TEK'notat.
den nyeste viden og teknologi

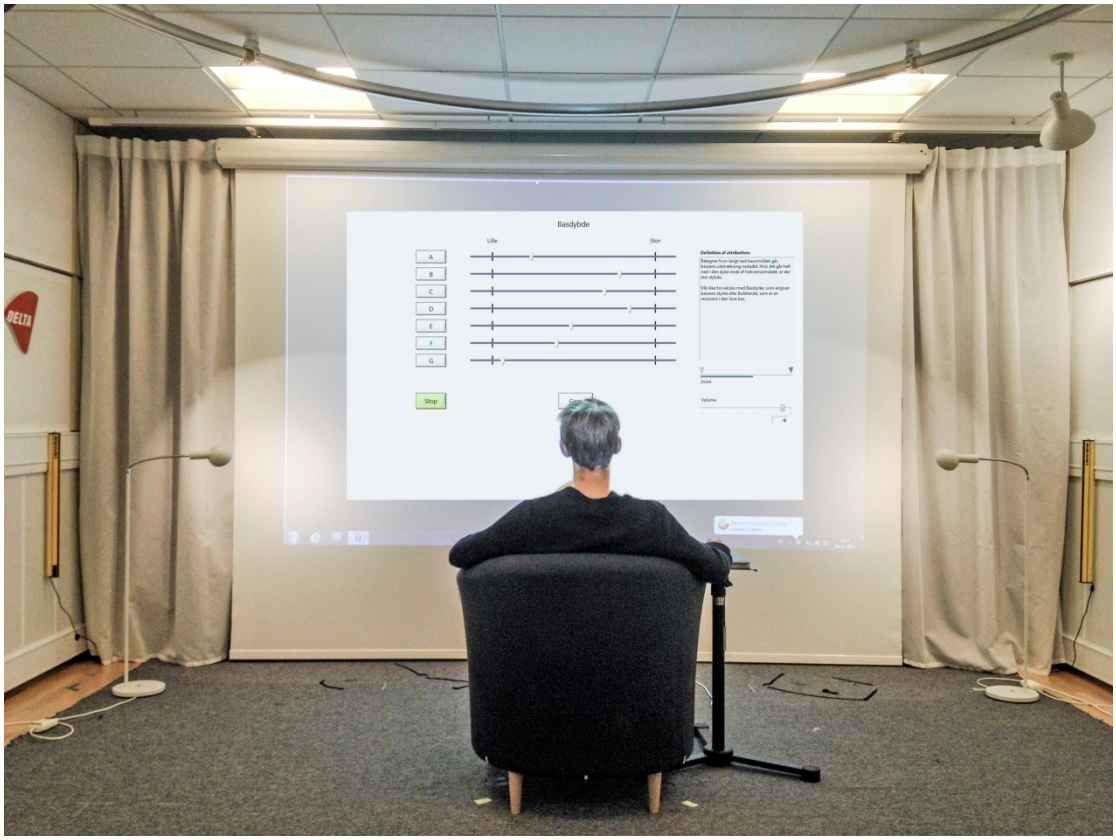
Marts 2016

TN11

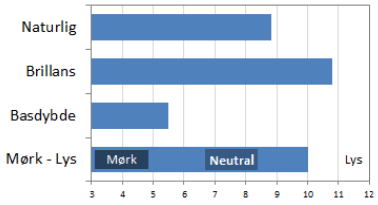
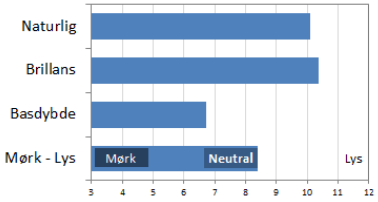
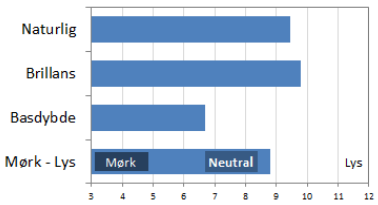
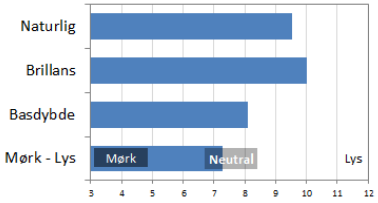
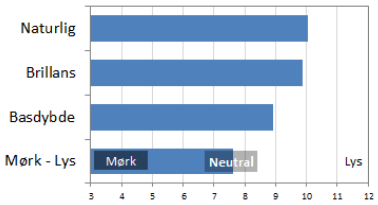
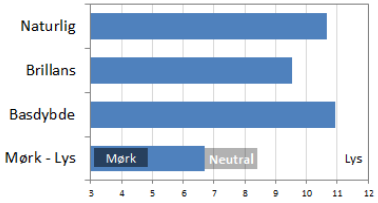
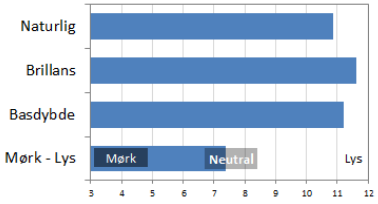
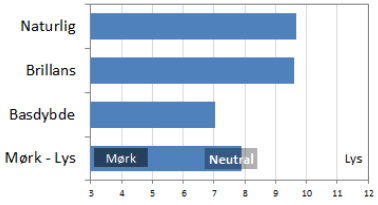
Objektive lyttetests af audioprodukter

- et værdifuldt redskab til produktudvikling og forbrugerinformation

Torben Holm Pedersen



DELTA
Venlighedsvej 4
2970 Hørsholm
Danmark
Tlf. +45 72 19 40 00
delta@delta.dk
delta.dk

Produkt	Tekniske data	Lydprofil
	Argon 6340 (399,-/stk.) Frekvensgang (-3dB) : 80-20.000 Hz Følsomhed: 84 dB (1W/1m) BxHxD : 14,8 x 23,9 x 16,5 cm Volumen: Ca. 6 l. Vægt: 2,3 kg	
	DALI ZENSOR 1 (999,-/stk.) Frekvensgang (-3dB): 53-26.500 Hz Følsomhed: 86,5 dB HxBxD : 27,4 x 16,2 x 22,0 cm Volumen: Ca. 10 l. Vægt: 4,2 kg	
	Bowers & Wilkins 686 S2 (1.799,-/stk.) Frekvensgang (-3dB) : 52-22.000 Hz Følsomhed: 85 dB BxHxD : 16,0 x 31,5 x 22,9 cm Volumen: Ca. 11 l. Vægt: 4,9 kg	
	Scandyna MiniPod MK3 (1.999,-/stk.) Frekvensgang (-3dB) : 55-22.000 Hz Følsomhed: 91 dB BxHxD : 21 x 34 (44 på ben) x 20 cm Volumen: Ca. 14 l Vægt: 2,3 kg	
	Bowers & Wilkins 685 S2 (2.399,-/stk) Frekvensgang (-3dB) : 52-22.000 Hz Følsomhed: 87 dB BxHxD : 19,0 x 34,5 x 32,4 cm Volumen: Ca. 21 l. Vægt: 6,8 kg	
	DALI OPTICON 2 (2.999,-/stk.) Frekvensgang (-3dB): 59-27.000 Hz Følsomhed: 87 dB BxHxD : 19,5 x 35,1 x 29,7 cm Volumen: Ca. 20 l. Vægt : 7,8 kg	
	Bowers & Wilkins CM1 S2 (3.299,-/stk.) Frekvensgang (-3dB): 50-28.000 Hz Følsomhed: 84 dB BxHxD : 16,5 x 28,0 x 27,6 cm Volumen: Ca. 13 l. Vægt: 6,7 kg	
	DALI MENUET (3.699,-/stk.) Frekvensgang (-3dB): 59-25.000 Hz Følsomhed: 86 dB HxBxD : 25,0 x 15,0 x 23,0 cm Volumen: Ca. 9 l. Vægt: 4,1 kg	

Indhold

Resumé 4

Baggrund 5

Ingredienser i en objektiv lyttetest 6

Lytterum 6
Måleopstilling 7
Ankersystemer 8
Lyttere 9
Attributter 9
Programmateriale 9
Niveaukalibrering 10
Forsøgsdesign og dataopsamling 10
Statistisk analyse. 11
Sensorisk varedeklaration 12

Eksempel på en lyttetest 13

Testopstilling 13
Lyttetesten 14
Højttalere 15
Resultater 16
Sammenligninger 19
Interessante sammenhænge 21
Modeller 23

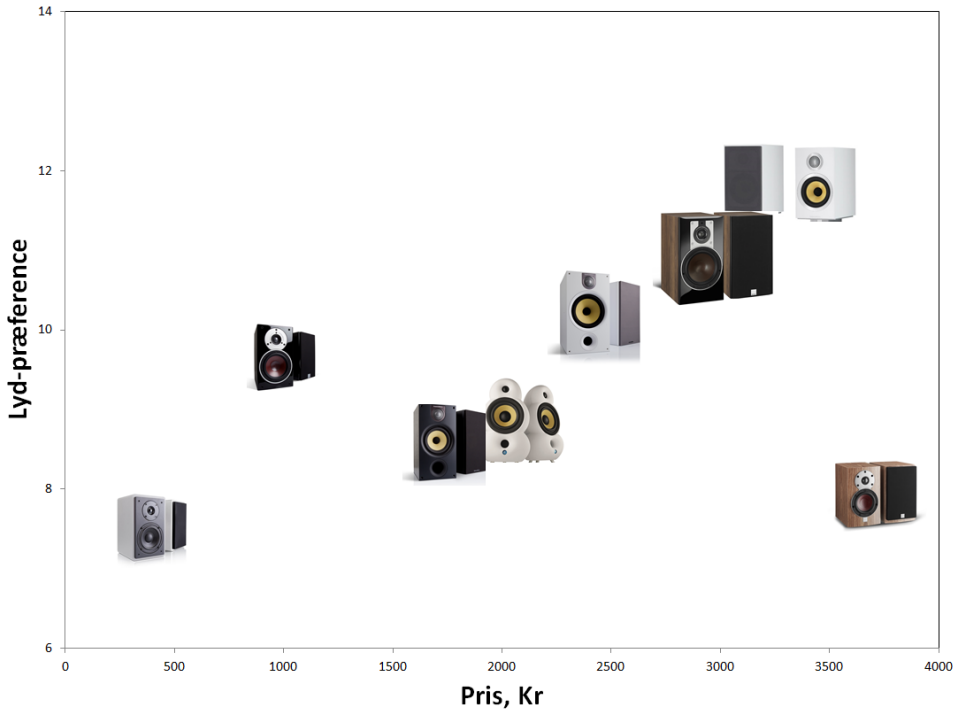
Appendiks 1 - Lydhjulet 26

Appendiks 2 – Alternative for lydprofiler 27

Rapporten kan downloades fra DELTA på
senselab.madebydelta.com/about/publications

Resumé

Dette TEK'notat viser resultater af en lyttetest, og beskriver hvad der kræves, for at man kan få en objektiv karakteristik af den opfattede lyd.



Notatet beskriver, hvad der skal til, for at man kan udføre pålidelige og reproducerbare lyttetests. Der er krav til lytteforholdene, fx lytterum og højttalerpositioner og der kræves et panel af trænede lyttere. Bedømmelserne bør foretages som en blindtest med produkterne i tilfældig rækkefølge og der skal være veldefinerede ord til beskrivelse af deres karakteristika. Man skal bruge forskelligt programmateriale (fx musik) og alle systemerne skal justeres til samme lydstyrke.

En sådan test giver til gengæld en objektiv karakteristik af lyden fra systemerne, som fx kan bruges ved produktudvikling eller som forbrugeroplysning som et nyttigt supplement til de tekniske data.

Anvendelsen af disse principper er demonstreret ved en test af 8 kompakthøjttalere, hvor hvert testresultat var en middelværdi af i alt 40 bedømmelser.

Resultaterne fra lyttetesten kan ses på side 2. Lyttetesten gav nyttige og supplerende oplysninger i forhold til de tekniske data. Den viste fx, at der var en tydelig forskel på den oplevede basdybde, selv om den tekniske angivelse af systemernes nedre grænsefrekvens var stort set ens.

