WHITE PAPER

dCS EXPANSE

Enhancing the Headphone Experience with *dCS* Expanse

dCS White Paper Series: #001
Released December 2020

info@dcsltd.co.uk
@*dCS*onlythemusic
www.dcsltd.co.uk

dCS
ONLY THE MUSIC

Enhancing the Headphone Experience with *dCS* Expanse

The increasing numbers of music listeners moving toward headphone listening has resulted in an imbalance between the way in which performances are monitored during recording (mainly on loudspeakers) and listened to on playback (increasingly on headphones). Although this imbalance has been understood for many decades, at *dCS* we felt that none of the existing technologies developed to address this were worthy of inclusion in our products.

The challenge involves not only electroacoustic factors but also psychoacoustic elements. When we listen to loudspeakers the signal from each loudspeaker is actually heard by both ears. The acoustic modification of the sound on the way to each ear is represented by Head-Related Transfer Functions (HRTFs). A straightforward approach based on HRTFs does improve the direction of the sound, but it also reduces the apparent amount of reverberation in the recording. It is the reverberation that gives us a sense of distance, space, and naturalness.

While artificial reverberation is an acceptable artistic tool in the recording studio, we do not consider it appropriate for the task at hand. The *dCS* approach is to work in sympathy with the recording, preserving the reverberation and tailoring our processing to the correlation between left and right signals.

Listening to digital music through a *dCS* system is in some ways a staggering experience. With *dCS* Expanse, headphone listeners can enjoy all of the hallmarks of *dCS* playback – precision, detail, immersive and engaging sound – along with a unique experience that stays true to the original recording.

Headphones and the dawn of stereo

When we think of the early days of stereo, we probably think of Alan Dower Blumlein, whose patent application describing almost every element of the technique was lodged in 1931. But the roots of stereo go back a whole 50 years before that, to the first-ever stereo broadcast – which was heard on headphones. The brilliant inventor and pioneer aviator Clement Ader placed several widely-spaced pairs of microphones along the footlights at the Paris Opera and fed the audio from them to a listening booth at the Paris International Exhibition of Electricity in 1881.

There, a handful of listeners at a time, standing at wall-mounted telephones, could hold a pair of earpieces to their ears and experience live stereo from the Opera. *Scientific American* reported, "Every one who has been fortunate enough to hear the telephones at the Palais de l'Industrie has remarked that, in listening with both ears at the two telephones, the sound takes a special character of relief and localization which a single receiver cannot produce." Stereo broadcasting by wire didn't catch on, probably due to the requirement for two pairs of wires to subscribers' homes, although Ader's *Théâtrophone* cable music service provided live mono broadcasts in Paris for some years and the idea was replicated in other cities such as London.

But while the first stereo was designed to be heard on headphones, this was not the case when stereophonic sound came to the living room in the late 1950s – the first stereo vinyl LPs were released in 1957-8. Here, the expectation was that you would listen seated in a "sweet spot" with left and right loudspeakers in front of you, forming an isosceles triangle with the listener, the ideal angle between the speakers at the listening spot being 60 degrees. Several different recording techniques were, and are, used, ranging from a simple coincident pair of microphones to extensive multitrack recording systems, and the effect experienced by the listener varies according to the technique in use.

Some techniques, such as the sum-and-difference technique using a single-point pair of microphones, are great for capturing ambience, reverberation and the sound of a recording environment such as a concert-hall, and often deliver a feeling of depth to the recording as well as exhibiting remarkable image stability that is less dependent on listener position. More common techniques such as multitrack mixing, where each sound is individually located in the soundstage from left to right depending solely on the relative amounts of signal fed to left and right channels, tend to produce a soundstage that occupies a straight line between the loudspeakers and can be thrown out if you move from your central listening position.

The common factor in all these techniques, however, is that when stereo recordings were made – particularly in the heyday of stereo vinyl – they would have been monitored almost exclusively on loudspeakers. Headphones might have been used to refine the positions of instruments within the soundstage or to check aspects of a mix, but they were the exception rather than the rule. In more recent years, as more and more musicians have preferred to record at home and in their own studios, headphone monitoring has become more common. But even today, you can expect that the majority of popular music mixes will have been monitored on loudspeakers, even if on smaller, "near-field" monitors; and in the case of classical music, loudspeaker monitoring is still very much *de rigueur.*

At the same time, the development of the personal stereo over the past four decades – from the first Sony *Walkman* onwards – has seen more and more people listening on headphones: from tiny earbuds for use with phones and personal music players on the go, to audiophile-quality units, in-ear or on-ear, for serious hi-fi listening at home. Meanwhile, for space and convenience reasons, the amount of loudspeaker listening, even at home, has reduced – with the exception of in-car audio.

