An Interactive Real-Time System for the Control of Sound Localization

Author(s): Marina Bosi
Source: *Computer Music Journal*, Winter, 1990, Vol. 14, No. 4 (Winter, 1990), pp. 59-64
Published by: The MIT Press
Stable URL: https://www.jstor.org/stable/3680791

Marina Bosi
Center for Computer Research in Music and Acoustics (CCRMA)
Department of Music, Stanford University
Stanford, California 94305 USA
mab@ccrma.stanford.edu

An Interactive Real-time System for the Control of Sound Localization

One of the most significant parameters in the perception of music is related to the spatial characteristics of sound. Aside from the use of directionality in the reproduction of music in sound systems of homes, movies theatres, and concert halls, composers have shown interest in exploitation of the spatial dimension of sound since as early as the sixteenth century. We have precedents in works by composers such as Gabrieli, Berlioz, and, after the turn of the century, by Varèse in *Déserts,* and by Stockhausen in *Gesang der Jünglinge* (Stockhausen 1961). The landmark after the advent of the electroacoustic music is probably Chowning's *Turenas,* in which the composer uses graphic techniques to control the path and speed of sound sources as well as reverberation in the surrounding space.

In this paper I will discuss the implementation of a tool that allows the composer to take advantage of the spatial control of sound through a portable, interactive real-time system having adequate graphics and computational ability.

Sound Localization: Acoustics and Psychoacoustics

In order for localization to occur, the listener needs to define the angular location of the sound source as well as the distance of the source relative to the listener. Our ability to localize sounds is influenced by many factors. On the one hand, the perception of sound direction based only on auditory information depends on the mechanism adopted by our brain to decode the sound stimuli. A striking example is the lateralization phenomenon. On the other hand, the characteristics of the listening space (e.g., shape, size, absorption coefficient) and the characteristics of the sound source are directly related to the physics of the event. To give the illusion of spatial sound images, we need to bring the psychological and physical aspects of the acoustical event into relation with the technology and the system used to reproduce the sound.

In this paper I will consider a two-dimensional model whose angular location is defined by azimuth (horizontal plane). From previous works on localization, we know that the azimuthal localization cues—interaural time difference (ITD) and interaural intensity difference (IID)—depend on the spectrum of the sound source (Fedderson et al. 1957). For wavelengths shorter than the interaural distance, IIDs are dominant cues, while for wavelengths longer than the interaural distance, ITDs in arrival time of features of the sound waveform are dominant cues. It is possible to simulate the auditory system cues through headphones (Kendall and Rodgers 1981). However, since they are related to the precise location of the listener, it is unrealistic to try to reproduce these cues in a concert situation, where the precise location of the listener is unpredictable.

With regard to the distance cue, solving the wave equation in a free field, we know that the amplitude of the variation of the sound pressure is inversely proportional to the distance, r, between the source and the listener. Furthermore, for large distances ($r \geq 100$ m), we need to take into account the attenuation of the signal in the air that depends on the frequency of the source, as well as temperature, humidity, and pressure. (See Table 1.) This introduces an exponential decay of the form e^{-kr}. We have chosen to reproduce this data from Harris (1962), since erroneous data from Kuttruff (1973) has been widely distributed in the computer music field: e.g., Moorer (1979).

In enclosed environments, intensity differences alone are only a relative cue to distance perception. It is the contribution of the reverberant energy that determines our ability to make distance discrimi-

Computer Music Journal, Vol. 14, No. 4, Winter 1990,

Fig. 1. Loudspeaker configuration.