Notatet viser også, hvordan resultaterne kan præsenteres på forskellig vis. Man kan fx give forbrugerne oplysning om højttalernes lyttemæssige karakter i form af en lydprofil, som kan bruges som en sensorisk varedeklaration.

En sammenligning mellem højttalernes pris og DELTA's ekspertlytteres præferencevurderinger viser - med enkelte undtagelser - at der er en sammenhæng mellem pris og præference. Husk dog at præference er subjektiv i modsætning til testens øvrige attributter. Forbrugernes præferencer er ikke nødvendigvis de samme.

Forfatter: Torben Holm Pedersen
Titel: Objektive lyttetests af audioprodukter.
Firma: DELTA, SenseLab

Baggrund

Lyttetests kan bruges i rigtig mange sammenhænge. De kan fx bruges til at undersøge, om datakomprimering af lydfiler giver kvalitetstab, om kopibeskyttelse indlejret i en lydfil kan høres, om støjundertrykkelse i en mobiltelefon forringer talen eller taleforståeligheden, om det ene eller det andet høreapparat er bedst i baggrundsstøj, i hvor høj grad forskellig slags støj fra vindmøller er generende osv.

Når det gælder lyttetests på audiosystemer, er der også forskellige muligheder. Det kan fx være din personlige prøvelytning af forskelligt udstyr (fx højttalere), som du overvejer at anskaffe, det kan være anmeldelser i aviser og magasiner, hvor en enkelt eller et par personer har lyttet til audioprodukter, det kan være et udviklingsteam, der lytter til nye produkter, og så kan det endelig være mere systematiske tests udført som blindtests med trænede lyttere under kontrollerede forhold. I dette TEK'notat vil vi beskæftige os med sidstnævnte, fordi sådanne tests kan give objektive og reproducerbare resultater med en veldefineret nøjagtighed.

I TEK'notat nr. 7 'Karakteristik af lyden fra audioprodukter'[1] blev der redegjort for relationerne mellem fysiske målinger (fx målinger af lydtryk, frekvenskarakteristikker m.m.) og lyttetests, som kan være objektive (perceptuelle målinger udført med trænede lyttere) eller subjektive (præferencetests, fx med forbrugere). Det blev også givet eksempler på, hvorfor tekniske data og produktanmeldelser ikke alene giver et fuldgyldigt billede af den opfattede lyd. Endelig blev der redegjort for, hvorledes man udvælger og tester ord (attributter), som kan bruges til at give en veldefineret karakteristik af de produkter, der bedømmes. Det blev vist, hvorledes disse attributter kunne organiseres i et lydhjul, som er et vigtigt element ved udførelsen af objektive lyttetests, og som også kan bruges til kommunikation af produktkarakteristika dels ved produktudvikling og dels mellem producent, salgsled og forbrugere. (Lydhjulet er gengivet i Appendiks 1).

I dette TEK'notat vil vi vise, hvordan objektive og sammenlignelige lyttetests udføres, samt give et eksempel på resultaterne fra en sådan test udført i samarbejde med forretningskæden Hi-Fi Klubben, der har stillet højttalere til rådighed for testen.

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

Ingredienser i en objektiv lyttetest

En objektiv lyttetest er en test, som ikke vedrører personlige præferencer. Resultaterne er middelværdier af bedømmelser foretaget af et antal lyttere (sædvanligvis mere end 10) af veldefinerede attributter. For at en lyttetest kan føre til reproducerbare og veldefinerede resultater, er der en række ingredienser, der skal indgå:

- Lytterum med neutral påvirkning af lyden (efterklang) og lav baggrundsstøj
- Måleopstilling og -udstyr til direkte sammenligning i en blindtest
- Ankersystemer (faste referencer) for at gøre resultaterne sammenlignelige med andre tests
- Udvalgte og trænede lyttere
- Veldefinerede attributter
- Kritisk programmateriale, afspillet ved samme lydstyrke fra alle systemer
- Randomiseret forsøgsdesign, dvs. hvor systemerne præsenteres i tilfældig rækkefølge,
- Effektiv dataopsamling
- Statistisk analyse
- Præsentation af resultater.

Lytterum

Hvis det drejer sig om lyttetests på højttalere (passive, aktive, streaming, bluetooth…), skal man have et lytterum, som ikke farver lyden. Det skal være højttalerne, der bedømmes, og ikke indflydelsen fra rummet. Efterklangstiden skal være lav og ligge inden for specificerede grænser. Specielt skal efterklangstiden være ens ved alle frekvenser (dog tillades en svag hævning ved lave frekvenser). Dernæst skal baggrundsstøjen være lav, så uvedkommende lyde ikke forstyrrer lytningen. Ideelt set bør standarderne EBU Tech 3276 og ITU-R BS.1116-3 til lyttetests på multikanals-højttalersystemer opfyldes. DELTA SenseLabs lytterum, se Figur 1, opfylder disse. Dvs. at der kan udføres test af både mono-, stereo- og surroundsystemer (5.1 til 22.2).

Figur 1.
En lytter bedømmer højttalere under en blindtest i DELTA's lytterum. Resultatet af testen er gennemsnittet af et antal lytteres bedømmelser. Lærredet med brugerfladen er akustisk transparent, og højttalerne er skjult bagved. Bøjlen i loftet er til ophængning af loftshøjttalere til multikanals-surroundsystemer.

Hvis det drejer sig om test med hovedtelefoner, er det tilstrækkeligt, at baggrundsstøjen ligger under det krævede, typisk kan man her bruge mindre lyttebokse, se Figur 2.

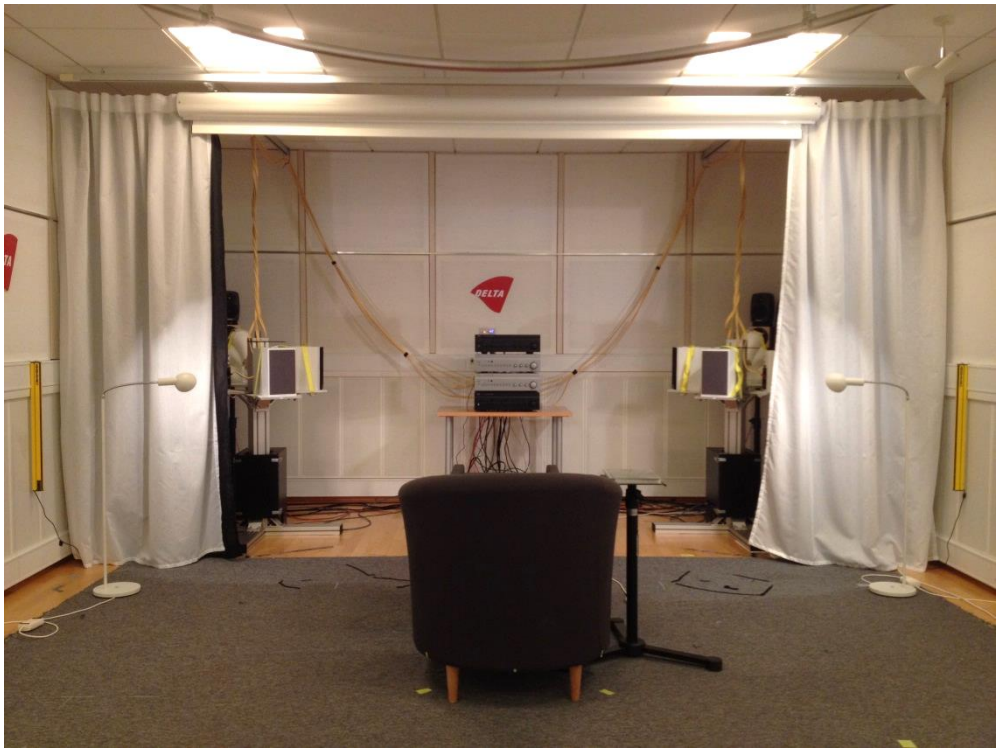
Figur 2.
Den ene af SenseLabs to lyttebokse til tests med hovedtelefoner. Baggrundsstøjen inde i boksen er meget lav (< NR 15).

Måleopstilling

Når lyttetesten skal udføres som en blindtest, ved lytteren ikke, hvad der lyttes på og kan ikke se højttalerne. Højttalerne skjules bag et akustisk transparent lærred eller gardin. I DELTA's lytterum udnyttes lærredet som brugerflade for test-softwaren, så der ikke forekommer uønskede lydrefleksioner o.l. fra en computerskærm foran lytteren, se Figur 1.

Der er et særligt problem ved lyttetest med højttalere: Hvis højttalerne står forskellige steder i rummet, vil lyden farves forskelligt, alt efter hvor højttalerne står. Det kan løses ved at bytte rundt på højttalerne under testen, så alle højttalere efter tur er placeret på alle pladser. Det er en besværlig og langsommelig proces.

For at kunne udføre hurtige og effektive lyttetests har DELTA i stedet valgt at konstruere nogle drejestandere, som sørger for at de højttalere, der lyttes til, altid er i samme position. Drejestanderne, se Figur 3 og 4, kan næsten lydløst og på ca. 1 sekund bringe højttalerne i stilling. Den hurtige skiftetid uden forstyrrende støj er vigtig, da menneskers akustiske hukommelse er kort.

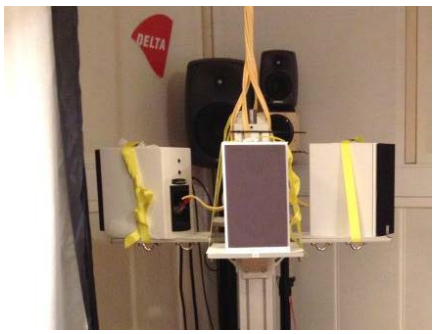


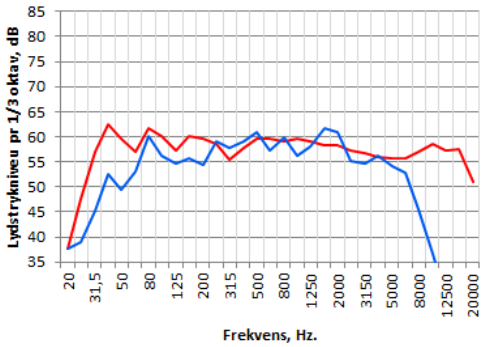
Figur 3.
Lytterummet med de to drejestandere, som gør, at højttalerne altid står i samme position i rummet, når der lyttes på dem.

En video hvor drejestanderne demonstreres kan ses på senselab.madebydelta.com/services/listening-tests.

Ankersystemer

Resultaterne af en lyttetest er ikke upåvirket af den sammenhæng, den bliver udført i. Hvis man tester mobiltelefoners indbyggede højttalere til musikafspilning sammen med en lille middelmådig reolhøjttaler, vil denne blive bedømt til at have en høj basstyrke. Hvis samme reolhøjttaler bliver bedømt sammen med Hi-Fi gulvhøjttalere, vil den blive bedømt til at have en lavere basstyrke. Hvis man ønsker at kunne sammenligne resultater fra forskellige lyttetests, er det nødvendigt, at der er nogle gennemgående systemer, der kan tjene som faste holdepunkter, såkaldte ankersystemer.





Figur 4.
Foto af en drejestander. I baggrunden ses de tre ankerhøjttalere der er med i tests for at skabe sammenlignelighed. Til højre er vist frekvenskarakteristikker målt i lyttepositionen af det øvre (rød) og nedre (blå) anker.

Der vælges ankersystemer, som for den pågældende produktkategori ligger henholdsvis højt og lavt for de fleste attributter. Ankersystemerne bliver i hver test bedømt på lige fod med de øvrige systemer,

og hvis bedømmelsen af dem har ændret sig siden foregående test, kan resultaterne korrigeres for denne ændring, så resultaterne fra de to tests bliver sammenlignelige.

