Proceedings of the Institute of Acoustics

THE VIRTUAL ROOM: DERIVING ACOUSTIC CHARACTERISTICS BY MODELLING

E de Geest and C F McCulloch MIOA

VPR, LMS International, Leuven, Belgium

SUMMARY

This paper deals with modelling rooms and other spaces using geometrical acoustics. Methods based on ray-tracing (more precisely, beam-tracing) are described, with details of various techniques which enhance the quality and accuracy of the results. Both coherent and incoherent sound sources, and backgrond noise, can be considered. Surface treatments can be both absorptive and diffusive. Localised models of individual sources using finite element and boundary element vibro-acoustic models are also mentioned, and it is shown that these modelling methods are not well-suited to be adapted to whole-room modelling, other than at very low frequencies.
The assessment of speech intelligibility using the results of beam-tracing models is described, together with a process which produces stereo auralizations of the sound at a selected position.
Various examples show verifications of the approaches which are presented and the effects of room constructions and treatments.

1. INTRODUCTION

It is increasingly of interest to know the acoustic performance of 'designed spaces' at the earliest possible stage in their design. This enables one to identify and (we hope) eliminate problems during design, rather than by (usually, expensive) trouble-shooting later on. This applies equally to auditoria, public spaces such as atria, shopping malls, railway stations and airport buildings, and to smaller spaces such as recording and listening rooms, home theatres, and vehicle interiors. Many different criteria may be applied in the assessment of a design. Using computer modelling, with an appropriate combination of software and computer, these criteria can be derived from a model, or a 'virtual room', long before physical construction. The computer modelling method must be sufficiently robust to give results with value to the design process, early enough to be of benefit, and integrated with other design methods such as CAD.

2. MODELLING THE SOURCE

2.1 Finite Elements and Boundary Elements

It is self-evident that acoustic sources such as loudspeakers are vibro-acoustic devices, in which vibrational energy (typically, controlled by some electrical signal) is converted into sound waves. The devices themselves are usually of a similar order of size as the wavelength at the higher frequencies for which they may be used, so that diffraction and wave phenomena are apparent. For modelling the structural parts (membrane/cone, surrounds, enclosure...) structural finite elements (FE) are an effective tool. For modelling the acoustic field in and around the device, acoustic finite elements are not satisfactory, because the field is infinite (at least partly) and acoustic boundary elements (BE) are more suitable. The structural FE model and acoustic BE model are then linked to provide two-way interaction, with the structural vibrations altered by the presence of the air or other fluid.

An example is shown in Figure 1. The results of the analysis provide distributions of pressures, propagating energy (intensity) maps, the influence of different vibration modes of the speaker cone or enclosure, and directivity diagrams. An important point to note is that such models and results are almost always deterministic and coherent.

2.2 Other source models

Using an FE or BE acoustic model to simulate a complete space, with the loudspeakers and other sources, is either worthless or of limited use, as explained below. Therefore, in these models, the sources have to be modelled in a 'lumped' way: a coherent source such as a loudspeaker may be modelled as a point source, with a given directivity, orientation, power and phase, which may vary with frequency. The detailed results of models such as those described in 2.1 above, or data derived from measurements, can be loaded into a suitable model, to provide a more precise description of the source in the room model.

3. MODELLING THE ROOM

3.1 Are Finite Elements and Boundary Elements practical for room models?

If FE or BE acoustics can be used to model a loudspeaker effectively through all its frequency range, as is generally the case, it begets the question: Can we use the same method to model the whole room (and even multiple sources within the room?

