Loudspeaker Phase, Straight And True

KEITH HOWARD EXAMINES THE WHOLE CONCEPT OF PHASE RESPONSE IN LOUDSPEAKERS, REFERRING TO PAST ATTEMPTS AT CONTROL AND CORRECTION, AND PREDICTING THE FUTURE

With the growing use of digital signal processing (DSP) in active loudspeakers, the long and somewhat tortuous history of linear-phase loudspeakers has reached an important juncture. At last linear phase can be achieved, using DSP, without the limitations inherent in attempting it with the passive or even active crossovers of old. It's some decades since linear phase was a headline issue with which audiophiles were widely familiar. So this is a good moment to examine the subject afresh, with particular attention to the what, why and how: What is linear phase? Why does it matter? How can it be achieved?

What Is Linear Phase?

Audiophiles are used to thinking of distortion as what is properly termed nonlinear distortion, where extra frequency components – harmonic products and intermodulation products – are added to a signal passing through an electrical circuit or more particularly a transducer such as a loudspeaker. But there is another form of distortion, linear distortion, in which no such cluttering of the frequency spectrum of the signal takes place.

Linear distortion occurs when the device in question has a non-flat magnitude versus frequency response, or indeed a nonlinear phase versus frequency response. The former we're used to seeing displayed as a conventional frequency response; the latter, historically, is much less frequently graphed in equipment reviews or in manufacturers' literature because it has by convention been considered unimportant, on the contestable basis that – within certain limits, of course – the human ear is largely 'phase deaf'.

Nevertheless linear distortions are classified as distortions because they modify the appearance of complex waveforms, an effect conveniently illustrated using the square wave as a probe signal (Fig. 1). The spectrum of a square wave comprises a cosine fundamental plus an infinite series of odd-order cosine harmonics of progressively decreasing amplitude. Because the ideal square wave has unconstrained bandwidth, no audio device can truly reproduce it, as all have inherent bandwidth restrictions (something which must always be borne in mind when assessing square wave behaviour). But provided that the device is linear-phase over its operating bandwidth, and the square wave fundamental frequency is towards (but not too close to) the low frequency end of that bandwidth, the reproduced waveform will be visually recognisable as a square wave, even if not possessed of infinitesimal rise time and perfectly square corners.

As an example, let's assume the square wave depicted in Fig. 1 is at 300Hz. Reproduced by a typical audio amplifier – with a low frequency limit frequency of well below 20Hz and a high frequency limit of above 20kHz – the waveform wouldn't be perfectly square but pretty close. Contrast this (Fig. 2)

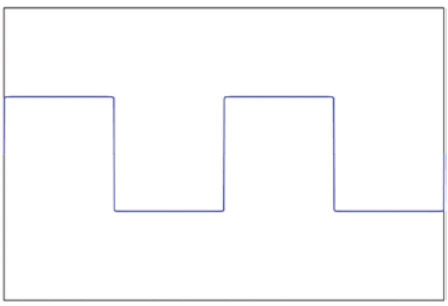

Fig. 1. The square wave, commonly used as a probe signal for phase distortion

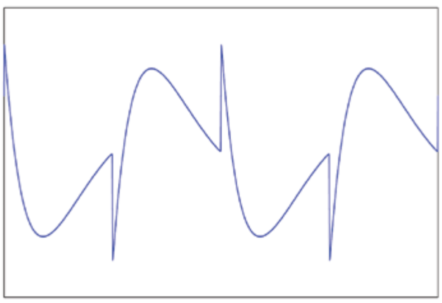

Fig. 2. A square wave, at the crossover frequency, as reproduced by an otherwise perfect loudspeaker having a fourth-order Linkwitz-Riley crossover

