

feature

understanding
monitors

monitors

demystified

PART 1: COMPROMISE & APPROXIMATION

Monitor speakers affect nearly all the decisions we make when recording and mixing — yet most of us know very little about how they are designed, and why they sound the way they do. In the first of a new series, **Phil Ward** explains what goes into the design of typical passive nearfields, and the effects they can have on what we record.

If you stop to think about it — though few people ever do — monitors are a fantastically important part of our studios. You use these lumps of chipboard, plastic, paper and glue on every recording you make, and they colour your sound more fundamentally than pretty much anything else. Perhaps the lack of consideration they are normally given is because monitors are so simple to use. They rarely need adjustment, have no software, no internal memory, no processor, no user manual (well, at least, not one you can find), never need an update and probably haven't even got a power socket.

Once a month or so you'll probably see, and no doubt even read, a monitor review in this very magazine. Typically, the reviewer will describe the design of the product, perhaps write a little about the engineering propaganda being disseminated by the manufacturer and then move on to describe some of the subjective audible characteristics that the monitor 'imprints' on recordings.

This imprinting process is analogous to another situation involving a piece of equipment many of us have in our studios, which is also called a monitor, rather confusingly; the screen on your computer. If you do any graphic design and/or computer-based illustration, you'll know how easy it is to be caught out by the colour distortions that screens add to everything you draw. It's especially frustrating if you also use a printer. The screen and printer colours are very unlikely to match and neither will be nominally 'accurate'. However, in the case of colour-matching, there are calibration procedures, software patches and set-up routines that, if you have the time and patience, can be used to minimise the difference between input and output.

There's nothing of this kind with monitor speakers, however, even though the basic problem



— matching an input signal with an output — is similar. So why are speakers different? How is it that a 'simple' pair of speakers can imprint such audible characteristics on music? What are the mechanisms at play, and given the existence of these mechanisms, how do speaker designers decide how their products should sound? Do the specifications that manufacturers publish have any value? And perhaps most importantly, is it possible to be smarter in our choice and use of monitors so that we can minimise, or at least better understand, the contribution they make to the sound of our recordings? Well, yes it is, and with only the analysis of a typical, smallish nearfield monitor for a safety net (and no, I'm not going to tell you which one it is, that wouldn't be fair), I'm going to make an attempt over the next couple of issues to shed a little light on the dark, mysterious world of monitor speakers and what horrible things they might be doing to your lovingly crafted music.

The Art Of Compromise

It's no fun being a speaker designer — your entire professional career is devoted to making the least awful sounds you can out of beautiful ones. Give speaker designers the perfectly recorded sound of a Stradivarius, a Martin, a Steinway or a voice and deep inside they'll know the very best they can do is compromise and approximate (slice a speaker designer and, like Blackpool rock, you'd probably find those two words printed right through the middle). The compromises start as soon as a new product is conceived and they arrive from three directions: first, the inconvenient fact that our ears are cleverer at listening than speakers are at speaking; second, the more inconvenient fact that users tend not to have unlimited space or cash; and third, the even more inconvenient fact that, as yet, nobody has developed a technique for telling the laws of physics who's boss. However, because they are the fundamental villains of the piece, I'll start this month with the laws of physics.

Relatively speaking, music is a wide-bandwidth beast with a wide dynamic range. A speaker is limited in both bandwidth and dynamic range (the latter perhaps to a far greater degree than many realise). When speaker designers try to widen both the bandwidth and the dynamic range of their products, the laws of physics exact a heavy penalty and the listener hears it happening. Following is an explanation of the classic low-frequency bandwidth versus dynamics trade-off, illustrated by a few acoustic measurements from the typical nearfield monitor I borrowed to write this piece.

Pass The Port

Like many speakers we can all name (there are pictures of several dotted around this piece in case your memory needs jogging), our No-Name Acoustics nearfield comprises a couple of drive units mounted on the front surface of an 'airtight' box. The box is there to suppress the output from the rear of the bass driver so that, at low frequencies, it doesn't cancel the output from the front. In addition to this (and providing somewhere for you to put your soft-toy studio mascots), the box also fundamentally defines the low-frequency limit for the system, as the 'stiffness' of the air inside makes it harder and harder for the drive unit cone to move as the output frequency falls (the cone displacement required to generate constant acoustic power increases exponentially as you move down the frequency spectrum).

