[2] Paul Kabbash, William Buxton and Abigail Sellen, "Two-Handed Input in a Compound Task," *Proceedings of CHI '94*, pp 417-423, 1994.

[3] S. Sidney Fels and Geoffrey E. Hinton, "Glove-TalkII: Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls," *IEEE Transactions on Neural Networks*, Vol. 9, No. 1, pp. 205-212, 1998.

[4] S. Sidney Fels and Geoffrey E. Hinton, "Glove-Talk: a neural network interface between a data-glove and a speech synthesizer," *IEEE Transactions on Neural Networks*, Vol. 4, No. 1, pp. 2-8, 1993.

[5] Tinsley A. Galyean, "Sculpting: An Interactive Volumetric Modeling Technique," *ACM Computer Graphics*, Vol. 25, No. 4, (SIGGRAPH '91, Las Vegas, 28 July - 2 August 1991), pp. 267-274, 1991.

[6] Axel G. E. Mulder, "Getting a GRIP on alternate controllers: Addressing the variability of gestural expression in musical instrument design," *Leonardo Music Journal*, Vol. 6, pp. 33-40, 1996.

[7] Axel G. E. Mulder, "Hand gestures for HCI," *Technical Report, NSERC Hand Centered Studies of Human Movement project*, Burnaby, BC, Canada: Simon Fraser University, 1996. Available through the WWW at http://www.cs.sfu.ca/~amulder/personal/vmi/HCI-gestures.htm

[8] Axel G. E. Mulder, "Virtual Musical Instruments: Accessing the Sound Synthesis Universe as a Performer," *Proceedings of the First Brazilian Symposium on Computer Music*, (Caxambu, Minas Gerais, Brazil, 2-4 August 1994, during the 14th Annual Congres of the Brazilian Computer Society), pp. 243-250, Belo Horizonte, MG, Brazil: Universidade Federal de Minas Gerais, 1998. Available through the WWW at http://www.cs.sfu.ca/~amulder/personal/vmi/BSCM1.ps.Z

[9] Axel G. E. Mulder, "Human Movementq Tracking Technology," *Technical Report, NSERC Hand Centered Studies of Human Movement project*, Burnaby, BC, Canada: Simon Fraser University, 1994. Available through the WWW at http://www.cs.sfu.ca/~amulder/personal/vmi/HMTT.pub.html

[10] Miller Puckette, "FTS: A real time monitor for multiprocessor music synthesis," *Computer music journal*, Vol. 15, No. 3, pp. 58-67, 1991.

[11] Franc Solina and Ruzena Bajcsy, "Recovery of parametric models from range images: the case for superquadrics with global deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 2, February, pp. 131-147, 1990.

[12] David Wessel, "Timbre space as a musical control structure," *Computer music journal*, Vol. 3, No. 2, pp. 45-52, 1979.

Objects in a Virtual Space: A Comparative Analysis beetween Image and Sound Spatial Representation and Synthesis

S. NATKIN
CEDRIC, CONSERVATOIRE NATIONAL DES ARTS ET METIERS
292 rue St Martin 75141 Paris Cedex 03, France
natkin@cnam.fr

1. Goal of the research

This paper is a comparative analysis of the spatialisation process in virtual space simulation both from the visual and audio point of view. It is the starting point of a research project of the CEDRIC laboratory multimedia team. The main purpose of this resarch is to find design principles for multimedia synthesis tools, that is to say tools able to create and manage multimedia virtual objects. The users of such tools should be able to define relationship between sound and visual characteristics and to play with sound and visual perception phenomena.

The spatialisation process is the last step in the synthesis of a 3D scene (Figure 1). It is mainly the computation of the object interactions, considered as light and sound sources or reflectors, in a virtual space. Comparing audio and image spatialisation principles has several practical goals:

- It is a step in the analysis of adequacy of each classes of algorithm to different classes of applications (from movies post processing to virtual reality).
- It allows to specify software and hardware components efficient in both fields.
- It provides a reference for the definition of audio and image objects descriptors for virtual scene specification languages (such as VRML (vrml) or JAVA3D (Java))
- More generally it is a first step into the creation of a multimedia synthesis tool.