In general the answer is 'no', but there are exceptions. The reasons for this inability of the FE or BE methods to model the whole space all relate to the coherent, deterministic nature of the modelling process: the computer model involves a discretisation into elements, so the primary results are 'spatial samples' of the interacting, interfering, waves. The waves are sinusoidal at one frequency, so the 'sampling frequency' in space needs to be around seven nodes per wavelength, or six linear elements, to capture the waves and interactions adequately. Consider the mesh of a simple cuboid room shown in Figure 2. The elements are approximately 0.5m long, so the maximum frequency for reasonable accuracy is 113Hz, for a sound speed of 340m/sec (340/0.5x6=113). The model can therefore only cover the very lowest frequencies of the audible range. If we wished to double this frequency we would have to halve the element size, quadrupling the number of elements and increasing the calculation time by ten or twenty (it may become several hours for a useful number of frequencies). For realistic rooms with dimensions of several metres in each direction, it rapidly becomes impossible to make workable models. The acoustic modal density can become very high, with much overlapping of modes, calling into question the purely-deterministic approach of FEM and BEM. Even for very small interior spaces such as vehicle cabins, there is an upper limit of 300-500 Hz.

Furthermore, FE or BE models use coherent sources: some sources in the room may be incoherent. There may also be significant diffusion of sound waves (re-radiation of some incident sound in arbitrary, random, directions) at certain surfaces. Neither of these phenomena can be handled, or at least only with considerable complications, in FEM or BEM.

3.2 Ray tracing

The above difficulties can be overcome using a method which combines deterministic calculations of sound propagation and 'statistical' data for surface properties. The method is ray tracing (or more accurately beam tracing) and is shown in Figure 3. A point source is formed by an array of conical or triangular beams. The spherical dispersion of waves from a monopole point source is handled by the continuously-increasing beam width. At each boundary surface of the space being modelled, a ray strikes that surface and loses energy, defined by the absorption coefficient of the struck surface. The remaining energy is reflected specularly. The process continues, up to some user-defined order of reflections. The distribution of acoustic energy in the modelled space is mapped by placing a grid of receiver points ('field points'): each time a beam passes near enough to capture a field point within its cross-section, an echo is identified and its contribution is stored in a histogram of energies and arrival times. Significant individual echoes can also be stored, to create an Echogram, as shown in Figure 4.

4. THE RESULTS FROM RAY-TRACING

4.1 Stationary sound

The results of a ray-tracing calculation are firstly the echograms/histograms at each receiver point, in other words the impulse responses. From these, by backwards (Schroeder) integration, stationary sound values due to a steady-state source are derived. If the receivers form a grid linked by elements, contour maps of these results can be plotted, such as SPL broadband, with/out weightings, SPL at specific frequencies, NC, NR, *etc*. If several sources are defined, their results can be superimposed, following the appropriate summation rule for coherence or incoherence (see later). Frequency-response spectra at any selected point can also be produced.

4.2 Derivatives of the impulse response

As well as the stationary sound results from integrating the whole echo/histogram, more complex quality parameters can be derived. The majority of these use formulae which are based on ratios of integrals of the impulse response over different time windows (for instance C80). Others use the decay rate of the impulse response (*eg* RT_{10} ... RT_{60}: note that these are position-*dependent* RTs. Position-*independent* RT can be calculated very rapidly, without the need for a grid of receivers, with Sabine and Eyring formulae using the geometry alone and using a Statistical method based on a ray-tracing to find the mean free path and its variance. The latter finds the 'visibility' of surfaces to the main reverberant volume and weights their absorptive effects accordingly). Spatial awareness of the sound can be assessed using Lateral Efficiency, which takes into account the arrival directions of each echo (beam) as well as their times and energy levels.

To improve the accuracy of these results, calculation parameters, such as the numbers of rays per source (*ie* solid angle of each beam) and the maximum order of reflections to which a beam is traced, can be refined. The left-over energy in the echogram tail (*ie*, the residual energy of each ray at the point where its tracing is stopped) can also be added-back to the results at each receiver, distributed according to correction factors derived from statistical room acoustics.

