of music for tape, but obviously the same techniques can be used to perform live electroacoustic music. None of this work is my own in any important sense. All the research leading to the more interesting of these techniques has been and is being done by others. I will illustrate a few very simple techniques which I consider of great importance for the diffusion of sound, but more important, I wish to speak as a composer about ways of thinking about preparing sound for diffusion. A short bibliography at the end of this article will lead the interested reader to more detailed information.

("Diffusion" is, of course, not the proper English translation for the French word of the same spelling. I don't seem to find an idiomatic expression signifying "sound projection" and the like, so, for want of a better word, I shall continue to speak of sound diffusion in this text.)

Why am I as a composer interested in including spatial information in my electroacoustic music? The first reason is to give sonic clarity to polyphonic structures. If I can localize musical structures within more or less clearly defined imaginary planes or spaces, the listener can follow more easily complex textures with several independently evolving voices. Here the spatialization is related to classical techniques of orchestration, where choices of instruments and their doublings are made primarily to achieve clarity of the musical line and not in the first place for reasons of timbre. A second reason for including spatial information in the composition itself is that spatial elements like depth, distance and proximity can have important metaphorical significance and can thus contribute importantly to the expressivity of our music. There is a third reason, which I call anecdotal and which does not interest me at all. Here I include all the dramatic uses of space where the sound zooms around in order to amaze us or tell little stories.

The first model of spatialization I would like to discuss is the classical two-dimensional stereo model. In this model, the musical space is considered to be a room of some size, opening through two windows (the loudspeakers) into a smaller room where the listener is, and about whose acoustics the composer knows nothing. Here the composer shapes the sound in the large room and simulates in two channels its arrival at the loudspeaker-windows. This model seems to be quite primitive, but it is already much better at giving a sense of depth to the sound that the usual amplitude panning (see Figure 1).

Amplitude panning is the absolute worst way to try to convey a sense of space. Its greatest failing is that panning a monophonic signal across two loudspeakers by changing the relative amplitudes of the two speakers leaves the signal phase-locked in both speakers, an impossible situation in the real world. In addition, it is difficult to avoid the famous "hole-in-the-middle" phenomenon. Quadraphonic or multi-channel sound systems which only pan amplitude between speakers may succeed in enlivening the sonic image but can never provide coherent spatial imagery of a sound.

What are the basic elements of the physical model? Returning to Figure 1, what is the difference between a sound at point A and the same sound at point B?

Let us imagine first the walls of the large room are completely absorbent and that we only hear the direct sound. The two main differences will be that the perceived amplitudes of the two signals will vary in as a function of the distances of their imagined sources and that the arrival times of the signals at the two speaker/windows will vary. What is the amplitude variation? One would think that the amplitude would decrease inversely with the square of this distance, but we have found in informal listening tests the relationship 1/d2.5 to be more satisfactory, for reasons we cannot explain. One could argue that linear amplitude change would work just as well, as long as one is not interested in modeling real spaces, but in fact, if one is interested in simulating linearly perceived movement of sound in space, a rule on the order of the one above must be used.

