TECH Document



March 2016 TN11

Objective listening test of audio products

- a valuable tool for product development and consumer information

Torben Holm Pedersen



DELTA Venlighedsvej 4 2970 Hørsholm Denmark Tel. +45 72 19 40 00 delta@delta.dk delta.dk

Product	Technical data	Sound profile			
	Argon 6340 (DKK 399 per unit) Frequency response (-3 dB): 80–20,000 Hz Sensitivity: 84 dB (1 W/1 m) W x H x D: 14.8 x 23.9 x 16.5 cm Volume: Approx. 6 I Weight: 2.3 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	DALI ZENSOR 1 (DKK 999 per unit) Frequency response (-3 dB): 53–26,500 Hz Sensitivity: 86.5 dB H x W x D: 27.4 x 16.2 x 22.0 cm Volume: Approx. 10 I Weight: 4.2 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	Bowers & Wilkins 686 S2 (DKK 1,799 per unit) Frequency response (-3 dB): 52–22,000 Hz Sensitivity: 85 dB W x H x D: 16.0 x 31.5 x 22.9 cm Volume: Approx. 11 I Weight: 4.9 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	Scandyna MiniPod MK3 (DKK 1,999 per unit) Frequency response (-3 dB): 55–22,000 Hz Sensitivity: 91 dB W x H x D: 21 x 34 (44 on legs), x 20 cm Volume: Approx. 14 I Weight: 2.3 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	Bowers & Wilkins 685 S2 (DKK 2,399 per unit) Frequency response (-3 dB): 52–22,000 Hz Sensitivity: 87 dB W x H x D: 19.0 x 34.5 x 32.4 cm Volume: Approx. 21 I Weight: 6.8 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	DALI OPTICON 2 (DKK 2,999 per unit) Frequency response (-3 dB): 59–27,000 Hz Sensitivity: 87 dB W x H x D: 19.5 x 35.1 x 29.7 cm Volume: Approx. 20 I Weight: 7.8 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	Bowers & Wilkins CM1 S2 (DKK 3,299 per unit) Frequency response (-3 dB): 50–28,000 Hz Sensitivity: 84 dB W x H x D: 16.5 x 28.0 x 27.6 cm Volume: Approx. 13 I Weight: 6.7 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			
	DALI MENUET (DKK 3,699 per unit) Frequency response (-3 dB): 59–25,000 Hz Sensitivity: 86 dB H x W x D: 25.0 x 15.0 x 23.0 cm Volume: Approx. 9 I Weight: 4.1 kg	Natural Brilliance Bass depth Dark-Bright 3 4 5 6 7 8 9 10 11 12			

Contents

Summary	
Background	5
Ingredients in an objective listening test	6
Listening room	6
Measurement set-up	7
Anchor systems	8
Listeners	9
Attributes	9
Programme material	9
Level calibration	10
Trial design and data collection	10
Statistical analysis	11
Sensory product information	12
Example of a listening test	
Test setup	13
Listening testing	14
Loudspeakers	
Results	16
Comparisons	
Interesting correlation	21
Models	23
Appendix 1- Sound Wheel	
Appendix 2 – Alternative for sound profiles	27

\mathbf{E}

The report can be downloaded from DELTA at senselab.madebydelta.com/about/publications

Summary

This TECH Document describes a listening test, and describes what is required to achieve an objective characteristic of the perceived sound.



The document describes what is required to carry out reliable and reproducible listening tests. There are for example requirements to the listening room and loudspeaker positioning, and a panel of trained listeners is also needed. The assessments should be carried out as a blind test with the products in random order, and there must be well-defined words with which to describe the products' characteristics. Different programme material should be used (e.g. music) and all of the systems must be set to the same loudness. Such a test will provide an objective characteristic of the sound from the systems, which can be used e.g. in product development or

as consumer information supplementing technical data. The use of these principles has been demon-

strated with a test of 8 compact loudspeakers, where each test result is an average value of a total of 40 assessments. The results from the listening test are provided on page two. The listening test provided useful and supplementary information in relation to the technical data. For example, it showed that there

Price, DKK

was a clear difference when comparing perceived bass depth, even though the technical data for the lower limiting frequency of each of the systems was more or less the same. This TECH Document also shows how the results can be presented, for example as a sound profile with the loudspeakers' sound characteristics, which can be used as sensory product information.