Lyttere

Hvis vi ser bort fra præferencetests, hvor lytterne bør vælges fra det relevante markedssegment, bør der bruges trænede lyttere med gode ører. DELTA's lyttepanel består af ca. 30 udvalgte[2] lyttere, hvis hørelse er testet til at være normal, og som derefter har fået en grundig træning. Ud over selve 'øretræningen' sikrer træningen, at lytterne forstår attributterne og bruger svarskalaer på en ensartet måde. Ekspertlytterne er specifikt trænet i de attributter, der er relevante for test af audiosystemer. Mange af ekspertlytterne har desuden en beskæftigelse eller interesse med relation til audiobranchen (musikere, komponister, Hi-Fi nørder, Hi-fi forhandlere, DTU-studerende i akustik, lyddesignere, lydteknikere…). Lytternes performance overvåges løbende for så vidt angår deres evne til at høre forskel på systemer og at kunne gentage egne bedømmelser i blindtest.

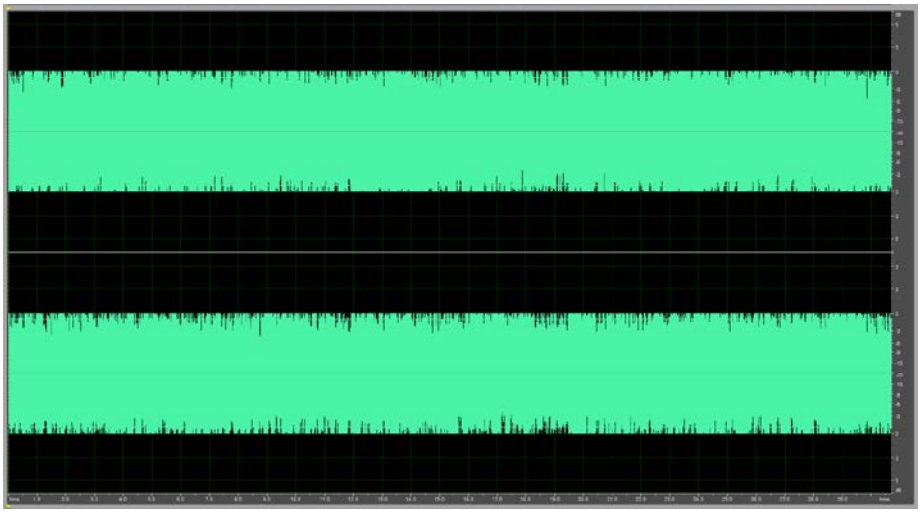
Attributter

Systemerne (fx højttalere) bedømmes på et antal attributter, der tilsammen udgør en lydprofil, som karakteriserer systemerne. Der er en række krav til gode attributter, som er nærmere omtalt i TEK'notat nr. 7[3]. Hovedkravene er, at de er entydige og veldefinerede, samt at de forstås på samme måde af alle lytterne. Ved lyttetests på audiosystemer benyttes som hovedregel attributter fra lydhjulet, se Appendiks 1. Normalt er det tilstrækkeligt med 6-10 attributter til at karakterisere systemerne. Attributterne kan vælges ud fra en pilottest, fx som de attributter, der bedst karakteriserer forskellene mellem systemerne. Alternativt kan de vælges af forsøgslederen eller kunden til lyttetesten.

Programmateriale

For at få belyst så mange af systemernes egenskaber som muligt, skal der benyttes flere forskellige programmaterialer (musikstykker, tale, lydeffekter…hvad der nu er relevant i den pågældende test). Programmaterialet vælges som regel blandt indspilninger i høj kvalitet med varierende dynamik og frekvensindhold. Mange moderne indspilninger amplitudekomprimeres, se Figur 5, og sådanne optagelser er ikke egnede til at give et nuanceret billede af de testede systemer. Evt. måles materialet igennem mht. komprimering og frekvensindhold for at sikre, at det ikke er programmaterialet, der sætter grænser for bedømmelsen. Typisk bruges nogenlunde ensartede passager fra musikstykkerne med en varighed omkring 20 sekunder.

[2] Se: Søren Vase Legarth & Nick Zacharov: Assessor selection process for multisensory Applications. Audio Engineering Society 2009, Munich, Germany.
[3] Udviklingen og valideringen af attributterne er også omtalt i: Torben Holm Pedersen & Nick Zacharov: The development of a Sound Wheel for Reproduced Sound. Audio Engineering Society 2015, Warsaw, Poland.



Figur 5.
Et eksempel på amplituden (y-aksen) af et musikstykke som funktion af tiden (x-aksen). Mange indspilninger er som denne, amplitude-komprimerede, så de ikke er særligt velegnede til at teste systemers dynamiske egenskaber.

Niveaukalibrering

De opfattede produktegenskaber er ikke uafhængige af lydstyrken. Hvis der eksempelvis er to forholdsvis ens højttalere med i en test, viser erfaringen, at den højttaler, der spiller højest, vil blive foretrukket. Derfor er det vigtigt, at lydstyrken fra alle systemerne er den samme under testen. Efter at være spillet til, indstilles alle systemerne derfor inden testen, så de har samme styrke. Normalt foretages denne kalibrering ved at justere systemerne til samme A-vægtede lydtrykniveau i lyttepositionen ved afspilning af et særligt støjsignal (lyserød støj).

Forsøgsdesign og dataopsamling

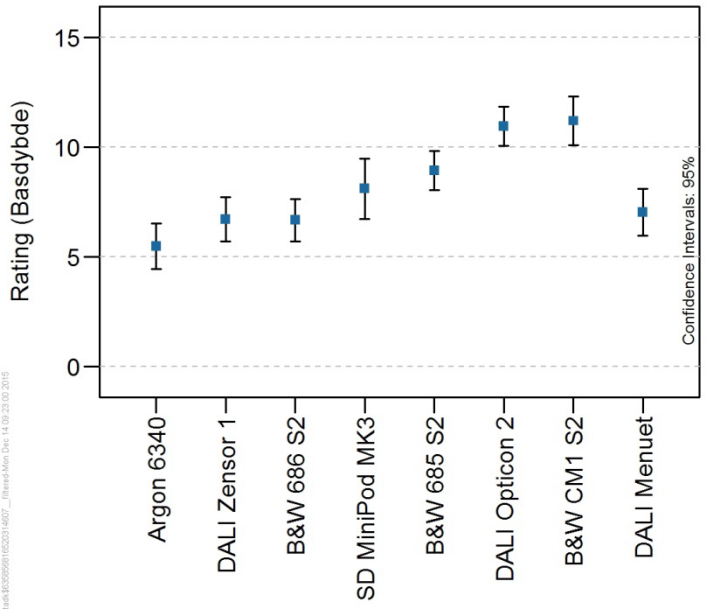
Med mindre man direkte ønsker at undersøge indflydelsen af produktdesign og mærke (brand value), skal lyttetesten udføres som en randomiseret blindtest. Dvs. at testen udføres og designes på en sådan måde, at lytteren ikke ved, hvad der lyttes til. Produkter, attributter og programmateriale præsenteres derfor i en ny tilfældig rækkefølge for hver lytter. Vigtigheden af dette kan illustreres ved, at selv erfarne lyttere, der efter en test præsenteres for resultatet, kan bliver overraskede over, hvor svært det kan være at høre forskel, selv på produkter som de ellers var helt sikre på, at de tydeligt kunne høre forskel på.

For at den randomiserede test kan afvikles hurtigt og effektivt med korte (< ca. 1 sekund) intervaller mellem de systemer, der skal bedømmes, er det nødvendigt med et automatiseret system til præsentation af stimuli (dvs. kombinationer af systemer og programmateriale) og til opsamling af lytternes svar. DELTA har udviklet flere systemer til dette formål. Et af systemerne er specielt udviklet til lyttetest på højttalere, så det også styrer drejestanderne under lyttetesten. Det betyder i praksis, at når opstillingen er gjort klar og niveaukalibreret, kan testen afvikles fuldautomatisk i det tempo, der passer den enkelte forsøgsperson. De mest nøjagtige bedømmelser opnås, når systemerne sammenlignes attribut for attribut. Et eksempel på brugerfladen for en sådan test er vist i TEK'notat nr. 7. Et andet system, SenseLabOnline, muliggør afvikling af lyttetests over internettet, og bruges fx til afvikling af test med støjundertrykkelse i mobiltelefoner på flere sprog.

Statistisk analyse

Den statistiske analyse bruges som det første til at overvåge datakvaliteten - fx ved at tjekke om lytterne kan skelne mellem forskellige stimuli, og om de kan reproducere deres egne bedømmelser i en blindtest[4]. Selv med trænede lyttere under optimale lytteforhold er der forskelle i de individuelle bedømmelser.

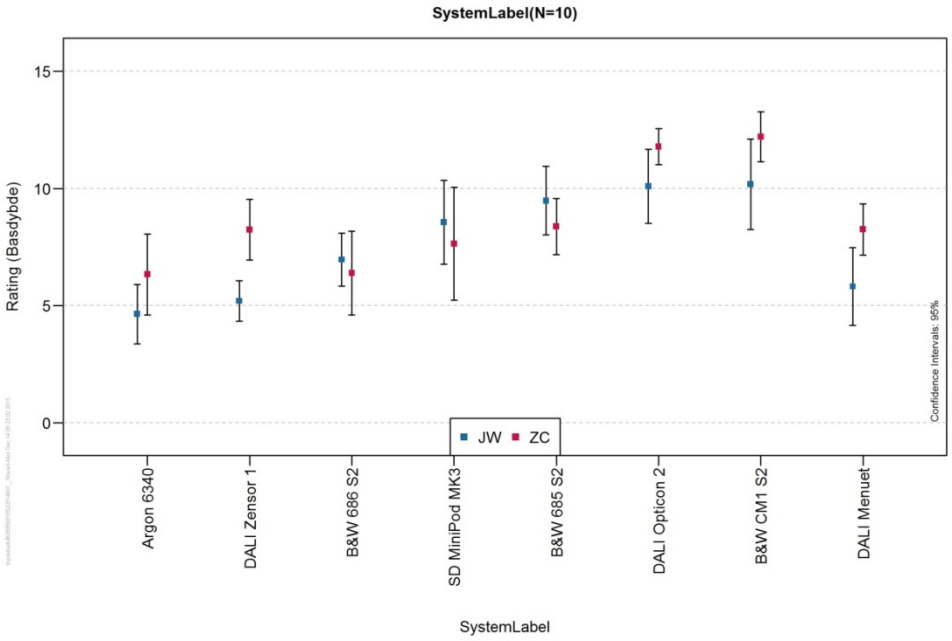
Dernæst vil det primære formål med statistikken ofte være at beregne middelværdier og de tilhørende nøjagtigheder, se fx Figur 6. Her giver 95 % konfidensintervallerne en indikation af, hvilke systemer (i dette tilfælde højttalere) der er signifikant forskellige. Man kan som tommelfingerregel gå ud fra, at hvis konfidensintervallerne ovelapper for to produkter, er det ikke sikkert, at de er signifikant forskellige mht. den pågældende attribut. Den statistiske analyse kan dog også levere en egentlig test af, hvilke produkter, der reelt er forskellige og hvilke, der ikke er det, ubestemthederne i testen taget i betragtning.



Figur 6.
Et eksempel på et resultat fra en lyttetest på højttalere med 10 lyttere mht. attributten 'Basdybde'. De blå punkter angiver middelværdier af alle lytteres bedømmelser af to forskellige musikstykker afspillet på højttalerne. De lodrette intervaller angiver 95 % konfidensintervallerne.

Hvis man vil grave lidt dybere kan en statistisk variansanalyse (ANOVA analyse) vise, hvilke forskelle i resultaterne, der kan henføres til forskelle mellem systemerne (fx forskellige højttalere), forskelle som skyldes programmaterialet eller forskelle, som skyldes lytterne samt eventuelle vekselvirkninger mellem disse faktorer. Et eksempel på det sidste kan ses på Figur 7, hvor især bedømmelsen af DALI Zensor 1 afhænger af det afspillede musikstykke.

[4] DELTA har udviklet et specielt værktøj til dette, 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. Der bruges dog også værktøjer fra programmet PanelCheck, http://www.panelcheck.com



Figur 7.
De samme resultater, som vist på Figur 6, men her opdelt på de to musikstykker JW og ZC, der blev lyttet til.

