



Rapport

Demonstratorprojekt støttet af Forsknings- og Innovationsstyrelsen

System Audio - Q113

Lyttetest med hi-fi-højttalere

SenseLab 016/13

Projekt nr.: FR21300-11

Side 1 of 30

2 annekser

14. november 2013

DELTA

Venlighedsvej 4

2970 Hørsholm

Danmark

Tlf. +45 72 19 40 00

Fax +45 72 19 40 01

www.delta.dk

CVR nr. 12275110

Denne side er blank



Demonstratorprojekt støttet af Forsknings- og Innovationsstyrelsen

System Audio - Q113

Lyttetest med hi-fi-højttalere

Journal nr.
SenseLab 016/03

Sagsnr.
FR21300-11

Vores ref.
SVL/THP/CVO/ilk

Resume

DELTA SenseLab har udført en systematisk blindtest af 3 højttalere (kompakt-, gulv- og væghøjttaler) som er udviklet af System Audio i fællesskab med fagbladet Ingeniørens læsere samt en tidligere model SA2K. Testen var led i en metodeudvikling og skal ikke ses som et færdigt testparadigme. Testen blev udført i SenseLabs lytterum med lav efterklangstid og der blev brugt tre udvalgte musikeksempler til testen.

Højttalerne blev testet efter en metode ”Komparativ lytning med reference”, hvor lytterne sammenlignede hver af testhøjttalerne med en referencehøjttaler. Denne metode har forskellige fordele, men er måske ikke så følsom for forskelle mellem testemnerne, som en direkte sammenligning af højttalerne.

I alt 25 lyttere deltog i testen, 15 fra SenseLabs lyttepanel og 10 interesserede lyttere fra Ingeniørens læserskare. Lytterne skulle angive deres præference og bedømme 7 forskellige attributter. Resultaterne fra hele gruppen indeholdt en del ”statistisk støj”. Efter en kvalifikation af lytterne med henblik på deres evne til at skelne mellem stimuli, til at reproducere deres bedømmelser og til at være i overensstemmelse med resten af panelet blev 11 lyttere udvalgt og det var nu muligt at se tydelige forskelle på resultaterne af de testede højttalere.

Baseret på bedømmelserne af de 7 attributter blev højttalerne karakteriseret i profildiagrammer, så deres egenskaber let kan sammenlignes. Lydkarakteristikken for hhv. kompakhøjttaleren og væghøjttaleren skilte sig ud, medens SA2K og gulvhøjttaleren ikke blev vurderet signifikant forskellige.

Præferencetesten viste, at kompakhøjttaleren blev foretrukket frem for væghøjttaleren. Gennemsnitspræferencen af kompakhøjttaleren var den højeste, men den var ikke signifikant bedre end for SA2K og for gulvhøjttaleren, ligesom væghøjttaleren ikke er signifikant mindre foretrukket end de to sidstnævnte, selv om den har den laveste middelværdi. Der var ikke signifikante forskelle på præferencerne for højttalerne for forskellige musikeksempler.

Alt i alt må det konkluderes, at selv om vi her har afprøvet en ny metodik med indirekte sammenligning, har vi påvist signifikante forskelle mellem højttalerne. Når resultaterne er opnået ved en blindtest, kan man stole på, at det virkelig er lydkarakter og lyd kvalitet, som har gjort udslaget. Den systematiske forsøgsteknik har også gjort det muligt at give information om, hvilke forskelle der reelt er signifikante.

DELTA, 14. november 2013



Christer P. Volk
SenseLab



Torben Holm Pedersen
SenseLab



Indhold

1. Baggrund.....	5
2. Formål	5
3. Generelt om lyttetest	6
3.1 Krav til lyttetest	6
3.2 Valg af testmetode	7
4. Testemner.....	8
5. Testmetode	8
5.1 Test paradigme.....	8
5.2 Eksperimentelt design.....	11
6. Testopstilling.....	11
6.1 Faciliteter og udstyr	11
7. Attributter.....	12
8. Lydeksempler og kalibrering af lydstyrke.....	14
8.1 Lydeksempler.....	14
8.2 Kalibrering af lydstyrke	14
9. Instruktion og træning af assesorerne.....	15
10. Valg af assessorer	15
11. Resultater	20
11.1 Profiler af lydkarakteristika	20
11.2 Præference	22
12. Konklusion	24
13. Anneks A - Faciliteter og udstyr	26
13.1 Lytterum	26
13.2 Udstyr	26
13.3 Opstilling	27
14. Anneks B - Instruktion til lytterne	28