4.3 Speech intelligibility

As well as the position-dependent room acoustic parameters derived from integrals of the impulse response, it is also possible to use the echo/histogram data to derive speech intelligibility measures, in particular Speech Transmission Index (STI) [1] [2] and RaSTI. The information content of the speech exists at low frequencies (0.25–32Hz) but is carried by modulating the audible signal, mainly in the mid-frequency bands. The loss of intelligibility of speech can be due to reverberation, interference and background noise. 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.
Hence, by using MTFs to analyze the superimposed impulse responses of all coherent sources, driven by the basic speech signal, and adding-in the effect of incoherent ('masking') sound, STI can be calculated, on a frequency-dependent basis (*eg* at each octave band centre).

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

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

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:

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

The STI is then an averaged, normalised, value of the MTF. Using Equation (2) however increases calculation time a little and data storage a great deal. (RaSTI, which is more often found in experimental work because it can be more-easily measured, is the average of the STI-values at 500 to 2000Hz).

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 are computed by their own beam-tracing sequences.

4.4 Auralization

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.

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. RAY-TRACING AT LOW FREQUENCIES

5.1 Phase Ray-tracing: room modes and frequency responses

The basic assumptions in the ray-tracing approach were that sound waves can be approximated by rays or beams, as carriers of energy. Therefore diffraction and interference are not automatically included, unlike the wave-equation-based approach of FEM or BEM. However, the path length from source to receiver is known, as well as any time delay (or phase shift) at the source and at reflections, so it is possible to derive, <u>for each frequency</u>, a phase corresponding to the arrival time of each echo. The echoes can then be added together (*eg* to give a stationary sound) taking this phase into account and using complex, vector, summation. The basic incoherent summation:

$$p_n \quad = \quad \sum_{i=0} \sum_{k=0} A_{ik}^2 \quad = \quad \sum_{i=0} \sum_{k=0} \frac{\rho c W}{4\pi d_{ik}^2} \prod_m (1 - \alpha_m) \qquad (3)$$

becomes:

$$p_n \quad = \quad \sum_{i=0}^{n} \sum_{k=0}^{N_i} A_{ik}^2 \ \exp(j\phi_{ik}) \ \exp(-jkd_{ik}) \qquad (4)$$

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 m^{th} wall reflection, and for the phase effects: ϕ_{ik} = phase shift at wall reflections plus initial phase at source, k = wavenumber (=ω/c, ω = circular frequency, c = speed of sound).

Using this approach enables the calculation of room characteristics at low freqencies, through and below the transition region F_{cutoff} to $4F_{cutoff}$, which is otherwise taken as the limit of ray-tracing, defined by the Schroeder cut-off frequency $F_{cutoff} = 2000(RT_{60}/V)^{-1/2}$ (approximately, using MKS units). For instance, a small room, 7x5x3metres, with RT_{60} = 0.5sec, has F_{cutoff} = 138Hz. The transition region is the frequency range in which modal overlap becomes significant.

Using Phase Ray-tracing, it is then possible to derive not only highly-position-dependent frequency responses, at low frequencies, but also indications of room modes (*ie* standing wave interference patterns).

5.2 Comparisons with FEM and BEM

The results of room calculations with Phase Ray-tracing have been compared with those from FEM and BEM models. The main functional differences are that usually the ray-tracing model calculates much faster, and only requires a geometry model not a complete mesh grid. The actual results have in many cases been very similar, which is remarkable given the completely-different methodologies. This gives confidence in the phase ray-tracing approach. Figure 5 compares room modes computed with FEM and Phase Ray-tracing on the 10m long room illustrated earlier, and Figure 6 compares frequency response functions for another room. It must be kept in mind however that the FEM and BEM approaches become increasingly inaccurate with increasing frequency, whereas the Phase Ray-tracing approach is the opposite, and requires increasingly-long calculations for very low frequencies, with refined parameters (high reflection order, narrow beams, ...) especially if mean absorption is low (*ie* long RTs).