This trend has resulted in a potential imbalance between the way in which performances are monitored during recording (mainly on loudspeakers) and listened to on playback (increasingly on headphones).

If our fundamental intention with high-fidelity music reproduction is to hear as close as possible to what the production team in the studio heard when they played back the mix and said, “That’s the one” – which seems a fairly non-controversial goal – then we have to address this imbalance. *How do we make listening on headphones sound like monitoring on loudspeakers?*

The headphone challenge

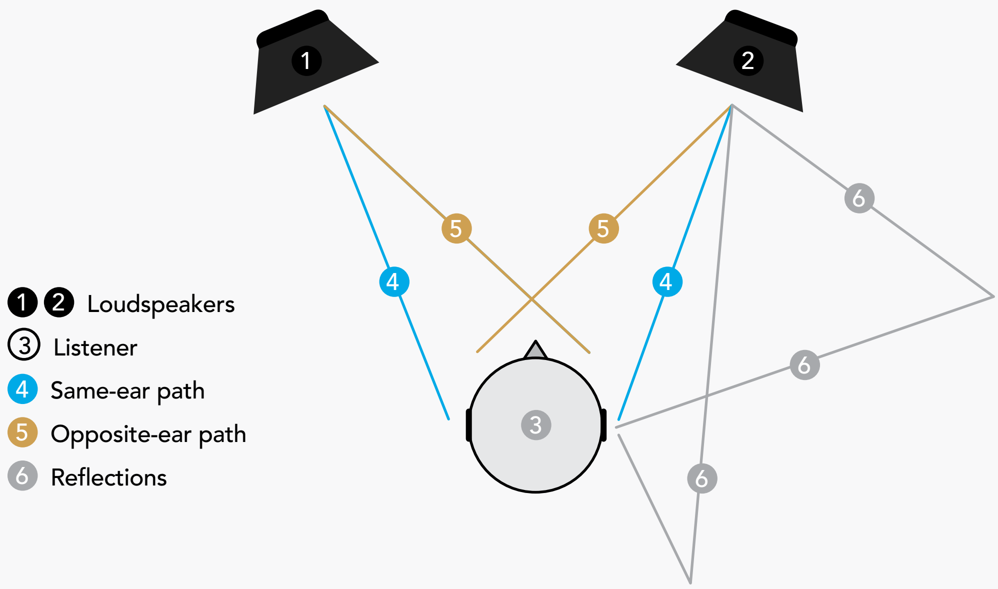
That was the challenge that we sought to address when we developed *dCS* Expanse headphone processing and it is not a simple challenge, involving as it does not only electroacoustic factors but also psychoacoustic elements – such as the effect of being isolated from the sound of the listening room, something that deeply affects how we hear playback on loudspeakers.

Several manufacturers and researchers, stretching back to the early days of home stereo, have offered solutions to this question and the results vary widely in both principle and effectiveness. Here at *dCS*, we felt that none of the existing technologies of which we were aware were worthy of simply licensing for inclusion in our products. We had to find a better way – and as is common practice at *dCS*, we developed our technique largely from scratch.

First of all, we have to determine what those factors affecting the translation of speaker-monitored recordings to headphone listening actually are. How do the two differ?

Subjectively, the first, and major, thing you notice about listening to an original stereo loudspeaker-tailored recording on headphones is that the stereo soundstage image is *inside your head*, following a kind of “arch” between left and right ears, where a sound source that’s equally loud in left and right ears will apparently be positioned in the middle of the top of your head. This is very definitely *not* what you hear on loudspeakers. In addition, the frequency balance experienced on headphones differs from that experienced when listening to loudspeakers. We also have to consider the effect of the many reflections reaching our ears from the walls, floor and ceiling of the listening room: such effects are always present, and listening to a pair of loudspeakers without such effects (in an anechoic chamber, for example) sounds strange and unnatural. Yet the acoustics of a real listening room with loudspeakers could be *anything*.