1. Baggrund

Ole Witthøft, grundlægger og udviklingschef af System Audio i Roskilde, har siden primo februar 2012 blogget om højttalerprojektet Q113 på fagbladet Ingeniørens webside (<http://ing.dk/blogs/hojttaler-q113>). Ideen med projektet var at udvikle en højttaler i fællesskab med frivillige blandt læzerskaren af bloggen. Bloggen havde det formål at udvikle en efterfølger til kompakthøjttaleren SA2K (Master edition) fra år 2000. Ultimo april 2013 var der tre konkurrerende prototyper klar til at blive bedømt¹ af bloggens læsere til et åbent hus arrangement hos System Audio. De tre prototyper er:

- En kompakt højttaler, Q113 Evolution
- En væghøjttaler, Q113 Revolution Wall
- En gulvhøjttaler, Q113 Revolution Floor

Højttalerene blev her sammenlignet ved en lyttetest, hvor de synlige prototyper blev bedømt af deltagerne i forhold til SA2K højttaleren. DELTA tilbød derefter at udføre en systematisk blindtest af højttalerne efter mere statistisk anerkendte metoder.

DELTA har som Godkendt Teknologisk Serviceinstitut (GTS institut) fået bevilget midler fra Videnskabsministeriet til etablering af et Competence-center for reproduceret lyd. Det er tanken, at centret skal kunne tilbyde traditionelle og nye elektroakustiske målinger og sensoriske tests, dvs. objektive (perceptuelle) lyttetest med trænede lyttere og præference-målinger med forbrugere (affektive test). DELTA har som et demonstratorprojekt i denne udvikling udført en blindtest af højttalernes karakteristika, hvor de vigtigste hørbare forskelle på højttalerne blev bedømt samt en præferencetest. De udførte tests er et led i en metodeudvikling og skal ikke ses som et færdigt testparadigme.

Rammerne for testen er fastlagt i samarbejde med System Audio

2. Formål

Formålet med testen var at afprøve et nyt testparadigme for blindtest af højttalere, med henblik på at kunne sammenligne med resultaterne fra lyttetesten fra åbent hus arrangementet (spørgeskemaundersøgelse efter guidet lytning på de synlige og navngivne højttalere).

¹ Detaljerne i metoden til evaluering kan læses i blogindlægget fra d. 18. april 2013: <http://ing.dk/blog/kom-til-aabent-hus-med-hoejttaler-q113-157984>.

3. Generelt om lyttetest

3.1 Krav til lyttetest

Når man betragter testen som led i en metodeudvikling, er der tale om følgende problemstilling:

En kvalificeret objektiv lyttetest til karakterisering af lyden kræver seks forskellige ingredienser og udføres med trænede og kvalificerede lyttere:

1) **Et lytterum**

DELTA SenseLabs lytterum opfylder standarderne EBU Tech 3276 og ITU-R BS.1116-1 til multikanals højttaler lyttetest. Det betyder bl.a. lav baggrundsstøj og lav efterklangtid. Se nærmere beskrivelse i afsnit 13.1 Anneks A.

Konklusion: Rummet opfylder de krav, man kan stille til en kvalificeret lyttetest.

2) **Et lyttepanel**

SenseLab har et trænet panel, der bruges til lyttetesten på Q113. Det består af 30 udvalgte og trænede lyttere, hvis performance vi overvåger løbende. Der er bl.a. akustikstuderende, tonemestre, musikere og andre, der arbejder professionelt med lyd, men også "almindelige mennesker" med gode ører. Denne test udføres med 15 af disse og vi supplerer med 10 interesserede lyttere fra Ingeniørens læserskare. Vi må antage, at disse har træning i lytning gennem deres interesse, men det kan vi ikke forudsætte.

Konklusion: Vi antager, at vi har generelt kvalificerede lyttere, men de er ikke trænet specifikt i dette domæne, de anvendte attributter og den anvendte metodik. Det er derfor vigtigt at vurdere resultaterne med henblik på at sikre kvalificerede bedømmelser.

3) **Udstyr og software til blindtest**

Vores lyttetest afvikles automatisk med software, som præsenterer lyd-lydeksempler i tilfældig orden til alle lyttere. Lytternes bedømmelser foretages på en brugerflade, som i dette tilfælde projekteres op på et lærred fra en projektor uden for rummet. Herved undgås akustiske refleksioner fra en dataskærm fysisk placeret i rummet.