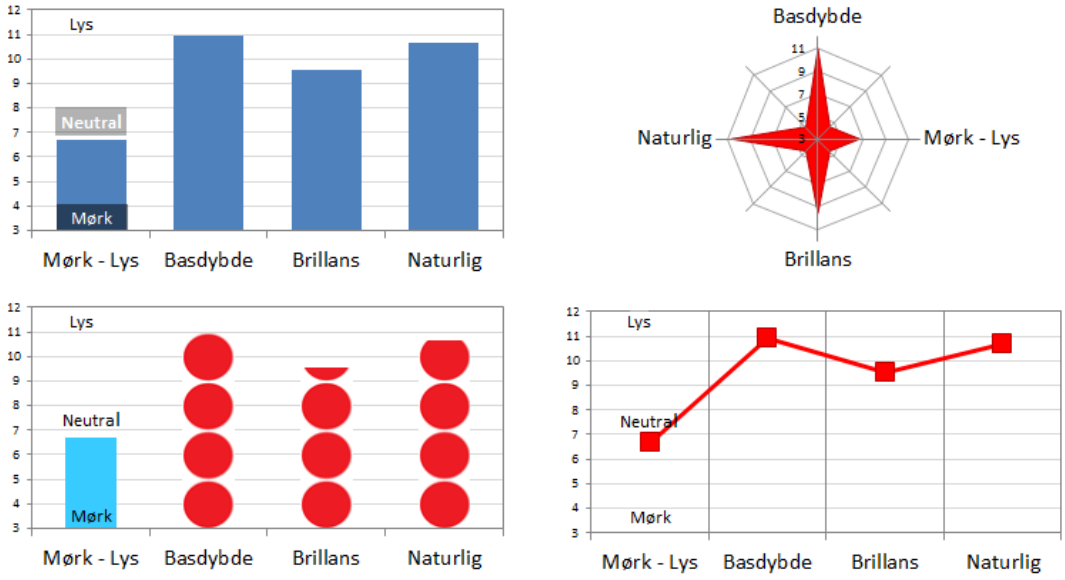
Sensorisk varedeklaration

Der er mange muligheder for at præsentere resultaterne fra sensoriske tests. Eksemplerne i det foregående afsnit er en teknisk præget og nøjagtig information om testens udfald. De henvender sig mest til fagpersoner, der arbejder med fx produktudvikling, markedsføring o.l.

Men resultaterne kan også udnyttes til at give forbrugerne en information om produkternes opfattede karakteristika, som ikke opnås gennem den tekniske varedeklaration. Et par eksempler, der angiver karakteren af øl, er vist nedenfor.

Figur 8.
To eksempler på sensoriske skalaer, der angiver karakteren af øl.

Tilsvarende kan lydprofilerne fra lyttetesten illustreres, og nogle muligheder er vist på Figur 9. For overskuelighedens skyld er det valgt kun at vise 4 attributter.



Figur 9.
Eksempler på mulige fremstillinger af en højttalers lydprofil. Eksemplet er for DALI Opticon 2.

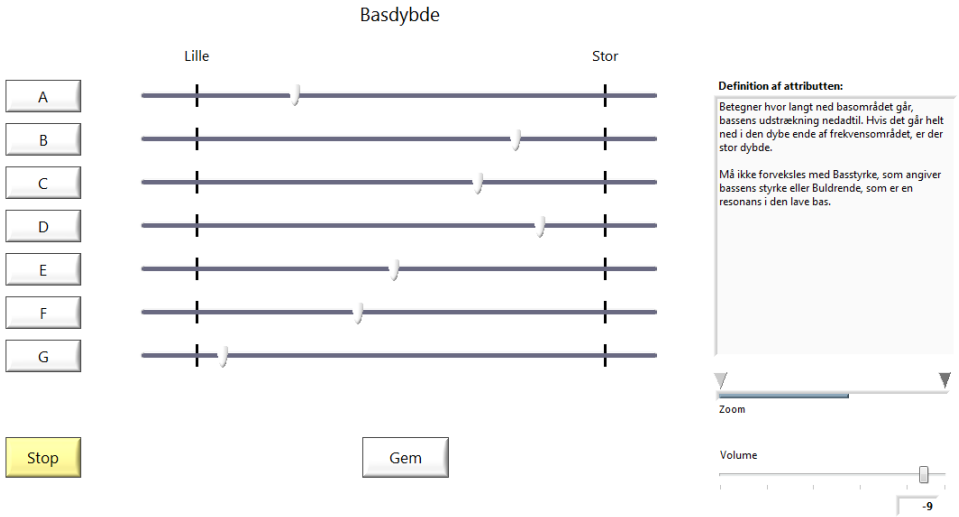
På side 2 er vist et eksempel på, hvorledes lydprofiler fra lyttetesten kan anvendes sammen med de tekniske data. I Appendiks 2 er vist andremuligheder for at illustrere lydprofilerne, og de skulle gerne give et indtryk af, hvilke typer, der bedst egner sig til oversigter.

Eksempel på en lyttetest

De tekniske data angiver på forskellig vis, hvad højttalere kan yde og tåle. Hvis man vil vise, hvordan højttalerne lyder, har man brug resultaterne af en lyttetest. Den her beskrevne lyttetest omfattede 8 stereosæt af kompakthøjttalere i prisklassen 399-3699 kr. pr. stk. fra Hi-Fi Klubben. Til denne lyttetest blev 6 objektive attributter udvalgt, og hvert resultat repræsenterer gennemsnittet af 10 trænede lytteres bedømmelser af højttalernes egenskaber.

Testopstilling

Testen foregik i DELTA's lytterum, som har neutral efterklang, der ikke farver lyden fra højttalerne, se Figur 1. Højttalerne var anbragt på en drejestander, så de højttalere, der blev lyttet på, altid var i den samme position i rummet, se Figur 3. Hvis højttalerne stod på forskellige faste pladser, ville rumakustiske fænomener kunne farve lyden forskelligt, alt efter hvor højttalerne stod.
Under testen vises bedømmelsesskalaerne på et akustisk transparent lærred (Figur 10). Dette skjuler højttalerne, så lytterne ikke kan se, hvad de lytter til.



Figur 10.
Eksempel på brugerfladen under lyttetesten. Lytteren kan vælge, hvilket system der skal lyttes til med knapperne A-G. Hvilke højttalere der "gemmer sig" bag knapperne varierer hen gennem testen. Bedømmelsen afgives ved at flytte skyderne, i dette tilfælde mellem 'Lille' og 'Stor Basdybde'.

Højttalerne blev drevet af NAD C356 BEE forstærkere. Inden testen blev niveauet fra hvert højttalersæt justeret til 70 dB(A) målt i lyttepositionen ved afspilning af lyserød støj. Højttalerne blev 'spillet til' i en time med dette signal.

Der blev lyttet på følgende musikstykker:
- Jennifer Warnes (JW): "Bird on a wire" fra CD'en Famous Blue Raincoat
- Zhao Cong (ZC): Moonlight on Spring River fra DALI CD 3.

Lyttetesten

Bedømmelserne blev foretaget af 10 af DELTA's udvalgte og trænede ekspertlyttere. Det eneste lytterne vidste om højttalerne bag lærredet var, at det var kompakthøjttalere.
Ved en prøvelytning med 8 af lytterne på 4 af højttalerne, blev det inden selve testen valgt, hvilke af lydhjulets (se Appendiks 1) karakteristika (attributter), der bedst beskrev forskellene mellem højttalerne, se Tabel 1.

Blindtesten blev afviklet fra en PC og bedømmelserne blev afgivet på ustrukturerede linjeskalaer fra 'Lidt' til 'Meget' hhv. 'Lille' til 'Stor' for de fem af attributterne og 'Mørk-Neutral-Lys'. Der var en bagvedliggende numerisk skala fra 0 til 15 (med en præcision på 0,1), som lytterne ikke kendte til.
Testen blev gennemført med én lytter ad gangen. Under testen var rækkefølgen af højttalerene og de attributter, der blev bedømt, forskellig for hver person. Der blev lyttet to gange på hvert af de to musikstykker (også i tilfældig rækkefølge) for hver attribut.
Testresultaterne bygger på i alt 1920 bedømmelser (eksklusiv ankerhøjttalere og 'Præference').

Attribut	Definition og beskrivelse
Brillans	Diskantudstrækning. Skala: – *Lidt:* Som at høre musik gennem en dør. Indelukket, sløret eller mat. – *Meget:* Krystalklar gengivelse, stort diskantområde med luftig og åben diskant. Lethed, renhed og klarhed med plads til instrumenter. Tydelighed i de øvre frekvenser uden at være 'Skarp' eller 'Skinger' og uden forvrængning.
Basdybde	Betegner hvor langt ned basområdet går, bassens udstrækning nedadtil. Hvis det går helt ned i den dybe ende af frekvensområdet, er der stor dybde. Må ikke forveksles med 'Basstyrke', som angiver bassens styrke, eller 'Buldrende' som er en resonans i den lave bas.
Punch	Angiver om anslagene på trommer og bas bliver gengivet med slagkraft, næsten som om man kan mærke slaget. Evnen til ubesværet at håndtere store styrkeudsving uden kompression (kompression høres som lydstyrkeudsving, der er mindre end man ville forvente af den opfattede originale lyd).
Mørk-Lys	Betegner balancen mellem bas og diskant. Skala: – *Mørk:* Overvægt af bas. Enten kraftig bas eller svag diskant. – *Neutral:* Bas og diskant høres lige kraftigt, der er balance i gengivelsen. Det gælder også hvis både bas og diskant er lige svage, eller hvis bas og diskant er lige meget for kraftige. Hvis det medfører, at mellemtonen er fremhævet eller tilbagetrukket, bedømmes dette under mellemtonestyrke. – *Lys:* Overvægt af diskant. Enten kraftig diskant eller svag bas. Årsagen til om klangen er mørk eller lys kan udledes af bedømmelserne af bas- og diskantstyrke.
Rumlig præcision	Kan de enkelte instrumenter og stemmer placeres tydeligt og adskilt i det rumlige i lydbillede? Hvor præcist står de enkelte lydkilder i rummet? Hvis de enkelte lydkilder bliver utilsigtet brede eller flyder ud, er præcisionen lav.
Naturlig	Gengivelsen lyder troværdig. Akustiske instrumenter, stemmer og lyde, lyder som i virkeligheden. Lyden svarer til lytterens forventning til originallyden uden nogen klangmæssig eller rumlig farvning eller forvrængning, 'Intet tilføjet – intet mangler'. Lydbilledet står klart i rummet og bringer dig nær den originale lydoplevelse.

Tabel 1.
De af lydhjulets attributter, der blev udvalgt til denne test.

Højttalere

I alt 8 kompakthøjttalere fra Hi-Fi Klubben var med i testen, se Tabel 2.

Fabrikat, Model	Pris kr./stk.	Frekvensgang (- 3 dB)	Størrelse (BxHxD) cm	Brutto rumfang, l
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

Tabel 2.
Oversigt over de højttalere, der var med i testen. De tekniske data stammer fra Hi-Fi Klubbens hjemmeside.

For at skabe sammenlignelighed med kommende tests var desuden tre ankerhøjttalere med i testen: Én der bedømmes højt på alle attributter, én der bedømmes lavt og én imellem de to.

Resultater

I tabellen nedenfor er resultaterne fra lyttetesten vist på skemaform. I Figur 10 og 11 er resultaterne vist grafisk.

Tabel 3.
Hvert af resultaterne i tabellen er en middelværdi af i alt 40 bedømmelser (10 lyttere, 2 gentagelser og 2 musikstykker).

Bedømmelsesskalaerne går fra 0 til 15. Nøjagtigheden er ca. +/- 1 enhed.