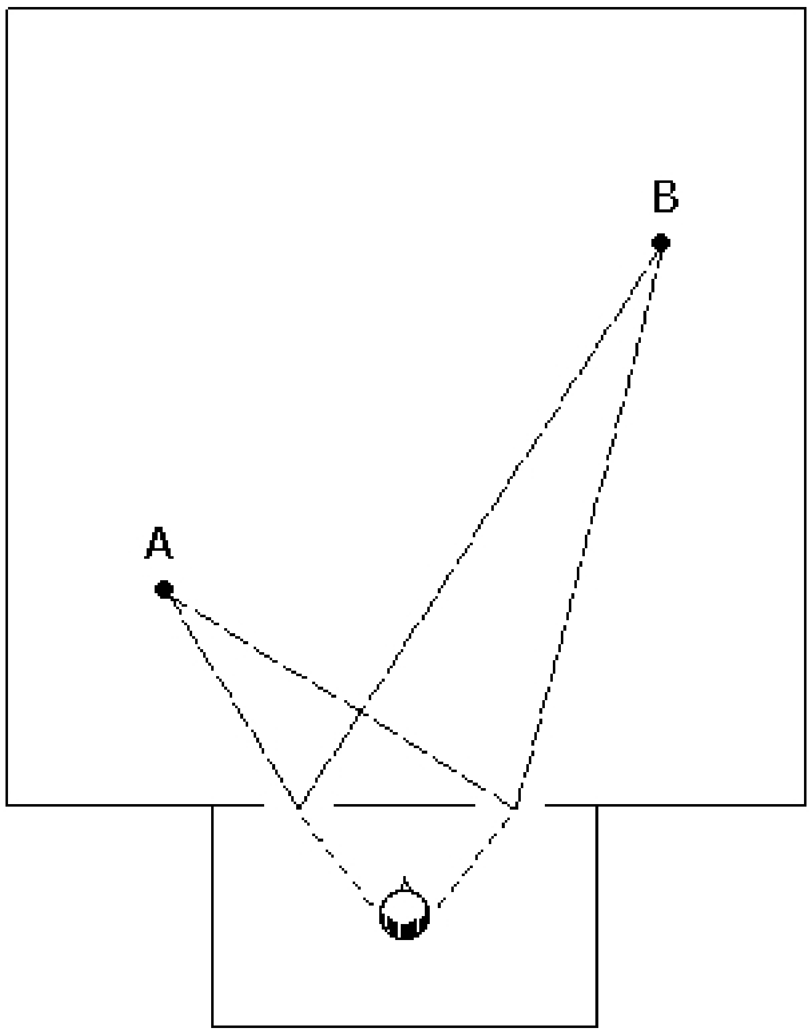

Figure1 . The virtual two-dimensional space used for simple simulations of placement and movement of sound. The sound from the larger space is heard through two "windows" (loudspeakers). Nothing is assumed about the acoustics of the listening space. The illustration shows how the direct sound from points A and B arrives at the "windows" with different angles of incidence and different time delays.

The calculation of the arrival time differences is very straightforward: one calculated the distance of the sound from each of the two speaker/windows in meters and divides these distances by the speed of sound, 340 m/s. Calculating the arrival times may seem like an unnecessary pedantry, but except when the sound is midway between the two speakers, delaying the signal by the appropriate amount in each loudspeaker decorrelates the two channels. In my experience, the phase decorrelation is the most important single perceptual aspect of spatialization.

I think this very simple model should be the very barest minimum of treatment used by any composer doing electroacoustic music. Amplitude panning should be strictly forbidden, and the signals in the two loudspeakers should always be decorrelated with relation to each other. I give in an appendix to this article two Csound instruments to realize this primitive spatialization, one for simply positioning sounds in space, the other for moving a sound from one point to another.

How can this physical model be improved? There are some easy additions which make it considerably better. The first is to correct for the hole in the middle (here a non-hole in the middle) by adding a third channel to both other channels. The second, more important correction is to add reverberation. Because the amplitude of reverberation depends primarily on the amplitude of a sound in a room but not on its distance, the ratio of direct to reverberated sound is an important perceptual cue for a sound's distance. A simple change will not scale reverberation, a more complex model will scale it in proportion to distance, but with much less decrease for distance than is the case for the direct sound. It is important that the reverberation be decorrelated between the two channels. A third, very easy correction is to correct for the absorption of high frequency energy by the air. Here the more distant the sound from the speaker/windows, the less high frequency energy it contains at the windows. There is real data on how to do this, but we do a very primitive and doubtless exaggerated correction to calculate the cut-off-frequency of a low pass filter in Csound. The final important improvement to this physical model is less straightforward than the others. It is to calculate the first reflections from the walls of the imaginary room based on the sound's position in two- or three-dimensional space and to reverberate each reflection separately and slightly differently. The reflections will reach the speaker/windows at different times and with different amplitudes and spectra as a function of the distance traveled by each reflection, but the reverberation should be sent in equal measure (but decorrelated!) to both channels. This simulation works well for rooms of moderate size where the first (and possibly second) reflections arrive at the speaker/windows during the first 100 milliseconds or so, but calculating first and second reflection data for moving sounds slows things down enough that I do not make use of this simulation.