6. THE INFLUENCE OF SOURCE CHARACTERISTICS AND ROOM TREATMENTS

6.1 Coherent sources

It has already been explained how coherent sources can be modelled with a phase-related summation of the echoes arriving at a receiver. Figure 7, upper left and right, shows RaSTI (used because it provides a single, averaged criterion) in a reverberant room with a coherent source. The self-induced echoes of the source produce interference and degrade the intelligibility. Adding a second coherent source improves intelligibility in the regions nearby the new source, and also elsewhere if the paths from both sources are similar (Figure 8, upper left). If the second source has a large time delay, this has a destructive effect, except very near the second source, where its own emissions dominate (Figure 8, upper right). If the second source is incoherent, its effect is purely destructive ('masking'). Background noise is likewise destructive, but not localised (Figure 9).

6.2 Wall absorption

With a single coherent source, it is self-destructive of STI due to its own echoes. Therefore, adding absorption (even due to the occupants) has some benefit in improving STI (Figure 7, upper and lower left). With multiple, coherent, sources, the effect is more marked, but what happens is in essence that the whole environment is being made much more damped, and each source has its own region of influence and little effect on the neighbouring sources: the room is 'dead' and the response at any point will be almost totally due to the loudspeaker system.

6.3 Wall diffusion

6.3.1 Local geometry

If a coherent source causes self-destructive loss of intelligibility in some parts of a room due to its own echoes, adding diffusive or dispersive devices should improve the situation. The dispersion can be achieved by either 'gross' geometrical devices, like corrugated side-walls, or by 'local' diffusive devices like wall treatments or diffuser panels. Figure 5 illustrates the geometrical effect: the room volume is almost constant, so the theoretical modes based on geometry remain the same (see FEM model modes in the upper part of Figure 5) but the response to a specific source changes due to the reorientation of reflections and the elimination of the parallel side walls and corresponding 'flutter' (see Phase Ray-tracing results in the lower part of Figure 5).

6.3.2 Diffusion coefficients

If the parallel walls are retained, but are treated with a diffusing material, the improvement in STI is more consistent, because it is not frequency- and position-dependent as it can be with geometrical changes like corrugations. The calculation principles of diffusing materials are that the energy of an arriving beam is not only partly absorbed and partly reflected specularly, but a fraction is also reflected diffusely:

$$P_{diffuse} = d.P_{reflected} = d\,(1-\alpha)\,P_{incident} \qquad (5)$$

where d is the diffusion coefficient. New rays are initiated on the surface, based on a random-number process using d. These new rays can afterwards follow purely specular paths or can generate further 'diffuse rays' if they again meet a diffusing material. The influence of adding diffusion (with no change in α) on the parallel sides of the example shown before, is shown in Figure 7 (left versus right).

6.4 Sound transmission and diffraction

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).
It is also possible to model diffraction over 'barriers' (for instance, the break-out of noise from an open arena) by using an algorithm based on path-length differences (a form of inversion of the barrier insertion formula of Maekawa *et al*). An enhancement of this, uses the phase technique already described, making the results narrowly frequency-dependent, as is needed for tonal or coherent sources.

6.5 Efficiency of calculations

As has already been mentioned, there is an inevitable trade-off between accuracy ('fineness' of the model, beam size, reflection order,...) and calculation time and data storage. Calculation time and data-storage requirements are also affected by the choice of the approximate or the general formula (1) or (2) for STI. 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. Likewise, arbitrary combinations of sources can be superposed. This means that the time-consuming part of the process does not have 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.

7. CONCLUSIONS

Ray-tracing (or more precisely, beam-tracing) provides an effective method for room modelling. Many different phenomena can be included in the model: source directivity and frequency-dependence; material characteristics and their frequency-dependence and if necessary their diffusive nature; phase/time relationships between sources, frequency-dependent if necessary; and special effects such as through-panel transmission and diffraction.

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.