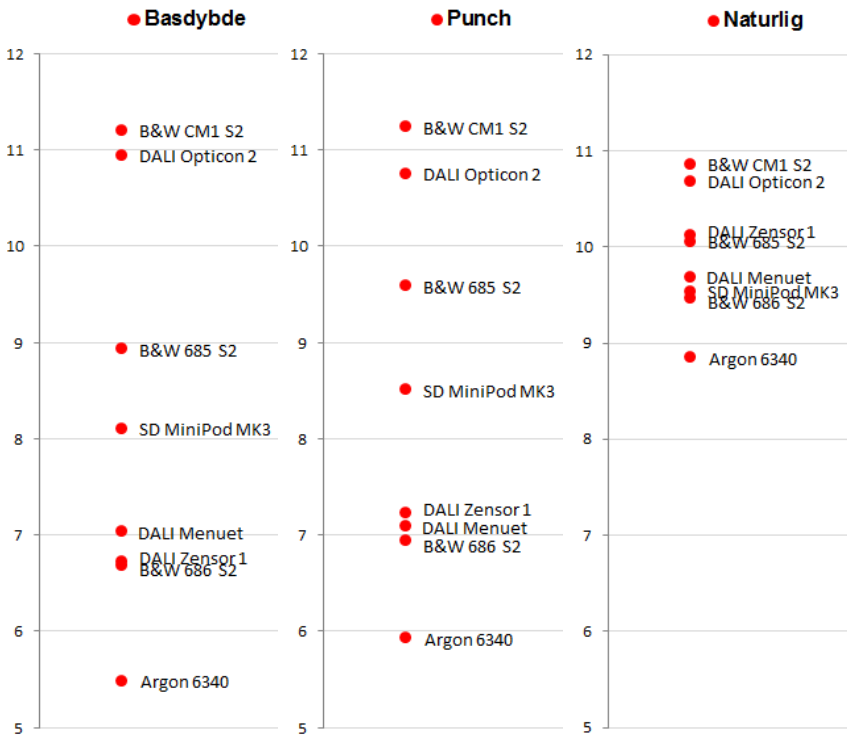
For 'Mørk-Lys' svarer den lave ende til Mørk og den høje ende til 'Lys'. En neutral gengivelse ligger ved 7,5 på 'Mørk-Lys' skalaen.

Systemer	Basdybde	Punch	Naturlig	Rumlig præcision	Brillans	Mørk-Lys
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

Basdybde er sammen med Punch den attribut, som mest markant adskiller systemerne. Argon 6340 har mindst basdybde. Herefter følger en gruppe med DALI Zensor 1, B&W 686 S2 og DALI Menuet. Lidt over igen ligger MiniPod MK3 og B&W 685 S2, og højest ligger DALI Opticon 2 og B&W CM1 S2, som ikke er signifikant forskellige. Det viste sig generelt, at bedømmelserne med musikstykket Zhao Cong (ZC) gav anledning til lidt højere bedømmelse af 'Basdybde' end musikstykket med Jennifer Warnes (JW), og det var mest udpræget for DALI Zensor 1, se Figur 8.

Punch. De samme bemærkninger som ovenfor er også gældende for Punch.

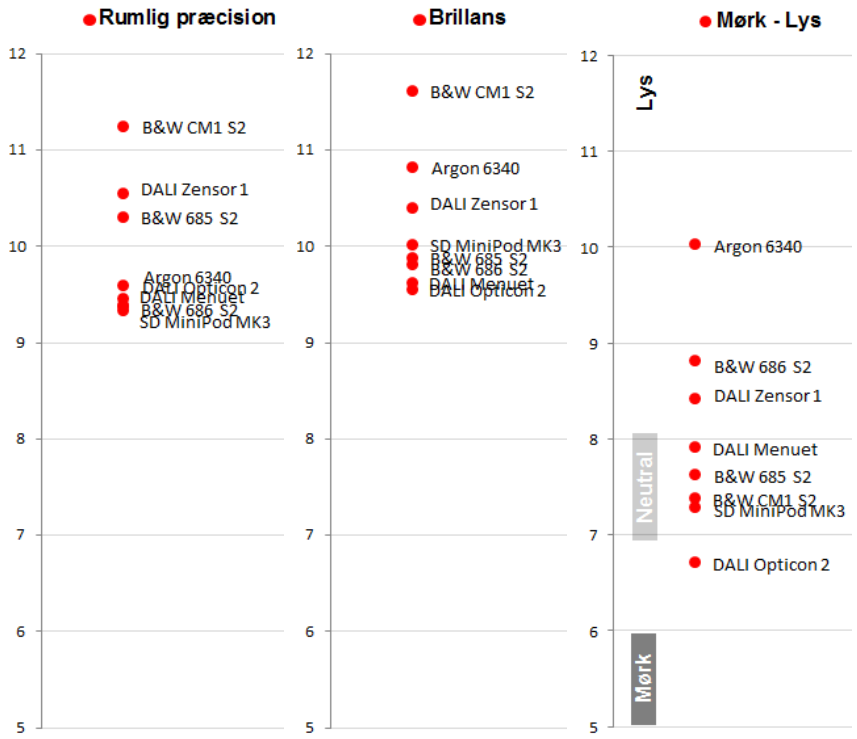
Naturlig. Der er relativt små forskelle mellem systemerne. Specielt bemærkes, at DALI Opticon 2, og B&W CM1 S2 ligger signifikant højere end Argon 6340.



Figur 11.
Resultaterne for 3 af de objektive attributter i blindtesten.

Hvert datapunkt er gennemsnittet af i alt 40 bedømmelser.

Bedømmelsesskalaen går fra 0 til 15. Nøjagtigheden er ca. +/- 1 enhed.



Figur 12.
Resultaterne for 3 af de objektive attributter i blindtesten.

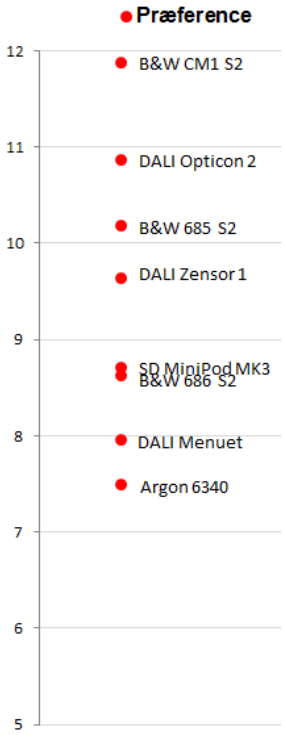
Hvert datapunkt er gennemsnittet af i alt 40 bedømmelser.

Bedømmelsesskalaen går fra 0 til 15. Nøjagtigheden er ca. +/- 1 enhed.

Rumlig præcision: Der er relativt små forskelle mellem systemerne. B&W CM1 S2 ligger højest, men er dog ikke signifikant forskellig fra DALI Zensor 1 og B&W 685 S2.

Brillans: Systemerne er ret ens. Det ses, at det eneste system, som ligger signifikant over gennemsnittet (på 10,2), er B&W CM1 S2. Bedømmelserne på de to musikstykker faldt ret ens ud for hvert system.

Mørk – Lys: Denne attribut er lidt speciel, dels fordi det neutrale punkt ligger midt på skalaen og dels fordi den i vidt omfang kan forudsiges ud fra 'Basdybde' og 'Brillans'. Det ses, at Argon 6340, DALI Zensor 1 og B&W 686 S2 ligger lidt til den lyse side, og de ligger også lavt mht. 'Basdybde'. DALI Opticon 2 ligger en smule under neutral, men ikke signifikant lavere end B&W CM1 S2.



Figur 13.
DELTA's ekspertlytteres præferencer en blindtest, hvor de kun vidste, at de lyttede på kompakthøjttalere.

Præference: Præference er i modsætning til de øvrige attributter en subjektiv attribut. Den anførte 'Præference' gælder således for DELTA's ekspertlyttere, og er ikke nødvendigvis repræsentativ for andre lyttere. Når det er sagt, er der alligevel et par interessante forhold. Lyden fra DALI Opticon 2 og B&W CM1 S2 er de to mest foretrukne. CM1 ligger højest, men forskellen er ikke signifikant. Herefter følger DALI Zensor 1 og B&W 685 S2. Lyden fra Argon 6340 er den mindst foretrukne af de testede højttalere, men den ligger dog stadig midt på skalaen.

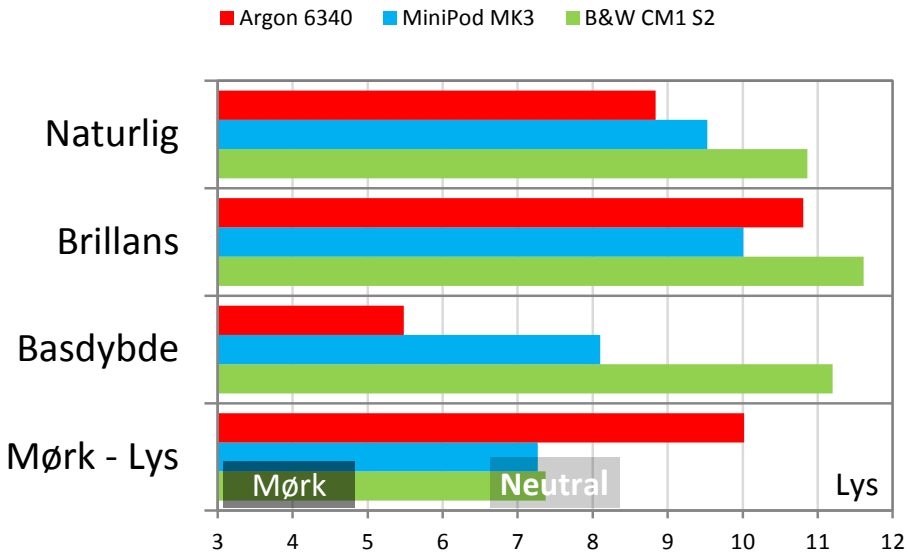
Det er også interessant at bemærke, at der i denne blindtest med få undtagelser er en ret god sammenhæng mellem eksperternes præferencer og prisen på højttalerne, se Figur 20. DALI Zensor 1 udmærker sig ved at ligge pænt over den generelle sammenhæng og DALI Menuet ligger under. Det ses også, at præferencerne til en vis grad hænger sammen med 'Basdybde' og dermed med højttalernes rumfang. Da DALI Menuet er en lille højttaler kommer den således lidt til kort i dette selskab.

Som nævnt er det ikke sikkert, at forbrugerne har samme præferencer som DELTA's lyttepanel i en dobbelt blindtest, hvor de jo heller ikke har set højttalerne. Det er kun forbrugerne selv, der kan afgøre, hvad de foretrækker.

Sammenligninger

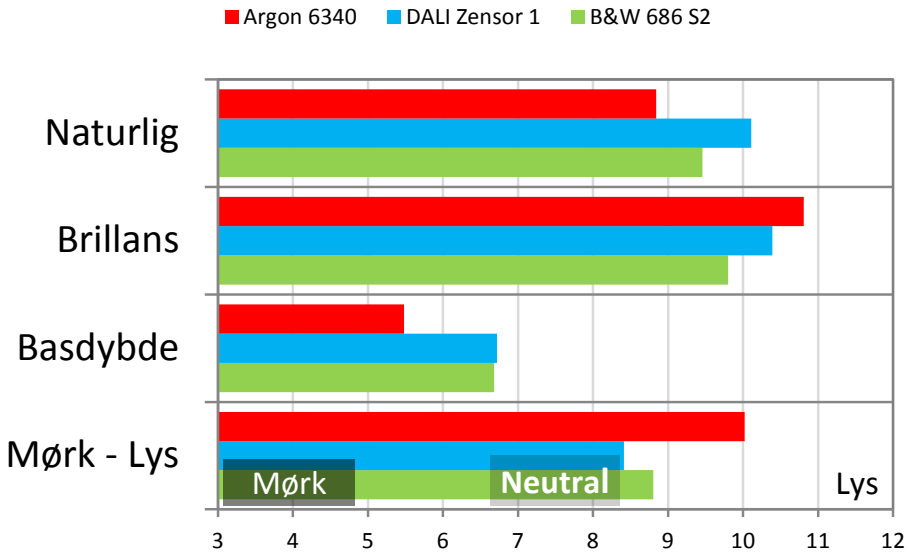
Forskellige former for præsentationer giver mulighed for at sammenligne højttalerne indbyrdes. I det følgende er givet nogle eksempler.

Det ses af Figur 14, at lyttetesten er i stand til at vise signifikante forskelle mellem systemerne.