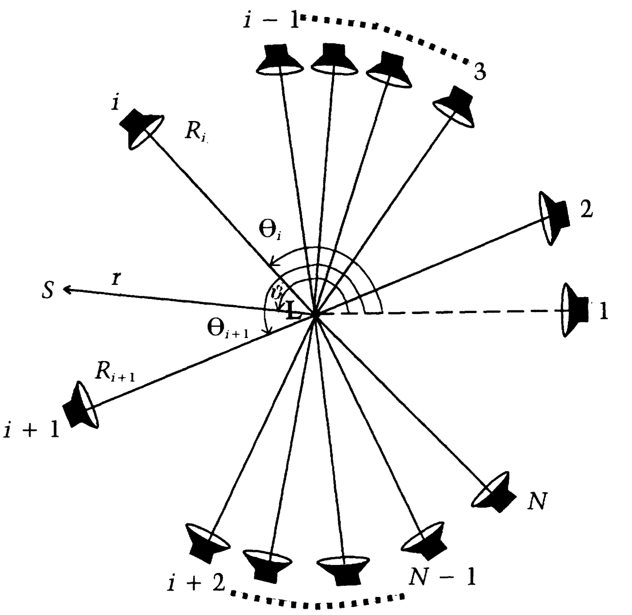


Table 1. Attenuation constant of air (m^{-1}) at 20°C and normal atmospheric pressure

	Frequency (Hz)				
Humidity	*2000*	*3200*	*4000*	*5000*	*6400*
40%	0.0032	0.0052	0.0072	0.0112	0.0172
50%	0.0028	0.0044	0.0060	0.0092	0.0144
60%	0.0028	0.0040	0.0056	0.0076	0.0120
70%	0.0028	0.0036	0.0052	0.0068	0.0100

nations (Sheeline 1982). The reverberant energy, which is the total energy minus the direct source energy, supplies the room information, giving general cues as to size, shape, and material construction. Considering the early reflection paths of the sound source, we can get a more accurate picture of the acoustical phenomena in the room. We might expect that the reflected sound represents a potential source of confusion in the localization of the sound source; however, the reflected energy returned to the auditor within the following 50 ms is integrated with the direct energy into the impression of a single fuller, appropriately localized sound. The accurate localization of a sound in spite of seemingly conflicting cues from reflections is sometimes called "law of the first wavefront" and sometimes the "precedence effect" (Wallach, Newman, and Rosenzweig 1949; Haas 1951). When the sound energy is distributed between speaker pairs a phantom sound image can be perceived at a different location than the physical sound sources. If two loudspeakers with the same acoustic outputs are placed at equal but opposite angles from the listener on either side of the median plane, the listener perceives an apparent single image located in the midway position between the speakers. Moreover, delay of the signal to one speaker or change of the distance of one speaker from the listener will cause the same type of illusion (Gardner 1973). If the intensity applied to the loudspeaker pair is not the same for each speaker, this can cause the illusion of an angular displacement in the sound image.

Overview of the Model

Our goal is to attain control over the spatial characteristics of musical sound through a system that allows real-time interaction and that would be portable and flexible enough to be used in many different situations (see Moore 1983). We need a general way to obtain reasonable-sounding values for the processing algorithm. The model described here is based on the following elements: the position of the sound source within the illusory space relative to the listener, and the number of loudspeakers, and their locations, as well as the characteristics of the listening space, such as shape, size, absorption.

The configuration of the variable number of loudspeakers can be arbitrarily chosen in order to reproduce any possible "real" configuration (for example, an open-air concert situation where the loudspeakers are placed in a semicircle in front of the public, the classical quad configuration, or any other precise location).

Consider the system configuration of Fig. 1. We want to simulate a source located at point S. We choose to represent point S in polar coordinates,

r, θ, where the origin is at the listener L and θ is the angle relative to speaker 1. The listener is surrounded by N loudspeakers numbered counterclockwise. The location of each speaker i can be represented by the doublet (R_i, Θ_i).

We know that in a free field for small distances, the intensity I relative to a sound source that irradiates omnidirectionally is

$$I = \frac{I_0}{r^2},$$

where I_0 is the intensity at the distance $r_0 \equiv 1$. For $r \leq r_0$ our system will deliver the maximum sound intensity level corresponding to the intensity I_0. For large distances we need to consider the additional factor in the dependence of the intensity on the distance r.

$$I = \frac{I_0}{r^2} e^{-kr},$$

where the constant k depends on the absorption of sound in the air. At fixed humidity, pressure, and temperature values, the absorption coefficient k increases with the increasing frequency of the signal. The absorption in the air, whose effect is negligible for small distances (<100 m from Table 1), acts like a lowpass filter, highly attenuating the high frequencies for large distances.