The primary result from the calculation is an impulse response echogram, from which many other results can be derived: measures of received sound level, acoustic quality and speech intelligibility. These can be mapped over any area, as required, and examined in detail at specific locations. Processing of the impulse response, including directional information for the rays, enables a binaural synthesis and convolution with source signals to produce auralizations, which add the subjective concept of the 'virtual room' 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 communications*, 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

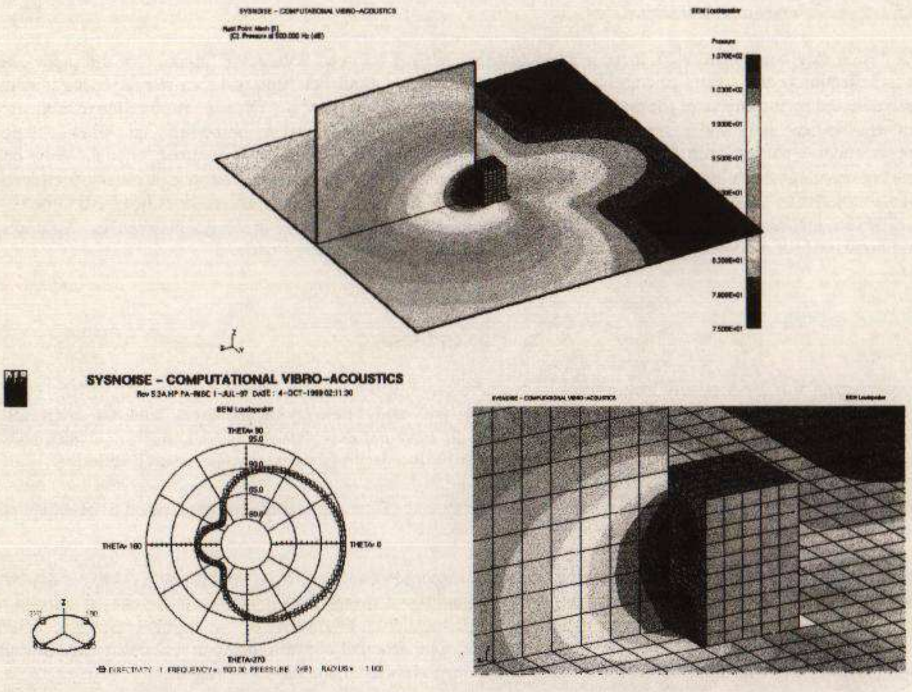

FIGURE 1: Loudspeaker FEM and BEM model: above, field response; below left, directivity diagram

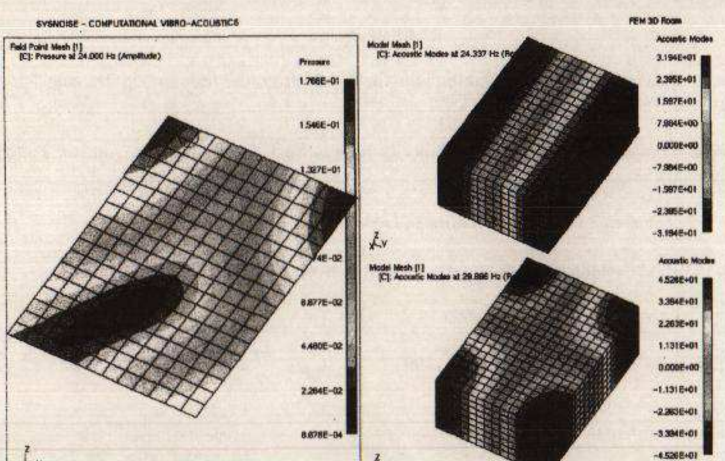

FIGURE 2: FEM model of room: left: response at 24 Hz ; right: modes at 24 and 30 Hz