The designers of the No-Name have chosen to reduce the bandwidth-limiting effect of the box by employing 'reflex loading' — a hole in the box extended by an internal tube in an arrangement often known as a 'port'. Reflex loading extends the low-frequency bandwidth of a speaker by adding a 'helper' resonance to the system. At really low frequencies, the 'slug' of air inside the tube simply pumps

backwards and forwards out of phase in response to movement of the bass driver, and so contributes nothing to the acoustic output. But at the port's resonant frequency (which is defined by the cross-sectional area and length of the tube, and the volume of the box), the slug moves in phase with the driver, adding significant extra acoustic output and reducing the movement of the drive unit cone. Sounds like a free lunch, doesn't it? But, as ever, there's no such thing.



Here's why not. Firstly, the extension of low-frequency bandwidth produced by reflex loading comes at the expense of dynamic accuracy, as the amplifier now not only has to control the movement of the drive unit cone but also the air in the tube. Secondly, reflex loading causes the system to display rapid change of phase with frequency, and if you express phase change as time you can see that a reflex-loaded speaker effectively adds a delay to low frequencies. That time delay can be expressed as a distance (ie. the speed of sound multiplied by time — more on this later) with the result that when speaker designers choose reflex loading they are also effectively choosing to move the bass player's fundamentals back three metres or so. And you thought kicking him was the only way to do that...

Thirdly, as soon as you increase the sound level to a point where the air passing through the port becomes turbulent (as any substance does eventually when put through a pipe at a certain speed — it's those dastardly laws of physics again), the air becomes a non-linear mess. At best, the port will make some odd farting noises, at worst, it effectively stops working at all. However, you can make the air continue to flow in a linear, non-turbulent fashion at higher sound levels by designing the exit surface of the port in a flared shape. Unsurprisingly, the manufacturers of the No-Name Nearfield have taken exactly this approach (see the 'Load Of Balls?' box on page 194 for another way of reducing port turbulence).

Fourthly, with or without generous flaring on entry and exit, a reflex port is fundamentally non-linear. The acoustic impedance (ie. resistance to movement) of the big wide world on the outside of the box is obviously not the same as that *inside* the box, so the flow dynamics as the air in the port rushes outward are *not* the same as when it rushes inwards. As a result, the average air pressure inside the box will drift away from the nominal atmospheric pressure outside, and the bass driver's voice coil will take on an average offset away from its nominal rest position. This offset can be relatively innocuous and result only in a slight increase in low-frequency harmonic distortion, but if the driver happens to have non-linearities in its magnet system that cause an offset in the same direction, they can be completely disastrous as the cone suddenly 'locks out' at one end of its travel. I've seen this happen in a reflex-loaded bass guitar cabinet fitted with a very well-regarded American bass driver. It's quite an impressive trick, and boy does it quieten the bass player...

Grasping The Graphs

That gives you an idea of the theoretical compromises involved in reflex loading. Now let's see how they affect the No-Name Nearfield in practice, by examining its frequency response graphically.

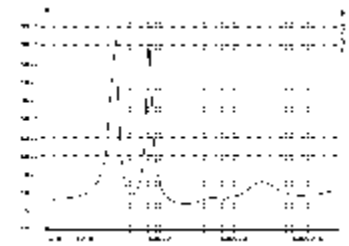
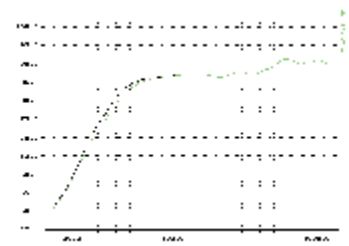
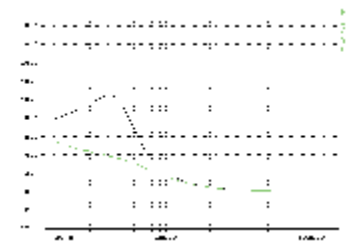
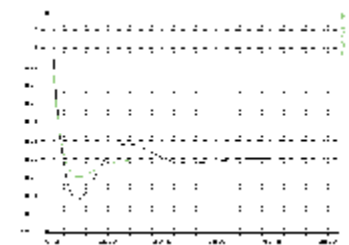
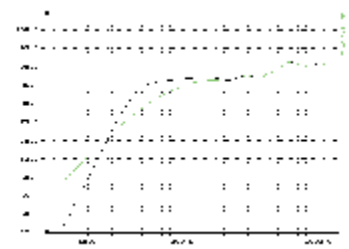
Figure 1, shows the actual low-frequency response of the No-Name Nearfield — the black line shows the sum of the output of the driver and the output of the reflex port from 20Hz to 1kHz,

plotted as output level in dB against frequency in Hz. The designers have chosen to give the No-Name a slightly overdamped response, where serious roll-off commences below around 55Hz (a bass guitar's bottom E has its fundamental at 44Hz). I wrote 'chosen', because tweaking the various parameters of box volume, port resonance, and drive unit allows an almost infinite number of different responses to be achieved; for example, responses that maximise bandwidth at the expense of response accuracy, time-domain behaviour and power handling at one end of the scale and, at the other end, responses that use the port rather more subtly — more perhaps as an aid to power handling (by reducing the amount the cone moves, or excursion, as designers like to call it) than to bandwidth extension. The green overlay in Figure 1 shows how the low-frequency response of the No-Name would look if it were not a reflex system, but a portless, closed box. This perhaps isn't quite fair, because the parameters for the No-Name box and driver were presumably chosen in the knowledge that a port was going to be used, but the overlay does illustrate the bandwidth extension offered by the port (as a closed box, the No-Name is 3dB down at around 100Hz compared to its ported self), together with its characteristic fast roll-off.