As we noted before, sound parameters dependent on interaural time differences, hence dependent on the precise position and orientation of the listener head, are uncontrollable when music is produced through a multiple channel loudspeaker system. Each adjacent speaker pair, i and $i + 1$, defines a portion of the plane between the directions Θ_i and Θ_{i+1}. When the source is located in that portion of the plane, we can simulate the angular location by distributing the source intensity between the speaker pair and applying a zero intensity to all the other speakers. The intensity coming from each speaker direction is determined by the ratio of the angular position of the source relative to the speakers and the angular separation of the two speakers (Chowning 1971).

The fraction of the intensity supplied by each speaker pair i, $i + 1$ will be

$$p_i = \frac{\Theta_{i+1} - \theta}{\Theta_{i+1} - \Theta_i} = 1 - p_{i+1}$$

$$\Theta_i < \theta < \Theta_{i+1} \qquad i = 1, 2, \ldots, N - 1$$

$$p_N = \frac{2\pi - \theta}{2\pi - \Theta_N} = 1 - p_1$$

$$\Theta_N < \theta < 2\pi \qquad i = N,$$

where (R_i, Θ_i) defines the position of the loudspeaker i and $\Theta_1 \equiv 0$. The first formula indicates that when the sound source is located at $\theta = \Theta_i$ $(i < N)$ the fractional intensities are $p_i = 1$ and $p_{i+1} = 0$, showing that the intensity I is entirely delivered by the speaker i.

As the source moves toward the speaker $i + 1$, the amount of intensity delivered by the speaker i decreases linearly with the angle θ, while the amount of intensity delivered by the speaker $i + 1$ will increase linearly with the angle θ up to the point where $\theta = \Theta_{i+1}$ when $p_i = 0$ and $p_{i+1} = 1$ (i.e., the intensity is entirely delivered by the speaker $i + 1$). In the case of the speaker pair N and 1, we are in the boundary conditions expressed by the second formula. In this case, since we are "wrapping around," we need to add 2π to the angle Θ_1, but the same considerations as above are valid.

To take into account that the speakers in general will be at different distance from the listener we also need to multiply the source intensity by the factor

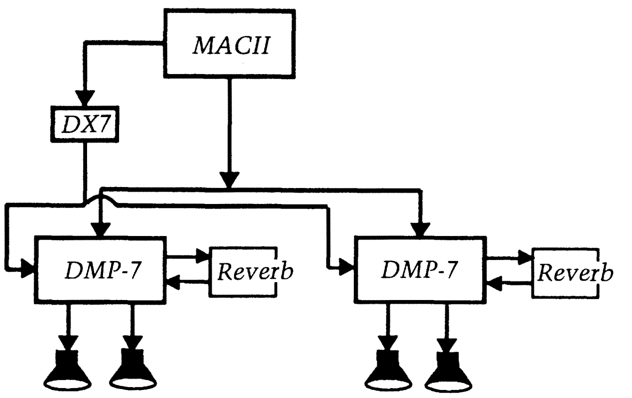
$$C_i = \left(\frac{R_i}{R_{max}}\right)^2$$

Notice if one speaker is farther away than another we make that speaker louder to correct for it, and vice versa.

A variation in the position r of the source will affect its frequency through the Doppler effect. We need to consider the corrected frequency

$$f' = f \frac{c}{c + \hat{r} \cdot \vec{v}_s},$$

Fig. 2. System schematic.



where f is the frequency of the motionless source, c is the speed of the sound in the air, $\hat{r}$ is the unit vector in the radial direction, and $\vec{v}_s = d\vec{r}/dt$ is the velocity of the sound source. We choose the origin of our reference system where the "ideal" listener is located; therefore the component of the source velocity along the unit vector $\hat{r} \equiv \vec{r}/r$ will be negative when the source moves toward the listener (i.e., the frequency will increase) and positive when the source moves away (i.e., the frequency will decrease).