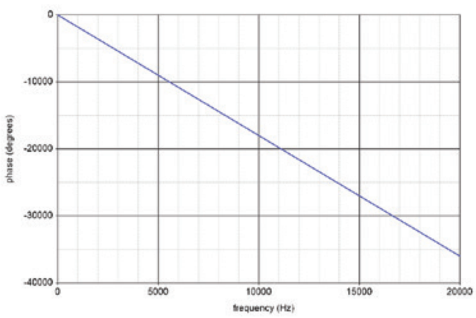

Fig. 3. Phase delay versus frequency for a circuit that introduces a frequency-independent time delay of 5 milliseconds

KEITH HOWARD

with the square wave output of a loudspeaker having a fourth-order Linkwitz-Riley crossover at 300Hz. Modification of the waveform (but not the frequency response) here is gross, due entirely to the nonlinear phase behaviour of the crossover network.

As this example suggests, the effects of nonlinear phase versus frequency response are principally a loudspeaker issue, but not entirely. Many amplifier designers will tell you that for best sound quality LF roll-off must be pushed down to 1Hz or so and HF roll-off up to around 100kHz to avoid the phase effects associated with narrower limits. The latter view isn't universally held, mind you, not least because the phase effects of a high-pass (low frequency) roll-off are considered worse than that of a low-pass equivalent.

Why 'linear phase' rather than 'flat phase'? The ideal, neutral frequency response is, at least notionally, flat – why shouldn't it be the same with phase?

A simple example will explain this. Imagine a 'black box' circuit that introduces a frequency-independent time delay of precisely 5 milliseconds (0.005s). At 20Hz this is equivalent to a phase delay of 36 degrees (a tenth of a wavelength); at 200Hz to 360 degrees (one wavelength); at 2kHz to 3600 degrees (10 wavelengths); and at 20kHz to 36,000 degrees (100 wavelengths). If we plot a graph of phase delay versus frequency for this circuit (making the phases negative to represent delay, and using a linear frequency axis), it looks like Fig. 3: a straight line.

This is the origin of the expression 'linear phase'. If a circuit or transducer has a phase delay directly proportional to frequency, this is equivalent to it introducing a constant, frequency-independent time delay. Such a circuit, because it delays all frequencies by exactly the same time interval, perfectly preserves the waveform of complex signals. Whereas if phase delay versus frequency is not linear, the waveform of complex signals will be altered – the circuit or transducer has introduced *phase distortion*.

In many circuits and transducers, magnitude versus frequency and phase versus frequency behaviour are intimately related and can be calculated one from the other. Devices that behave in this way are termed *minimum phase*. Audio amplifiers are minimum phase and so too are loudspeaker drive units. Sufficiently far away from their inherent LF and HF roll-offs, minimum-phase devices display almost linear-phase behaviour. Unfortunately the same is not generally true of multi-way loudspeakers: their drivers may be minimum-phase but their crossover networks are typically not – they display *all-pass* behaviour, with flat frequency response but nonlinear phase response. So a flat frequency response through crossover does not indicate blameless phase behaviour – far from it, as we saw in Fig. 2.

Loudspeakers actually have two distinct ways in which they introduce large phase distortions. The all-pass behaviour of many crossover networks is one of them; the other is their inherent low frequency roll-off, which even in the largest speakers occurs well above the sub-5Hz corner frequency favoured in amplifiers. In all the furore that surrounded linear phase loudspeakers 40 years ago (prompted by Bang & Olufsen's launch of its Beovox Uni-Phase range in the late 1970s), nobody said much about the LF phase problem – but no loudspeaker can ignore it if it's to be genuinely linear-phase.

When loudspeaker phase distortion is discussed and graphed, what's normally shown is not a plot of phase delay versus frequency (like Fig. 1), but rather a plot of group delay versus frequency (group delay being the local slope of the phase delay versus frequency curve expressed as a time). Converted to group delay, Fig. 1 would be a horizontal line of value 5ms. Non-constant group delay versus frequency indicates phase distortion.

Why Does Linear Phase Matter?

This question has an obvious answer: phase distortion, as we've seen, alters the signal waveform. To prevent this, and preserve high signal fidelity, we need to use loudspeakers which are not significantly phase distorting. But this absolutist approach ducks the important question: is the phase distortion introduced by conventional loudspeakers audible? If not then it's manifestly a waste of effort to endow them with linear-phase behaviour.