Most importantly, when we listen to music on headphones, the left and right signals are coupled directly to our left and right ears respectively. When we listen to loudspeakers, it’s no longer quite so simple. First, the signal from each loudspeaker is actually heard by *both* ears. The signal from the left speaker, for example, reaches the left ear first and then a little later reaches the right – it’s also a tiny bit quieter. Parts of our head get in the way of the signals reaching the right ear, effectively reducing the level at certain frequencies. Sound is reflected off the head and the torso, affecting the sounds that reach the ear. Even the pinnae – the fleshy bits of the outer ear – affect what we hear and do so in ways that vary from one individual to the next. The changes in frequency, phase, amplitude and impulse response caused by these factors are called *Head-Related Transfer Functions* or HRTF.



At the very least, this means that at the heart of any headphone-listening loudspeaker emulation system will be a "cross-feed" of some sort: the need to feed a portion of the left signal to the right ear, and vice-versa. But overall, we need to consider all the factors that affect how we should generate headphone signals from loudspeaker-oriented recordings and what they contain.

Some of these factors are associated with how we, as humans, experience and localise sound sources that we hear. There are three primary mechanisms that we use for stereo localisation, and these are:

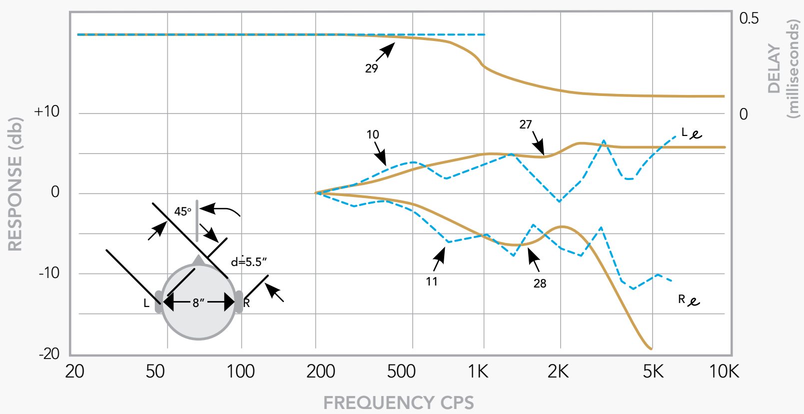
- The difference between the levels a sound exhibits at the left and right ears ("inter-aural level difference" or ILD). This is the mechanism used to create the stereo sound-stage in recordings mixed from multitrack – the most common type of recording.
- The difference in *arrival time* of a sound between left and right ears (of the order of hundreds of microseconds) – also known as inter-aural time and phase difference, or ITD. It is actually possible to localise a sound anywhere in a stereo soundstage by using phase differences alone, even if the levels are identical in the two channels ("phase-shift panning"): both level and phase are important in localising a sound source. Distinct from ITDs is the Precedence Effect, which finds that when two identical sounds are presented in close succession (a matter of a few milliseconds apart) they will be heard as a single acoustic event. This is related to the "Haas Effect" – the finding that humans localise sound sources in the direction of the first arriving sound despite the presence of a single reflection from a different direction.
- *Reverberation:* the way in which an original sound creates echoes, reflections and diffusions as it bounces off surfaces and objects within the audible field and beginning after we have heard the sound directly from its source.

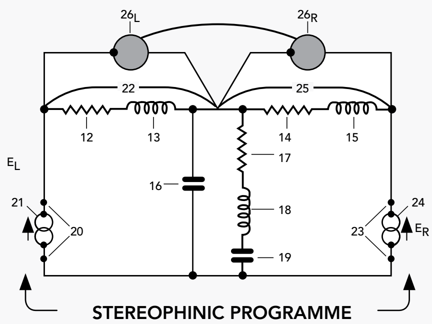
It makes sense to try to emulate these mechanisms, although none of them are without issues. We might try to emulate ITDs by delaying a sound between the left and right ear feeds, for example – but the delay could cause interference in the case of an original centrally-placed signal, creating a comb-filtering effect – but such an effect also occurs in nature. Factors like this need to be allowed for as much as possible.