A comparison of the loudspeakers' prices and DELTA's expert listeners' preferences shows that there is a general correlation between price and preference. However, it should be noted that preference is subjective in contrast to the test's other attributes.



Author: Torben Holm Pedersen Title: Objective listening test of audio products Company: DELTA, SenseLab

Background

Listening tests can be used in many contexts. For example, to examine whether data compression of sound files results in quality loss, whether copy protection embedded in a sound file can be heard, whether noise suppression in a mobile phone compromises speech quality or speech intelligibility, or whether a particular type of hearing device is best with background noise, or determining to what degree different sounds from wind turbines cause discomfort.

With regard to listening tests on audio systems, there are also different options. For example, it may be a personal listening test of different equipment (e.g. loudspeakers), which you consider purchasing, it may be reviews in newspapers and magazines, where one or two people have listened to the audio products, it may be a development team listening to new products, and it may finally be more systematic testing, carried out as blind tests with trained listeners under controlled conditions. In this TECH Document, we will focus on the latter, because such tests can provide objective and reproducible results with a well-defined precision.

In TECH Document no. 7 'Perceptual characteristics of audio'¹, we looked at the relationships between physical measurements (e.g. measurements of sound pressure, frequency characteristics, etc.) and listening tests which can be objective (perceptual measurements carried out by trained listeners) or subjective (preference tests, e.g. with consumers). The document also provided examples of why technical data and product reviews do not produce a complete picture of the perceived sound. Finally, the document explained how to choose and test words (attributes), which can be used to provide a well-defined characterisation of the products which are being assessed. It was shown that these attributes could be organised into a sound wheel, which is an important element when carrying out objective listening tests, and which can be used for communicating product characteristics, for product development, or for communication between the manufacturer, sales unit, and consumers. (The sound wheel is depicted in Appendix 1).

In this TECH Document, we will show how objective and comparable listening tests are carried out, and we will provide an example of the results from such a test, which was executed in collaboration with the retail chain Hi-Fi Klubben, who provided loudspeakers for the test.

¹ http://senselab.madebydelta.com/about/publications/

Ingredients in an objective listening test

An objective listening test is a test that does not relate to personal preferences. The results are the average values of assessments made by a number of listeners (usually more than 10), using well-defined attributes. For a listening test to be reproducible and for it to have well-defined results, a number of ingredients must be included:

- A listening room, with a neutral effect on the sound and a low background noise
- A measurement setup and equipment for direct comparison in a blind test
- Anchor systems (fixed references), so that the results are comparable with other tests
- Selected and trained listeners
- Well-defined attributes
- Critical programme material, played with the same loudness from all of the systems
- Randomised test design, i.e. the systems are presented in random order
- Efficient data collection
- Statistical analysis
- Presentation of results

Listening room

In the case of listening tests of loudspeakers (passive, active, streaming, Bluetooth ...), a neutral listening room must be used. It is the loudspeakers that should be assessed, not the influence of the room. The reverberation time must be short and lie within specified limits. In particular, the reverberation time must be the same at all frequencies (however a weak increase is permitted at low frequencies). Next, the background noise must be low, so that unintended sound does not interfere with the listening test. Ideally, the standards EBU Tech 3276 and ITU-R BS.1116-3 for listening tests of multichannel loudspeaker systems should be complied with. DELTA SenseLab's listening room, see Figure 1, complies with these. Thus, tests of mono, stereo, and surround systems (5.1 to 22.2) can be carried out there.



Figure 1

A listener assesses loudspeakers during a blind test in the DELTA listening room. The result of the test is an average of the total number of listener assessments. The screen with the user interface is accustically transparent, and the

acoustically transparent, and the loudspeakers are hidden behind. The bar in the ceiling can be used to suspend loudspeakers for multichannel surround systems.



If the test is to be carried out using headphones, it is sufficient that the background noise is less than required. Typically, in these situations a smaller listening booth is used, see Figure 2.

Figure 2 One of SenseLab's two listening booths for tests using earphones. The background noise inside the booth is very low (< NR 15).)

Measurement setup

When the listening test is executed as a blind test, the listener does not know what he is going to hear and cannot see the loudspeakers. The loudspeakers are hidden behind an acoustically transparent curtain or screen. In DELTA's listening room, the screen is used as a user interface for the test software, so no undesirable sound reflections or similar occur from a computer screen in front of the listener. See Figure 1.