Konklusion: Metodikken til afvikling af en randomiseret blindtest er i orden, men en ny metode med en referencehøjttaler (se afsnit 5.1) er ikke afprøvet før.

4) **Egnede musikeksempler**

Lydeksemplerne skal udfordre de karakteristika/attributter der bedømmes. Man kan jo f.eks. ikke bedømme basgengivelse af et musikudsnit med en piccolofløjte osv.

Konklusion: Vi er nødt til at se på musikeksemplerne som en variabel, der har indflydelse på resultatet.

5) **Relevante, objektive og veldefinerede attributter**

Højttalernes karakteristika kan bedømmes i forskellige dimensioner, der findes mange måder at beskrive lyden på. SenseLab er ved at udvikle et generelt sæt af attributter til højttalertest. Arbejdet er ikke færdigt, så vi har valgt attributter ud fra



vores erfaring og prøvelytninger snarere end fra et veldokumenteret sæt af attributter.

Konklusion: Det er ikke sikkert, at de valgte attributter udtømmende beskriver forskellene mellem højttalerne og det er ikke sikret, at alle attributter er relevante.

6) **Statistisk viden og værktøjer**

Den statistiske analyse sikrer adskillelse af de mange faktorer i lyttetesten, således at indflydelsen af højttalere, musikeksempler, lyttere og rækkefølge kan vurderes separat. Således sikres at grundlaget og signifikansen for de konklusioner, som udledes fra testen, er på plads og kan retfærdiggøres.

Konklusion: De metoder, der tages i anvendelse, skal kunne afsløre, hvilke data der er valide og signifikante.

3.2 **Valg af testmetode**

Når en lyttetest skal udføres, skal der træffes en række valg, som skal sikre et passende kompromis mellem på den ene side at give relevante, repræsentative og signifikante resultater og på den anden side sikre, at testen ikke bliver en uoverskuelig og alt for tidskrævende affære.

Valget af testmetode til sammenligning af højttalere blev baseret på paradigmerne beskrevet i DELTA's interne dokument om lyttetestprocedurer², afsnit 9: Listening test paradigms. Metoderne fra dette dokument er:

- Komparativ lytning
 - Parvis sammenligning
 - Komparativ lytning med reference
- Intensiv monadiske lytning

Kombinationerne af de mange variable i en test kan let få den til at svulme op. Det gælder især for metoder med parvis sammenligning. Ideelt set ville vi gerne, for hver af de 4 højttalere, teste f.eks. fem programmaterialer (f.eks. 3 forskellige musikgenrer samt mandlig og kvindelig tale) tre forskellige lydstyrker (baggrunds niveau, mellemstyrke og høj lydstyrke) to forskellige placeringer (ved væg og frit i rummet). Endvidere kunne man nemt finde ti karakteristika eller attributter, man ønskede bedømt. I en såkaldt fuld-faktorielttest, vil det betyde, at hver lytter skulle foretage $4 \times 5 \times 3 \times 2 \times 10$ bedømmelser, dvs. i alt 1200 bedømmelser, som hver typisk tager et halvt minut, dvs. ca. 10 timers lytning pr. person. Hvis man yderligere ønskede at sammenligne alle 12 (inkl. rækkefølge) kombinationer af to højttalere, ville det blive helt uoverskueligt.

² Torben Holm Pedersen: Listening test procedures. På daværende tidspunkt benyttet i ver. 4.

For at testen ikke skal blive for omfattende, har vi fravalgt parvis sammenligning, som formentlig er den testtype, der giver den bedste opløsning og derved kan afsløre de mindste forskelle mellem højttalerne.

En anden mulighed er det man kan kalde ”intensiv monadisk lytning”, hvor man lytter intensivt til hver højttaler i f.eks. en halv time og bedømmer hver af de valgte attributter uden sammenligning med andre højttalere. Det kræver dog meget af lyttere og programmateriale og giver formentlig mindre opløsning end den parvise sammenligning.

Da en sammenligning af de forskellige Q113 højttalere er relevant, har vi valgt en mellemting mellem de nævnte to testtyper nemlig komparativ lytning med en reference.

Her sammenlignes hvert stereosæt (placeret optimalt i forhold til lyttepositionen og rettet mod denne i henhold til standarden ITU-R BS.1116-1) med et sæt referencehøjttalere med en fast placering i rummet. Referencehøjttalerne virker som ”anker” for bedømmelserne og udgår fra bedømmelserne i sidste ende.

