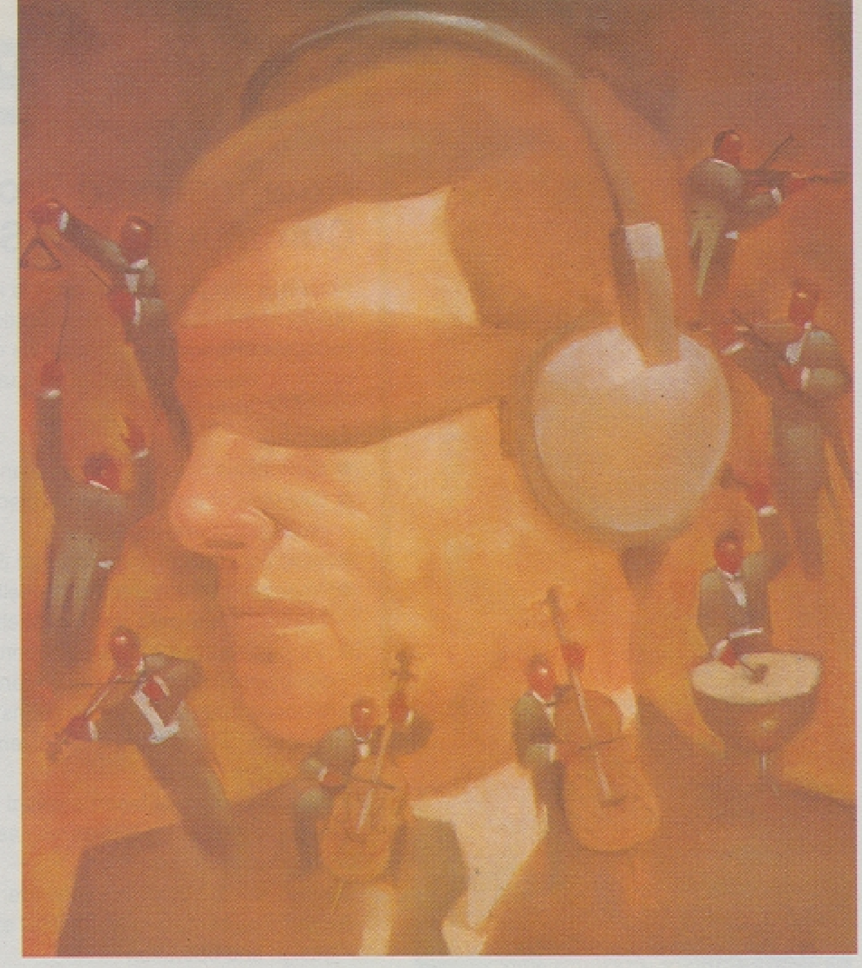


Stereo from all angles V

Loudspeakers are the most difficult aspect of stereophonic reproduction. While existing microphones are not ideal, they have reached a standard that cannot be criticised by listening through today's loudspeakers. John Watkinson explains why stereo requires more accurate speakers and presents a new measurement technique.

It is axiomatic that the accuracy of the stereophonic imaging system needs to be at least as good as the accuracy of the human ear, or deficiencies will be heard. A complete knowledge of the human auditory system will allow suitable quality criteria to be set down so that systems can be designed to a suitable performance.

An earlier part of this article set showed that the ear is a lossy device because it exhibits masking. Not all of the presented sound is sensed.

If a lossy loudspeaker were designed to a high standard, the losses may be contained to areas that are masked by the ear, and then that loudspeaker would be judged transparent. Douglas Self has introduced the term 'blameless' for a device whose imperfections are undetectable.

This is good news for the loudspeaker designer because the ear has finite accuracy in frequency, time and spatial domains. This means that a blameless loudspeaker is not just a concept. It could be made real by the application of sufficient rigour.

Although sufficiently complete knowledge of the human auditory system exists, the loudspeaker industry has, with a few exceptions, failed to act on it. Instead, it delivers commoditised products at low cost with a correspondingly low performance that has not improved in years.

Audible defects are introduced into the reproduced sound in frequency, time and spatial domains, giving the loudspeaker a kind of character which is best described as a signature or footprint.

The stereophonic sound field at the listener carries information in frequency, time and spatial domains. Unless sufficient accuracy exists in each domain there will be a loss of realism. While loudspeakers with an adequate frequency response are relatively easy to find, a sufficiently accurate time and spatial response is much more elusive.

The legacy loudspeaker is designed solely to have a flat on-axis frequency response, because traditionally that was all that was thought necessary. The time and spatial domains were neglected almost totally.