There is a particular challenge when executing listening tests with loudspeakers: If the loudspeakers are positioned in different locations in the room, the sound will be coloured differently, depending on their positioning. This can be solved by rearranging the loudspeakers during the test, so that all of loudspeakers are listened to from the same positions. It is a difficult and slow process.

To execute the test faster and more efficiently, DELTA has instead chosen to build some rapid speaker spinners, which ensure that the loudspeakers that are listened to, are always in the same position. The spinners, which can be seen in Figure 3 and Figure 4, are almost silent and take around one second to move the speaker into position. The short changeover time and lack of interfering noise, is very important because people's acoustic memory is very short.



To see a demonstration of the speaker spinners in actions, visit: senselab.madebydelta.com/services/listening-tests.

Anchor systems

The results of the listening test are not unaffected by the setting in which they are carried out. If you test mobile telephones' built-in loudspeakers for listening to music together with a small average bookshelf speaker, the latter will be assessed as having a high bass strength. If the same bookshelf speaker is assessed together with Hi-Fi floor speakers, it will be assessed as having a lower bass strength. In order to compare the results from different listening tests, it is necessary that there are some recurring systems, which can be used as fixed reference points – so-called anchor systems.



Anchor systems are chosen, which for the applicable product category lies respectively high and low for most attributes. The anchor systems in each test are assessed on an equal basis with the other systems and if the assessment of them has changed since the last test, the results can be corrected according to this change, so that the results from the two tests are comparable.

Figure 3

The listening room with two speaker spinners: the loudspeakers are always in the same position in the room when they are listened to.

Figure 4

The speaker spinner in front. The three anchor loudspeakers which are used to create comparison between tests can be seen in the background. The frequency characteristics are shown on the right, measured in the listening position of the upper (red) and lower (blue) anchor.

Listeners

A part from preference tests, where the listeners should be chosen from the relevant market segment, trained listeners with good ears should be used. DELTA's listener panel consists of 30 selected² listeners, whose hearing is tested to show that it is normal, and who then undergo thorough training. In addition to the actual "ear training", the training ensures that the listeners understand the attributes and use response scales in a uniform manner. The expert listeners are specifically trained in the attributes, which are relevant for testing of audio systems. Many of the expert listeners are also employed or have an interest in the audio sector (e.g. musicians, composers, Hi-Fi enthusiasts, Hi-Fi dealers, students studying acoustics at the Technical University of Denmark, sound designers, and sound technicians). Their performance is monitored on an ongoing basis, to check their ability to hear the difference between systems and to be able repeat their own assessments in blind tests.

Attributes

The systems (e.g. loudspeakers) are assessed on a number of attributes, which together constitute a sound profile that characterises the systems. There are a number of requirements for good attributes, which are described in detail in TECH Document no. 7.³ The attributes are required to be clear and well defined, and must be interpreted in the same way by all the listeners. In general, when carrying out listening tests of audio systems, attributes from the sound wheel are used. See Appendix 1. Normally 6-10 attributes are sufficient to characterise the systems. The attributes may be selected based on a pilot test, e.g. the attributes that best characterise the differences between the systems. Alternatively, the test leader or the customer may choose a set of attributes for the listening test.

Programme material

To throw light on as many properties of the systems as possible, several different programme materials must be used (pieces of music, speech, sound effects whatever is relevant to the test in question).

In general, the programme material is selected from high-quality recordings with varying dynamics and frequency content. Many modern recordings compress amplitudes (see Figure 5). Such recordings are not suitable to provide a varied picture of the tested systems. If required, the material can be measured with regard to compression and frequency content, to ensure that it is not the programme material which sets the limits for the assessment. Usually, uniform passages from pieces of music are used, which have a duration of about twenty seconds.

² Cf. Søren Vase Legarth & Nick Zacharov: Assessor selection process for multisensory Applications. Audio Engineering Society 2009, Munich, Germany.

³ The development and validation of the attributes is also covered in: Torben Holm Pedersen & Nick Zacharov: The development of a Sound Wheel for Reproduced Sound. Audio Engineering Society 2015, Warsaw, Poland.



Figure 5

An example of amplitude (Y-axis) of a piece of music as a function of time (X-axis). Many recordings are, like this, with compressed amplitude, thus they are not very suitable for testing a system's dynamic properties.