4. Testemner

I testen indgik et stereosæt af følgende fire højttalere:

- A. kompakhøjttaler SA2K (Master edition) fra år 2000
- B. kompakhøjttaler, Q113 Evolution, prototype
- C. Væghøjttaler, Q113 Revolution Wall, prototype
- D. Gulvhøjttaler, Q113 Revolution Floor, prototype

5. Testmetode

5.1 Test paradigme

Der udføres en profil- eller attributtest til karakterisering af lyden fra hver højttalertype samt en præferencetest. Begge tests udføres som blindtests, hvor hvert stereosæt på skift sammenlignes med et sæt referencehøjttalere med en fast placering i rummet. Referencehøjttalerne virker som ”anker” for bedømmelserne. Der ligger ikke i ordet ”reference”, at disse højttalere er noget specielt eller er særligt gode i forhold til testhøjttalerne. Referencehøjttaleren har en jævn frekvensgang i et frekvensområde, der matcher testemnerne og udgår af bedømmelserne i sidste ende. Optimalt set skal referencen helst ikke afvige for meget fra testemnerne mht. de valgte attributter. Dette vil medvirke til at sikre fornuftig skalabrug, hvor vurderingerne ikke klumper sammen i den ene ende af skalaen på grund af for store forskelle.

Der lyttes på skift og efter lytternes eget valg på referencehøjttalerne og testemnerne. Omskiftningen mellem reference- og testhøjttaler sker ”on the fly” uden andre hørbare effekter end de forskelle, som højttalerne giver anledning til (fade-out + fade-in sekvens



<100ms). Forskellen mellem referencehøjtaleren og testhøjtaleren for de aktuelle attributter angives som svar. Denne metodik forhindrer, at resultatet påvirkes af rumakustiske forhold sammenlignet med f.eks. parvis sammenligning, hvis testhøjtalere med forskellig placering i rummet sammenlignes direkte.

Da referencehøjtalere og testhøjtalerne under hele testen har samme placering i rummet, udgår også rummets indflydelse på de relative resultater. Dette paradigme vælges desuden også af praktiske hensyn, da metoden kræver få flytninger af højtalere og ikke er så tidskrævende som alternativerne.

Højtalerne Genelec 8050A bruges som reference, og SA2K samt de tre Q113 prototyper sammenlignes under testen på skift med denne reference. Da forsøgslederen manuelt skal flytte en ny testhøjtaler hen i testpositionen mellem hver lytterunde, er dette ikke en dobbelt-blindtest.

En ulempe ved denne metodik er, at testemnerne ikke sammenlignes direkte, hvilke sandsynligvis gør testen mindre følsom overfor små forskelle testemnerne imellem.

5.1.1 Præferencetest – subjektiv vurdering af lydkvaliteten

Indledningsvis udførte lytterne en præferencebedømmelse. Denne blev lavet ved hjælp af et af DELTA udarbejdet Labview program³, hvori testen afvikles som en separat test inden attributtesten.

5.1.2 Profilerings – objektiv karakterisering af forskelle i lydbilledet

Der blev udvalgt et antal attributter, som skønnes at kunne karakterisere de mest udprægede forskelle mellem højtalerne (se afsnit 7). For hver højtaler blev det bedømt, i hvor høj grad denne besad den pågældende attribut sammenlignet med referencehøjtaleren. Det blev valgt at benytte *Semantic differential* teknikken⁴ til evalueringerne. Her vurderes alle de udvalgte attributter, se afsnit 7 (i forhold til referencen) på samme skærm billede og assessorerne vælger selv rækkefølgen af vurderingerne.

