Speaker Testing / Evaluation

By Paul Mac – Freelance Technical Writer – Oct 2015

Playing Fair

It this goes some way to explaining why subjective evaluation - critical listening - is a valuable tool in choosing a monitoring loudspeaker. However, it is important that any evaluations and comparisons should be fair and consistent. Variations between the listening conditions of systems under consideration should be minimized; language and criteria should be consistent; and features such as physical appearance and brand should, as far as possible, be discounted from judgements on the quality of sound.

Below we will look at several aspects of subjective loudspeaker testing and at the art of 'critical listening'. We will look at setting up loudspeaker tests - both in an ideal situation and the more usual and less-than-ideal situations, plus the issue of blind listening. We will also look at the AES20 standard on the subjective evaluation of loudspeakers, which outlines a scoring system, plus a set of criteria and terminology for evaluation - an invaluable guide to focusing your critical listening.

Good, critical, subjective evaluation of loudspeakers is a kind of measurement for the real world. It's important that you go about it in a way that doesn't skew the result to the point where you make a bad decision. That is, listening should be part of choosing a loudspeaker, but it should be critical, consistent, and fair to be valid.

Blind Testing

Blind listening tests are exactly what they sound like. The idea is that the listener cannot see which loudspeaker he or she is listening to and, therefore, cannot make judgements based on the look, brand, or previous performances of a loudspeaker. This takes away any preconceptions a listener might have of brands, and judgements will be more valid.

Normally, blind testing is performed using a screen between the listener and the speakers - visually opaque but acoustically transparent screens can be constructed from material such as speaker grille fabric.

A development of the idea of blind testing is 'double blind' testing, which is the preferred testing method of AES20. Double blind testing requires both the listener and anybody else involved in the test (test administrator / switcher) to be unaware of the speaker selection. Thus, a third party is required to set the test up.

Don't get confused with ABX testing, where sources or equipment are chosen randomly and a log is kept of test results and then analyzed to ensure the judgements are 'statistically valid' To perform valid ABX testing with loudspeakers the listener would have to be unaware not only of what brand the loudspeaker was, but which of the loudspeakers (A, B, C etc.) in the test were being selected. This would only be possible if every speaker in the test could be put in the same position, otherwise, the listener would be aware they had moved to a specific position to perform the test. Rotating multi-speaker stands have been suggested as a solution to this...

One other way of getting, at least, close to the ABX ideal would be to repeat the complete test procedure for particular criteria several times, but each time have a third party move the loudspeakers under test to a new, random position. Inconsistency in scoring for particular criteria would 'null' that test.

Some people do frown on blind testing for various reasons, but these objections normally come from the HiFi world, which relies on a certain amount of brand/golden ears myth and mystery to sustain aspects of the market's business. That is, it is often true that more expensive equipment sounds better - but that's not necessarily because it sounds better! The increasing popularity of ABX testing of codecs and audio formats, however, has focused attention on the idea that maybe there is more bias attributable to the listener than many would like to believe...

Practical Approach

While a perfect room, a flawless double blind testing system, and lots of time would make an ideal scenario for evaluating loudspeakers, unfortunately, this situation is rarely available or possible. The facility for well-managed blind testing at retail outlets, for instance - or even good positioning of loudspeakers and desirable acoustics at the same outlets - is rare. Also, studios are often limited in both space and appropriate facilities for mounting several loudspeakers and managing these kinds of tests.

This means that testing procedures and conditions might have to be compromised. If blind listening is paramount, this might involve blindfolding a listener rather than providing a screen. If blind listening is not possible or desirable, a listener should, at the very least, use critical listening techniques, familiar material, and level/position matching of products to be evaluated.

Speaker & Room

A 'bad' room can seriously compromise a loudspeaker's performance and thus render any judgement invalid. Here are some headline precautions that should help when setting up a loudspeaker test...

Balanced Treatment
It is often wrongly assumed that an abundance of absorption is a good 'fix-it' for a bad room, but this can often make the situation worse. Absorptive and diffusive treatments will often only work over specific bandwidths and in specific room positions. Too much high-frequency absorption will dull a room - having the same effect has a low pass filter. Some treatment manufacturers offer design services and advice, and it's worth taking advantage of these when you can.

Don't Forget Diffusion
Absorption is often the first port of call for acoustic treatment, but diffusion (reflection in random directions) is just as important – it can help prevent coherent reflections by 'mixing up' the reverberant field.