Level calibration

The perceived product properties are not independent of sound level. If for example, there are two relatively uniform loudspeakers included in a test, experience has shown that the loudest playing loudspeaker will be preferred. Thus, it is important, that the sound level of all of the systems is kept the same during the test. Therefore, after being preconditioned ("burned in"), all of the systems are adjusted before the test so that they have the same level. Normally, this calibration is done by adjusting the systems to the same A-weighted sound pressure level in the listening position by playing a special noise signal (pink noise).

Test design and data collection

Unless you directly want to study the influence of product design and brand value, the listening test should be carried out as a randomised blind test. That means, the test is carried out and designed in a manner that ensures the listener does not know which product is being listened to. Therefore, products, attributes, and programme material are presented in a new random order for each listener. The importance of this can be illustrated by the following: even the most experienced listener who after a test is presented with the result, can be surprised as to how difficult it is to hear the difference, even with products they were certain that they could hear a clear difference between.

To enable the randomised test to be carried out quickly and efficiently with short (< approx. 1 second) intervals between the systems that are being assessed, it is necessary to automate the system for presenting the stimuli (i.e. combinations of systems and programme material) and for registering the listener's responses. DELTA has developed several systems for this purpose. One of the systems has been specially developed for listening tests on loudspeakers, so that it also controls the speaker spinners during the listening test. In practice, this means that when the setup is ready and the sound level calibrated, the test can be executed fully automatically, at a tempo that suits the listener. The most precise assessment is achieved when the systems are compared attribute by attribute. An example of the user interface for such a test is shown in TECH Document no. 7. Another system, SenseLabOnline, makes it possible to execute listening tests over the internet, and is used for example, to execute testing with noise suppression in mobile phones in several languages.

Statistical analysis

The statistical analysis is used first to monitor the data quality – e.g. by checking whether the listeners can differentiate between different stimuli, and whether they can reproduce their own assessments in a blind test.⁴ Even with trained listeners under optimal listening conditions, there are differences in the individual assessments.

Secondly, the primary aim with the statistics is often to calculate the mean values and the associated precisions. For example, see Figure 6. Here, 95 % confidence intervals provide an indication of which systems (in this case: loudspeakers) have a significant difference. A general rule of thumb can be used; if the confidence intervals overlap for two products, it is not certain that there is a significant difference in relation to the attribute in question. However, the statistical analysis can also supply a formal test of which products really are different and which are not, with the precision in the test taken into consideration.



If you want to go into more detail, a statistical variance analysis (ANOVA analysis) can show which differences in the results can be ascribed to a difference between the systems (e.g. different loud-speakers), differences which are due to the programme material or differences due to the listeners, and any interactions between these factors. An example of the latter can be seen in Figure 7, where the assessment of DALI Zensor 1 depends in particular on which piece of music is played.

Figure 6

An example of a result from a listening test of loudspeakers using 10 listeners with respect to the attribute Bass Depth. The blue dots state the average value of all of the listeners' assessments of the two different pieces of music played on the loudspeakers. The vertical intervals state 95 % confidence intervals.

⁴ DELTA has developed a special tool for this, called eGauge: Lorho, G., Le Ray, G., & Zacharov, N., Measuring experience and expertise in audio, The 10th Pangborn sensory science symposium, Rio de Janeiro, Brazil, 2013. However, tools from the programme PanelCheck are also used, http://www.panelcheck.com



Figure 7

The same results as shown in Figure 6 (Bass Depth), but here divided between the two music pieces, JW and ZC which are listened to.

Sensory product information

There are many options for presenting the results from sensory tests. The examples in the preceding section are technical in nature and give precise information about the test's outcome. They are aimed primarily at professionals who work with product development, marketing, etc.

But the results can also be utilised to provide the consumers with information about the products' perceived characteristics, which are not achieved through the technical product information. Two examples showing the characteristics of beer are shown below.



Similarly, the sound profiles from the listening test can be illustrated, and some options are shown in Figure 9. To make things manageable, only four attributes are shown.

Figure 8 Two examples of sensory scales providing the characteristics of beer. The scales in the right picture are: Sweet – Bitter and Light - Strong



Figure 9 Examples of possible ways to present a loudspeaker's sound profile. The example is for DALI Opticon 2.

On page two in this TECH Document there is an example of how sound profiles from the listening test can be used together with the technical data. Other ways to illustrate the sound profile are shown in Appendix 2, and ideally this should give an impression of which types are best suited as overviews.