To summarize, given the position of the speakers, R_i, Θ_i, and, at each instant, the position of the source, r, θ, and its velocity, $\vec{v}_s$, we can determine at each instant the corrected frequency, f', and the relative intensity values delivered by the active channels i and $i + 1$:

$$I_i = IC_i p_i = \frac{I_0}{r^2} e^{-kr} \left(\frac{R_i}{R_{max}}\right)^2 \frac{\Theta_{i+1} - \theta}{\Theta_{i+1} - \Theta_i},$$

$$I_{i+1} = IC_{i+1}\, p_{i+1} = \frac{I_0}{r^2} e^{-kr} \left(\frac{R_{i+1}}{r}\right)^2 \frac{\theta - \Theta_i}{\Theta_{i+1} - \Theta_i}.$$

We may easily extend this algorithm to the case of more than one sound source. Let's say we have $j = 1, 2, \ldots, M - 1$, M sources represented by the points r_j, θ_j. We would apply the results above to each source and superimpose at each instant the intensities $I_{j,i}$ and $I_{j,i+1}$ for each channel pair, i and $i + 1$, activated by every source j.

The reverberation plays an essential role in the perception of distance. In our model, we consider the reverberant energy to be spatially uniform; therefore the ratio of reverberant to direct sound energy reaching the listener will increase as the distance r between the sound source and the listener increases. The reverberation time of a listening space is directly related to the decay rate of the reverberant energy, and it depends on the characteristics of the physical space such as volume and material construction. The early reflections, which are separated by a time interval greater than 50 ms from the direct sound, vary with the distance and are primarily a function of the proximity and the characteristics of the room boundaries.

Implementation

Figure 2 shows the system implementation for a single source ($j = 1$) and four channels ($i = 1, 2, 3, 4$). A Macintosh II sends MIDI control signals to a Yamaha DMP-7 digital mixer, which provides modules for the spatial processing algorithm. MIDI control signals are sent as well to a synthesizer source (Yamaha DX7 at present). Once the position of the speakers is selected, as the user draws a sound trajectory on the screen with the mouse, the information about the position, r, θ, of the sound source with respect to the listener and the velocity, $\vec{v}_s$, are detected. The intensity of the source, $I(r)$, is computed and the active channel pair is determined based on the angular location of the source. We now can determine the normalized intensity supplied by each speaker in the active channel pair:

$$I'i = [I\ C_i\ p_i]/I_0,$$

$$I'_{i+1} = [I\ C_{i+1}\ p_{i+1}]/I_0,$$

where I_0 is the maximum intensity (i.e., the intensity when the distance between the source and the listener is $r = r_0 = 1$); C_i is the factor that takes into account differences in distance between source and speakers; and p_i is the panning factor.

The value I' for each speaker is then mapped to a MIDI number in the range 0–127 according to the DMP-7 fader control function. We have measured this function and found that the amplitude is linearly related to the MIDI number (the intensity is

Fig. 3. Measured correspondence between MIDI numbers and amplitude and intensity (normalized values).

Fig. 4. MIDI mapping of the intensity values.

Fig. 5. User interface.