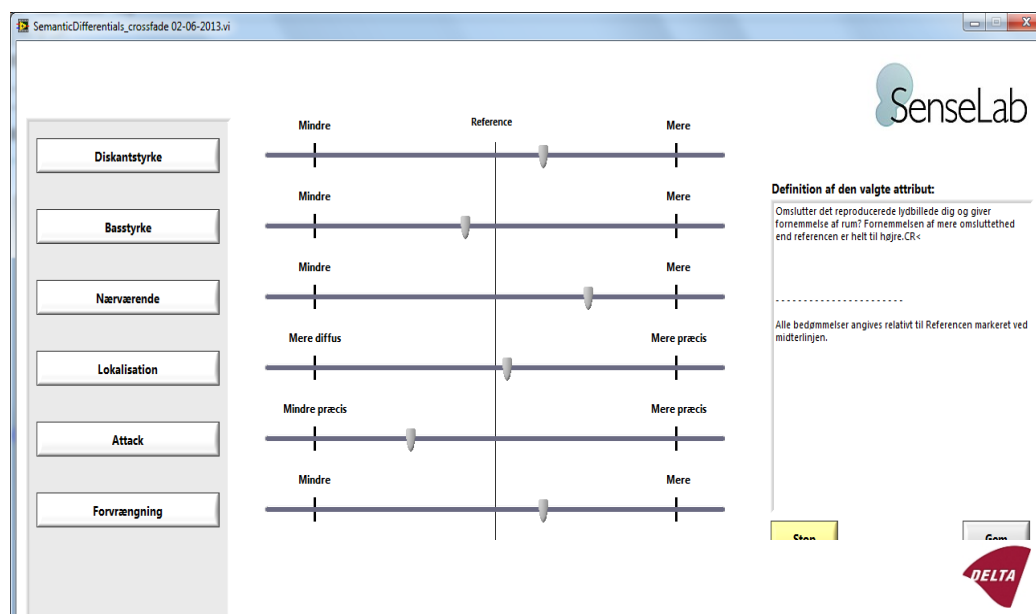
Denne metodik er velegnet, når der ikke er adgang til direkte sammenligning af testemner, og giver assessorerne god mulighed for at vurdere delelementerne i helhedsindtrykket.

³ Der blev benyttet et Labview-program udarbejdet af Søren Legarth, DELTA: *SemanticDiff SA test v2-v4*. Kun få ikke-betydelige ændringer fra v2 til v4.

⁴ Semantic differential teknikken er f.eks. beskrevet i: S., Joel L. Stone, Herbert, *Sensory Evaluation Practices*, Second edition. San Diego, United States: Academic Press, Inc., 1993.

En ulempe ved den konkrete implementering af metoden var, at alle attributter nødvendigvis skulle vurderes for alle musikstykker. Dette satte store krav til valget af musikudsnit. I denne test indgår f.eks. et musikudsnit (Bach) med en sopran, som ikke indeholder noget videre bas, hvorved basstyrke er svær at vurdere, hvilket giver statistisk støj i resultaterne.

Brugerfladen for testen er vist nedenfor:



Figur 1

Brugerfladen til profil- eller attributtesten til karakterisering af lyden. Bedømmelsen af hver attribut afgives ved at stille skyderne i den ønskede position. Den valgte attribut er defineret i øverste højre hjørne og med play-knapperne kan man skifte mellem reference- og testhøjttaler. Hvis man synes, at en speciel del af musikeksemplet er særligt egnet til bedømmelsen af den aktuelle attribut, kan der zoomes ind på den del.

Der er tastaturgenveje, så man ikke behøver at se på skærbilledet, når man f.eks. ønsker at skifte mellem testhøjttaleren og referencehøjttaler. For hver af de syv attributter vist til venstre foretages en sammenligning med referencehøjttaleren. Det angives, hvor meget mere eller mindre testhøjttaleren besidder den aktuelle attribut end testhøjttaleren. Den valgte attribut er nærmere defineret i boksen til højre (se afsnit 7). Hvert lydeksempel er ca. 15 sekunder langt. Når alle attributter er bedømt for en given højttalertype, holder lytteren en kort pause uden for lytterummet, mens en ny højttaler placeres i samme position som den foregående. Rækkefølgen af højttalerne er forskellig for forsøgspersonerne.

5.2 Eksperimentelt design

Rækkefølgen af de fire højttalesystemer er for hver assessor randomiseret i et balanceret design baseret på 24 assessorer. Da 25 assessorer deltog i testen, er rækkefølgen af testemnerne for den første assessors test ikke del af designet, men valgt arbitrært til ABCD. Som konsekvens er denne kombination benyttet to gange i testen.

For at spare tid er der ikke gentagelser af alle kombinationer af højttalesystemer og musikudsnit (lydeksempler), men for at kunne vurdere assessor præstationen er bedømmelserne gentaget med ét af musikudsnittene - et Jennifer Warnes nummer (se afsnit 8). Dette udsnit er valgt, forbi det blev vurderet til at kunne benyttes til vurderinger af alle eller hovedparten af de udvalgte attributter.