A second model which I would like to consider briefly is the psychoacoustic or virtual acoustical model. The goal of this model is to represent the entire three-dimensional space surrounding the listener by transforming one's signal in accordance with what we know about the psychoacoustic cues for distance, azimuth and elevation of the sound source. This model functions best when listened to over headphones, and so it is not really of interest when we are talking about sound diffusion, but let us look briefly at the components of the model anyway.

Besides the cues for distance (amplitude and arrival time difference), the psychoacoustic model essentially adds cues for direction. These cues are spectral in nature. For listening with headphones, the model simulates interaural intensity differences and interaural time differences for each ear, which was not necessary with the physical model described above. The spectral cues depend upon the filtering effect on the signal of the outer ear, the shoulders and the upper torso. There is considerable measurement data of such head-related transfer functions (HRTF's) available (data for the azimuth cue has been incorporated into the program SoundHack, for instance), and modifying the signal with these functions helps a great deal in creating the illusion of a three-dimensional surround sound field. Although the simulation of actual HRTF's is computationally expensive, modeling only the perceptually most salient features (like the migration of the "8-Khz notch" as a function of perceived height) is not terribly difficult with the tools available in Csound. It is important to "directionalize" not only the direct sound but also as many of the early reflections of the sound as possible. A conscientious model calculates and "directionalizes" as many of these reflections as possible and reverberates each one separately.

Finally, and very important for virtual reality situations, the position of the listener's head should be monitored and the entire sound field rotated and skewed to compensate for any head movement.

Obviously, the virtual acoustic model is not practical for sound diffusion in concert, on the one hand because the fine details calculated for the sound as it arrives at the listener's ear get lost in the acoustics of the particular room in which one listens, on the other because of the dependence of the sound image on head position. Nonetheless, the incorporation of head-related transfer functions and the calculation of early reflections before actual reverberation can both be realized in the physical model described above and can contribute greatly to giving the sound a three-dimensional quality.

The final model, and the one that interests me most for future work, is the full surround-sound model where the sound field itself is stable and consistent over a large part of the listening space. There are several models to choose from here, whose technologies are both proprietary and in the public domain, and there is no reason why future research cannot lead to many more satisfactory solutions for surround sound.

The model I am best acquainted with, and the one I know how to realize in Csound, always an important criterion for my own work, is the ambisonic representation of three-dimensional sound. I first encountered ambisonics in 1979, when I used a Calrec microphone to record the tape part of my piece Aber die Namen der seltnen Orte und alles Schöne hatt' er behalten for Baritone, five instruments and tape. I placed 16 loudspeakers at various locations in the Espace de Projection at IRCAM, played my 16-track tape through them and recorded them with the Calrec ambisonic microphone, which preserved information about direction and distance. When played back through the Calrec decoder, the results were remarkable: the positions of the loudspeakers, their distances and directions were captured with great precision by one microphone. Since 1979, information about ambisonic theory has become more generally available, and it has become possible to use the techniques for the preparation of tapes without using the Calrec microphone. I cannot give a account of ambisonic theory here – there is plenty of information available – but I would like to summarize those features of ambisonics which make it interesting for me as a composer.

Ambisonics is an extension of the M-S stereo recording technique developed by Alan Blumlein in the 1930's. Ambisonics requires four channels to transmit information about sound pressure levels in three perpendicular planes as well as the sound pressure level of the omnidirectional signal (three channels suffice if only the full 360-degree horizontal plane is required). The information can be very easily synthesized. When played back, the four ambisonic channels are sent through simple psychoacoustic filters (shelf filters) and put through a simple amplitude matrix to combine the signals. One of the interesting things about ambisonics besides its very robust sound field is that the information in the four channels is only translated into information for a specific number of loudspeakers at definite locations at the very last moment. In principle, the same ambisonically coded four-channel tape can serve for diffusion in a situation which only reproduces the horizontal plane with four loudspeakers, in a situation reproducing full periphony with six loudspeakers, or a situation with periphony played by 18 (or any other number of) loudspeakers. The only requirement is that the decoder used know how to convert the signal for the set-up desired. I have written Csound programs which allowing placing a monophonic sound in ambisonic space, providing the four-channel information for the ambisonic coding and decoding, and although I have not yet used the programs myself, a student of mine wrote an entire opera's worth of ambisonic music using them, and the results sound very good. I shall continue to work in this direction, although I do not for a moment suggest that ambisonics is the only interesting technique for achieving three-dimensional surround sound. With the advent of the Digital Versatile Disk (DVD) format, it will be possible to store four channels and more of information on CDs for the same amount of music as today. Ambisonic decoders may well be included in many CD players as standard equipment. This will provide a very robust way for composers to have full compositional control over the entire periphonic sound field and to reproduce the sound field in concert with great fidelity under a great number of different conditions.