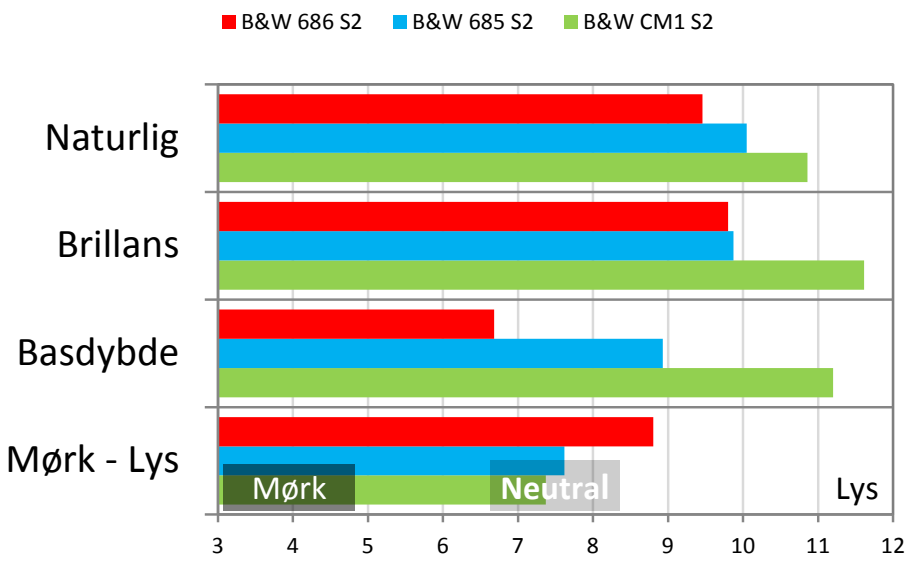
Figur 14.
Sammenligning mellem de systemer, der på de fleste attributter i lyttetesten scorede hhv. lavest, i midten og højest.

I Figur 15 er vist en sammenligning af de tre systemer med de laveste priser: Argon 6340 til 399 kr., DALI ZENSOR 1 til 999 kr. og Bowers & Wilkins 686 S2 til 1.799 kr. Det ses, at man får mere basdybde og en mere neutral klangbalance ('Mørk-Lys') ved de lidt dyrere systemer.



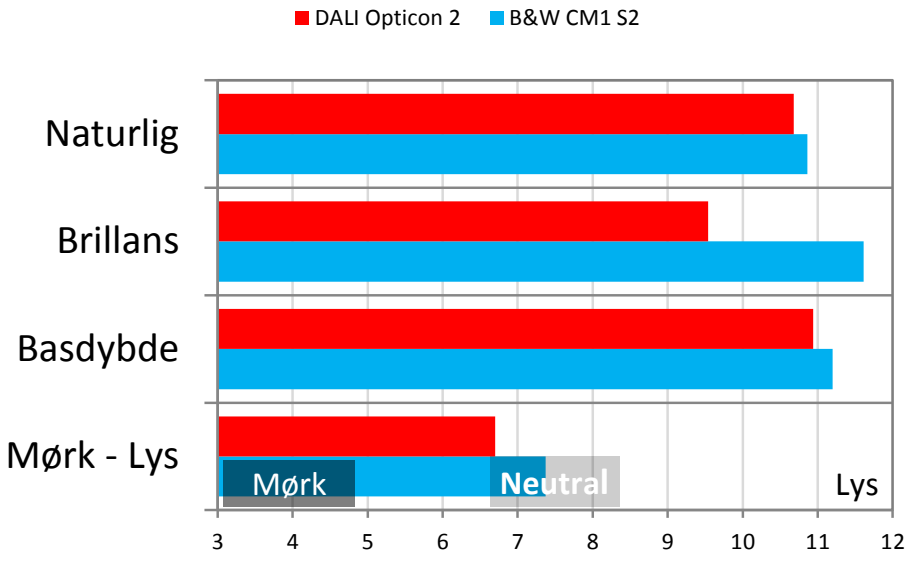
Figur 15.
En sammenligning af de tre billigste systemer.

Af Figur 16 ses, at B&W har formået at fordele højttalernes lyttemæssige kvaliteter svarende til deres priser.



Figur 16.
En sammenligning af testens tre B&W systemer.

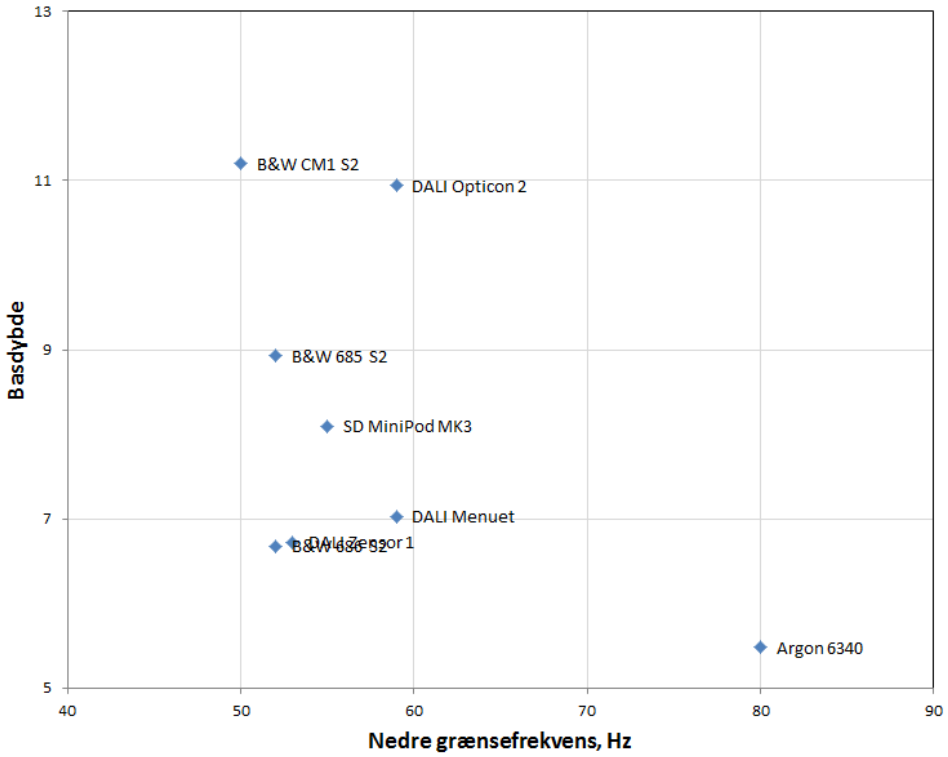
Endelig ses af Figur 17, at forskellene mellem testens to dyreste systemer, DALI OPTICON 2 til 2.999 kr. og Bowers & Wilkins CM1 S2 til 3.299 er meget små, når der ses bort fra 'Brillans'. B&W højttaleren har den højeste 'Brillans'. Hvis man foretrækker en brillant lyd er B&W CM1 S2 et godt bud, men hvis man foretrækker en mindre brilliant og lidt blødere lyd er DALI OPTICON 2 et bedre valg.



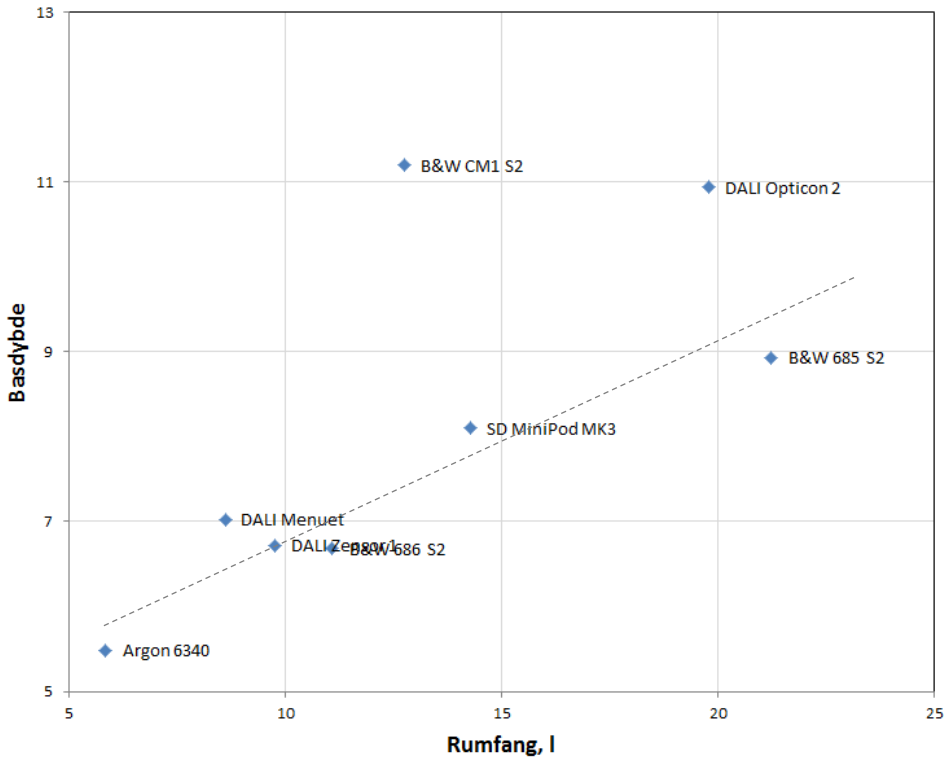
Figur 17.
En sammenligning af 2 af testens dyreste systemer.

Interessante sammenhænge

Figur 18 viser, at der generelt ikke er en sammenhæng mellem den oplevede basdybde og de tekniske data for nedre grænsefrekvens. Fire højttalere, der alle har næsten samme nedre grænsefrekvens på 50-53 Hz, bliver bedømt til at have meget (og signifikant) forskellig 'Basstyrke'. Det samme gælder DALI Menuet og DALI Opticon 2, hvis basdybder bliver bedømt meget forskelligt, og begge er opgivet til at have samme nedre grænsefrekvens på 59 Hz. Selv om det ikke er en ny erkendelse, at man ikke kan bedømme en højttalers lyd ud fra de tekniske data, er det alligevel interessant at se det demonstreret med en veldefineret lyttetest. Det er altså mere fornuftigt at tage udgangspunkt i en lyttemæssig bedømmelse end de tekniske data, hvis man vil vide, hvordan en højttaler lyder.

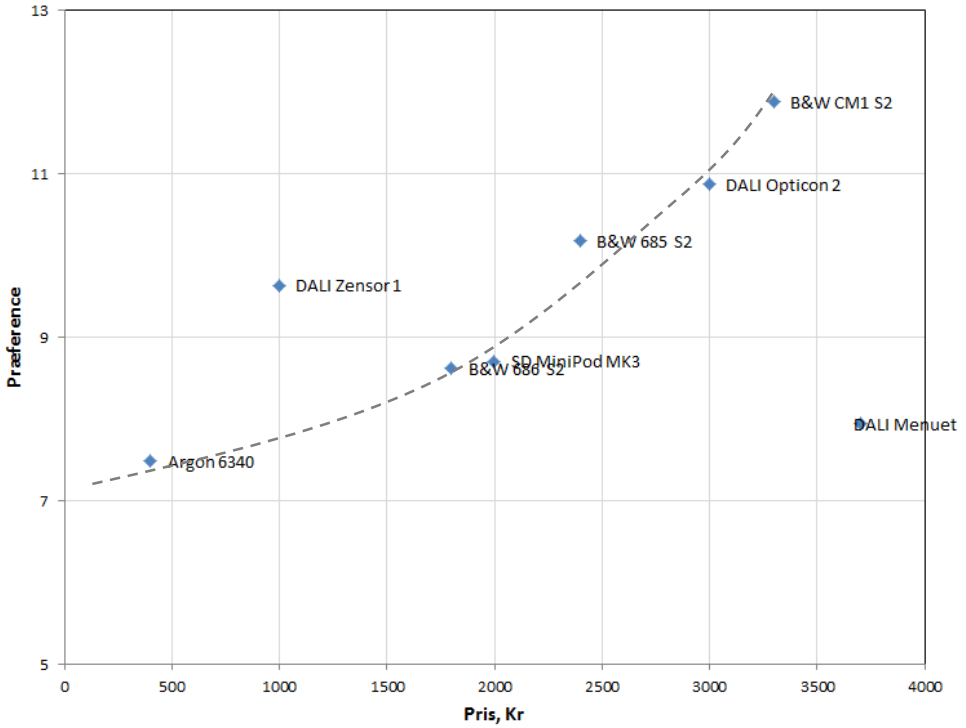


Figur 18.
Sammenstilling af højttalernes tekniske data for nedre grænsefrekvens og testens bedømmelser af 'Basdybde'.