Fig. 3

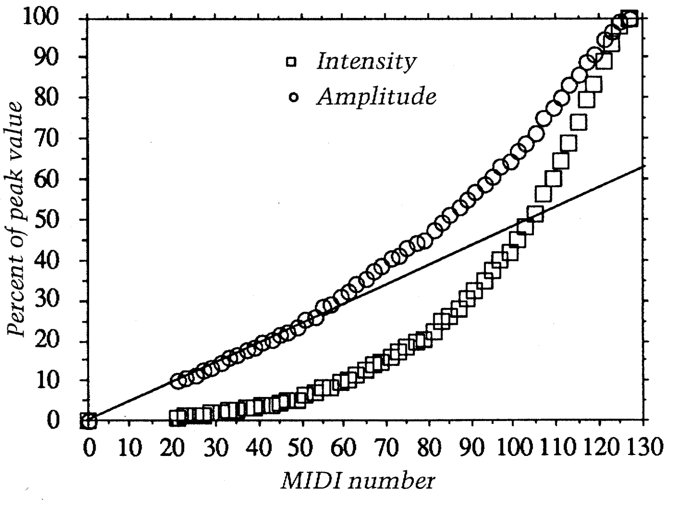


Fig. 5

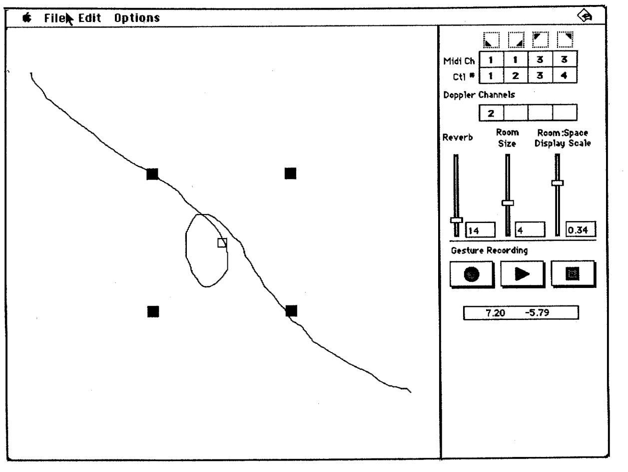
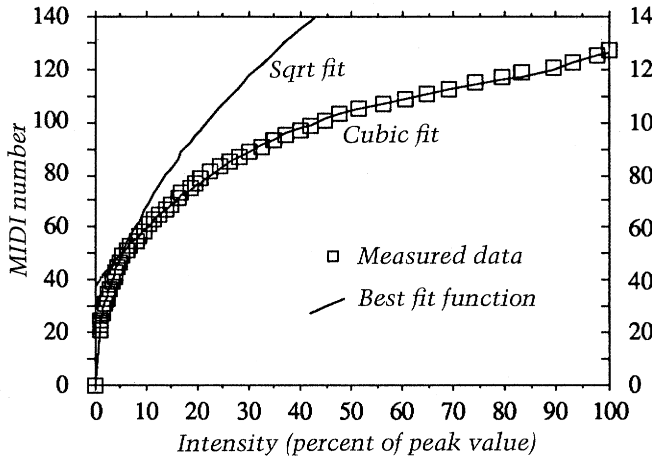


Fig. 4



quadratic) for MIDI numbers in the range 0–50 (Fig. 3).

$$\text{MIDI number} \equiv I'' = a\sqrt{I'}$$

for I' in the range of 0–10% of I_0, where least squares fitting yields $a = 21.45$. Beyond this range there is no such simple relation, so we resort to a polynomial fit. We find cubic order (Fig. 4) is enough.

The velocity of the sound source determines the Doppler correction factor g to the value of its frequency, where $g = c/(c + \hat{r} \cdot \vec{v}_x)$. If we limit the speed of the sound source $\hat{r} \cdot \vec{v}_s$ in the range 0, $c/2$, then the frequency of the sound source varies in the range $f - f/2$, $f + 2f$, (i.e., plus or minus an octave). This is mapped to a MIDI number affecting the pitch bend function of the synthesizer (Loy 1985).

The reverberation module used (the DMP-7 reverberation room effect at present) allows modification of the reverberation time and the setting of an initial delay that depends both on the source position and the listening space shape. A control over the early reflection patterns that depends on the size and shape of the room and the source position is also implemented.

The user interface is shown in Fig. 5. As a sound path is drawn on the screen (which corresponds to the illusory space), the displacement of the sound in the acoustical environment specified by the user can be heard. The control panel allows the user to select the amount of reverberation and the room size, as well as the display scale. The source position (x- and y-coordinates in meters) within the illusory space is indicated at each instant at the bottom of the control panel. It is possible to record sound trajectories and to create and edit files of recorded trajectories. The loudspeaker configuration within the space can be varied by pressing the option key while dragging the loudspeakers (indicated by black squares on the screen) with the mouse.