A Very Brief Introductory Bibliography

Begault, D. 1991. Challenges to successful implementation of 3-D sound. Journal of the Audio Engineering Society 39(11), 864-870. (Head-realated transfer functions)

Chowning, J. 1971. The simulation of moving sound sources. Journal of the Audio Engineering Society 19, 2-6.

Computer Music Journal 19:4–Winter 1995. Contains special sections on Sound Localization and Sound Spatialization with good articles by Gary Kendall, David Malham and Andrew Myatt.

Kendall, G. S., W. Martens & S. L. Decker. 1989. Spatial reverberation: Discussion and demonstration. In Mathews, M. V. & J. Pierce. Current directions in computer music research. Cambridge, Mass., pp. 65-87.

Malham, D. G. 1995. Basic ambisonics. http://www.york.ac.uk/inst/mustech/3d_audio/ambsnlct.htm

Moore, F. R. 1989. Spatialization of sound over loudspeakers. In Mathews, M. V. & J. Pierce. Current directions in computer music research. Cambridge, Mass. , pp. 89-103.

Moorer, James A. 1985. About this reverberation business. In Roads, C. & J. Strawn. Foundations of Computer Music. Cambridge, Mass.

Appendix

The following are two Csound programs for placing and moving sound in an imaginary 2-dimensional space.

```
; putsound.orc
; this places a sound at a specified point in two-dimensional space

; p4        soundin file number
; p5        size of one side of room in meters
; p6, p7    x and y coordinates of the point between 0.0 and 1.0
;(x=0 all left, x=1 all right, y=0 all front, y=1 all back)
; p8        distance of the loudspeakers from one another in meters

sr    =     44100
kr    =     441
ksmps =     100
nchnls =    2

instr 1

isize       =     p5
ix          =     p6 * isize
iy          =     p7 * isize

ivelair     =     340         ; velocity of sound in air
ilspeaker   =     (isize - p8) / 2  ; position of left speaker on x-axis
irspeaker   =     isize - ilspeaker ; position of right speaker on x-axis
ifactor     =     1.618

; calculate distance of point from each speaker
idlsq =     (ilspeaker - ix) * (ilspeaker - ix) + (iy * iy) ;Pythagoras
idrsq =     (irspeaker - ix) * (irspeaker - ix) + (iy * iy)
idisl =     sqrt(idlsq)
idisr =     sqrt(idrsq)

;calculate time delay to each speaker
idelayl     =     idisl / ivelair
idelayr     =     idisr / ivelair

; get term for amplitude correction
iterml      ipow  idisl, 2.5
itermr      ipow  idisr, 2.5
iterml      =     1 / iterml
itermr      =     1 / itermr

; get cut-off points for low-pass filter
ifilterl    =     50000 / (ifactor * idisl)
ifilterr    =     50000 / (ifactor * idisr)

; get some sound
a1          soundin     p4

aleft delay a1, idelayl       ; the delayed signal to left speaker
aright      delay a1, idelayr       ; the delayed signal to right speaker
```

```
;correct amplitude for distance
aleft       =     aleft * iterml
aright      =     aright * itermr

; correct for loss of high frequencies through air absorption
aleft       tone  aleft, ifilterl
aleft       tone  aleft, ifilterl
aright      tone  aright, ifilterr
aright      tone  aright, ifilterr

      outs1 aleft
      outs2 aright

endin
```

```
; putsound.sco
; this places a sound at a specified point in two-dimensional space

; p4        soundin file number
; p5        size of one side of room in meters
; p6, p7    x and y coordinates of the point between 0.0 and 1.0
;(x=0 all left, x=1 all right, y=0 all front, y=1 all back)
; p8        distance of the loudspeakers from one another in meters

;                 p4    p5    p6    p7    p8

i1    0     2     99    20    .3    .8    5
i1    +     2     99    20    .8    .3    5