How we handle reverberation requires special treatment. Reverberation levels appear reduced when listening on headphones, yet the ratio of direct to reverberant sounds is one of the most important audible cues that give us a sense of distances. We tend to focus on the direct sound, but in fact over half the sound we hear when listening to a pair of loudspeakers is reverberant sound. Thus it's important to maintain the apparent amount of reverberation. But you can't simply add artificial reverb: it may be fine in some cases (such as a studio multitrack mix) but it could sound very artificial on a coincident-miked classical concert in a concert-hall. Remember those "DSP Effects" settings on 80s hi-fi systems? No, we never used them either. What we need instead is what we might call "reverberation recovery" – getting the best *representation* of the existing reverberation, whether it's from a concert hall or studio reverb processor; retaining it rather than losing it. Significant aspects of reverberant sounds lie in the *difference* between left and right signals, which is one reason Blumlein-style sum-and-difference coincident-mic recordings are so good at capturing acoustic environments.

Fundamentals of loudspeaker-to-headphone transcoding

Work on the conversion of stereo recordings "intended for" loudspeaker listening so that they can deliver a similar experience on headphones was first performed early-on in the history of consumer hi-fi. Back in 1961, Ben Bauer at CBS in the United States published an Audio Engineering Society paper that described a method of signal conversion between loudspeaker and headphone – and the other way round. The technique became the subject of a patent a couple of years later. However, while Bauer's approach has a good deal to teach us, there are several factors to be aware of.





Diagrams from Benjamin Bauer's 1963 patent

First, it needs to be borne in mind that the kind of stereo experience he was trying to recreate with a loudspeaker-based recording played back on headphones was the kind of listening experience you would achieve by recording with a dummy head – in other words a *binaural* experience. He wanted a conventional loudspeaker-monitored recording to sound, on headphones, as if it had been recorded with a dummy head. This may not be entirely desirable in the context of modern recordings, most of which have very little in common with a dummy-head experience. In addition there are shortcomings to Bauer's approach that, while difficult or impossible to avoid at the time, could today be addressed with modern Digital Signal Processing.

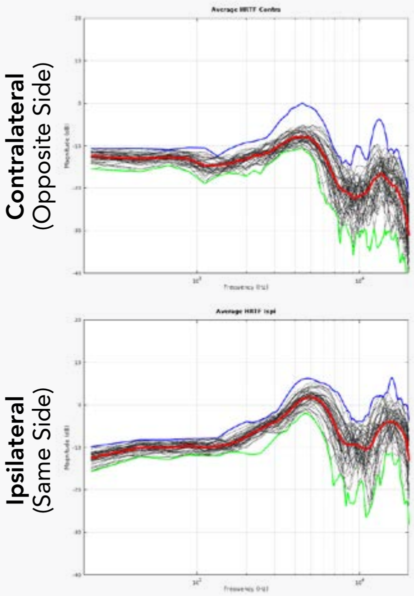
Bauer's technique addresses the need for overall equalisation to compensate for some of the characteristics of headphone listening, and the need for a *cross-feed between left and right channels* to take the sound out of the listener's head. However, there are other requirements that Bauer's technique does not address, and areas where his approach appears to be incorrect. He doesn't address the loss of *reverberation* on headphones, which can significantly affect the sense of space in a recording. In addition, Bauer's method of generating a cross-feed signal introduces a delay at low frequencies only, which may be undesirable. The cross-feed signal may work better if it extends to higher frequencies, and if alternative degrees of bandwidth limiting are used. And more recent research calls into question Bauer's values for the delay in the cross-feed. And there's more.