The black curve on Figure 2 shows the impulse response of the No-Name over the same frequency range — in other words, a plot of the excursion of the speaker cone against time when a fast click or impulse is put through it. If you're surprised that the movement of the cone is expressed on the left-hand axis in Volts, don't be — speakers work because a voltage put through a moving-coil driver creates movement, after all. The negative voltages simply indicate that the cone is behind its usual rest position (which is equal to 0V).

Now you know what the graph means, can you see how much difficulty the speaker has in stopping after the impulse is put through it? The characteristic frequency of the ringing overhang is equal to the resonant frequency of the reflex port. If a bass player, say, were to play a note at that frequency and stop instantaneously (we can all dream...) the speaker would add that resonant tail to his playing (in fact he wouldn't need to be playing a note at exactly the port resonance, just somewhere near it would do). The No-Name Nearfield actually has a pretty well-behaved and well-damped port, but in extreme circumstances, where a highly resonant port has been chosen, the port resonant frequency overhang can begin to affect the accuracy with which a speaker reproduces pitch. It can make the pitch of things sound slightly hazy, if not definitely out of tune. And you thought it was that bass player trying his fretless... Again, there's a green overlay in Figure 2 showing the time-domain behaviour of the No-Name with the port blocked up. Notice how much better the cone stops moving?

Figure 3 is a graph of delay in milliseconds plotted against frequency in Hertz. The black curve illustrates the added time delay of the No-Name at low frequencies (this is known as the 'group delay' and is phase change expressed as time, as explained earlier). Again, a green overlay shows what happens without the port. Comparing the black line with the



green, and taking a specific example, you can see from the graph that a bottom E at 44Hz with the port occurs over 10 milliseconds later than without it. Since the speed of sound is 330 metres per second, this delay is equivalent to moving bottom E back over three metres (as mentioned earlier). However, the real-world specific audibility of such low-frequency time delay effects is a subject of much discussion and argument among speaker folk and hi-fi reviewers. I sit firmly on the fence, and I've included the group delay graph primarily to illustrate the complexity of the issue (and so that I could do that gag about the bass player's fundamentals). There are so many factors that influence the perception of low-frequency performance that it's pretty much impossible to identify one and be sure of its guilt. I think there's little doubt though that the choice of a low-frequency response that seriously distorts the dynamic, temporal and even pitch information present in the signal, all in the name of a little more bandwidth, is not the right one for a monitoring tool.

Perhaps the most interesting measurement on the low-frequency end of the No-Name Nearfield — especially in the context of a speaker's dynamic range failings — is illustrated in Figure 4. This graph of output level against frequency shows the differing frequency response at two different (but fairly low) drive levels. This time the green curve has nothing to do with the presence of the port or not — it simply represents the response of the ported No-Name at a higher drive level than the black curve. This shows the compression introduced predominantly by the port, as the two curves should be the same. The difference is only about 1dB from 30 to 60Hz but then the green curve's drive level wasn't particularly high either. At higher levels I would expect the port compression effect to become quickly more obvious. And if the No-Name didn't have a reasonably well-flared port exit I'd expect the port compression to become, if not specifically audible, then certainly a significant influence on the way compression was used at a mix.

Now, you probably read that paragraph and felt cheated because I said it was interesting. Bear with me — here's why. The No-Name was lent to me by its owner (brave man), and one of his subjective feelings about the product is that it sounds a little 'bass-light' until it's being driven reasonably hard. Now that looks to be at odds with Figure 4, which shows the low-frequency level *decreasing* as the product is driven harder. So what's going on? Well, it's hard to say exactly without carrying out a *lot* more measurements, but based on my experience with other monitors, I am in a position to speculate why the readings I have taken seem at odds with the owner's carefully-formed, long-term opinion. Firstly, it's well known that, far from being flat, the frequency response of the human ear is level-dependent at both frequency extremes. Secondly, the actual situation is far more complex than can be revealed by one or two measurements. It's quite possible, for example, that after initially falling with increasing level, the response of the No-Name develops a resonant peak as its port starts to misbehave more seriously. Or perhaps the resonant low-frequency energy that the speaker begins to generate as other compression effects and non-linearities begin to unfold is perceived as the missing bass. Whatever the underlying mechanisms for the subjective feeling, measurements of the No-Name reveal that it has a pretty well-behaved and sensible low-frequency response. So, rather than drive the speakers into non-linear behaviour in order to feel they are working properly, it might be better to change their position in the room (nearer to rear wall or corner) so that they sound 'right' at lower levels, which produces more linear behaviour. In turn, this 'linear' solution is likelier to result in monitoring that more accurately reflects the recording and the mix.