This issue is the subject of long-standing controversy in high-quality audio. The academic evidence – as opposed to the anecdotal evidence – mostly suggests that the phase distortion introduced by typical passive crossovers is barely audible, if at all. The late Siegfried Linkwitz (co-inventor of the popular Linkwitz-Riley crossover alignment) answered the question to his own satisfaction by building an op-amp-based all-pass filter that introduced the same phase distortion as a fourth-order LR crossover, while having a flat magnitude response. Having performed listening tests *via* headphones with and without the filter in circuit, he concluded that phase distortion was inaudible [1].

John Vanderkooy and Stanley Lipshitz performed similar experiments, also using headphones, and concluded that "On normal music or speech signals phase distortion appears not to be *generally* audible [their italics], although it was heard with 99% confidence on some recorded vocal material" [2].

But as Dan Shanefield of Bell Labs wrote many years ago [3], "What we…need is an experiment that directly compares a phase-coherent loudspeaker with an incoherent one, keeping everything else identical, and playing music." When Shanefield wrote that in 1977 such an experiment wasn't at all easy to conduct but with the emergence of digital audio and DSP it is now relatively straightforward.

The importance of correcting low frequency phase distortion in loudspeakers, caused by the speaker's bass roll-off, became manifest in 1983 in a paper written by Laurie Fincham of KEF [4]. Exploiting the ability of digital recording to capture bass frequencies with an accuracy denied to analogue tape recorders (which have bass roll-off), KEF made an orchestral recording using a digital recorder and B&K *4133* pressure microphone to ensure extended low frequency response and minimum phase distortion. This and other recordings were then used as part of listening tests in which electronic equalisation was used to vary the speakers' bass corner frequency from normal (no equalisation) right down to 5Hz.

As Mike Gough, at KEF at the time, would later recall in *50 Years of Innovation in Sound* (the book celebrating the company's 50th anniversary): "Listening to a system that was truly flat down to 20Hz was weird, and not what you would expect. You didn't always hear more bass…but male voice lost all its chestiness and we began to realise that what you heard as chestiness, what you thought was an excess of bass, was actually the transient response of the bass roll-off… Move that resulting group delay…out of band and everything sounds much more natural."

KEF's 'fix' took the form of the KUBE (KEF Universal Bass Equaliser), which first launched in 1984. Correcting the loudspeaker's LF phase alone wasn't feasible using an analogue circuit, so the KUBE (with circuit implementation by Peter Baxandall) mimicked the original experiment by equalising the speaker's bass response to force the corner frequency – and attendant phase distortion – to below 20Hz. The benefit of doing this is shown in Figs. 4 and 5, which plot group delay versus frequency for closed-box and reflex-loaded speakers with progressively lowered corner frequencies.

Less than a decade later, Michael Gerzon and others involved in the abortive B&W digital room correction project addressed the issue digitally, allowing bass phase distortion to be corrected without resort to extending the bass response (with its inherent problem of bass overload). Describing the outcome in 1991 Gerzon wrote [5], "The subjective effect of phase compensation of the bass from loudspeakers is very marked, giving a much tighter and more 'punchy' quality, with greater transparency, and interestingly a subjective extension of bass response of a least half an octave."

Given these very different assessments of the significance of phase distortion, I first attempted a 'Shanefield experiment' in year 2000 using a pair of B&W *N-803*s, with the late Alvin Gold doing the listening to avoid any bias on my behalf. Phase distortion due to the crossovers and bass roll-off was corrected by pre-processing WAV files ripped from CD, and the original and processed files were then written to CD-R for easy comparison.