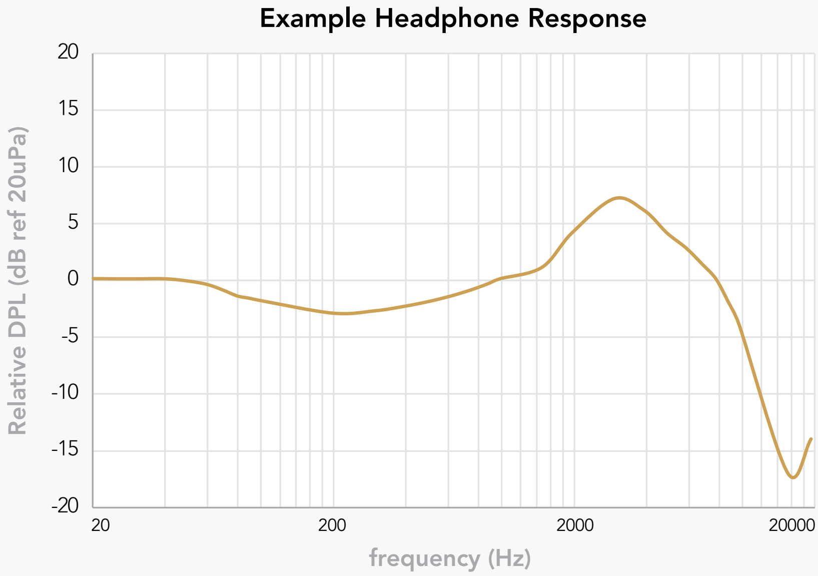
In the search for a better headphone transfer process, there are also areas where it may be best to leave well alone. As many who have worked with binaural recording systems will attest, the more closely you model the dummy head used for recording (for example adding realistic ears, modelling internal head densities and so on), the more likely you are to end up with a system which is exceptionally realistic for some people and works poorly for others. Trying to account for human head characteristics – ie HRTFs – *too specifically* in loudspeaker-to-headphone conversion may result in a similar effect so we have developed methods that work well for all individual HRTFs.

The question of reverberation remains a difficult one to address. We've noted that a great deal of the reverberant signal is *contained* in the difference between left and right channels, and particularly at lower frequencies. However, one aspect of cross-feeding between left and right channels is to reduce the difference signal, and although this can enhance localisation in the stereo sound-stage, it tends to reduce the perceived amount of reverberation, and thus the perception of apparent distance.

Another factor that has to be taken into consideration is the fact that modern headphones are designed to present a range of audio experiences with a wide range of source material and listening environments while being fed simply from the original stereo signal; different headphones differ in a number of characteristics, not least overall equalisation, where, for example, a manufacturer may arrange for their headphones to give a "warmer" sound by rolling off some HF or boosting at lower frequencies. A loudspeaker-to-headphone conversion system has to function effectively whatever the headphone manufacturer's design decisions, and whatever the listener's purchasing decisions based on what they like to hear.

Good Headphones already take into account the frequency component of Head Related Transfer Functions



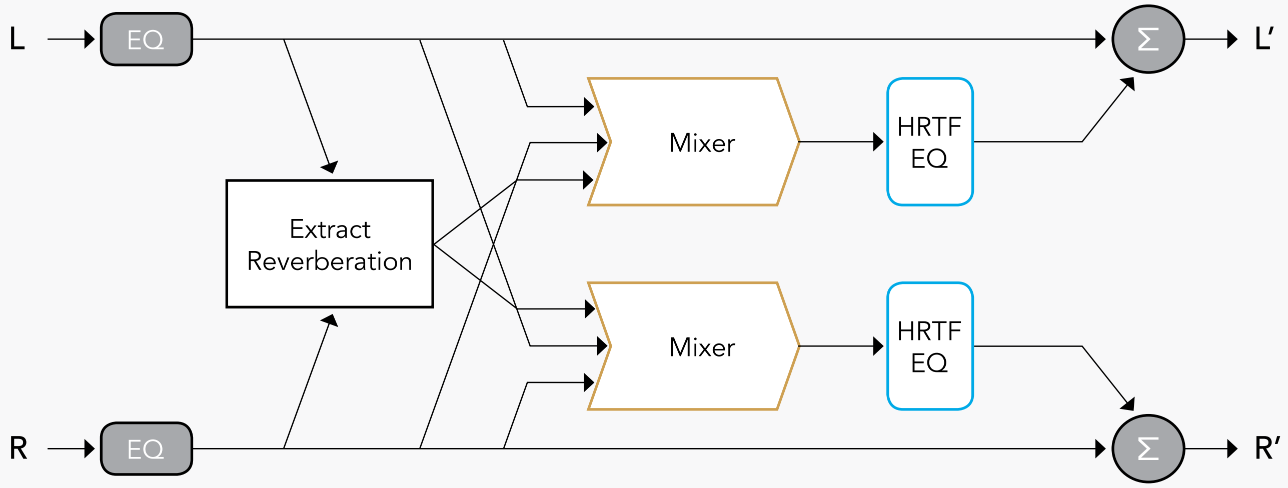


The *dCS* approach

When *dCS* determined to address the question of how to convert a stereo signal recorded with loudspeaker monitoring for the best result on headphones, all these factors were taken into account, including overall equalisation of the path, reverberation preservation, and optimising the cross-feed characteristics. All signal processing with *dCS* Expanse is carried out entirely in the digital domain.