A fixation with the frequency domain

A scant knowledge of communications theory will reveal that a sine wave has no bandwidth and carries no information, and this must also be true if that sinewave is carried as sound. In fact the majority of information in sound is carried in transients, and a fixation with the frequency domain is of no help in getting the transient response accurate enough.

What is needed is a balanced design where frequency, time and spatial characteristics are afforded sensible proportions of design effort. The critical domains will be considered briefly.

The requirement for linear addition of an unlimited number of signals puts a stringent requirement on the distortion characteristics of the whole system from microphone to speakers. Non-linearity in a mono system will result in intermodulation products, but some of these may be masked by the signal because signal and harmonics emanate from the same single speaker.

In a stereo system, intermodulation between signals due to sound sources at different locations may cause distortion products at other locations in the virtual image. Attentional selectivity – the 'cocktail party effect' explained in the second of these articles* – allows the ear to detect sounds in different places that would be masked if they were coincident. It follows that the distortion criteria for stereophonic equipment must be more stringent than for mono.

The normal criterion for frequency response in mono is relatively undemanding, but this is also irrelevant for stereo. In the November issue, I illustrated the difference in level needed between speakers to produce a given location of the sound source. That figure is reproduced here as **Fig. 1**.

Listening angles

If any angular accuracy in the image is required, then the function shown in that figure must define the accuracy to which the frequency responses of the two speakers must track. If we want an angular accuracy of, say, 5° the tracking has to be within an eighth of a decibel.

As we are considering geometry, the decibel is not an appropriate unit here and it is better to use linear units. The frequency response tracking tolerance then becomes about 1.5%, which is feasible with care.

In practice the listener will not be exactly on the forward axis of both speakers, and may even subtend a slightly different angle to each. Under these conditions, the directivity of the speakers becomes critical.

The criterion must be that for angles likely to be subtended by the listener, the frequency response must be as accurate as it is on axis. This aspect of speaker design is badly neglected and leads to the requirement that the listener sit in a precise 'sweet spot'. In fact the existence of a small sweet spot is evidence of a poor loudspeaker.

A spatially-accurate virtual image can only be obtained if the loudspeaker acts as a point source. This is because the size of the original sound source is convolved with the acoustic size of the speaker.

Audio and optics

There is a direct analogy here with optics, as both are concerned with image reproduction. **Figure 2a)** shows that in optics the image is reduced in resolution because it is convolved with the spatial impulse response of the lens. Figure **2b)** shows that the acoustic image is convolved with the spatial impulse response of the loudspeakers. In both cases the ideal is an impulse response comprising a singularity or narrow spike. In a lens, this requires infinite aperture; in a speaker this requires zero acoustic size, which is a

*August 1999

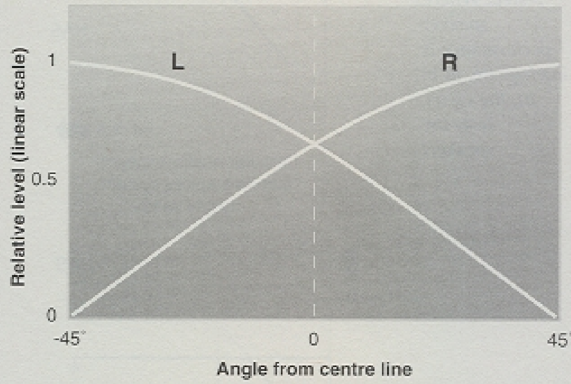
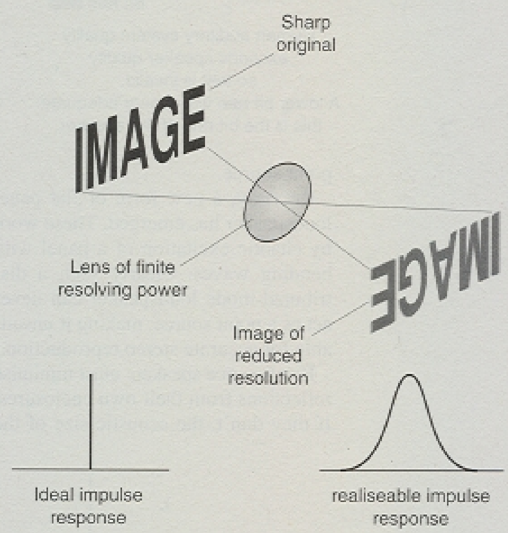


Fig. 1. Apparent position of the virtual sound source is a function of the level difference between the channels. This defines the characteristics that a microphone must have.