At the end of the listening Alvin concluded, "On the whole I was impressed by the consistency of the improvements [with phase correction], and in every case they were improvements. The slightly emphatic quality of the *N-803*'s treble… was replaced by a more refined, less obvious and more transparent quality, and…..the bass became less blurred and boomy. The midband changed less but there were still subtle yet significant changes in the same direction, leading typically to enhanced separation between instruments. The abiding impression was of a system that was a little more natural and integrated, with (at its best) a greater sense of dynamic freedom and refinement. The differences were never gross, and often they were quite subtle, although in no case could they have been dismissed as so subtle as to be negligible."

I've had the opportunity to perform similar comparisons on a number of occasions since, and anyone owning a KEF *LS50 Wireless* can perform it, albeit without phase correction of the LF roll-off, *via* a switch in the accompanying App. My reactions broadly align with Alvin's: although the differences are not night and day, linear-phase loudspeaker behaviour adds to the listening experience in ways which anyone concerned to achieve the highest fidelity will recognise and welcome.

In his recent review of the Kii *Three* (*HIFICRITIC Vol12 No4*), Martin Colloms was in no doubt of the benefits that linear phase brings. "I cannot imagine a better demonstration of the audibility of correct time alignment; in particular the fabled linear phase promise." Anecdotally, then, there is little doubt that the phase distortion of typical passive loudspeakers is audible, and that linear phase brings clearly perceivable benefits.

How can Linear Phase be Achieved?

Although various schemes have been proposed for achieving linear-phase crossover filtering with passive networks or subtractive filtering, most modern passive and active loudspeakers – those without DSP – continue to introduce significant phase distortion through crossover ('significant' meaning a readily visible change is visible in the waveform of complex signals such as music). A few loudspeaker designers use first-order slopes to achieve a linear-phase crossover, but this raises practical issues – such as the inadequate suppression of bass and midrange drivers' out-of-band resonances – which most designers are not prepared to accept. So they prefer higher-order crossovers even though they are phase-distorting.

Bang & Olufsen's filler driver technique has never, to my knowledge, been used by anyone else, despite its patent protection having long expired. This allows higher-order crossovers to be used but requires an additional phase-correcting 'filler' driver having a bandpass response centred on the crossover frequency. Fig. 6 illustrates the low-pass, high-pass and filler driver responses for a phase-linear third-order system using Butterworth low-pass and high-pass slopes.

The practical issues of achieving linear-phase crossover behaviour with conventional passive or active crossovers has prevented linear-phase loudspeakers becoming the norm. But linear-phase crossover behaviour is almost trivially easy to achieve if those analogue filters are replaced by digital filters. Two different methods can be used: either phase-linear FIR (finite impulse response) low-pass and high-pass sections; or IIR (infinite impulse response) low-pass and high-pass filters, which are not linear-phase, plus an FIR phase-correcting filter.

The former is more adaptable; the latter (used in the Kii *Three*, for instance) has the advantage of being more computationally efficient. Whereas the length of an FIR filter increases inversely with corner frequency – so that if 200 coefficients are required for a 3kHz low-pass filter, 2000 will be necessary for the equivalent filter at 300Hz – the length of a given IIR filter is independent of corner frequency. For instance, a fourth-order Linkwitz-Riley low-pass filter requires just 20 multiply and accumulate operations when conventionally realised as four biquad filters. But in the IIR case the length of the FIR phase-correcting filter does, of course, scale with inverse frequency.

As well as making it easy to banish crossover phase distortion, DSP also allows the LF phase issue to be addressed. The first method of doing this is the digital equivalent of the KUBE: equalisation is applied to extend the bass response, and thereby push significant phase distortion to the lower reaches of the audible range or beyond. The advantage of doing this digitally is that the process can be smart, reducing the bass extension when overload threatens. (This is the approach taken in the Kii *Three*.)

The Gerzon approach – of leaving bass extension alone and fixing the phase directly – is feasible too, but it demands a correcting filter with long pre-response and, hence, considerable latency, both of which features are reason for caution. Perhaps a combination of the two approaches may prove to be optimal.