Reverb Time
The 'RT60' (time taken for an audio impulse to decay by 60dB) recommended by AES20 is just under half a second, relating to a "normally furnished domestic listening room." However, many control rooms are designed to a shorter RT60 than this. RT60 is normally chosen for the type or function of the listening

environment the speakers are destined for. As a rule of thumb, make sure that the room you're testing in is 'comfortable' but not too dead.

Room Modes
Room modes are natural notches and peaks at particular positions and frequencies and are a function of wavelength and room dimensions/geometry. You should always try to minimize the effect of these modes, but it is also wise to know where these modes are and avoid, for example, positioning speakers and/or listeners where they will be unduly affected.

Speaker Positioning
A good rule of thumb is to position speakers at least a meter away from reflective surfaces such as walls, desks, and so on. However, be aware that some manufacturers might recommend something different for their specific models.

Avoid Symmetry
If you arrange speakers symmetrically within a symmetrical room, you will end of with 'coherent interference', which will cause notch filtering and other undesirable effects. Avoid 'coherence' between the position of the speakers and the geometry of the room.

The Triangle Position
For optimal stereo positioning, form an equilateral triangle between left and right speakers, and the listening position.

Heavy Speaker Stands
Generally speaking, a good speaker stand is a heavy, solid speaker stand. The strategy here is to use that weight to lower the resonant frequency of the stand, and the floor it's standing on, so there are no 'third-party' resonances in the audio spectrum. 'Decoupling' products are available, though be aware that the effectiveness/quality of these devices on a variety of surfaces is variable.

The Right Height
The height at which speakers are mounted is important. The center of the speaker (normally between the woofer and the tweeter in two-way systems) should be in line or slightly above the ears of a listener in the on-axis listening position.

There's a big difference between listening and 'critical listening. You need to ensure the conditions are right, and that your ears are focused on the right thing...

All About The Room

A loudspeaker can only be as good as the room it's in. Ensure best possible room performance and speaker positioning. Have a look at the 'Speaker & Room' section for an introduction to room considerations. To spot any big problems, you could start by checking some test tones (sweeps and noise) to listen for obvious anomalies such as big frequency losses, tonal imbalance, and phase problems.

Focus
Focus on one aspect of loudspeaker performance at a time, and repeat, in the same way a camera lens can focus on a single object and adjust 'depth-of-field' to exclude or blur other things. AES20 is an

excellent base for identifying the aspects to focus on (see the AES20 introduction). You should also repeat sections of the listening material that aid this focus - it will help confirm your opinion.

Choose Your Subject

Use appropriate listening material. A good stereo recording is required for imaging tests; full-frequency material such as pipe organ + orchestra or high intensity film sequences are good for spectral tests. In some cases, it might be appropriate to use test tones such as sweeps or pink noise to illuminate aspects of a loudspeaker's performance such as spectral balance.

Know Your Subject

Use familiar listening material. Unfamiliar material will become familiar over the period of testing and so opinion and/or scores will be inconsistent.

Keep levels consistent!

Louder speakers are consistently judged 'better'. All speakers under the test should be checked for equal output level at the listening position before evaluation. AES20 recommends using pink noise and an A-weighted SPL measurement, but any valid loudness measurement is preferable to none. Using the pink noise will give you a steady, wide bandwidth level so you can reliably compare different speakers. 85dB A-Weighted SPL is a popular reference level for normal listening. You should also set low and high levels for the purposes of dynamics evaluation.

Please bear in mind that the European directive covering noise exposure specifies an SPL (Sound Pressure Level) of no more the 85dB A (A-weighted measurement) over a period of eight hours. Short louder bursts are allowable provided that the 85dB average over an eight-hour period is maintained. The upper limit for exposure to an impulse sound is 137dB.

Measured Response

When comparing several pairs of speakers, you should be able to ensure that the distance and angle to the listener is consistent. One way of doing this is to line speakers up in staggered pairs and to move the listener to the correct position. This is important for both time and level consistency.

Single Channel Tests

For some tests, it may be advantageous to listen to one speaker only for further focus on a single criterion. Exclusively spectral and dynamic tests can be done with a single channel. Tests with a time-based/multichannel requirement include those with imaging and environmental focus. Please be aware that 'mono'ing' a stereo feed for this purpose may not be a good idea as mono phase issues in the recording might be interpreted as spectral deficiencies.

Switch Smooth

Make switching between speakers on test as smooth, easy, and fast as possible. Noisy or delayed switchovers can adversely affect judgement. Use mute groups/automation or snapshots on a mixing console, or a dedicated switching box to make it easier.

