Proceedings of the Institute of Acoustics

HEARING THE NOISE YOU WANT, BY DESIGN: GEOMETRICAL ACOUSTICS FOR SPEECH INTELLIGIBILITY AND AURALIZATION

C F McCulloch MIOA (1) and E de Geest (2)

(1) Dynamic Structures & Systems Ltd, Sheffield
(2) Virtual Prototype Refinement, LMS International, Leuven, Belgium

SUMMARY

Methods for modelling acoustic behaviour in enclosed spaces, open and 'mixed' situations, have been implemented in software based on beam-tracing methods. The methods enable effective predictions of the speech intelligibility of coherent sources, mixed with coherent and/or incoherent reverberations and incoherent unwanted, background sources. The usual measure of speech intelligibility is Speech Transmission Index, which is calculated from the results of the numerical model using a modulation transfer function method.
Three-dimensional sound propagation, barrier effects and noise transmission through partitions (into or out of buildings) can also be modelled.
The methods are applicable to the analysis of entertainment, public-address and emergency systems, as used in transportation infrastructure such as railway and underground stations, airports, public buildings, retail environments and sports and other auditoria, as well as the transmission of unwanted noise from different sources into the environment.
As well as the quantitative measures used in the assessment of a design, a method for auralization of speech or other signals in the predicted situation is presented, enabling a more subjective assessment.
The paper discusses the theoretical basis, modelling approaches and practical techniques used. Some examples illustrate typical phenomena encountered in modelling the performance of PA and other systems.

1. INTRODUCTION

The intelligibility of speech is of increasing interest and concern in the acoustic design of transport infrastructure, such as stations and airports, as well as in other public places, such as shopping malls, sports stadia and entertainment venues. As well as the intelligibility of public announcements or broadcast material, intended to contribute positively to the occupants' experience or to inform them, there is nowadays often a very important specification requirement for safety and evacuation instructions to be provided verbally, and possibly to be controlled to specific zones within a larger space or series of spaces. The inherently reverberant nature of many of the spaces, means that the design of a public-address sound system can be difficult, especially when it often conflicts with architectural or aesthetic demands.

Much of the physical layout of a design (and indeed finishes and other details) may be fixed at an early stage in a project, so traditionally the acousticians arrive late in the day, with many constraints on their installation and little chance to influence other design features to improve the acoustics. However, there is now the possibility to design-in better acoustics, by making predictive calculations at an earlier stage. Numerical models, using geometrical acoustics, can be integrated with other CAD tools and can be updated at intervals as the design progresses. This paper is concerned with some of the methods used in such models, in particular the beam-tracing method and procedures for calculation of Speech Transmission Index, and the creation of auralizations of the actual environment to enable subjective assessment as well as objective rating.

2. RAY- AND BEAM-TRACING METHODS FOR GEOMETRICAL ACOUSTICS

2.1 General Features

Ray- or beam-tracing uses the basic assumption of geometrical acoustics, namely that the propagation of sound can be considered as the dispersion of energy, spreading spherically (from a point source) in straight lines and reflecting specularly whenever it strikes a surface. (Refinements to take account of diffusion and diffraction phenomena will be discussed later). The basic condition for this assumption is that the objects in the sound field have typical sizes an order larger than the acoustic wavelength, or that the density of acoustic modes within (partially) enclosed spaces is high. Another way of saying this, is that the frequency range is medium to high. For the normal audible frequency range, this places many constructions and even vehicle interiors within the scope of the ray- or beam-tracing method, except perhaps for small spaces such as very small rooms or vehicle cabins at low ('boom', mode) frequencies.

The beam-tracing method can be preferred over the ray-tracing method for several reasons, in particular its improved accuracy in determining arrival times of echoes and its closer conformity to the classical rules for echo-arrival density (number of echoes per unit time) versus time from impulse, without having any significant penalty in calculation time. The propagation of a beam (typically a triangular pyramid or a cone) is shown in Figure 1.