As an example, Figure 2 is an informal representation of a virtual scene defined in a object programming environment (Jacobson 1993). This scene contains two objects: a wall and a man. Each object is defined as a new instance of a generic class, which includes some audio and image specifications of the object (source or reflector for example). The object "man" has two "animation" methods one four the sound and one for the image. This allows, for example, an animation such that the man's image leaves the scene in one direction and the sound leaves the scene on another one.

The image spatialisation process is well defined and a classical survey can be found in (Foley 1997). There have also been numerous papers on sound spatialisation and its relation to psycho acoustic (see for example (Blauert 1996) (Jot 1997), (Jot 1998) or (Julien 1994)). To our knowledge a very little work have been done in the comparison and the synthesis of these two fields. This paper is a first step in this direction. It is organized as follows: In the next section we give a more precise definition of the spatialisation process, the third section is devoted to the comparison of the physics and perceptual similarities and differences used in the design of sound and image synthesis algorithm. The last section presents a classification of algorithms in both domains. The presentation of this paper will be illustrated with several sound and video examples.

2. The spatialisation process

Numerous parameters related to sound and light propagation define objects in a virtual scene. An object is first defined by its current shape and position, then it can be considered either as a sound (resp. light) source or reflector. The absorption and reflexion characteristics of a reflector must be described either by a formal model or using a recorded representation (we call such a representation a texture in the sequel). The specification of a source object includes the directionnality and the signal. For example a light source can be defined just by a three component (RVB) constant light spectrum and the diffusion solid angle. The walking man of figure 1 can be considered as an omnidirectionnal sound source, which signal is determined by a recorded sound object (i.e. a sound file walking.aiff). To complete the scene description one must add a model of the medium propagation model (wave celerity and absorption as a function of the spectrum). This medium specification is generally implicit for light (except for foggy weather scene) and explicit for sound.