6. Testopstilling

6.1 Faciliteter og udstyr

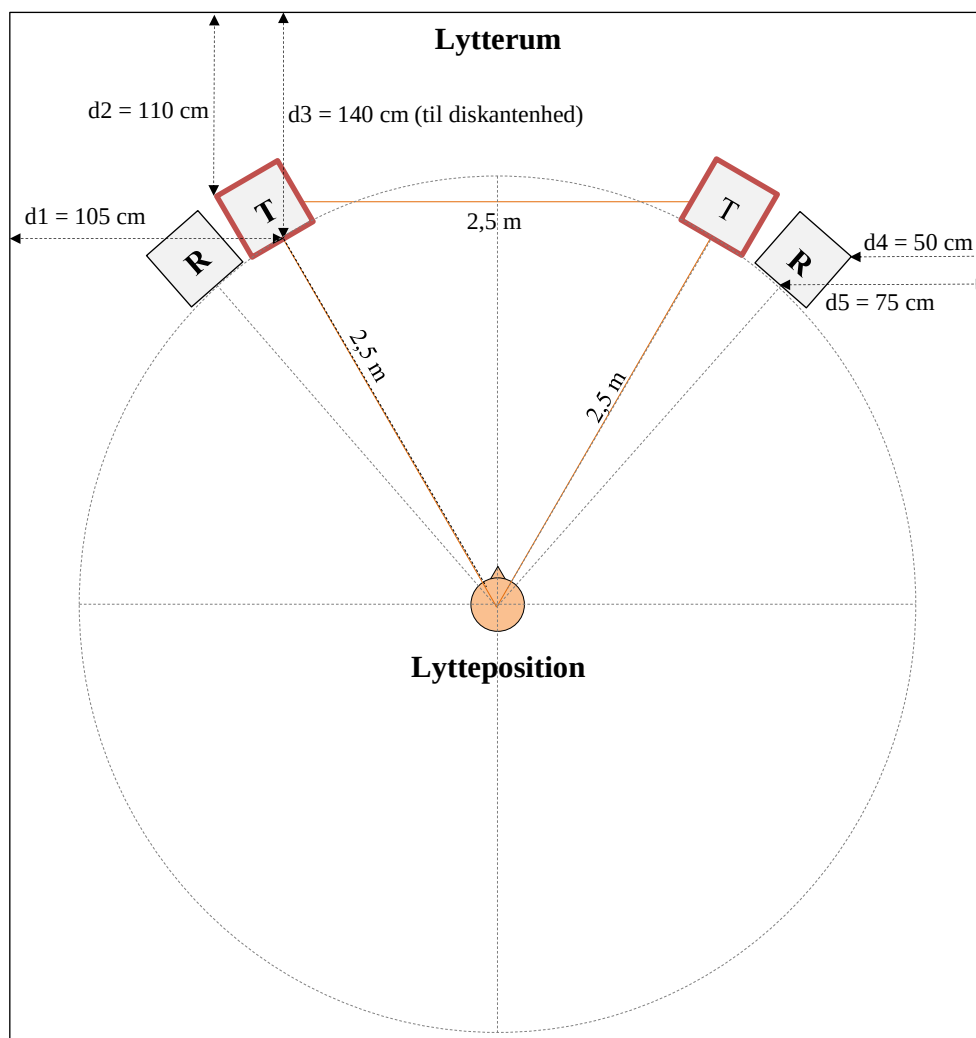
Opstillingen af højttalere i lytterummet er skitseret i Figur 2, hvor T er positionen for det højttalesæt der testes og R er referencehøjttalesættet. Kompakt- og væghøjttalere var monteret på standere.

Til at flytte højttalerne til og fra testpositioner blev der brugt transporthunde (MDF-plader med drejelige hjul) med en lille stabel af løse MDF-plader, som enten testhøjtteren (gulvhøjtteren) eller højtterstanderen står på. Antallet af plader er justeret for hvert sæt stereohøjttere, således at diskantenheden er placeret i 110 ± 2 cm højde over gulvet og dermed er i samme horisontale plan som den gennemsnitlige lytters øre.

For at sikre samme position af testemnerne i testposition T, er der lavet en ”docking station”, som positionerer transporthundene i den korrekte afstand og vinkel. Referencehøjtternes placering ændres ikke under lyttetesten.

Imellem højttalerne og lyttepositionen, er ophængt et lag tyndt sort akustisk transparent gardin bag et hvidt gardin af samme materiale. Bag højttalerne er opsat sort stof af samme materiale. Denne kombination sikrer at lytteren ikke kan se højttalerne. De aktive referencehøjtternes grønne diode på fronten blev dækket med tape.