Example of a listening test

The technical data states the loudspeakers' performance and what they can withstand. If you want to show how the loudspeakers sound, you need to use the results of a listening test. The listening test described in this section covers eight stereo sets of compact loudspeakers sold by Hi-Fi Klubben within the price range DKK 399–3,699 per item. Six objective attributes were selected for this listening test, and each result represents the average of 10 trained listeners' assessments of the loudspeakers' characteristics.

Test setup

The test was carried out in DELTA's listening room that has a neutral reverberation, which means it does not colour the sound from the loudspeakers. See Figure 1. The loudspeakers were set on a speaker spinner, which ensured that the loudspeakers always were in the same position when listened to in the room. See Figure 3. If the loudspeakers were placed in different fixed positions, the spatial-acoustic phenomenon could colour the sounds differently, depending on where the different positions of the loudspeakers were.

During the test, an assessment scale on the acoustically transparent screen is shown (Figure 10). This hides the loudspeakers, so the listener is unable see what is being listened to.

Figure 10

An example of the user interface during the listening test. The listener can choose which system shall be listened to, using the buttons A-G. The loudspeakers that are "hidden" behind the buttons vary during the test. The assessment is given by moving the sliders, in this case between "A Little" and "A lot" Bass Depth.



The loudspeakers were driven by NAD C356 BEE amplifiers. Before the test was carried out, the sound pressure level for each loudspeaker was adjusted to 70 dB(A), measured in the listening position by playing pink noise. The loudspeakers were preconditioned with this signal for an hour.

The following two pieces of music was listened to:

- Jennifer Warnes (JW): "Bird on a wire" from the CD 'Famous Blue Raincoat'

- Zhao Cong (ZC): Moonlight on Spring River from DALI CD 3.

The Listening test

The assessments were carried out by 10 of DELTA's selected and trained expert listeners. The only thing that the listeners knew about the loudspeakers were that they were compact loudspeakers. In a pilot test with eight of the listeners on four of the loudspeakers, it was decided prior to testing which sound wheel's (see Appendix 1) characteristics (attributes) would best describe the differences between the loudspeakers. See Table 1.

The blind listening test was executed from a PC and the assessments used unstructured line scales from "A little" to "A lot" respectively, and from "Small" to "Large" for five of the attributes, and "Dark-Neutral-Light". There was an underlying scale from 0-15 (with a precision of 0.1), which the listeners did not know about.

The test was carried out by one listener at a time. During the test, the sequence of loudspeakers and the attributes that were assessed were different for each person. The two musical pieces were listened to twice (also in random sequence) for each attribute.

The test results are based on a total of 1920 assessments (excluding anchor loudspeakers and 'Preference').

	Attribute	Definition and description
Table 1. The sound wheel's attributes, which were chosen for this test.	Brilliance	 Treble or high frequency extension. Scale: A little: As if you hear music through a door, muffled, blurred, or dull. A lot: Crystal-clear reproduction extended treble range with airy and open treble. Lightness, purity, and clarity with space for instruments. Clarity in the upper frequencies without being sharp or shrill and without distortion
	Bass Depth	Denotes how far the bass extends downwards. If it goes down in the low end of the spectrum, there is great depth. Should not be confused with Bass strength, which indicates the strength of the bass or 'Boomy' which relates to resonances in the lower bass region. Scale: A little - A lot
	Punch	Specifies whether the strokes on drums and bass are reproduced with clout, almost as if you can feel the blow. The ability to effortlessly handle large volume excursions without compression (compression is heard as level variations that are smaller than one would expect from the perceived original sound). Scale: A little – A lot
	Dark- Bright	 Denotes the balance between bass and treble. Scale: Dark: Excessive bass. Either loud bass or weak treble. Neutral: Bass and treble are perceived equally loud, there is a balance in the reproduction. This also applies if both bass and treble are equally weak or if the bass and treble both are too loud. If it leads to prominent or soft midrange this is assessed by the midrange strength. Bright: Excessive treble. Either loud treble or weak bass. The cause for the sound being dark or light can deduced from the assessments of bass strength and treble strength.
	Spatial precision	Can the individual instruments and voices be clearly positioned and separated in the spatial sound image? How precisely do the individual sound sources stand in the room? If the individual sound sources unintentionally are wide or broadened out, the precision is low.
	Natural	Sounds reproduced with high fidelity. Acoustic instruments, voices and sounds, sounds like reality. The sound is similar to the listener's expectation to the original sound without any timbral or spatial coloration or distortion, "Nothing added - nothing missing". The soundstage is clear in space and brings you close to the perceived original sound experience. Scale: A little – A lot

Loudspeakers

In total, eight compact loudspeakers from Hi-Fi Klubben were included in the test. See Table 2.