Figur 19.
Sammenstilling af højttalernes brutto-rumfang og testens bedømmelser af 'Basdybde'.

Figur 19 viser, at basdybden generelt stiger med bruttorumfanget, dog har DALI Opticon 2 og B&W CM1 S2 en større basdybde end man skulle forvente ud fra deres rumfang.



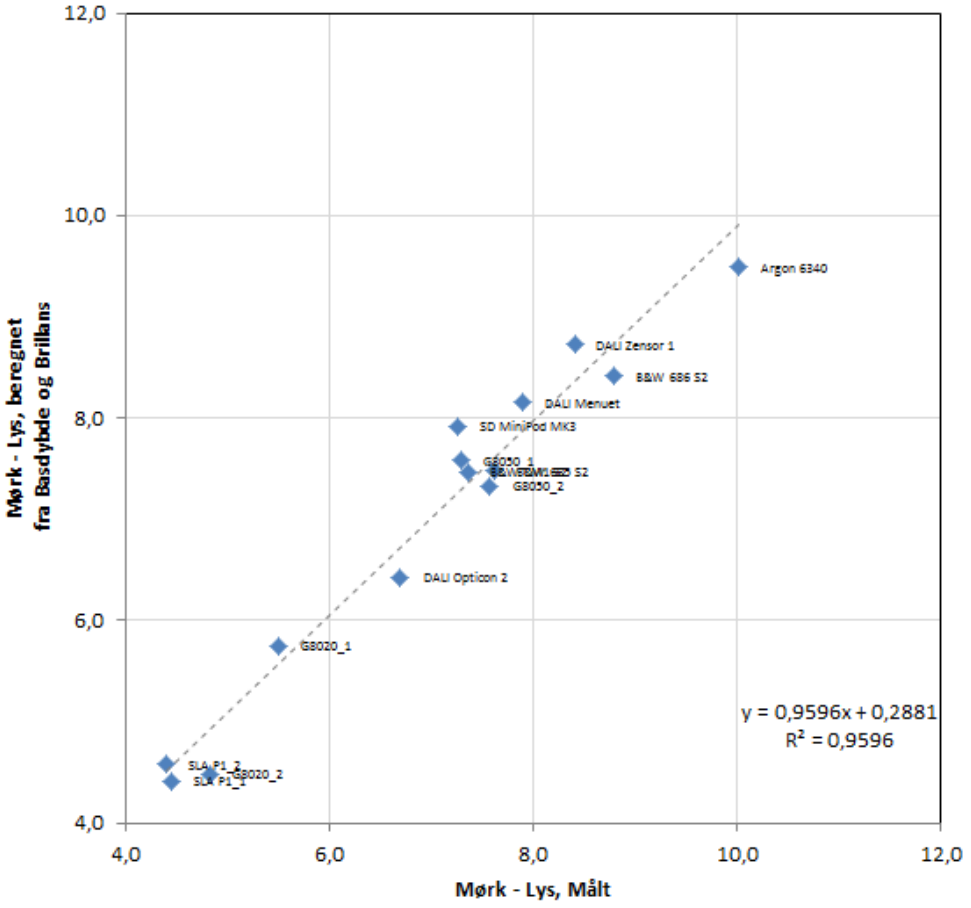
Figur 20.
En sammenstilling højttalernes pris og DELTA's ekspertpanels præference-bedømmelse af lyden.

Figur 20 viser, at der generelt er en sammenhæng mellem pris og præference. Lyden fra DALI Zensor 1 har dog en høj præference i forhold til prisen, mens den meget lille DALI Menuet har en lavere. Som nævnt tidligere er de anførte præferencer et gennemsnit af DELTA's lyttepanels præferencer i en blindtest, og dette kan afvige fra forbrugernes præferencer.

Modeller

Med data fra lyttetesten kan man opstille forskellige modeller. Modellerne gælder for den aktuelle test. Om de har en mere generel gyldighed må selvfølgelig efterprøves ved hjælp af lyttetests med andre højttalere.

Af definitionen på attributten 'Mørk-Lys', se Tabel 1, kan man nemt få den ide, at denne attribut må kunne forudsiges ud fra bedømmelser af bas og diskant. Det er faktisk også tilfældet. Det ses af Figur 21, at der en fin, næsten én-til-én sammenhæng mellem de bedømmelser af Mørk-Lys, der var resultatet af lyttetesten, og de værdier, der er beregnet ud fra en model, der bygger på bedømmelser af 'Basdybde' og 'Brillans'.

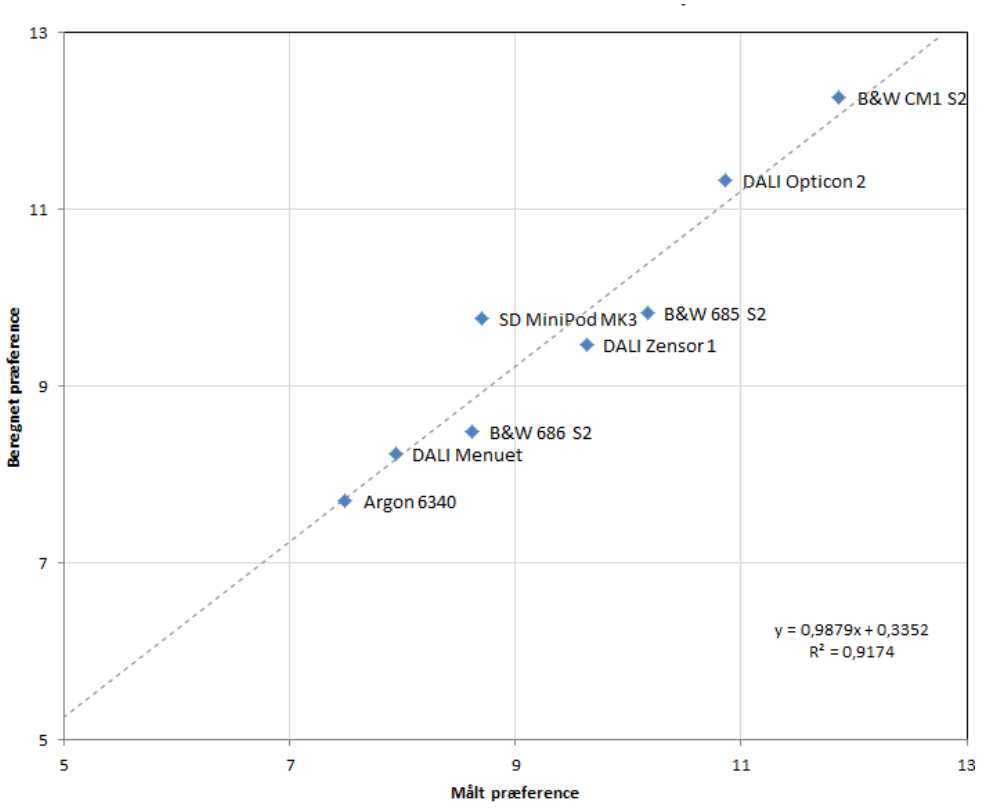


Figur 21.
Resultater fra lyttetesten af bedømmelser af 'Mørk-Lys' (x-aksen) sammenstillet med resultaterne af en model, der beregner 'Mørk-Lys' ud fra 'Basdybde' og 'Brillans' (y-aksen).

Det er også interessant, hvis man kan forudsige forbrugernes præferencer. I den store sammenhæng afhænger forbrugernes præferencer jo af produktet som helhed samt de påvirkninger, de er udsat for gennem salgsmateriale m.m.

Ikke desto mindre er det også interessant, hvis man kan opstille en model for lytternes præferencer alene for lyden af de forskellige højttalere.

Figur 22 viser resultaterne af beregninger udført med en model, der baserer sig på resultaterne af 'Basdybde' og 'Brillans' i første, anden og tredje potens. Det ses, at man med ret god nøjagtighed kan opstille en sammenhæng for 'Præference' ud fra bedømmelserne af de to nævnte attributter for denne test.



Figur 22.
Resultater fra lyttetesten af 'Præference' (x-aksen) sammenstillet med resultaterne af en model, der forudsiger 'Præference' ud fra 'Basdybde' og 'Brillans' (y-aksen).

Konklusion

Notatet har belyst, hvilke ingredienser der skal til for at udføre pålidelige og reproducerbare lyttetests, der giver stabile resultater med en kendt nøjagtighed. Ud over specificerede lytteforhold og højttaler-positioner skal der et antal trænede lyttere til. Bedømmelserne foretages med veldefinerede attributter i en blindtest med randomiserede præsentationer af produkterne. Desuden skal man bruge forskelligt programmateriale og systemerne, der testes, skal indstilles til nøjagtigt samme afspilleniveau under testen.

Hvis man gør det, kan man få en objektiv karakteristik af lyden fra systemerne, som fx kan bruges i forbindelse med produktudvikling eller til forbrugeroplysning som supplement til de tekniske data. Hvis testens resultater skal kunne sammenlignes med andre tests under tilsvarende forhold, skal der også være nogle gennemgående systemer, ankerhøjttalere, med i testen.

Anvendelsen af disse principper er demonstreret i en test, der var udført som en dobbelt blindtest, hvor de 10 lytterne kun vidste, at de lyttede på kompakthøjttalere. Hvert af testresultaterne var en middelværdi baseret på i alt 40 bedømmelser (10 lyttere, 2 gentagelser og 2 musikstykker) for hver attribut og for hver højttaler. Attributterne blev bedømt på skalaer fra 0-15, og nøjagtigheden i testen var ca. +/- 1 enhed. Nøjagtigheden kan øges ved at øge antallet af trænede lyttere.

Lyttetesten gav supplerende oplysninger i forhold til de tekniske data. Der var for eksempel en tydelig forskel på den oplevede basdybde og den tekniske angivelse af systemernes nedre grænsefrekvens. Resultaterne viste også klare og signifikante lyttemæssige forskelle mellem de billigste og dyreste systemer.

Det er illustreret, hvordan resultaterne kan præsenteres og sammenlignes på forskellig vis, som kan give forbrugerne vejledning om højttalernes lyttemæssige karakter, en sensorisk varedeklaration.