Højtteropstilling og benyttet udstyr kan ses i afsnit 1 – Anneks A.



Figur 2

Skitse af højttaleropstillingen til lyttetesten. R betegner referencehøjttaleren og T betegner højttaleren under test.

7. **Attributter**

Lyden fra højttalere kan beskrives med mange forskellige attributter. Til en given lyttetest udvælger man derfor de mest relevante, dvs. dem der bedst kan bruges til at beskrive forskellene mellem de produkter der testes.

Attributterne til testen blev valg på baggrund af uformelle lyttesessioner af forsøgslederne. De blev udvalgt med fokus på at beskrive de tydeligste forskelle mellem Q113 højttalerne. Ydermere blev der lagt vægt på at udvælge attributter, som det blev vurderet, at panelets assessorer ville have gode og ensartet forståelser af.

Følgende attributter blev valgt:

A. Præference

1. Basstyrke
2. Diskantstyrke
3. Nærværende
4. Lokalisation
5. Envelopment
6. Attack
7. Forvrængning

Diskantstyrke

Den relative styrke af diskanten, dvs. de lyse toner (høje frekvenser) fra f.eks. bækken, triangel og s-lyde (hvislelyde).

Basstyrke

Den relative styrke af bassen, dvs. de dybe toner (frekvenser) f.eks. mandestemmer, bas, stortromme. Må ikke forveksles med basområde, som angiver basens udstrækning.

Nærværende

Indtrykket af at et instrument eller en vokal træder frem fra lydbilledet og er til stede i rummet.

Lokalisation

Kan instrumenter og stemmer placeres tydeligt og adskilt i det rumlige lydbillede? Hvor præcist står de enkelte lydkilder i rummet?

Attack

Hvor god er højttalerens evne til at gengive transienter/impulslyde? Gengives instrumentanslag fra stortromme og bas præcist og distinkt? Er bassen stram og kontant eller er den blød og buldrende? Mindre præcis høres ved, at anslaget breder sig ud i tid og at klimakset i anslaget udviskes.

Forvrængning

Er der uren lyd eller mislyde som f.eks. hvislen, skratten, klipning eller forvrængning fra højttaleren? Hvis ingen forvrængning er hørbar, placeres svaret på midterlinjen svarende til Referencen.

Envelopment

Omslutter det reproducerede lydbillede dig og giver fornemmelse af rum omkring dig? Fornemmelsen af mere Envelopment (omsluttethed) end Referencen bedømmes til højre for midterlinjen.

Attributterne fordeler sig således mellem: 1. og 2. er klangfarve attributter, 3. er enten klangfarve (mellemtone) eller en rumlig attribut. 4. og 5. er rumlige attributter og 6. og 7. beskriver andre hørbare forskelle.

Alle attributter bedømmes separat for hvert højttalersæt og hvert musikeksempel.



8. Lydeksempler og kalibrering af lydstyrke

8.1 Lydeksempler

Lydeksempler betegner det programmateriale, der lyttes på, i dette tilfælde forskellige musikstykker. Til denne test har vi valgt følgende tre lydeksempler⁵, som vi vurderer tilsammen er egnede til at vurdere de valgte attributter:

Artist(er)	Album	Nummer
Jennifer Warnes	Famous Blue Raincoat	Bird on a Wire
<i>Begrundelse:</i> En optagelse af høj standard, som giver specielt mulighed for at bedømme attack, kildeadskillelse, rumlig gengivelse og lokalisation. [14 s udsnit]		
Karina Gauvin, Luc Beausejour	Bach: Little Notebook for Anna-Magdalena	Willst du dein Hertz mir schenken
<i>Begrundelse:</i> Ren sopran stemme med en sprød gengivelse af cembalo. Endvidere en god gengivelse af rum. [16 s udsnit]		
Paula Cole	This Fire	Tiger
<i>Begrundelse:</i> Nummeret er specielt ved at have en kraftig femstrenget basguitar, som går ned til 31 Hz og er specielt egnet til bedømmelse af basgengivelsen. [14 s]		

Tabel 1

De musikeksempler der blev anvendt til lyttetesten.

8.2 Kalibrering af lydstyrke

Højttalerne blev indledningsvis justeret til samme lineære lydtrykniveau i lyttepositionen ved hjælp af båndpasfiltreret (80-14.000 Hz) lyserød støj.