i1    +     2     99    20    .6    .7    5
i1    +     2     99    20    .1    .5    5
i1    +     2     99    20    .9    .1    5
i1    +     2     99    20    .1    .9    5
i1    +     2     99    20    .4    .6    5
i1    +     2     99    20    .5    .5    5
e

=============================================
```

```
; movesound.orc
; this moves a sound from point 1 to point 2

; p4        soundin file number
; p5        size of one side of room in meters
; p6, p7    x and y of point 1 from 0.0 (all left) to 1.0 (all right)
; p8, p9    x and y of point 2 from 0.0 (all front) to 1.0 (all back)
;(x=0 all left, x=1 all right, y=0 all front, y=1 all back)
; p10       distance of the loudspeakers from one another in meters

; in this program, all the calculations are made in a-time to avoid noise
; when the movement is fast or over a large distance

; the movement is only calculated in two dimensions

sr    =     44100
kr    =     441
ksmps =     100
nchnls =    2

instr 1

isize       =     p5
ixstart     =     p6 * isize  ; convert coordinates to meters
ixend       =     p7 * isize
iystart     =     p8 * isize
iyend       =     p9 * isize

ivelair     =     340               ; velocity of sound in air
imax        =     32768
ilspeaker   =     (isize - p10) / 2 ; position of left speaker on x-axis
irspeaker   =     isize - ilspeaker ; position of right speaker on x-axis
ifactor     =     1.618

; get the vectors from point 1 to point 2
ix    =     ixend - ixstart
iy    =     iyend - iystart

; get some sound
a1          soundin     p4

; a counter for the duration of the note
atime       phasor      1/p3

; the coordinates of the moving point
ax    =     ixstart + atime*ix
ay    =     iystart + atime*iy

; calculate distance of point from each speaker
adlsq =     (ilspeaker - ax) * (ilspeaker - ax) + (ay * ay) ;Pythagoras
adrsq =     (irspeaker - ax) * (irspeaker - ax) + (ay * ay)
adisl =     sqrt(adlsq)
adisr =     sqrt(adrsq)
```

```
;calculate and create time delay to each speaker
adelayl     =     adisl / ivelair
adelayr     =     adisr / ivelair
adum  delayr      1
aleft deltapi     adelayl     ; the delayed signal to left speaker
aright      deltapi     adelayr     ; the delayed signal to right speaker
            delayw      a1

; avoid taking the 2.5th power of the distance every sample
kdisl downsamp    adisl
kdisr downsamp    adisr
kterml      kpow        (kdisl), 2.5
ktermr      kpow        (kdisr), 2.5

;correct amplitude for distance
aleft =     aleft * (1 / kterml)
aright      =     aright * (1 / ktermr)

; correct for loss of high frequencies through air absorption
aleft tone  aleft, (50000 / (ifactor * kdisl))  ; quick and dirty
aleft tone  aleft, (50000 / (ifactor * kdisl))
aright      tone  aright, (50000 / (ifactor * kdisr))
aright      tone  aright, (50000 / (ifactor * kdisr))

      outs1 aleft
      outs2 aright

endin
```

```
; movesound.sco
; this moves a sound from point 1 to point 2

; p4        soundin file number
; p5        size of one side of room in meters
; p6, p7    x and y of point 1 from 0.0 (all left) to 1.0 (all right)
; p8, p9    x and y of point 2 from 0.0 (all front) to 1.0 (all back)
; p10       distance of the loudspeakers from one another in meters


;                 p4    p5    p6    p7    p8    p9    p10   p11

i2    0     10    99    10    .5    .1    .5    .9    3     1.5
e
```

This text appeared as “A poor man's techniques of sound diffusion“ in G. Bennett & F. Barrière (Ed), *Composition / Diffusion in Electroacoustic Music. Proceedings of the International Academy of Electroacoustic Music 1997.* Bourges 1998.