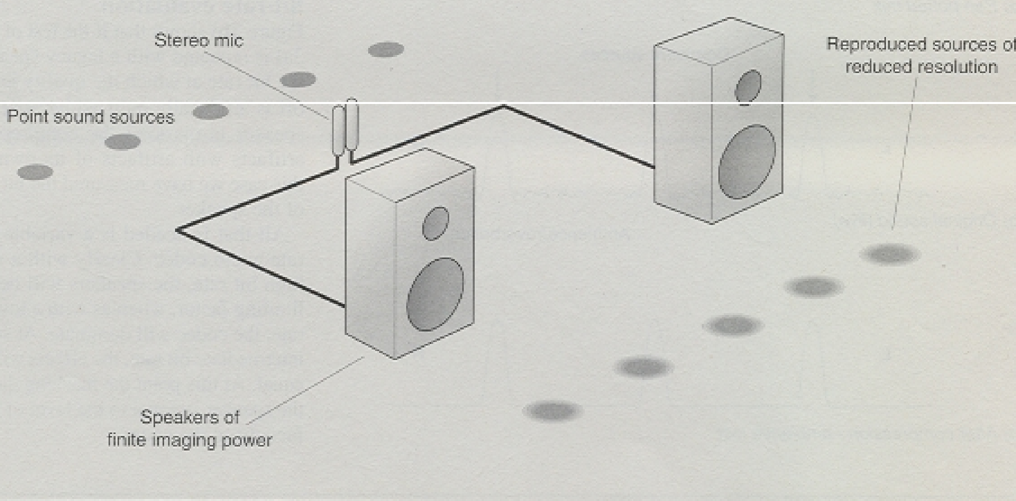
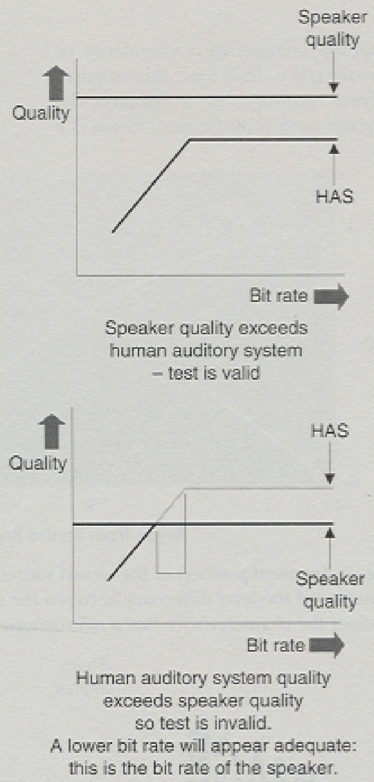


Fig. 2a) Optical systems have measurable imaging performance and can be designed for a given specification. In b), although the mechanism is the same, there is no standard unit of speaker imaging accuracy.

Fig. 3a) Where speaker performance is sufficient, the compressor is tested. Where speaker performance is inadequate, as in b), the codec tests the speaker!



point source.

Recently, a new form of flat panel loudspeaker has emerged. These work by chaotic excitation of a panel with bending waves. Clearly such a distributed-mode loudspeaker can never act as a point source, making it unsuitable for accurate stereo reproduction.

Point-source speakers must minimise reflections from their own enclosures. If they don't, the acoustic size of the speaker becomes the enclosure width. Unfortunately, powerful enclosure reflections are exactly what are provided by the legacy rectangular loudspeaker with sharp corners. These reflections are due to acoustic impedance changes.

If you could see the sound, you would double up with mirth at how ineptly it was being radiated.

A new type of measurement

Traditional loudspeaker measurements do not measure the spatial accuracy or acoustic size of the speaker. Another technique is needed to do that.

The stereo loudspeaker system can be modelled as an information channel of finite capacity which can actually be measured as an equivalent bit rate. A poor pair of loudspeakers will measure as having a low bit rate.

Psychoacoustic researchers have known for some time that poor quality loudspeakers give erroneous results when measuring hearing loss. Hearing loss is a reduction in information capacity of the ear and trying to measure this with a speaker of reduced capacity is not useful. Loudspeakers with a reputation for realism also give the highest scores when performing intelligibility tests on hearing impaired subjects.

In professional audio, the ability of an engineer to monitor sound quality can only be as good as the information capacity of the speakers used. When the speaker information capacity is limited, the presence of a defect in the signal source may go unheard and it may erroneously be assumed that all is well when in fact it is not.

Compression effects

Audio bit-rate reduction or compression is becoming popular, but it has some undesirable characteristics for stereo. It should follow from what has been said above that compression codecs can only meaningfully be assessed on speakers of adequate information capacity. It also follows that the definition of a high quality speaker is one which readily reveals compression artifacts.