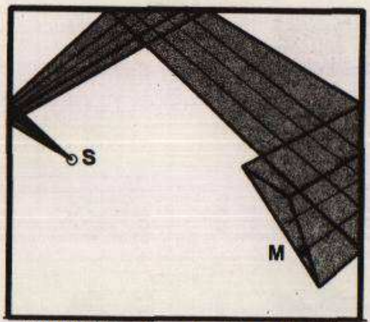

FIGURE 3: Beam-tracing principle

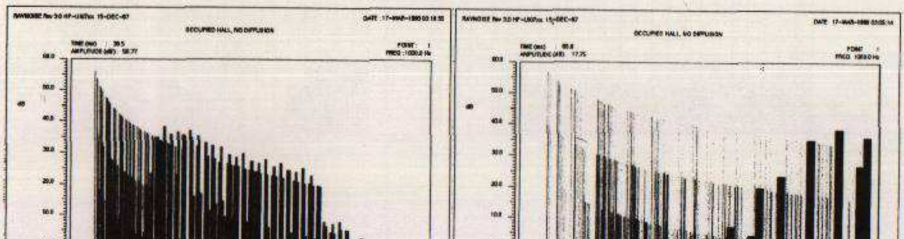

FIGURE 4 (a): Echogram at centre of occupied room, one coherent source, no diffusion (right: zoomed)

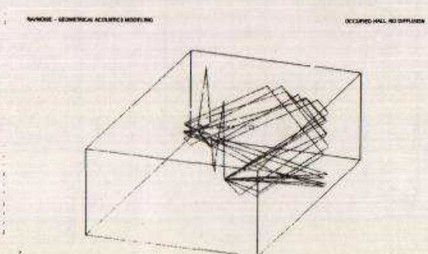

FIGURE 4(b): Ray paths (assoc. with Fig. 3)

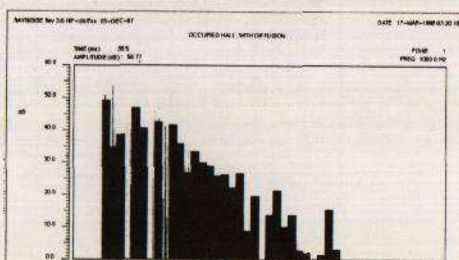

FIGURE 4(c): Echogram at centre of occupied room one coherent source, with diffusion added