Real-World Conditions

Make the listening conditions appropriate for your application. That is, try to conduct listening tests at the same levels and in the same kind of environment.

AES20, listening criteria, and choosing your listening material

AES20, the Audio Engineering Society (AES) standard for the subjective evaluation of loudspeakers, was first published in 1996 yet, anecdotally, is relatively unsung and probably under-used.

It describes a complete set-up, blind listening procedure, scoring system, and set of criteria for judging and comparing loudspeakers. It splits listening criteria up into seven different categories and a total of 29 different tests, each of which helps the listener focus on a very particular aspect of loudspeaker performance. For example, the category Sound-stage imaging includes Stage Width and Image Depth Localization amongst its nine criteria.

In many practical situations, it might not be convenient or desirable to follow AES20 to its full extent, but even if you use it just to focus your own critical listening on important aspects of loudspeaker performance it is a valuable tool. For example, you could try focusing your listening by section rather than by sub-section (although 'C.1.5.3 Transient Impact Or Punch' may well deserve its own section).

That said, larger-scale procurements, such as tenders for multiple installations, probably do work on a scale that makes full implementation of AES20 both practical and desirable where procedural credibility and accountability are especially important; and possibly where operators are having trouble coming to a consensus!

Here we have included a description of the main seven AES20 categories, specific sub-category examples, plus tips on listening for these and what listening material might be best suited to the job.

1) Spectral Uniformity
This focusses on tonal quality and balance, plus frequency range. You will probably benefit from doing much of your spectral testing with just a single speaker. If the spectral balance or any of the AES20 criteria change dramatically when you use two speakers, then this normally points to a problem with one of the speakers, a phase issue, or a speaker placement issue.

The Coloration criteria is an overall quality to the sound, whereas the other parts of this category are more focused.

The idea for Octave Balance is interesting and basically asks you to listen for those things that are normally associated with an octave-wide band, such as sibilance and intelligibility in voices, "boomy bass", and so on. Remember that a perceived increase in voice intelligibility, for example, can be because of an undesirable exaggeration in the mid-range. Useful program material for Octave Balance might be a small jazz ensemble with clean vocals so you can easily focus on the octave bands. Also, a simple spoken voice extract could help really focus on that very important range.

Bass/treble balance and frequency extension can be lit up with quite dense spectral material such as highly 'produced' pop or dance music, or action scenes and extreme weather scenes from film soundtracks.

2) Sound Stage Imaging
This focusses on accurate reproduction of the 'stage', which is the overall stereo image and its reproduction of width, depth, and height perception. AES20 defines it as the area containing all of the performers or sound sources... The location of the entire stage with respect to the listener. At first glance

it could be close to (3) Localization, but that is more specifically about absolute position of sources than the stage. Sound Stage Imaging is the overview - the image in its entirety. The Localization section focuses on absolute placement of sources.

As well as the basic front-back, left-right, up-down, and width criteria (as well as the overall 'stage' criteria), Sound Stage Imaging has the 'Image' section. Helpfully, this not only asks the listener to concentrate on position and extension, but on width of sources within the stage, such as the wide image of a orchestral string section compared to a narrow image of a soloist.

Obviously, these tests require a stereo set-up. You could, however, apply the same or similar criteria to a surround set-up if you felt that was necessary for you, but placement for multi-subject set-ups would be difficult.

For listening material, both sound stage imaging and localization require good stereo recordings of well-defined spaces. An simple recording of an opera, for example, would give you a wide stage to listen to, a slightly forward/low orchestra position, movement of the actors and ensemble to follow, and a room to limit the stage width and give clarity to your judgement. Other than that, any good stereo field recording, even stereo sound FX recordings intended for atmospheres in natural history and drama can be a big help here.

It is possible to do the imaging and localization tests without 'natural' stereo recordings, as long as the material has a well-defined stereo image and enough variation in placement to make judgements valid. Left/right width should be fine, though depth and height may not be particularly apparent on anything other than a natural stereo recording.

3) Localization
This section concentrates on the position and clarity of discrete sources in the stereo field. There are basic localization criteria (left-center-right and depth) as well as more relative ideas for listening focus. These include Image Separation, which identifies, for example, space between voices in an ensemble; and Open and Transparent, which answers the question of whether sources appear truly 'phantom' or whether they are tied to the speaker positions.