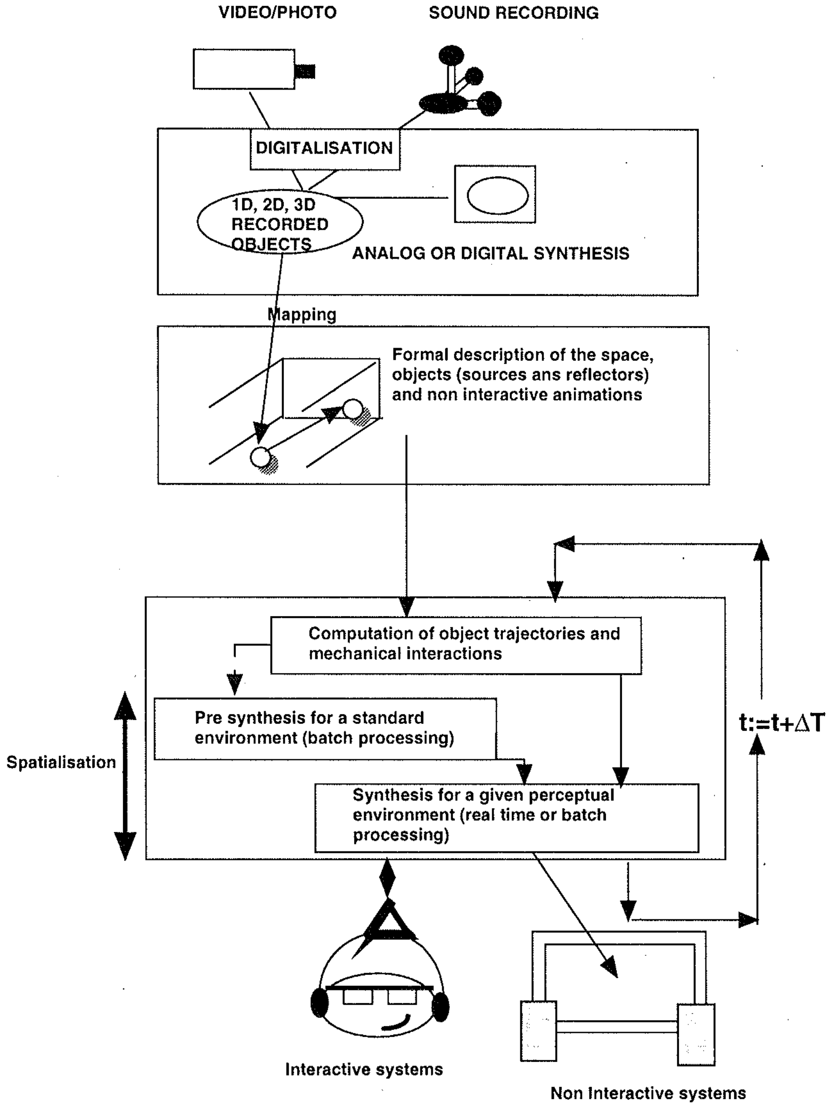
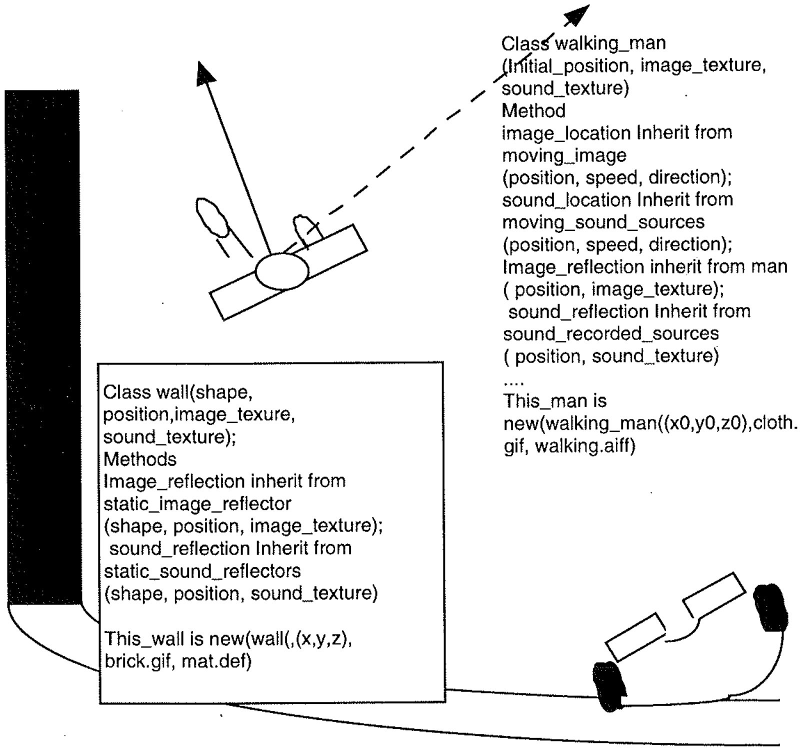

VIDEO/PHOTO
SOUND RECORDING
DIGITALISATION
1D, 2D, 3D RECORDED OBJECTS
ANALOG OR DIGITAL SYNTHESIS
Mapping
Formal description of the space, objects (sources ans reflectors) and non interactive animations
Computation of object trajectories and mechanical interactions
Pre synthesis for a standard environment (batch processing)
Synthesis for a given perceptual environment (real time or batch processing)
Spatialisation
t:=t+ΔT
Interactive systems
Non Interactive systems

Figure 1: Sound and image synthesis process



```
Class walking_man
(Initial_position, image_texture,
sound_texture)
Method
image_location Inherit from
moving_image
(position, speed, direction);
sound_location Inherit from
moving_sound_sources
(position, speed, direction);
Image_reflection inherit from man
( position, image_texture);
 sound_reflection Inherit from
sound_recorded_sources
( position, sound_texture)
....
This_man is
new(walking_man((x0,y0,z0),cloth.
gif, walking.aiff)
```

```
Class wall(shape,
position,image_texure,
sound_texture);
Methods
Image_reflection inherit from
static_image_reflector
(shape, position, image_texture);
 sound_reflection Inherit from
static_sound_reflectors
(shape, position, sound_texture)

This_wall is new(wall(,(x,y,z),
brick.gif, mat.def)
```

Figure 2: Specification of a virtual scene

The South East corner of figure 2 represents a spectator of the virtual scene which see and hear the virtual scene through a perception display. It can be a headset with goggles or loudspeakers and a screen. The spectator may be either passive or interacting with the virtual scene (at least by moving his head).