The initial result of the beam-tracing, for a given receiver position, is a set of echoes, each of which has an arrival time (due to the total path length from source to receiver) and an amplitude (due to the initial energy level of the beam and the energy absorbed at each intersection with a surface and absorbed along the path). These echoes are tracked in an echogram: for more compact data storage, a large part of the echogram data (say, for echoes with more than 5 or 10 reflections) can be accumulated into a histogram, where each 'bin' contains all the echoes arriving within a short time interval. By integrating the energy in the echo/histogram over certain longer time intervals after the first arrival (*eg* 0 to 80 msec, 80 msec to ∞, for C80) various sound quality measures can be derived, giving indications of the clarity of speech, music, *etc*. By using the directional information on beam arrivals, Lateral Efficiency (a measure of spaciousness) can also be found. The determination of STI, a better measure of speech intelligibility, is discussed later. The steady-state noise results are found by an integration of the echo/histogram:

$$p_n \quad = \quad \sum_{i=0}^{n}\sum_{k=0}^{N_i} A_{ik}^2 \quad = \quad \sum_{i=0}^{n}\sum_{k=0}^{N_i} \frac{\rho c W}{4\pi d_{ik}^2} \prod_m (1-\alpha_m) \tag{1}$$

where:
p_n = pressure at receiver, A_{ik} = amplitude of reflection *ik*, n = reflection (image) order, N_i = number of image sources of order *i*, d_{ik} = distance between image source *ik* and receiver, ρc = impedance of the air, W = power of the source, α_m = absorption at the mth wall reflection.

2.2 Phase Ray-tracing

In the preceding discussion, the source(s) and echoes combining together to give the predicted SPL and other derivatives of the echo/histogram, were assumed to be uncorrelated or incoherent. Thus, the contribution of each of the computed echoes or ray-paths is added just in terms of energy. However, if we consider that the path length from source to receiver is known, for any given frequency one can readily find a 'phase' of each arriving echo, taking phase shifts at the source (dependent on frequency and/or initial direction of the ray) and at wall reflections (possibly even dependent on angle of incidence) into account if necessary. Adding the pressures from each reflection vectorially (magnitude and phase) Equation (1) becomes:

$$p_n = \sum_{i=0}^{n} \sum_{k=0}^{Ni} A_{ik}^2 \exp(j\phi_{ik}) \exp(-jkd_{ik}) \qquad (2)$$

where: ϕ_{ik} = phase shift at wall reflections, plus initial phase at source; k = wavenumber (=ω/c)

This so-called phase ray-tracing method allows the frequency range of application of beam-tracing to be extended downwards, through the 'transition region' (usually defined by the Schroeder cut-off frequency F_s and 4 F_s) and enables standing-wave patterns and the like to be determined in small or larger spaces. Very good correlations with the results of boundary element and finite element acoustics models have been found.

2.3 Coherent and Incoherent Sources

Using Equation (2), one can add together not only the effects of the echoes from one source, but also many sources, if they are coherent: that is to say, they have a known phase relationship (equivalent, for any given frequency, to a known time delay versus some reference). Thus, for instance, one can consider the interaction of multiple loudspeakers in a public address system, taking the electronically-introduced time delays into account.

If some sources in a model are incoherent (*ie*, their phase is at any time random with respect to the reference) their effects can also be included, by a combination of the expressions of Equations (1) and (2). This enables the degradation of acoustic quality parameters due to unwanted noise ('masking') to be calculated.

2.4 Diffuse Reflections

Some acoustic problems, such as flutter echoes or lack of dispersion of sound input, may be addressed by using diffusers. The behaviour of a diffuser can be modelled in the beam-tracing process by the generation of new beams, carrying a part of the energy incident on the diffuser. The diffuse power from a given incident beam can be found as:

$$P_{diffuse} = d.P_{reflected} = d\,(1-\alpha)\,P_{incident} \qquad (3)$$

where: d is the diffusion coefficient.

The diffusion coefficient can often be very frequency-dependent, thus the proportion of incident power re-radiated diffusely can vary greatly over the frequency range of interest. The calculation of the diffuse radiation can be handled by generating new beams, radiating in random directions into the hemisphere above the diffusing surface. These new beams can then be traced in the same way as 'original' specular beams, so that the complete matrix of specular-specular, specular-diffuse, diffuse-diffuse and diffuse-specular energy transfers can be handled. However, the statistical nature of the procedure gives the diffuse re-radiation some uncertainty over its phase, so it should only be added to the correct histogram 'bin' for its arrival-time, and not considered as a distinct echo. There is also a certain penalty in terms of increased calculation time.

2.4 Sound Transmission

The transmission of sound <u>through</u> surfaces of a beam-tracing model can be handled by computing the incident intensity on one face of a surface ('panel') from the sound level at a nearby receiver, taking into account whether it is free-field or diffuse (or usually, some proportion of both) based on the echo/histogram data. This intensity, combined with a defined, frequency-dependent, transmission loss of the panel, and its area, is used to compute the power of the re-radiating source on the opposite face. Such sources can themselves be made coherent or incoherent, to reflect the nature of the incident field. Thus it is possible to assess the intelligibility and other measures, with the effects of transmitted sound, whether that be the speech signal itself (*eg* to assess privacy between rooms) or incoherent noise (*eg* break-in of environmental noise degrading the STI in a building or vehicle).