Non-ideal loudspeakers act like compressors in that the distortions, delayed resonances and delayed re-radiation they create conceal or mask information in the original audio signal. If a real compressor is tested with non-ideal loudspeakers, certain deficiencies of the compressor will not be heard. The spatial compression of non-ideal stereo loudspeakers conceals real spatial compression artifacts.

Lossy compression does not preserve the original waveform, but seeks to be blameless by placing the noises where they will be masked. This can be achieved at a high enough bit rate, beyond which no improvement in quality would be observed.

Naturally, one would want to carry out listening tests to see if this goal had been achieved. If blameless loudspeakers are used, the test is valid as shown in **Fig. 3a)**. As the bit rate increases, the quality levels off where the human auditory system is masking all of the compression artifacts.

However, the legacy loudspeaker is not blameless. When a loudspeaker has a signature, how are we to know that the speaker signature is not masking compression artifacts? When listening in series, as we must, on hearing a deficiency, how are we to determine whether this was the codec or the speaker?

Bit-rate evaluation

Figure **3b)** shows that if the test of Fig. 3a) is repeated with a legacy speaker, the bit rate at which the quality levels off is artificially reduced because the speaker is masking the compression artifacts with artifacts of its own. In this case we have measured the bit rate of the speaker.

All that is needed is a variable bit-rate audio codec. Clearly with a very high bit rate, the speakers will be the limiting factor, whereas with a low bit rate, the codec will dominate. At some intermediate bit rate, the effects will be equal. At this point the masking due to the speaker is equal to the level of artifacts from the codec.

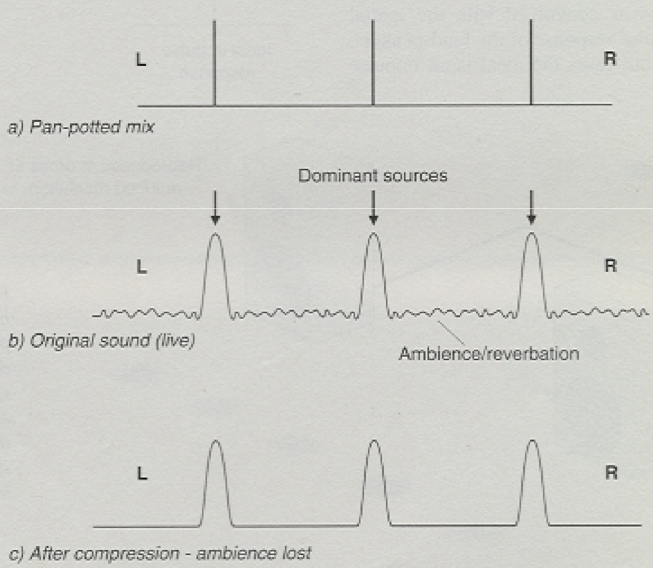


Fig. 4. In stereo, the result of bit-rate reduction is a loss of ambience.

At any lower bit rate, compression artifacts will become audible over the footprint of the speaker. The worse the information capacity of the speaker, the lower the bit rate at which the artifacts are audible.

As a result by simply varying the bit rate of a codec, it becomes possible to measure the effective bit rate of a pair of loudspeakers. The measured bit rate reflects the spatial or imaging accuracy in a way that other tests do not.

A virtual sound source from a panpot has zero width and on blameless speakers would appear as a virtual point source as in **Fig. 4a)**. As a result stereo reverb is added to panpotted mixes and this is audible between the point sources as in Fig. **4b**). A similar result is also obtained with real sources using a coincident pair of mikes. In this case the sources are the real sources and the sound between is reverb/ambience.

Figure **4c**) shows what happens when accurate speakers are used to assess some audio compressors. Even at high bit rates, corresponding to the smallest amount of compression, it is obvious that there is a difference between the original and the compressed result.

An absence of ambience and reverb

The dominant sound sources are reproduced fairly accurately, but what is most striking is that the ambience and reverb between is virtually absent, making the decoded sound much drier than the original.

However, upon reproducing such a stereo signal with the legacy square box speaker, the point sources have been spread by the speaker footprint so that there are almost no gaps between them, effectively masking the ambience as in **Fig. 5a**). This represents a lack of spatial fidelity.

Figure **5b**) shows what happens when legacy speakers are used to assess a compression system. The spatial smear increases the dominant source size to such an extent that the ambience between is inaudible. As a result if this ambience is removed by a compressor its loss will not be noticed. This is why poor speakers cannot be used to assess compressors.