Results and Future Directions

The goal of this work is to provide the composer with a tool to project the musical sound into an illusory acoustical space using a variable number of speakers. The model implementation, realized in a real-time implementation for the Macintosh II and MIDI devices through an interactive software package written in the C language, allows a sound trajectory to be drawn on the screen and the movement of the sound to be heard (complete with Doppler shift) in real-time. This system was tested in different acoustical spaces using a quadraphonic reproduction system, and listeners found that it produces convincingly localized sound images.

Although the initial implementation allows for simple room geometries, one of our goals for future work is to take more accurate account of the features of the acoustical space. For instance we would like to include multiple ray paths for each source. We also would like to replace the use of the digital mixer with a spatial processing software for a specialized DSP chip like the Motorola DSP56000. The model presented could readily be extended to three dimensions.

Acknowledgments

The author wishes to thank the Accademia Nazionale dei Lincei, Fondazione I. Veneziani in Rome for its support of this research and John Chowning for being such a helpful guide in the conception and realization of this project. Thanks to Max Mathews and Chris Chafe for all the suggestions while developing this work; to Earle Schubert, Max Mathews, and Richard Goldberg for reading early versions of this article and providing valuable comments; as well as to Jay Kadis and all the people at CCRMA for their help. Particular thanks to David Zicarelli without whose assistance this work would not exist.

References

Chowning, J. M. 1971. "The Simulation of Moving Sound Sources." *Journal of the Audio Engineering Society* 19:2–6.

Fedderson, W. E., et al. 1957. "Localization of High-Frequency Tones." *Journal of the Acoustical Society of America* 29:988–991.

Gardner, M. B. 1973. "Some Single- and Multiple-Source Localization Effects." *Journal of the Audio Engineering Society* 21:430–437.

Haas, H. 1951. "Über den Einfluss eines Einfachechos auf die Horsamkeit von Sprache." *Acustica* 1:49–58.

Harris, C. M. 1962. "Absorption of Sound in Air in the Audio-Frequency Range." *Journal of the Acoustical Society of America* 35:11–17.

Hartmann, W. M. 1983. "Localization of Sound in Rooms." *Journal of the Acoustical Society of America* 74:1380–1391.

Kendall, G. S., and C. A. P. Rodgers. 1981. "The Simulation of Three-Dimensional Localization Cues for Headphone Listening." *Proceedings of the 1981 International Computer Music Conference.* San Francisco: Computer Music Association, pp. 225–243.

Kuttruff, H. 1973. *Room Acoustics.* New York: Wiley.

Loy, D. G. 1985. "Musician Make a Standard: The MIDI Phenomenon." *Computer Music Journal* 9(4):8–25. Reprinted in C. Roads, ed. 1989. *The Music Machine.* Cambridge, Massachusetts: MIT Press. pp. 181–198.

Moore, F. R. 1983. "A General Model for Spatial Processing of Sounds." *Computer Music Journal* 7(3):6–15. Reprinted in C. Roads, ed. 1989. *The Music Machine.* Cambridge, Massachusetts: MIT Press. pp. 559–568.

Moorer, J. A. 1979. "About This Reverberation Business." *Computer Music Journal* 3(2):13–28. Reprinted in C. Roads and J. Strawn, eds. 1985. *Foundations of Computer Music.* Cambridge, Massachusetts: MIT Press. pp. 605–639.

Sheeline, C. W. 1982. "An Investigation of the Effects of Direct and Reverberant Signal Interactions on Auditory Distance Perception." Stanford Department of Music Report STAN-M-13.

Stockhausen, K. 1961. "Music in Space." *Die Reihe, Vol. 5: Reports, Analyses.* Bryn Mawr, Pennsylvania: Theodore Presser Co., pp. 67–82.

Wallach, H., E. B. Newman, and M. R. Rosenzweig. 1949. "The Precedence Effect in Sound Localization." *American Journal of Psychology* 52:315–336.

YAMAHA Digital Mixer Processor DMP-7 Operation Manual.