The spatialisation process computes the sound and the images as they are diffused through the perception display, considering the time evolution of the scene and the interactions between objects in the light and sound propagation. Figure 3 illustrate this computation, which is essentially composed of a "ligthning" algorithm (Foley 1997) and a projection algorithm according to the current location of the spectator. The last step (post processing) is the adaptation of the results computed to the perception display.

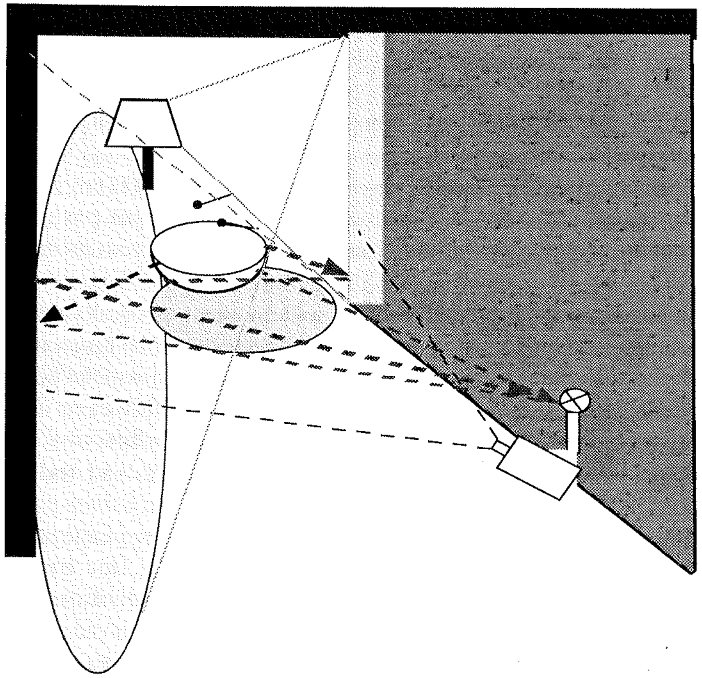


Figure 3: Illustration of the spatialisation process

3. Comparisons between sound and images spatialisation principles

Spatialisation relies on a subtle merging between physical and psycho perception principles to define computability efficient algorithms. At the most formal level, wave physics, the two phenomena can be considered as quite similar as they are described by the same kind of partial differential equations. This is the main reason which allows to use similar classes of model. For example optic and acoustic geometry use the same "Descarte's laws" of reflection which leads to the well known ray tracing class of algorithm (Whitted 1980),(Foley 1997),(Vorlander 1989).

But formal or numerical solution of an accurate physical model can not be obtained in a reasonnable amount of computation, even for rather simple dynamic scenes. Moreover, the goal of the spatialisation process as defined in the last section is not to simulate accurately the real world (as for example in virtual acoustic) but to built a "good" perception of the virtual scene according to the design goals and the diffusion display used.

Hence the spatialisation process uses characteristics of the perception of light and sound, which are quite different (sampling rate, directionnality, accuracy...) related to the wave propagation characteristics.

Many factors can be considered comparing the spatialisation algorithms. It concerns mainly:

- The difference between the celerity of waves (330 m/s vs 300 000 km/s).

Hence the spatialisation of an image does not depend on the past (light propagation is instantaneous) and the spatialisation of the sound is a time convolution process.

- The perception of waves attenuation and diffraction

Sound attenuation in a homogeneous medium must generally be considered. Light attenuation in a homogeneous «transparent» medium need not to be taken into account (except if needed or for special effects). Diffraction of sound objects is essentially considered through the variation of the celerity as a function of the spectrum. Diffraction in light propagation through transparent objects determines the computation of the ray direction.

- The wavelenghts related to the size of source and reflectors, and the accuracy of human discrimination (mm vs nanom).