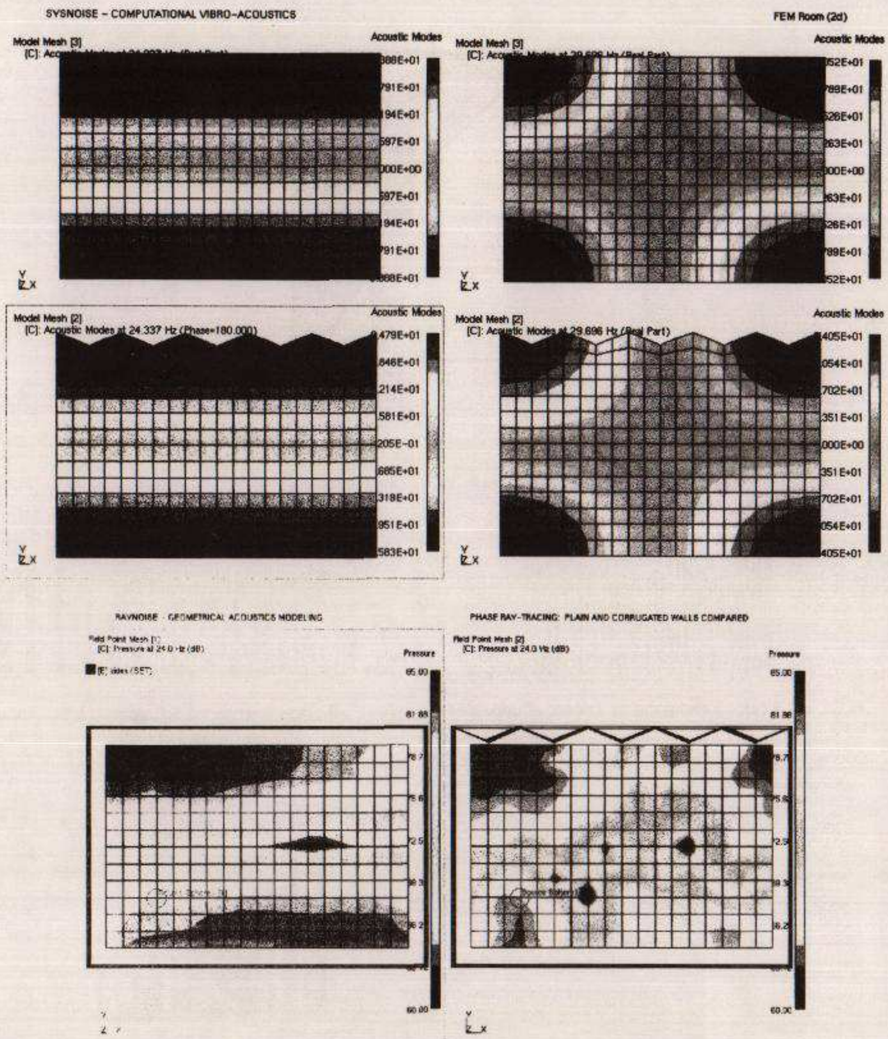

FIGURE 5: **Room modes from FEM (above) and 'response' at same frequency (24 Hz) from Phase Ray-tracing (below): - showing the effect of a corrugated side wall**

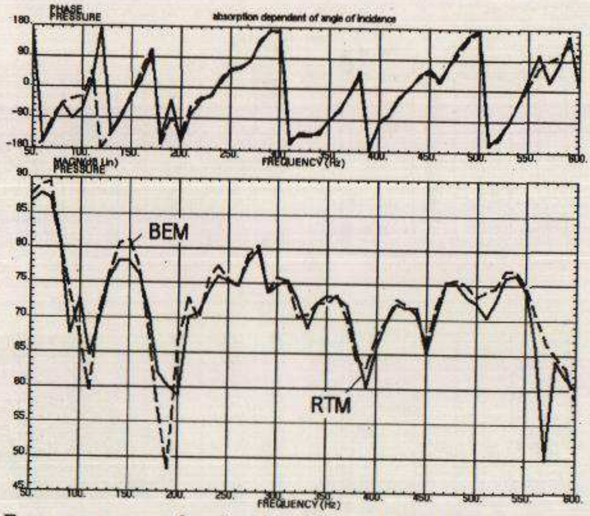

FIGURE 6: Frequency response functions from FEM and Phase Ray-tracing compared

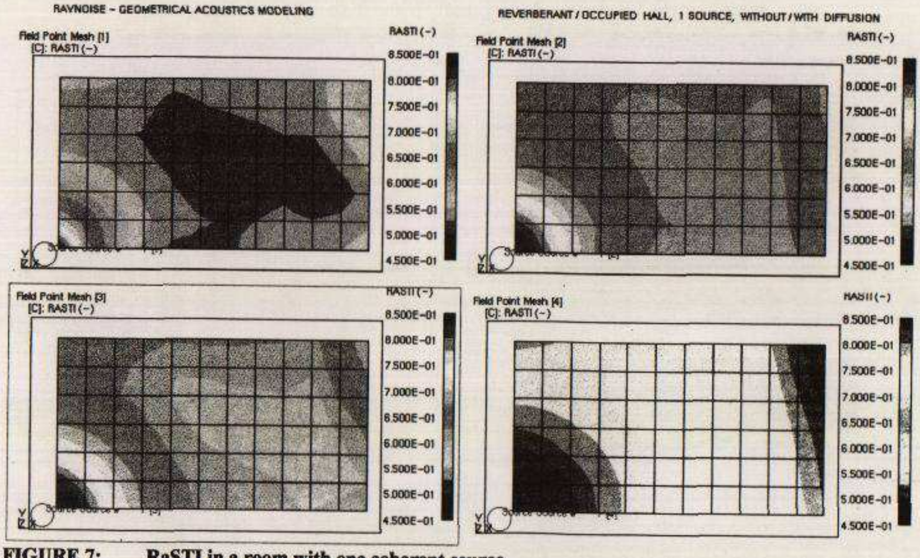

FIGURE 7: RaSTI in a room with one coherent source
(left) no diffusion (above) reverberant room (below) occupied room
(right) with side-wall diffusion (above) reverberant room (below) occupied room