3. SPEECH INTELLIGIBILITY

3.1 Definitions

The loss of intelligibility of speech is mainly due to reverberation, interference and background noise. Speech intelligibility (and the clarity of other signals such as music) can be assessed by simple ratios of integrals of echo energies over certain time windows, in the echo/histogram. However, the Speech Transmission Index (STI) [1] [2] is a better objective criterion. (Rapid STI, RaSTI, which is more often found in experimental work because it can be more-easily measured, is merely the average of the STI-values at 500 to 2000 Hz).

Speech is an auditive flux of which the spectrum varies continuously in time, and the individual sinusoidal components of the envelope of the speech signal need to be preserved adequately to maintain intelligibility. The Modulation Transfer Function (MTF) is a response curve which expresses the attenuation of the speech-signal envelope, due to the room (or the environment in general). The reverberant transmission from the source (or a set of coherent sources) and the degradation due to any sort of background noise, are all taken into account.

3.2 Formulae

If we assume that the reverberation process in the space follows a pure exponential decay, the MTF is given by:

$$\mathrm{MTF}(F) \quad = \quad [\,1 + (2\pi F \frac{T}{13.8})^2\,]^{-1/2} . [\,1 + 10^{-SNR/10}\,]^{-1/2} \qquad (4)$$

where: T = reverberation time (sec); SNR = ratio of background noise (dB); F = modulation frequency (Hz).

However, a more general form (without the exponential assumption) is to be preferred, which gives the MTF as the normalised Fourier transform of the impulse response, in the frequency range 0.25 to 32 Hz (wherein the information content of the speech is carried by waves of higher frequencies):

$$\mathrm{MTF}(F) \quad = \quad \frac{\left| \int_0^{\infty} e^{2\pi jFt} . p^2(t)\, dt \right|}{\int_0^{\infty} p^2(t)\, dt} . [\,1 + 10^{-SNR/10}\,]^{-1/2} \qquad (5)$$

The STI is then an averaged, normalised, value of the MTF.

3.3 Effects from Incoherent Sources and Background Noise

Incoherent sources and background noise both act to degrade the STI, by adding-in uncorrelated, random, elements, characterised by SNR in the formulae above. The main difference between incoherent sources and background noise is that the latter is a position-independent spectrum, whereas incoherent sources have localised effects which should be computed by their own beam-tracing sequences.

3.4 Modelling Methods and Calculations

The calculation of STI requires the echo/histogram data which result from a normal beam-tracing calculation. The number of beams (or conversely the beam angle) and the order of reflections influence the accuracy of the results, similarly to the other acoustic quality measures such as Définition, but are traded-off against calculation time (which is thus problem-dependent). Calculation time and data-storage requirements are also affected by the choice of the approximate or the general formula (4) or (5). An intelligent data-storage process has been built into a proprietary program [3] in which these methods are embodied: this allows the power levels, relative phases (time delay) and other parameters of multiple sources to be adjusted, *after* the main beam-tracing calculation has been completed, but *before* deriving the various results. This means that the time-consuming part of the process rarely has to be re-run, whilst (for instance) evaluating the effects of adjusting power levels and time delays between loudspeakers, or turning on/off (or up/down) certain sources of background noise.

4. AURALIZATION

4.1 The Binaural Impulse Response

The aim of auralization is to simulate the binaural listening experience at one or more positions in the model.

The impulse response data (*ie*, the echo/histograms) computed using the beam-tracing method are monaural. Therefore, they need to be converted to binaural (2-channel) data, related to the orientation of a listener and the behaviour of sound around the head. This is done using Head Related Transfer Functions, which may be generic (for a typical human head, ear pinnae, *etc*) although possibilities for rapid scanning of any person's head and the generation of individual HRTFs is under investigation by some researchers [4].

From the echo/histogram data, a causal filter is generated for each echo, using its transfer function, thereby taking into account source directivity, wall absorption, air absorption, and different path lengths. This is multiplied by the HRTF corresponding to the angle of incidence. An impulse response is derived using inverse FFT, and all impulse responses are summed, taking into account time delays due to path length differences and source data, giving the Binaural Impulse Response (BIR). If echoes have been stored in the histogram rather than individually (typically for higher-order reflections and for diffuse reflections) the angle-of-incidence and precise arrival-time data are not known, so no HRTFs are used and the histogram is sampled randomly to generate the causal filters.