The main aspects of the object characteristics used in the image spatialisation process depend rather simply on the wave spectrum. The RVB model, for example, relies on the filtering characteristics of human vision. Accurate sound spatialisation relies on the whole spectrum. We hear sound sources, reflectors are mainly perceived as contribution to a reverberation process. In counterpart we see essentialy light reflectors. So sound sources must be modelled precisely, sound reflectors can be defined from a more "global" point of view. Accurate lightening and shadowing of a 3D scene needs a precise definition of reflectors and sources may be modelled in a more approximate way.

Last but not least, one must consider the human anatomy which imply that we see at a given time a half space and hear the whole space.

4. Spatialisation algorithms

Spatialisation algorithms can be classified in term of complexity which increases with the precision of the physical model used as the algorithm basis. The same classification can be used in the audio and image processing fields. In this section we present a classification and discuss briefly each class of algorithm in both fields.

The simplest spatialisation process is just a mean value computation of the energy diffused in the scene. The mean value parameter is more related to a perceptual than a physical parameter. The ambiant lightning algorithm for example does not consider light sources but a homogeneous field in which each reflector has a given reflexion coefficient. More sophisticated models (Phong and Warn lightening algorithms (Bui-Tuong 1975), (Warn 1983)) consider the superposition of the ambiant light field and the direct effect of light sources. The complexity of this class of algorithm is proportional to the product of the number of light sources by the number of visible pixels. So the lightning algorithm follows a hidden surfaces computation which complexity is roughly proportionnal to the number of reflectors and which depends on the scene geometry and the spectator position. The computation must be repeated independantly for each image (25 to 30/s). As the light sources and the spectator position move slowly and the the light spectrum of each source is generally constant in the time, the computation can be optimized by an interpolation of an image from the previous one.

The corresponding sound spatialisation algorithm consists of the superposition of the direct sound and a statistical response of the room the later beeing synthesized by the use of basic reverberation algorithms (see [Jot 1997 2) for a survey and a presentation of new results). For example the IRCAM Spat processor combines the direct sound, the early reflexions and late reverberation. In this tool, the room is not necessarely defined by its geometry but by simple perceptual factors. The complexity is proportionnal to the number of sources and listeners, but is essentially dependant of the "memory" of the process which is inherent to the fact that the sound emitted by each source is a time function.

It is interesting to note that this class of algorithms were found by graphic computer scientists for the image and initially based on perceptual factors. In counterpart the sound algorithms were defined by acousticians and directly correlated to the physical analysis of the phenomenon.

The second class of algorithm relies on geometrical optic and acoustic. The well known ray tracing (Whitted 1980) (and image source in the sound field) model is the main basis principle of this class. Ray tracing is used in many other fields of simulation such as physic of particles or billiards. The algorithm determines the trajectories of sound and light rays from each source to each spectator. Rays are reflected on each reflector according to the Descarte's laws. At each reflection point the part of the signal which is absorbed and reflected must be computed. Hence in both fields the algorithm relies on a geometrical description of objects. This description can be rather simple in sound computation and must be very precise in the lightning algorithm. Light rays are generally geometrical lines, sound rays can have a "volume". The main difference between the two fields is related to the directionnality of the perception. This allows in image synthesis (Figure 4), to consider only the rays which start from the spectator's eyes and reach each monitor pixel, which is obviously impossible in audio spatialisation. Of course the difference on the memory of the two computations is the same than in the mean value class. This leads in the audio field to the source image algorithm (Vorlander 1989). Reflected rays are replaced by virtual sources situated behind the walls of the room. The contribution of each virtual source is delayed with respect to the original one (direct sound) by an amount of time which model the sound propagation. This simplification is efficient when the sources and the spectator have a static position. If the spectator moves, the contribution of each virtual source has to be periodically computed, if the source mooves both the position and the contribution of each virtual source must be computed for each source location.