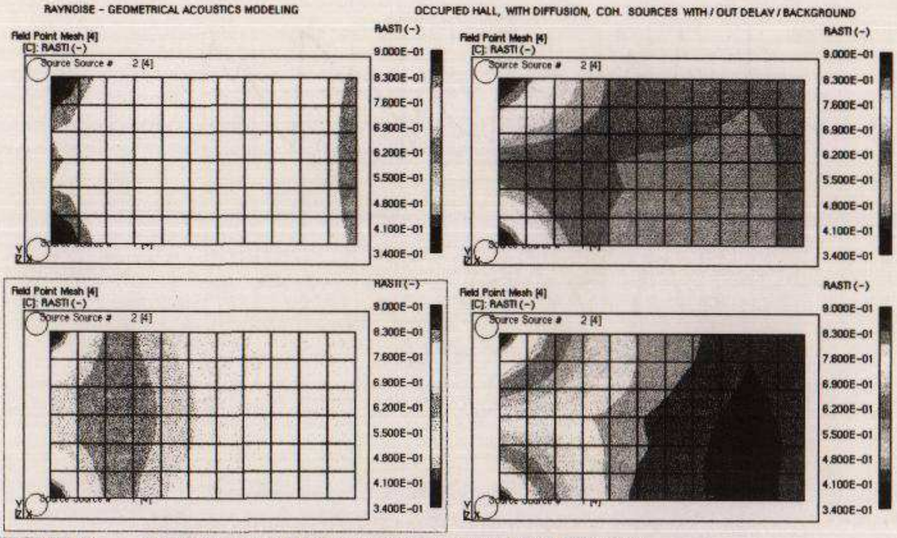

FIGURE 8: **RaSTI in a room with two coherent sources, with diffusion:**
(left) sources in phase (above) without (below) with added background noise
(right) source 2 with time delay (above) without (below) with background

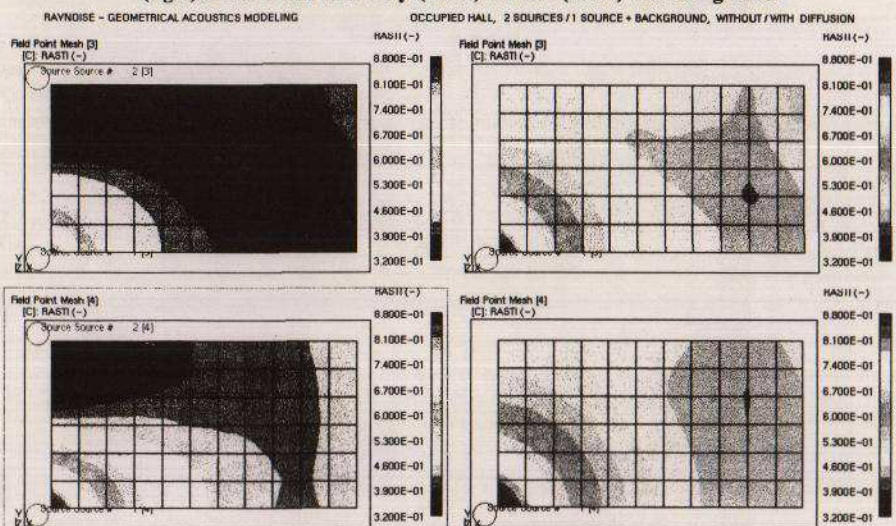

FIGURE 9: **RaSTI in a room with coherent and incoherent noise:**
(left) coherent and incoherent source (above) without (below) with diffusion
(right) coherent source and background noise (above) without (below) with diffusion