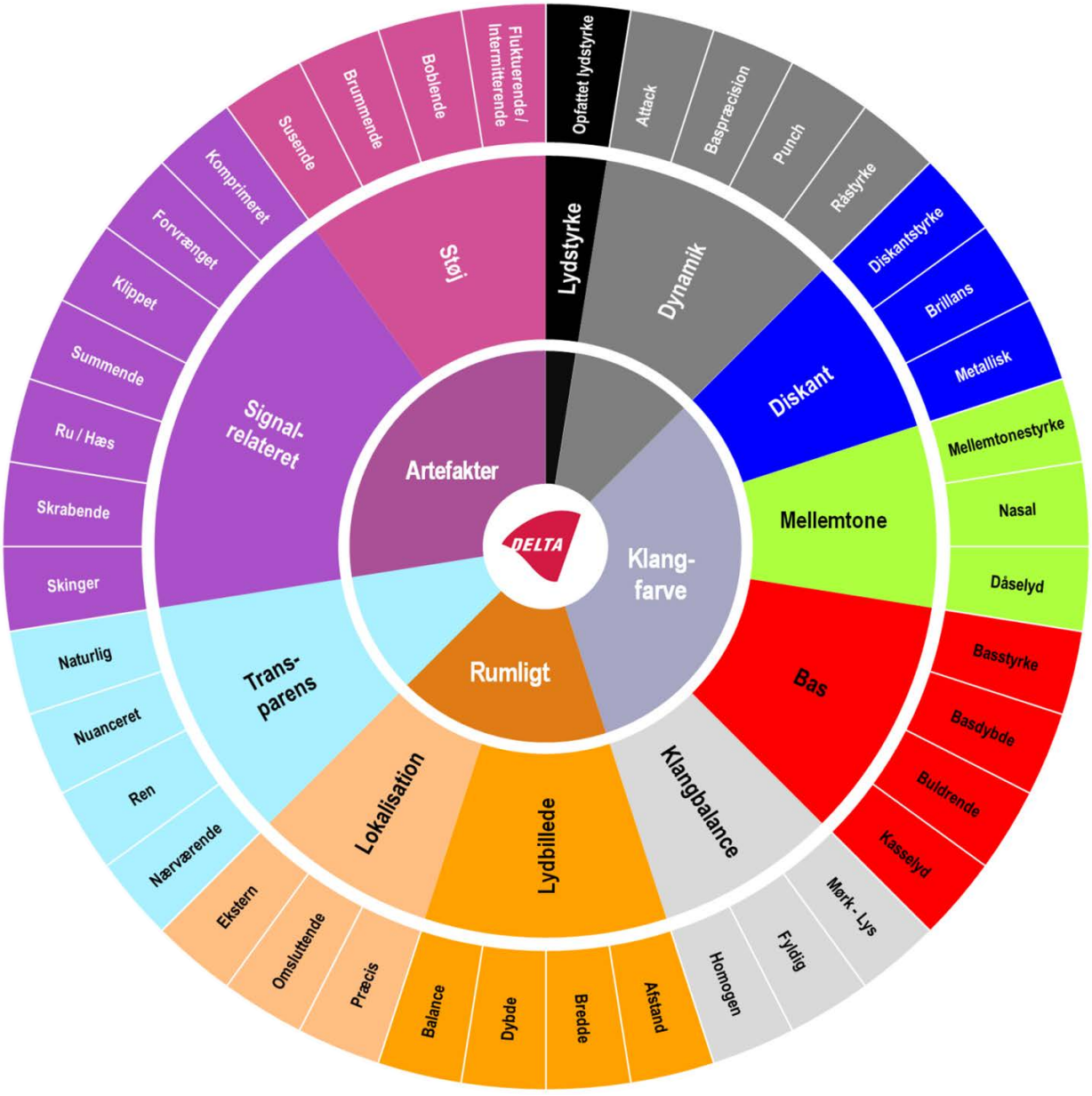
En sammenligning mellem pris og gennemsnittet af DELTA's ekspertlytteres præferencevurderinger af lyden viser, at der med enkelte undtagelser er en sammenhæng mellem pris og præference. Man skal dog huske på, at 'Præference' er en subjektiv attribut i modsætning til de øvrige attributter, og at forbrugernes præferencer ikke nødvendigvis er de samme som DELTA's ekspertlytteres præferencer. Desuden er der for forbrugerne ofte også andre produktegenskaber end lydkarakteren i spil.

Det kan konkluderes, at tekniske målinger alene ikke giver fyldestgørende oplysninger om lydkarakteren af audioprodukter. Ved at supplere de tekniske data med resultater af objektive lyttetests er det muligt at få en pålidelig karakteristik af den opfattede lyd.

Læs mere, besøg:
senselab.madebydelta.com.
Her finder du baggrundsartikler og rapporter.

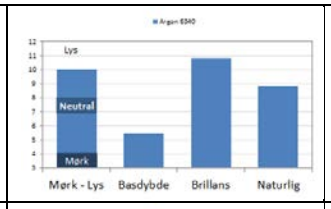
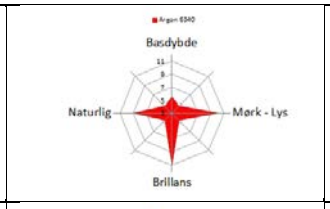
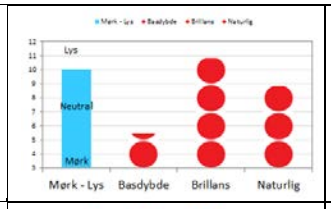
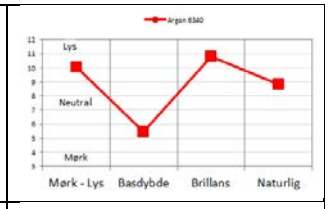
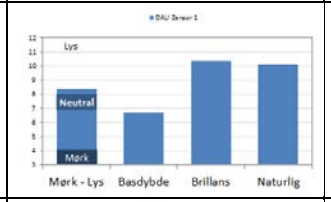
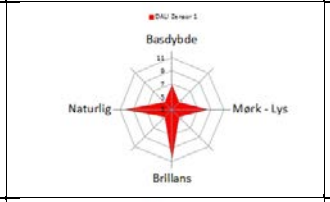
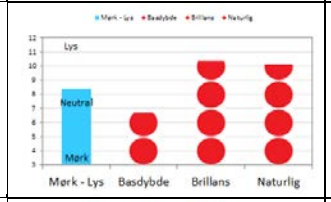
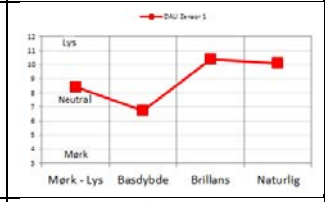
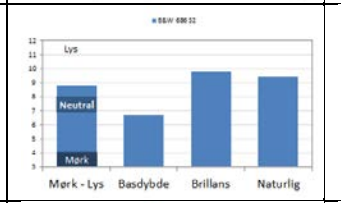
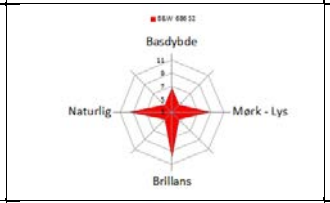
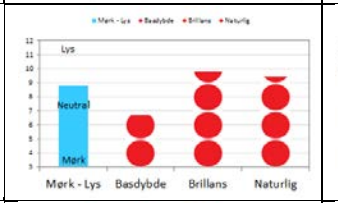
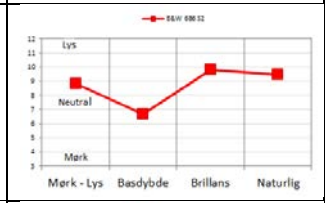
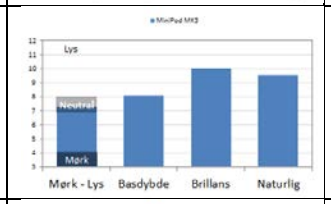
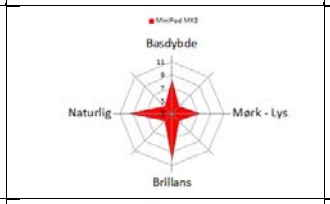
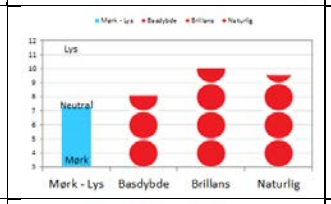
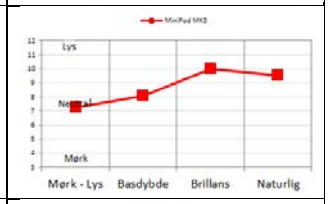
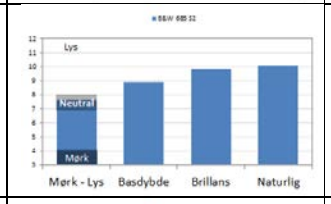
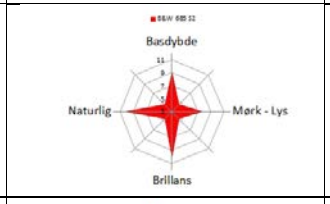
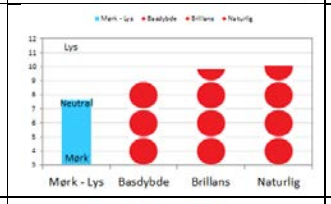
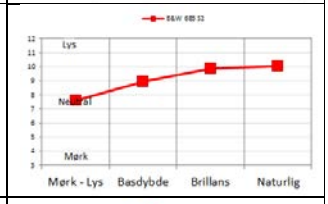
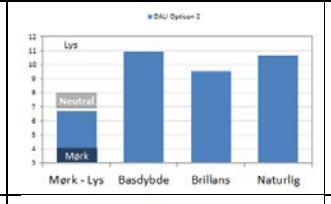
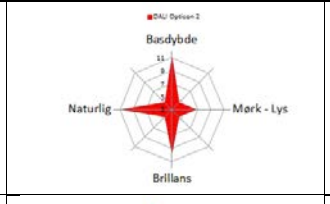
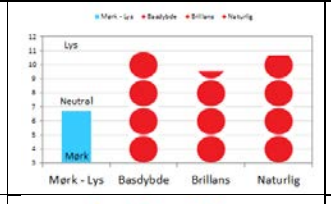
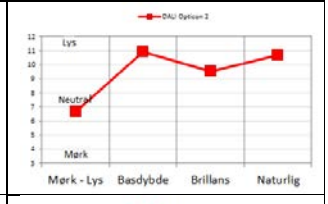
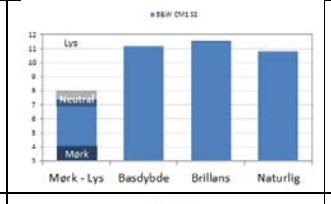
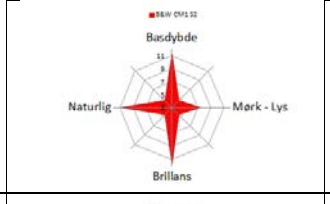
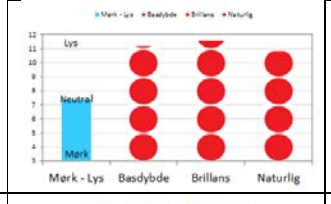
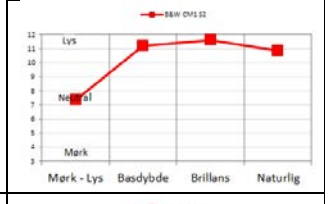
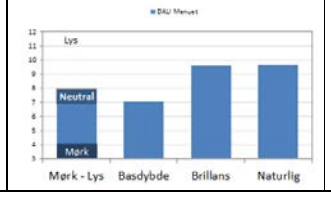
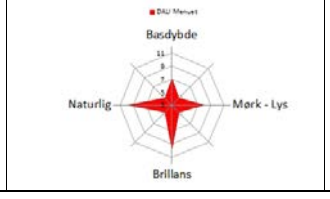
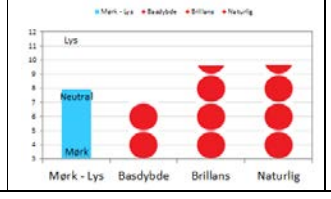
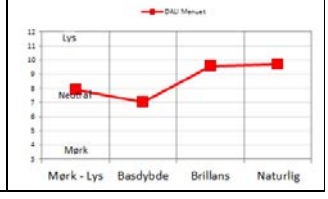
Appendiks 1 - Lydhjulet

Ved lyttetest med reproduceret lyd (audio og Hi-Fi) vælges attributterne fra nedenstående lydhjul. Alle attributter er nærmere defineret, så lytterne har samme forståelse for deres betydning, hvilket er en af forudsætningerne for objektive bedømmelser.



Lydhjulet med attributter for audio. For yderligere information og definition af attributterne, se TEK'notat nr. 7 på http://senselab.madebydelta.com/about/publications/

Appendiks 2 – Alternative for lydprofiler

Argon 6340				
DALI Zensor 1				
B&W 686 S2				
MiniPod MK3				
B&W 685 S2				
DALI Opticon 2				
B&W CM1 S2				
DALI Menuet				

Om TEK’notat

Hvis du vil vide mere om det emne, TEK’notatet beskriver, kan du besøge www.senselab.madebydelta.com hvor du kan finde baggrundsrapporter og artikler.

DELTA udgiver regelmæssigt rapporter i TEK’notat-serien for at kommunikere den nyeste internationale viden indenfor vores fagområder. Formålet er at understøtte en fremrykning af tidspunktet, hvor nye teknologiske landvindinger giver et forretningsmæssigt afkast til danske virksomheder.

Om DELTA

We help ideas meet the real world

DELTA er en selvejende højteknologisk virksomhed, der stiller viden til rådighed for alle. Som vores kunders strategiske partner sikrer vi optimal anvendelse af avanceret teknologi. Vi udvikler, tester, certificerer og rådgiver i alle faser af vores kunders produktudvikling. Vi har været teknologiens stifindere siden 1941, og det er vores vision, at Danmark skal være det bedste sted at udføre højteknologisk produktudvikling.