virtual production AUDIO

In television production, the use of synthetically-generated visual images is becoming well established. So far, most of the work has been carried out on the video picture generation, but it is also important to create a credible audio illusion. It is well known, from common observation of almost any badly dubbed television programme or cinema film, that discrepancies between the audio and visual spatial cues significantly degrade the illusion of reality.

The human hearing system and synthetic sound fields

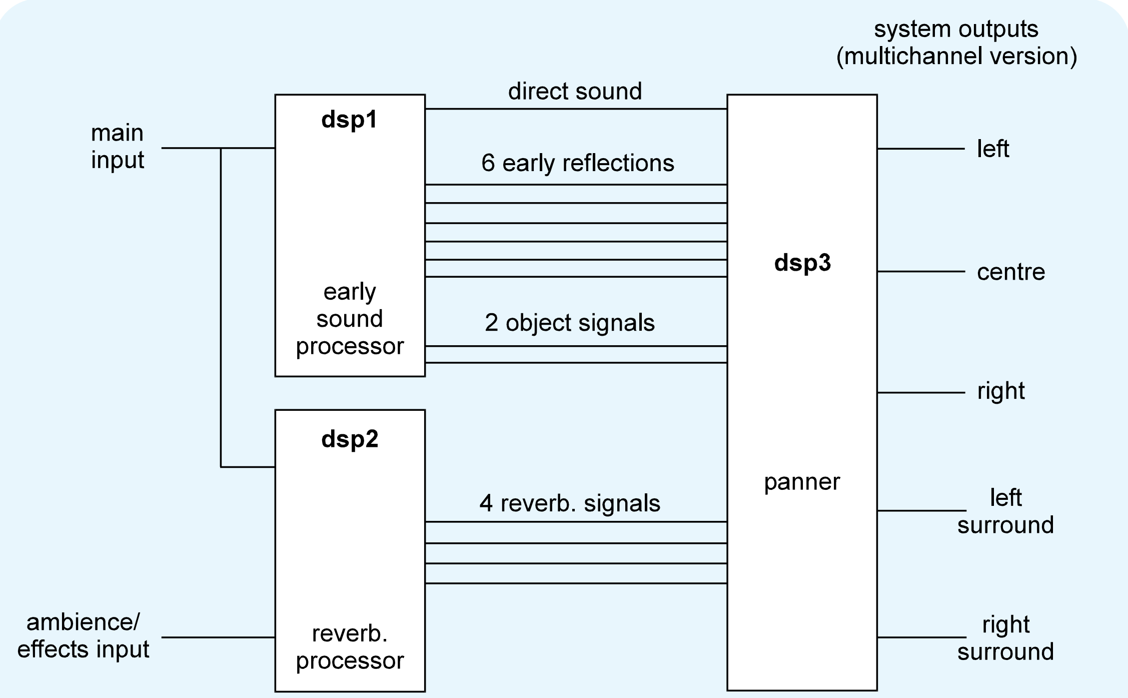
The human hearing system is well adapted to obtaining spatial cues from the acoustic environment. Evolution has ensured that maximum use is made of sound parameters and clues relating to relatively close sources, up to a time delay of about 30 ms. After that time, the sound is blended and heard as reverberation.

Sound, starting from a source, travels outwards in all directions at about 340 m/s. That uniform spreading proceeds for only a short time, until part of the sound wave strikes some object. What then happens is always complicated. The sound may be reflected, absorbed or diffracted by the object. It is impractical to create even a reasonably accurate model of anything but the most primitive acoustic space. It may well be unnecessary anyway. For the creation of a credible sound impression in a virtual production, a relatively simple synthesis will result in a convincing audio illusion, partly because it is assisted by the visual information.

At its simplest, the sound field reaching a listener can be divided into three main components, the direct sound, a small number of discrete early reflections and the overall reverberation.

The experimental audio system

The experimental system consists of a commercial digital signal processing (dsp) development platform, uses three, 40 MHz DSP56000 processors in an industrial, IBM-PC® type computer. The audio image processing is divided between the processors. One produces the direct and six delayed early signals and the second produces the four reverberation signals. The third combines these signals to produce the panned output signals to drive the loudspeakers.



Arrangement of dsp modules.

BBC

Research & Development

Direct and early sound

Geometric modelling of the direct and early sound is essentially trivial and self-evident. For the purposes of the experimental system, the space is assumed to be a 'rectangular' room. That gives six first-order surface reflections and makes the calculation of the discrete image locations trivial. No second-order early reflections are included. The ideal (3-D) model is modified to fit the 2-D reproduction paradigm by mapping the reflections from the walls, floor and ceiling onto the horizontal plane. The room geometry, source and listener locations are input control parameters. All of the seven discrete sound source images are mapped onto the reproduction loudspeaker layout by amplitude panning.

The direct sound and the early reflections are potentially subject to several different kinds of frequency filtering, for example the selective effects of air absorption over larger distances or non-uniform source directivity. Those filter functions are approximated by simple filters.

Reverberant sound

The reverberation processor is based on well-established principles employed by commercial artificial reverberation generators. The overall properties of the space are calculated from basic acoustic principles. The reverberation is assumed to be controlled by boundary surface absorption, uniformly distributed on all surfaces, together with air absorption. The required reverberation time is an input control parameter. The reverberation processor also has a second input port for ambience and effects.

Internal objects

The basic acoustic model represents only the interior surfaces of the empty room. One internal object is included in the model, either as an obstruction or as a reflector. The two additional signals required for this are generated by the direct/early dsp processor.

Movement

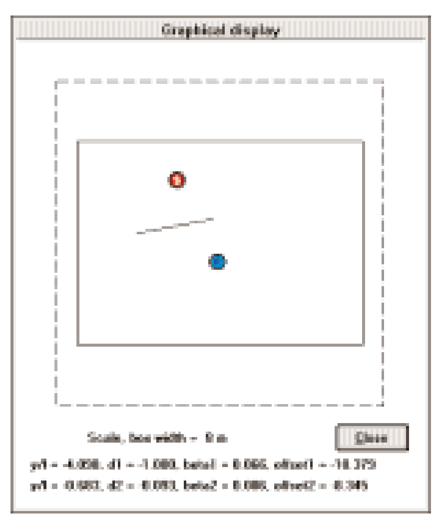
When such a source moves in a real room, large and rapid changes in level occur (due to interference patterns) but do so smoothly. In the case of the model, the basic audio sampling interval of 20 ms would give rise to step phase discontinuities, corresponding to every 7 mm of movement. It is necessary to interpolate the sampling interval give a finer resolution.

Network Interface

The audio virtual production system will eventually be controlled from the video system, by way of a source tracking system. A TCP/IP network interface and management system is included in the audio virtual production processor, to form a communication channel for control packets.

Manual Control System

For the purposes of testing and demonstration, a manual interface is included to allow input of room, source and receiver geometry and acoustic room characteristics. A simple graphical interface allows the source and receiver to be moved by the mouse cursor.



Graphical user interface for manual control.

Further information: BBC Research & Development Kingswood Warren Tadworth Surrey KT20 6NP United Kingdom
Tel: +44 (0)1737 832361 Email: info@rd.bbc.co.uk Website: www.bbc.co.uk/rd