	Manufacturer, Model	Price DKK per item	Frequency response (-3dB):	Size (W x H x D) cm	Gross vol- ume, I
- S -	Argon 6340	399	80–20,000 Hz	14.8 x 23.9 x 16.5	6
	DALI ZENSOR 1	999	53–26,500 Hz	16.2 x 27.4 x 22.0	10
	Bowers & Wilkins 686 S2	1,799	52–22,000 Hz	16.0 x 31.5 x 22.9	11
	Skandyna MiniPod MK3	1999	55–22,000 Hz	21.0 x 34.0 x 20.0	14
	Bowers & Wilkins 685 S2	2,399	52–22,000 Hz	19.0 x 34.5 x 32.4	21
	DALI OPTICON 2	2,999	59–27,000 Hz	19.5 x 35.1 x 29.7	20
	Bowers & Wilkins CM1 S2	3,299	50–28,000 Hz	16.5 x 28.0 x 27.6	13
	DALI MENUET	3,699	59–25,000 Hz	15.0 x 25.0 x 23.0	9

Table 2.

Overview of the loudspeakers that were included in the test. The technical data comes from Hi-Fi Klubben's website.

To create comparison with future tests, three anchor loudspeakers were also included in the test: One which is assessed high on all attributes, one which is assessed as low on all attributes, and one which is assessed somewhere in-between.

Results

The table below shows the results from the listening test in a table format. The results are shown in graphical form in Figure 10 and Figure 11.

Systems	Bass Depth	Punch	Natural	Spatial precision	Brilliance	Dark-Bright
Argon 6340	5.5	5.9	8.8	9.6	10.8	10.0
DALI Zensor 1	6.7	7.2	10.1	10.5	10.4	8.4
B&W 686 S2	6.7	6.9	9.5	9.4	9.8	8.8
MiniPod MK3	8.1	8.5	9.5	9.3	10.0	7.3
B&W 685 S2	8.9	9.6	10.1	10.3	9.9	7.6
DALI Opticon 2	10.9	10.8	10.7	9.5	9.5	6.7
B&W CM1 S2	11.2	11.2	10.9	11.2	11.6	7.4
DALI Menuet	7.0	7.1	9.7	9.4	9.6	7.9

Table 3. Each of the results in the table is a mean value of a total of 40 assessments (10 listeners, two repetitions

and two pieces of music).

The assessment scales range from 0–15. The precision is approx. +/- 1 unit.

For Dark-Bright, the low end of the scale corresponds to Dark and the high end of the scale corresponds to Bright. A natural reproduction of the sound lies at 7.5 on the Dark-Bright scale.

Bass Depth together with Punch is the attribute that separates the systems most markedly. Argon 6340 has the least Bass Depth. Next, comes a group with DALI Zensor 1, B&W 686 S2, and DALI Menuet. MiniPod MK3 and B&W 685 S2 are slightly higher, and the highest is DALI Opticon 2 and B&W CM1 S2, which are not significantly different. In general, the assessments with the musical piece Zhao Cong (ZC) resulted in a slightly higher assessment of Bass Depth than the musical piece with Jennifer Warnes (JW), and this was most evident with DALI Zensor 1. See Figure 8.

Punch The comments given above, also apply to Punch.

Natural There are relatively small differences between the systems. Note in particular, that DALI Opticon 2 and B&W CM1 S2 are significantly higher placed than Argon 6340.

Figure 11. The results of three objective attributes in the blind test.

Each data point is an average of a total of 40 assessments.

The assessment scale ranges from 0– 15. The precision is approx. +/- 1 unit.





Figure 12. The results of three objective attributes in the blind test.

Each data point is an average of a total of 40 assessments.

The assessment scale ranges from 0– 15. The precision is approx. +/- 1 unit. **Spatial precision**: There are relatively small differences between the systems. B&W CM1 S2 is the highest, however not significantly different from DALI Zensor 1 and B&W 685 S2.