4.2 Convolution of Impulse Response and Anechoic Signals

When the BIR is established, it can be convolved with (*ie*, used as a filter of) a suitable source signal such as an anechoic recording, using digital signal processing. This is done with an FFT method, so that end effects are treated with zero padding. [5] [6]

5. APPLICATIONS AND EXAMPLES

5.1 General Remarks on Applications

The methods presented here have been used on many practical examples. Typical applications include:

- determining the intelligibility of public announcements in rail stations, airports, stadia, *etc.*
- placement, power, directivity and relative time (phase) of loudspeakers in PA systems, for the above.
- designing cost-effective acoustic treatments for PA-related installations with excessive reverberation.
- assessing in-vehicle audio for entertainment and/or information purposes.
- objective measures and subjective assessments of acoustics and active and passive controls in auditoria *etc.*

5.2 Test Cases

A simple geometry is used to illustrate some of the principles described in this paper. The test case consists of a rectangular room, 25m x 15m x 7.5m (see Figure 4) with air at 20°C and 50% relative humidity. (Air absorption along the beam paths is also included). An omnidirectional (spherical) coherent source is positioned 1m in x- y- and z-directions from one corner. Results are calculated on a grid of receiver points ('field point mesh') at 1.5m height across the room. All calculations are based on the eight standard octave bands from 63Hz to 8kHz.

5.2.1 Reverberant case for the 'baseline' has all walls with low absorption coefficients (0.02, 0.02, 0.03, 0.04, 0.05, 0.07, 0.06, respectively) except the end wall behind the source is anechoic. Figure 2 (upper left) shows a contour plot of RaSTI. Considerable degradation can be seen in the central region, as soon as one moves away from the source. Since there is only one source, this is clearly due to the reverberant self-interference of later echoes, and in particular a 'flutter' problem from the parallel side walls: the echogram for the point in the centre of the room (Figures 3, 4) shows this. (Note that the narrow vertical lines are individually-stored echoes; the wider bars are the histogram 'bins' for later reflections and diffuse reflections; there is no diffusion in this case).

5.2.2 Added diffusion (coefficients 0.5, 0.5, 0.5, 0.6, 0.7, 0.65, 0.6, 0.5) on the side walls disperses a large part of the early as well as late echoes (Figure 4: note the many early diffuse-reflection bars) and improves STI over the whole region (Figure 2, upper right).

5.2.3 Added absorption (coefficients changed to 0.24, 0.28, 0.4, 0.78, 0.98, 0.96, 0.87, 0.8) over the whole floor also improves STI (Figure 2, lower left) but not so effectively as diffusion. The effect is due to reduced Reverberation Time. (Energy levels, SPL, also fall...). Finally, the combination of absorption across the floor and diffusion to disperse interfering discrete echoes, gives the best performance (Figure 2 lower right).

5.2.4 A second source in the other corner of the room will probably degrade the STI if it is incoherent with the original source, as will random background noise. In this case, the problem is interference (or 'masking') so diffusion is of relatively little value (compare Figure 6 upper and lower plots). Figure 6 (left) also shows that the interference from the incoherent source is more local (top left of room as shown, especially with diffusing walls) whereas the effect of background noise is spread throughout (Figure 6, right) but less due to the level chosen.

5.2.5 A second coherent source gives an improved STI (Figure 7, upper left, *cf* Figure 2, lower right with same wall conditions: note that the contour scales are different) when it is in phase with the original source. It also overcomes the problem of background noise, within the 'range' of the second source (Figure 7, lower left, *cf* Figure 6, lower right; again, note the scales). However, if the second source has a (wrong) time delay, in this case 100 millisec, it has a destructive effect: compare the right-hand and left-hand plots of Figure 7.

6. CONCLUSIONS

Geometrical acoustic modelling, using beam-tracing methods, offers a powerful tool for the assessment of acoustic performance of architectural, structural and vehicle designs, at the development stage of a project. Not only sound levels but also quality parameters can be predicted. Phase ray-tracing allows the frequency range and other aspects of the applicability of the method to be extended, in particular for arrays of coherent sources. Proper algorithms for the derivation of measures of speech intelligibility, such as STI, give a robust design analysis tool in which all positive and negative influences can be taken into account and their influence can be quantified. Auralizing the sound received in the 'virtual space' adds a useful subjective tool to the objective results.