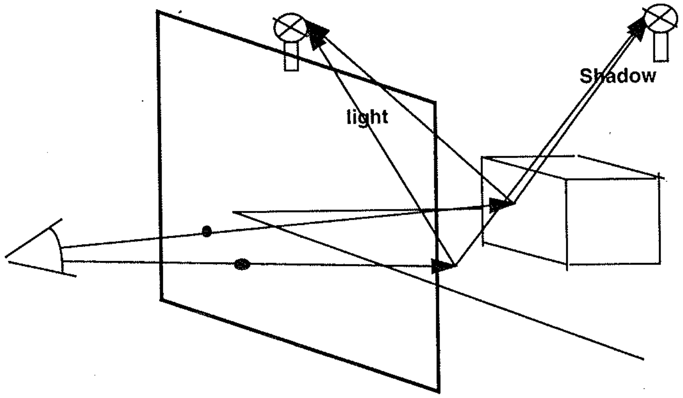


Figure 4: Recursive Image Ray tracing

This geometric class of algorithm is probably the one where the comparative analysis can lead to the most interesting results in term of common description of multimedia objects and software components. But, up to now, these algorithms are still too complex for a real time computation.

The last class of algorithm relies implicitely or explicitely to the numerical solution of the Dirichlet problem. We include in this class Radiosity algorithms, which replace the partial difference equations by a linear algebraic system which express the energy conservation and solves implicitely the Laplacian field. The algorithm are quite similar in audio and image synthesis and are used for "static" computation (still image in computer graphics, impulse response in virtual acoustic). So it is not yet usefull for the spatialisation process as defined in this paper.

5. Conclusion

This paper is a first comparative analysis of the spatialisation in audio an image synthesis. Numerous points have been omitted, such as the generation according to the reproduction display or a state of the art on virtual scene specification languages. As a position paper, we hope to have show the interest to this field of research, which is, to our point of view, the basis for the design “real“ multimedia tools.

Acknowledgements

This paper is the results of many discussions with Jean Marc Jot from IRCAM. I would like also to thanks Cecile Le Prado, composer, users of sound spatialisation tools and my wife, for her critical observations on the first draft of this paper, Pierre Cubaud and Jean Marc Farinone of CNAM for there usefull remarks and corrections.

6. References

Blauert 1996 Blauert J, Spatial Hearing: The Psychophysics of Human Sound Localisation, MIT Press, 1996

Bui-Tuong 1975 Bui-Tuong, Phong, *Illumination for Computer Generated Pictures*, CACM, 18 V6, June 1975

Cohen 1995 M. Cohen, E. Wensel, *The Design of Multidimensionnal Sound Interfaces*, Technical report 95-1-004, University of Aizu, February 1995

Foley 1997 J.D. Foley et als, *Fundamentals of Computer Graphics*, 4th Ed. Addison Westley, Reading 1997

Jacobson 1993 Ivar Jacobson, *Object Oriented Software Engineering*, Second Edition, Addison Westley 1993.

java http://www.javasoft.com/products/javamedia/3D/

Julien 1994 P. Julien, O. Warusfel, *Technologies et perception auditive de l'espace*, Cahiers de l'Ircam 1994, http://varese.ircam.fr/articles/textes/ Julien94/

Jot 1997 J.M. Jot, O. Warusfel, *Techniques, algorithmes et modèles de représentation pour la spatialisation des sons appliquée aux services multimedia*, IRCAM research report 1996, http://varese.ircam.fr/articles /textesJot97a/

Jot 1997-2 J.M. Jot, L. Cerveau, O. Warusfel, *Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model*, AES 103rd Convention, Sept 1997, New York, USA

Jot 1998J.M. Jot, *Real Time spatial processing of sound, music multimedia and interactive human-computer interface*, to be published in ACM Multimedia System Journal, special issue on audio and multimedia.

Vorlander 1989 M. Vorlander, *Simulation of the transient ans steady state sound propagation in rooms using a new combined ray tracing/image source algorithm.*, J. Acoust Soc Am. 86 (1) July 1989

vrml http://www.vrml.

Warn 1983 D.R. Warn, *Lithening Control for Synthetic Images*, Proc of SIGRAPH 83, In Computer Graphics 18 V3 July 1984.

Whitted 1980 Whitted T., *An improved illumination method for shared display*, CACM 23 (6), 1980