Tips For The Ported

All of the above might read like a character assassination of any speaker with a reflex port, but that's really not my intention. Thoughtfully and carefully conceived reflex loading (and I'd include the No-Name in that category) can work well, but the technique by its very nature has characteristics that it pays to be aware of. So, if you have a pair, how best to work with


reflex-loaded monitors? Maybe try a few of the following little experiments.

- Find out what frequency the port is tuned to. It'll be somewhere between 30Hz and 80Hz, and the smaller the speaker, the higher it's going to be. If the manufacturer publishes technical data including an impedance curve, the port frequency (for reasons I don't have space to go into here) is the minimum impedance value between the two low frequency peaks. Figure 5 is the impedance curve for the dear old No-Name, which reveals its port frequency to be around 43Hz. If you can't get hold of an impedance curve, play a sine wave patch from a keyboard at a reasonable level and just look at the movement of the speaker cone as you sweep the note up an octave and a half from low B to E. The point at which both the cone moves the least and there's a healthy breeze blowing from the port is the tuning frequency. Once you know the port frequency, you can be pretty certain that if your monitors are going to misbehave, that's where they'll do it. It's useful to know the nearest musical note to the tuning frequency, too, because you can then be aware that any consistent problems associated with that note may well be a monitor artifact and can perhaps be safely ignored (so no more worrying why that F sharp always sounds all boomy and slightly out of tune...).

A Load Of Balls?

Further to the subject of avoiding turbulence in speaker ports, the recently introduced Nautilus hi-fi speaker range from B&W features a technology they call FlowPort. FlowPort is an array of small hemispherical dimples set into the flared exit surface of the reflex port. The argument proposed is that the dimples work by helping the airflow remain non-turbulent "in much the same way as the remarkably similar dimples on the surface of golf balls". I've absolutely no idea if this really works or is significant, but the engineer (or was it a marketing man, I wonder?) who came up with a technology that ties a range of deeply stylish, very male-oriented and high-value hi-fi speakers to the similarly upmarket, male-oriented sport of golf deserves some sort of gong. That's marketing genius, that is.

- Have a good critical listen to some simple low-frequency material recorded at different levels. Maybe record a bass guitar or keyboard piece especially for the purpose. Does the quality and character of the sound change alarmingly with level? If it does, you can bring this knowledge to bear when you record, or more likely when you mix.
- Put a sock in it — literally. If you have a pair of passive reflex-loaded monitors you'll do no harm just having a listen to how they behave with the ports blocked. The measurements on the No-Name Nearfield illustrate the fundamental change that blocking the port can bring about on the low-frequency behaviour of the system. If you have a bass problem on a mix, blocking the port is as useful as trying a different monitor or even room — perhaps more so, because you're only changing one variable. Don't however, just listen to bass level — there's bound to be less bass with the port blocked. Listen instead to how the bass character changes with level, and to how clearly you can identify the pitch of the bass notes in each case — with port blocked *and* unblocked.

Next month, we'll explore more deep and dark speaker mysteries, this time further up the bandwidth, including the complete tosh that is the typical 'frequency response' curve. Now, where's my anorak got to...? 

Glossary

http://www.sospubs.co.uk/sos/regular_htm/glossary.htm

SOUND ON SOUND

Europe's No1 Hi-Tech Music Recording Magazine Sound On Sound

Go To SOS 
Website 
Homepage 

Media House, Trafalgar Way, Bar Hill, Cambridge CB3 8SQ, UK.

Telephone: +44 (0)1954 789888 Fax: +44 (0)1954 789895

Email: info@sospubs.co.uk Website: www.sospubs.co.uk

© 2000 Sound On Sound Limited. The contents of this article are subject to worldwide copyright protection and reproduction in whole or part, whether mechanical or electronic, is expressly forbidden without the prior written consent of the Publishers. Great care has been taken to ensure accuracy in the preparation of this article but neither Sound On Sound Limited nor the Editor can be held responsible for its contents. The views expressed are those of the contributors and not necessarily those of the Publishers or Editor.