Initially, the incoming stereo signal is passed through individual left and right processing stages that equalise the overall audio path and optimise the subjective height of the stereo image. The following stage is designed to preserve the reverberation content of the signal, because the main headphone cross-feed system encountered later on will tend to reduce the amount of reverberation. Remember, the reverberation is largely contained in the *difference* between the two channels: and the difference also controls the stereo width. So here, in a *dCS* innovation, the signal is effectively *widened* prior to the cross-feed system itself. This cross-feed stage, which represents another *dCS* innovation, is a hybrid system in which the cross-feed signal is delayed to simulate the effects of left signals being heard by the right ear and vice versa, the delay and frequency profile of the signal being based on the entirety of a large corpus of head-related transfer functions – nothing too specific that would favour some listeners and reduce the effect for others.

The stage also takes account of the level of correlation (similarity) between left and right signals. This results in some comb-filtering of fully-correlated and fully decorrelated signals – but this is an effect that we experience in the real world (and on loudspeakers).



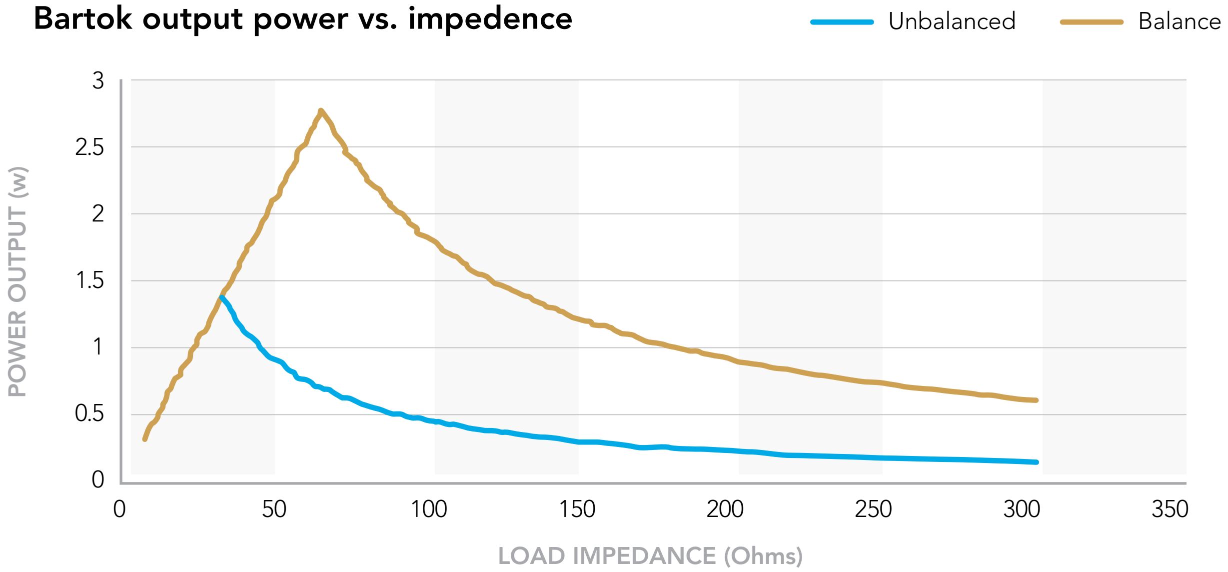
As a result of this processing, the stereo imaging is pushed out of the listener's head and instead is presented as a conventional stereo soundstage in front of the listener. The timbre of the sound and other psychoacoustic factors are adjusted, and the overall sense of space and reverberation is preserved.