Brilliance: The systems are almost identical. The only system that lies significantly above the average (at 10.2) is B&W CM1 S2. The assessments of the two pieces of music are almost identical for each system.

Dark–Bright: This attribute is slightly special, partly because the neutral point lies in the middle of the scale, and partly because it to some degree can be predicted from Bass Depth and Brilliance. The results indicate that Argon 6340, DALI Zensor 1, and B&W 686 S2 are positioned slightly more to the Bright side, and they are positioned low respectively in relation to Bass Depth. DALI Opticon 2 lies slightly below neutral, but not significantly lower than B&W CM1 S2.



5

Preference: In contrast to the other attributes, Preference is a subjective attribute. The stated Preference applies to DELTA's expert listeners, and is not necessarily representative for other listeners. Saying that, there are still some interesting findings. The sound from DALI Opticon 2 and B&W CM1 S2 are the most preferred. CM1 is positioned the highest, but the difference is not significant. Next comes DALI Zensor 1 and B&W 685 S2. The sound from Argon 6340 is the least preferred of the tested loud-speakers, however it still lies in the middle of the scale.

It is nevertheless interesting to note that in this blind test, with few exceptions, there is a very good correlation between the experts' preferences and the price of the loudspeakers. See Figure 20. DALI Zensor 1 is distinctive because it is positioned well above the general trend line, and the DALI Menuet is positioned below. It can also be seen that the preferences to a certain degree follow the Bass Depth and thus the loudspeakers' volume. Since DALI Menuet is a small loudspeaker, it comes up a little short here.

As stated earlier, it is not certain that the consumers have the same preferences as DELTA's listening panel in a double blind test, where no one sees the loudspeakers. It is only the consumers themselves who can decide what their preferences are.

Figure 13. DELTA's experts' preferences in a blind test, where they only knew that they listened to compact loudspeakers.

Comparisons

Different forms of presentations provide the opportunity to make comparisons between loudspeakers. Some examples are given below.

As indicated in Figure 14 the listening test is able to show significant differences between the systems.



Figure 15 shows a comparison of the three systems with the lowest prices: Argon 6340 at DKK 399, DALI ZENSOR 1 at DKK 999, and Bowers & Wilkins 686 S2 at DKK 1,799. It shows that the more expensive systems provide more bass depth and a more neutral timbral balance (Dark-Bright).





Figure 15. A comparison of the three least expensive systems.





Finally, Figure 17 shows that the differences between the two most expensive systems, DALI OPTI-CON 2 at DKK 2,999 and Bowers & Wilkins CM1 S2 at DKK 3,299, are very small, without taking Brilliance into account. B&W loudspeakers have the highest Brilliance. If you prefer a brilliant sound, B&W CM1 S2 is a good choice, but if you prefer less brilliance and slightly softer sound, then DALI OPTICON 2 is the better choice.



Figure 17. A comparison of the test's two most expensive systems.

Interesting relationships

Figure 18 shows that in general, there is no relationship between the experienced bass depth and the technical data for the low frequency limit. Four of the loudspeakers more or less had the same lower frequency limit of 50-53 Hz, yet they were assessed as having a large (and significant) difference in the perceived Bass Strength. The same applies to DALI Menuet and DALI Opticon 2, whose respective bass depth was assessed as being quite different, while the data stated that both had a lower frequency limit of 59 Hz. The fact that you cannot assess a loudspeaker's sound based on the technical data is hardly new, however it is still interesting to see it demonstrated with a well-defined listening test. In other words, if you want to know how a loudspeaker sounds, it is more sensible to use a perceptual assessment of a loudspeaker's sound based on a listening test rather than taking outset in the technical data.



Figure 18. Comparison between the loudspeakers' technical data for lower frequency limit and assessments of Bass Depth.



Figure 19 shows that the bass depth generally increases with gross volume. However, based on their volume, DALI Opticon 2 and B&W CM1 S2 have a bass depth larger than excepted.



Figure 20. A comparison of the loudspeakers' price and DELTA's expert panel's preference of the sound.

Figure 20 shows that generally, there is a correlation between price and preference. The sound from DALI Zensor 1 however, has a higher preference in relation to the price, while the very small DALI Menuet has a lower preference. As stated earlier, the stated preferences are an average of DELTA's listener panel's preferences in a blind test, and this may differ from the consumers' preferences.