MPEG and Dolby

The same effect is apparent to the same extent with both MPEG layer 2 and Dolby AC-2 coders, even though their internal workings are quite different. In retrospect this is not surprising because both are probably based on the same psychoacoustic masking model.

MPEG layer 3 fared even worse because the bit rate is lower. Transient material has a peculiar effect whereby the ambience would come and go according to the entropy of the dominant source. A percussive note would narrow the sound stage and appear dry but afterwards the reverb level would come back up.

If an opportunity arises to compare the same commercially available recording on CD and MiniDisc with accurate loudspeakers it will be obvious that the MD version is inferior. All of these effects largely disappear when the signals to the speakers are added to make mono which removes the ear's ability to discriminate spatially.

The value of subjective testing

Because of the phenomena described here, audio codecs have reached the market which produce audible artifacts even at high bit rates, despite exhaustive subjective testing. When one examines the results of any subjective compression test, it becomes clear that the type of loudspeakers used would have been those having the shortcomings mentioned above. As a result these subjective tests are invalid because the masking of the legacy speakers was masking the coder being tested.

While lossy compression may be adequate to deliver post-produced audio to a consumer with mediocre loudspeakers, these results underline that it has no place in quality stereo reproduction environment.

When assessing codecs, loudspeakers having poor diffraction design will conceal artifacts. When mixing for a compressed delivery system, it will be necessary to include the codec in the monitor feeds so that the results can be compensated. Where high quality stereo is required, either full bit rate PCM or lossless (packing) techniques must be used.

Audio codecs can only be developed fully if blameless monitoring loudspeakers are available. Such precision loudspeakers require no more than an appropriate degree of rigour during the design stage, along with some high-grade circuit design. But they have the advantage that there usually needs to be very little change between the prototype and the production phase.

Interesting conclusions

The ability to measure loudspeaker information rate allows different designs to be compared objectively. This leads to some interesting conclusions.

The measured bit rate of legacy loudspeakers is disturbingly low at about one tenth the bit rate of a CD. This means that 90% of the disc data is discarded in the speaker. Clearly if this is the case any further increase in recording format performance, such as increasing the sampling rate, must be a complete waste of time until attention is given to loudspeakers.

It is also interesting to be able to measure the noise floor of loudspeaker drive units. Most cheap drive units use ferrite magnets. As ferrite is a ceramic, it is an insulator and the magnetic field is free to move within the magnet due to the coil reaction.

Unfortunately this domain-wall movement, which was described by Barkhausen, is not linear. The result is a form of quantised flux modulation that causes audible program modulated noise on the audio. As far as I can see, it is impossible to obtain 16-bit resolution with a ferrite magnet. Conductive magnetic materials such as Neodymium and the older alnico and alcomax do not suffer from this problem. ■

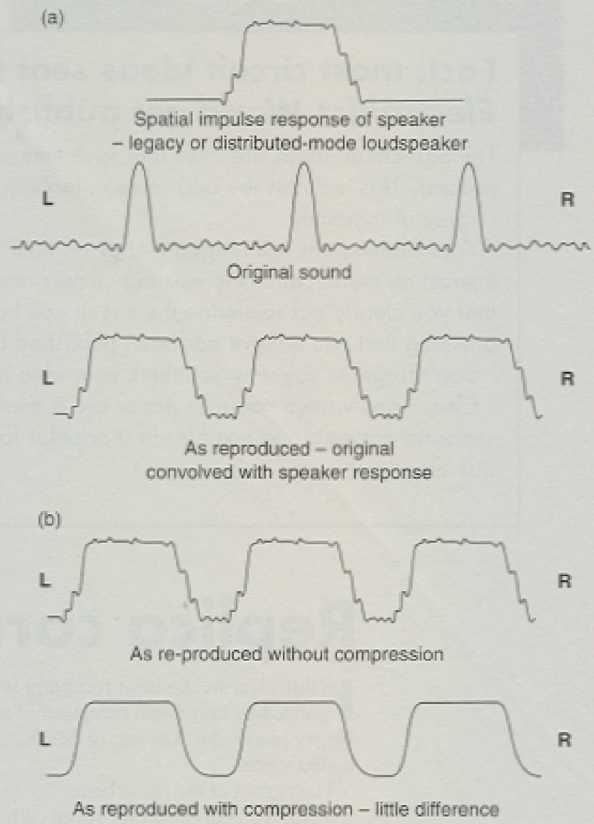


Fig. 5. In a), point sources have been spread by the speaker footprint so that there are almost no gaps between them, resulting in a lack of spatial fidelity. In b), legacy speakers are used to assess a compression system. Spatial smear increases the dominant source size to such an extent that the ambience between is inaudible.