Listening material in this category can be the same as for Sound Stage Imaging, but might also include more discrete recordings of, say, single performers moving around a stage, or small ensembles where individual instruments and their positions should be easier to focus on.

4) Ambience Reproduction
This focuses on the quality of the reproduction of ambience, reverb, and environment. It's a judgement of how well the speakers reproduce the space in which a recording was done or the space that the production recreates. For the most part, this requires the listener to focus not on the actual sound sources themselves, but rather on the sound of the environment - the early reflections, the reverberant tails, any low-end architectural rumble, and so on. A useful question to ask yourself here is 'can I visualize a space just by hearing someone play or sing in it?'

By definition, the listening material here should be a variety of high-quality recordings in a variety of spaces. The opera from the other two imaging categories would be useful, a small acoustics ensemble in a small room, a choral performance in a church, and so on – all these will help focus on ambient qualities. Also, film and drama situations can be enlightening here. The only thing to be wary of is

artificial reverbs. Some are very natural, but some are noisy, grainy, metallic, abrupt, or unbalanced - often intentionally. These can mask ambience very effectively and so you should be exceptionally familiar with such material to use it for this kind of testing.

The Direct–Reverb Rendition criteria is particularly interesting and good for focusing your listening. It asks you to listen for the ratio of direct or reverberant sound being generated by a source. Close sources should have higher direct content than sources further away, with a greater distinction between their direct and reverberant sound.

For the most part, the Ambience Reproduction tests would be done with a multi-channel set-up as this is required for reproduction of an environment. However, there might be instances where you will find examining the quality of a reverberant sound, for example, easier with a check on a single speaker.

5) Dynamics and Distortion
This focuses on high and low volume level performance, transients, and more. There are eight separate criteria in the Dynamics and Distortion section, most of which deal with quality at loud listening levels. However, you approach this, it is important to protect your hearing. The European limits on noise exposure are covered briefly in the 'Critical Listening Tips' section, but in any case, it is good practice to listen to high levels for as short a time as possible.

Basic criteria here include Distortion, Maximum Loudness, and Transient Impact or Punch. AES20 also asks the listener to check for compression artefacts at high levels, as well as maintenance of timbal quality. That is, if the audio becomes 'strained' or hardened at high levels then there this would be marked down.

One aspect that takes a little practice to identify and focus on is transient performance - the reproduction of fast attacks and rapidly changing levels. It's worth honing this skill as transient performance, in conjunction with good spectral performance, relates directly to a number of technical loudspeaker aspects. Material that might make this easier includes high quality live drum loop recordings. There are any number of these available on sample CDs, and the transient aspect, combined with the very revealing character of cymbals and snare drums, is ideal.

For checking that material isn't compressed (or at least doesn't show signs of compression) at high levels, it's a good idea to use listening material with minimal compression in the first place. Raw acoustic recordings are good here as ensembles can be too dense to hear subtle artefacts.

The AES20 Pianissimo Clarity aspect requires you to listen to material loud then soft, and judge how well timbral quality is maintained between the two.

Dynamics and distortion tests can be done on single speakers.

6) Listening Fatigue
This focuses on loudspeaker performance and its effect on the listener over longer durations. It's reasonably self-explanatory. The best advice here for listening material is to try and use a good representation of what you'd be listening to most in daily work. Whether it's dance, rock, dialogue, orchestral, jazz or anything else, you'll need to be confident that you can work on a loudspeaker for a number of hours without discomfort.

Obviously, this is a difficult aspect to judge on multiple systems as that would take a long time and valid comparisons are virtually impossible. For practical purposes, it's probably a good idea to whittle the choice down to two or even one before embarking on a fatigue test.

7) Robustness
Focuses on tolerance of listener position, or off-axis performance. This is best tested with stereo material from the Sound Stage or Localization sections. It involves moving up to half a meter from the ideal listening position to check that the imaging and ambience performance are relatively well maintained. It includes a seated/standing test and a head rotation test.

Highlights – Notes

Play fair!

1 - When evaluating multiple speakers make sure the SPL is matched between the speakers using Pink Noise.

2 – Make sure the same content is played on all the speakers being evaluated.

3 – Play multiple forms / sources of music that you are very familiar with and that will most likely be used on the speaker system.

4 – Don't listen with your eyes. If possible, do a blind test using a scrim or acoustically transparent cover.

5 – Move around the room. Listen to the speakers in different areas of the room.

 6 – Actively listen to the differences in the speakers. (highs, mids, lows)

7 – Take notes during the test.