Models

The data from the listening test can be used to generate different models. The models apply to the actual test. Of course, their general validity needs to be tested using listening tests of other loud-speakers.

Based on the definition of the attribute Dark-Light, see Table 1, you could easily think that this attribute can be predicted on the basis of the assessments of bass and treble. This is in fact the case. Figure 21 shows that there is a fine, almost one-to-one correlation between the assessments of Dark-Light, which were the results of the listening test, and the values that are calculated on the basis of a model, which is built on the assessments of Bass Depth and Brilliance.



Figure 21. The results from the listening test of the assessment of Dark-Light (X-axis) compared with the results of a model which calculates Dark-Light based on Bass Depth and Brilliance (Y-axis).

It would also be interesting, if the consumers' preferences could be predicted. In the general context, consumers' preferences depend on the product as a whole and the influences they are exposed to from the sales material, etc.

Nevertheless, it is interesting if a model can be made for the listeners' preferences, solely in relation to the sound of the different speakers.

Figure 22 shows the results of calculations carried out using a model which is based on the results of Bass Depth and Brilliance in the first, second, and third power. It shows that a very good and precise correlation with Preference can be made, based on the assessments of the two named attributes for this test.





Conclusion

The TECH Document has shown which ingredients are needed to carry out reliable and reproducible listening tests, which give stable results with a known precision. In addition to the specified listening conditions and loudspeaker positions, a number of trained listeners are required. The assessments are carried out using well-defined attributes in a blind test using randomised presentations of the products. In addition, different programme material should be used and the systems which are tested, must be adjusted precisely to play back equally loud during the test.

If all of this is carried out, an objective characteristic of the perceived sound coming from the systems can be achieved, which, for example in connection with product development or consumer information, can supplement the technical data. If the test results are to be comparable with other tests under corresponding conditions, there also has to be some recurring systems, i.e. anchor loudspeakers, included in the tests.

The use of these principles is demonstrated in a test, which was carried out as a double blind test, where the 10 listeners only knew that they were listening to compact loudspeakers. Each of the test results was a mean value based on a total of 40 assessments (10 listeners, two repetitions, and two pieces of music) for each attribute and for each loudspeaker. The attributes were assessed on a scale from 0–15, and the precision of the test was +/- 1 unit. The precision can be increased by increasing the number of trained listeners.

The listening test provided supplementary information in relation to the technical data. For example, it showed a clear difference between the perceived bass depth and the technical data for the systems' lower limiting frequency. The results also showed clear and significant differences in the perceived sound between the least and most expensive systems.

It has been illustrated how the results can be presented and compared in different ways, which can give consumers guidance about the loudspeakers' sound characteristics, in the form of sensory product information.

A comparison of the loudspeakers' prices and the average of the DELTA's expert listeners' preferences show that – with individual exceptions – there is a relationship between price and preference. However, preference is a subjective attribute in contrast to the other attributes, and the consumers preferences are not necessarily the same as those of DELTA's expert listeners. In addition, product properties other than sound characteristics often come into play with consumers.

It can be concluded that the technical measurements alone do not give complete information about the sound character of audio products. By supplementing the technical data with the results of listening tests, it is possible to give an objective characteristic of the perceived sound.



To learn more, visit: senselab.madebydelta.com. Background articles and reports are available on the website.

Appendix 1- Sound Wheel

In the case of the listening test with reproduced sound (audio and Hi-Fi), the attributes are selected from the sound wheel shown below. All of the attributes are defined in more detail, so that the listeners have the same understanding of their significance, which is a prerequisite for objective assessments.



The sound wheel with attributes for audio. For more information and definitions of the attributes, see TEK Document no. 7 at http://senselab.madebydelta.com/about/publications/



Appendix 2 – Alternative sound profiles

About TECH Document

To learn more about the subject described in this TECH Document, visit <u>www.senselab.madebydelta.com</u>, which has background reports and articles.

DELTA regularly publishes the TECH Document series to report the latest international knowledge in our specialist areas. The arm is to help bring about faster turnaround where the newest technological breakthroughs become commercially viable for Danish companies.

About DELTA

We help ideas meet the real world

DELTA is a self-governing high-tech company that makes knowledge available to everyone. As strategic partner to our clients, we ensure the optimum application of advanced technology. We develop, test, certify and advise in every phase of our customers' product development. We have been a technology pathfinder since 1941 and it is our vision that Denmark must be the best place to carry out high-tech product development.