REFERENCES

[1] T.Houtgast *et al*, 'Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function, Part I' *Acustica* 16 p60, 1980

[2] R.Plomp *et al*, 'Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function, Part II' *Acustica* 46 p75, 1980

[3] RAYNOISE ® Revision 3.0, LMS International, Leuven, Belgium, 1998

[4] Y.Kahana, ISVR, University of Southampton, *private communication*, 1998

[5] J.Martin *et al*, *J Acoust Soc Am*, 94 (6) 1993

[6] H.Kuttruff, *J Audio Eng Soc*, Vol 41, No 11, 1993

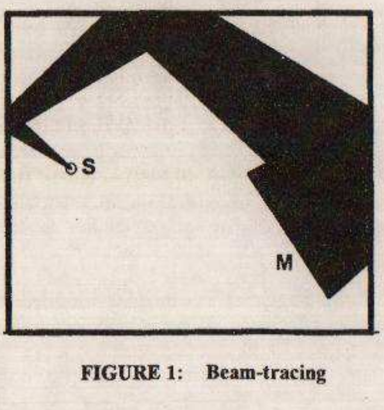

FIGURE 1: Beam-tracing

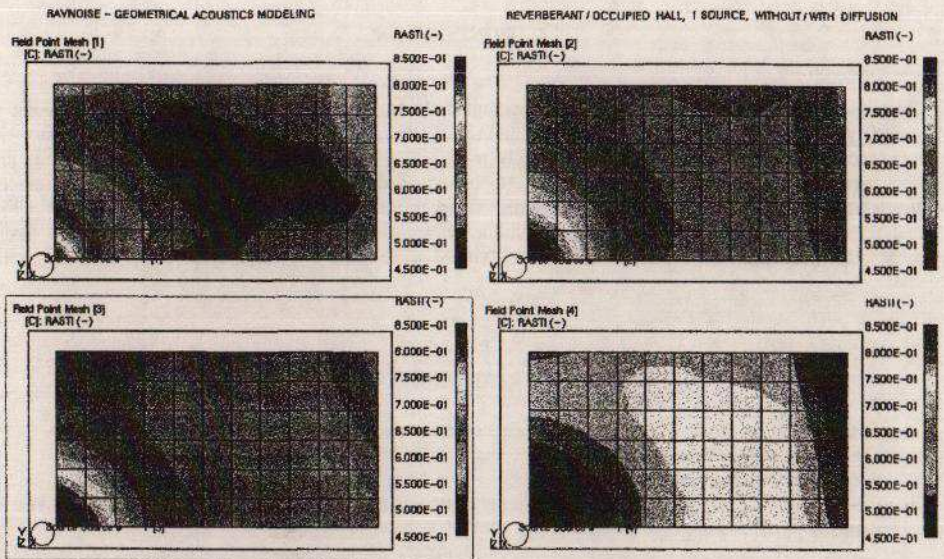

FIGURE 2: **RaSTI in a room with one coherent source**
(left) no diffusion (above) reverberant room (below) occupied room
(right) with side-wall diffusion (above) reverberant room (below) occupied room

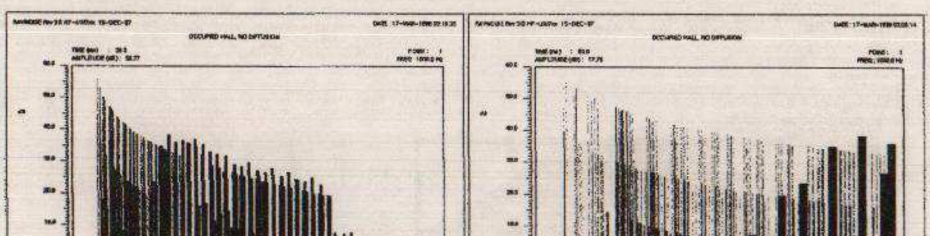

FIGURE 3: **Echogram at centre of occupied room, one coherent source, no diffusion (right: zoomed)**

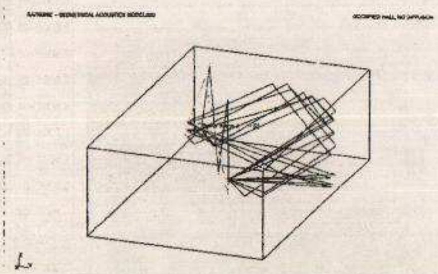

FIGURE 4: Ray paths (associated with Fig. 3)