Amplification for headphones

In the past, it has been common for *dCS* owners to use the existing line out stage on their unit to drive headphones, by making up a special cable to run from the outputs to the headphones. It actually does this job reasonably well – even though not designed for the purpose – with its low output impedance and low harmonic content and distortion, and its ability to drive a fair amount of current. When we designed the Bartók Headphone DAC, however, we decided to add a specially-designed headphone amplifier to drive a wider range of headphones (particularly low-impedance types, which wouldn't do so well on the existing line output), both balanced and unbalanced, directly, in addition to the standard line outputs. The design, which attempts to retain the essential character of the existing line output stage while adding the functionality described above, follows a dual-mono configuration, keeping the two channels completely separate. The amplifiers are powered by a combination of linear and switch-mode power supplies that are completely separate from those employed in other parts of the product, particularly necessary as quite high ground return currents are experienced with low-impedance headphones and these need to be kept well clear of the DAC and associated circuitry. A substantial toroidal transformer is employed in the linear PSU.

The signal is picked up from the DAC just before its line output stage and passes to a differential input stage with high common-mode rejection, minimising noise and crosstalk, and then to a gain-switching stage before passing to differential inputs on the power amplifiers. These are primarily, but not always, operated in Class A, and offer the ability to drive both high and low impedances optimally, from 33Ω upward.

Bartok output power vs. impedence



At 300Ω, the output in balanced configuration is 600mW, while the maximum output power in balanced operation is 2.8W with a 66Ω load. Power delivery is limited either by voltage (for high impedances) or current (for low impedances) as shown above.

When the output stage is running in Class A, the linearity is excellent, with harmonics 105dB or more below the fundamental at all practical load/level combinations. At around 150mW into 33Ω, the stage transitions from Class A to Class AB operation. This equates to an uncomfortably high volume level for all but the most inefficient of headphones. Beyond this, harmonics increase slightly in level to 95dB or more below fundamental at full voltage swing.

As noted earlier, one of the challenges of headphone amplifier design is the need to drive the wide range of possible loads adequately. High impedance headphones need to be driven with a large voltage swing (we provide 6.8V rms unbalanced / 13.6V rms balanced) whilst low impedance units require high current (we provide up to around 200mA rms). Output current is limited at this level for output stage protection purposes. In addition there is DC sensing protection circuitry, and the transformer temperature is monitored.

Beyond delivering adequate voltage and current, it is desirable to have a low output impedance. This enables the amplifier the control headphone drive units more accurately. The output impedance of our headphone amplifier in balanced mode is 100mΩ, much of which is accounted for by the unavoidable contact resistance of the connectors themselves.

The contact resistance of the unbalanced connector has another effect in that the common return connection necessarily and unavoidably creates a source of crosstalk – typically about 70dB with a 33Ω load. As headphones create an artificially-widened stereo image anyway, this isn't really a problem, but it's one area where the balanced connection is demonstrably superior owing to the separate return paths for each channel and the lower contact resistance of the connectors. We obtain a crosstalk figure in excess of -100dB at 20kHz via the balanced connection.

Two front-panel outputs are provided: a ¼in stereo jack socket for conventional common-ground headphones in which the ground return is commoned between the two sides of the headphone; and a 4-pin XLR for use with balanced headphones, where the two sides of the headphones are separately wired – a useful technique that allows double the voltage swing in the case of high-impedance headphones, with no ground return currents to contend with, and minimum crosstalk due to the lack of ground currents passing down a single connector.

Noise performance of the amplifier stage is insignificant in comparison to the (very low) noise of the preceding stages. Depending on impedance and efficiency, different headphones require different volume control settings. The low voltage drive required by low impedance headphones could necessitate large amounts of attenuation being applied at the (digital) volume control which would result in the output being unnecessarily noisy. As a result, stages of analogue attenuation are provided.

There is 10 dB of attenuation available in the DAC output stage, and another 20dB in the amplifier stage itself, making it possible to provide up to 30dB of attenuation so that the volume control can be used in the central part of its range whatever the characteristics of the headphones in use. A menu setting allows four level configurations: 0, -10, -20 and -30dB.

info@dcsltd.co.uk | @dCSonlythemusic | www.dcsltd.co.uk

Manufactured in the UK by Data Conversion Systems Ltd
Unit 1, Buckingway Business Park, Anderson Road, Swavesey, Cambridge, CB24 4AE