Steeper bass roll-offs generate worse phase distortion, which threatens to be a particular problem with filter-assisted bass reflex alignments. These have distinct performance advantages in other respects, so in Meridian's DSP speakers phase correction is achieved using a cascade of all-pass filters, a patented technique which Meridian calls EBA (Enhanced Bass Alignment). Readers wishing to know more about it should download patent WO 2014/106756, which has a complete description.

The Decade of Linear Phase

It's early days but the number of DSP-equipped active loudspeakers is slowly growing, and when they're as successful at redefining expectations as the Kii *Three* and the KEF *LS50 Wireless*, you have to suppose that we'll see more of them. The 2020s may well be remembered as the decade when linear-phase loudspeakers at last established themselves.

Of course, not everyone is yet ready to abandon the passive loudspeaker and the choice it allows in partnering amplification, but outboard DSP can achieve much of what inboard DSP can. We have already seen moves in this direction with Devialet's *SAM* and Linn's *Exakt* technologies. Expect to see other likeminded developments in the next few years.

References

1) S Linkwitz, 'Active Crossover Networks for Noncoincident Drivers', *J Audio Engineering Society*, Vol 24, No 1, January/February 1976

2) S P Lipshitz, M Pocock and J Vanderkooy , 'On the Audibility of Midrange Phase Distortion in Audio Systems', *J Audio Engineering Society*, Vol 30, No 9, September 1982

3) D Shanefield , 'Further experiments on phase audibility', *Wireless World*, p79, October 1977

4) L R Fincham, 'The Subjective Importance of Uniform Group Delay at Low Frequencies', AES 74^{th} Convention, October 1983; later published under the same title in the AES Journal, June 1985

5) M Gerzon, 'Digital Room Equalisation', *Studio Sound*, reprinted by B&W

6) E Baekgaard, 'Loudspeakers – The Missing Link', AES 50^{th} Convention, March 1975

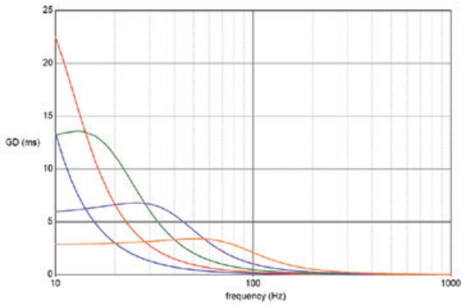

Fig. 4. Low-frequency group delay versus frequency for a closed-box loudspeaker of maximally flat (B2) alignment, showing the effect of progressively reduced corner frequency: 80Hz (orange trace), 40Hz (violet trace), 20Hz (green trace), 10Hz (red trace) and 5Hz (blue trace)

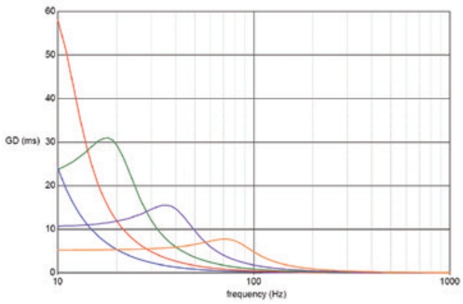

Fig. 5. Low-frequency group delay versus frequency for a vented-box (reflex-loaded) loudspeaker with maximally flat (B4) alignment, showing the effect of a progressively reduced corner frequency: 80Hz (orange trace), 40Hz (violet trace), 20Hz (green trace), 10Hz (red trace) and 5Hz (blue trace)

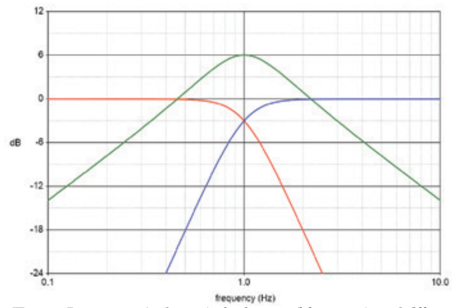

Fig. 6. Low-pass (red trace), high-pass (blue trace) and filler-driver (green trace) responses for a linear-phase 3rd-order Butterworth crossover