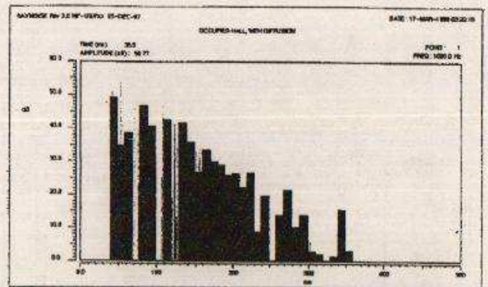

FIGURE 5: Echogram at centre of occupied room one coherent source, with diffusion

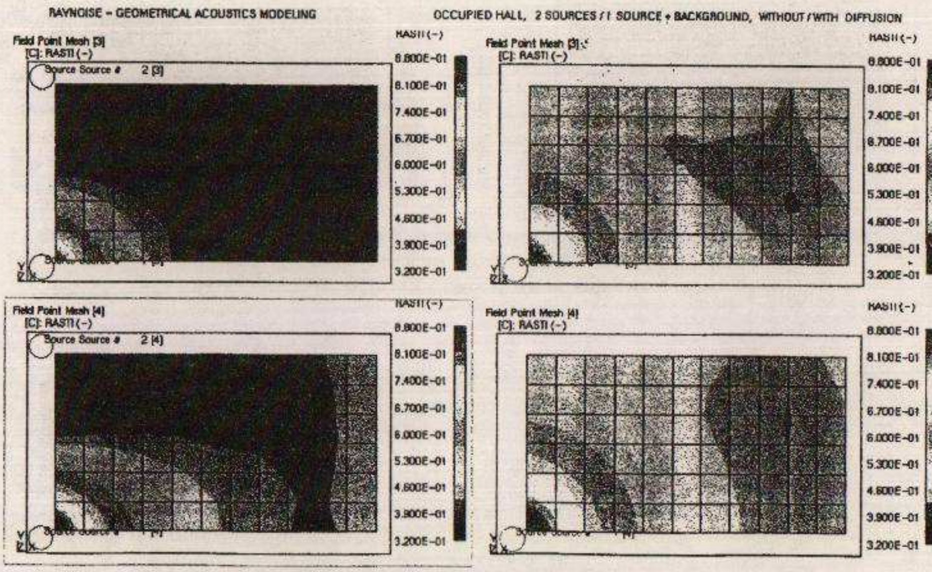

FIGURE 6: RaSTI in a room with coherent and incoherent noise:
(left) coherent and incoherent source (above) without (below) with diffusion
(right) coherent source and background noise (above) without (below) with diffusion

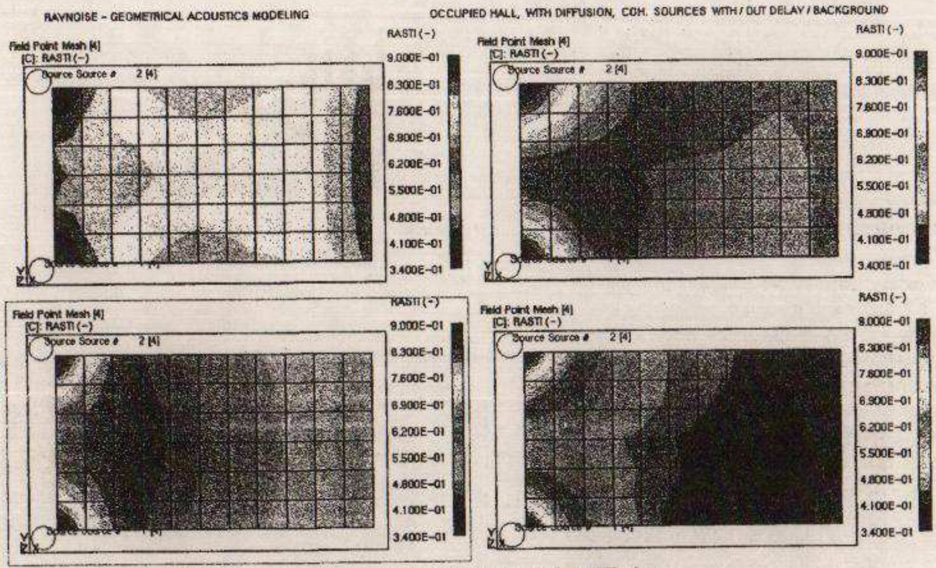

FIGURE 7: **RaSTI in a room with two coherent sources, with diffusion:**
(left) sources in phase (above) without (below) with background
(right) source 2 with time delay (above) without (below) with background