Herefter blev Paula Cole nummeret skruet op til lidt under det niveau, hvor der forekom hørbar forvrængning i bassen. De to andre numre er justeret til en lavere lydstyrke for at give variation i niveauerne. De resulterende niveauer i lyttepositionen målt med reference-højttaleren var som vist i Tabel 2.

⁵ Udvalgt blandt anbefalinger fra Geoff Martins blog, <http://www.tonmeister.ca/wordpress/>

	L_{Aeq}, dB	L_{Ceq}, dB	$L_{A,max,F}, dB$	$L_{C,max,F}, dB$
Jennifer Warnes	66	77	73	83
Karina Gauvin	74	75	84	86
Paula Cole	71	86	79	94

Tabel 2

Kalibreret lydtrykniveau med referencehøjttaleren for de tre musikudsnit - afrundet til hele dB.

9. Instruktion og træning af assessorerne

Instruktionen til testen kan findes i afsnit 14 - Anneks B, og indeholder 1) Instruktion i testen, 2) Instruktion i brugen af Labview-brugergrænsefladen og 3) attributbeskrivelserne.

Som træning til testen gennemføres en attributvurdering på ét sæt højttalere. Denne benyttes for at give assessorer kendskab til musikudsnit, attributter og brugergrænsefladen.

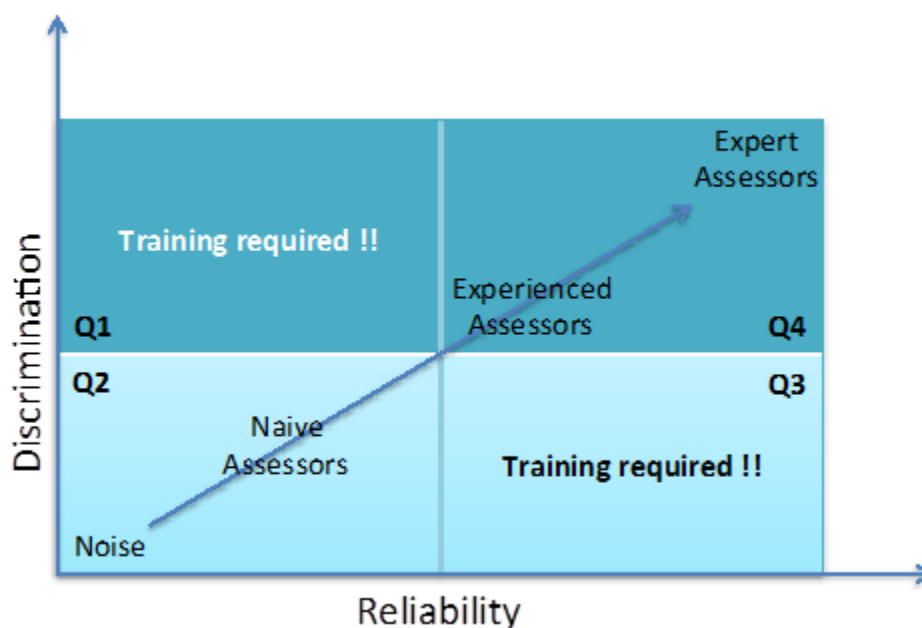
Desværre giver træningen ikke kendskab til spændet af lydkvalitet testemnerne imellem, hvilket kan føre til ”forsigtig” brug af skalaen, så yderpunkter ikke benyttes eller ikke benyttes i starten af testen. Sidstnævnte faktor kan lede til uensartet skalabrug igennem testen og gør muligvis normalisering af data nødvendig.

10. Valg af assessorer

Det er vigtigt at forholde sig til lytternes kvalifikationer i forbindelse med resultatbehandlingen. Ud fra resultaterne fra attributterne bedømmes lytternes evne til:

- at kunne diskriminere mellem højttalerne,
- at kunne gentage egne bedømmelser,
- at være i overensstemmelse med resten af panelet.

Det er vigtigt, at lytterne kan diskriminere mellem stimuli, men hvis de ikke samtidig kan reproducere sine bedømmelser ved gentagelser, falder troværdigheden af bedømmelsen. På den anden side kan man nemt opnå at reproducere sine bedømmelser ved at svare det samme hele tiden, men det giver ingen diskrimination mellem stimuli. For gode assessorer skal begge dele være til stede samtidig, se Figur 3.



Figur 3

Illustration af betydningen af evnen til at skelne mellem stimuli, "Discrimination", og evnen til at reproducere sine bedømmelser ved gentagelser, "Reliability".

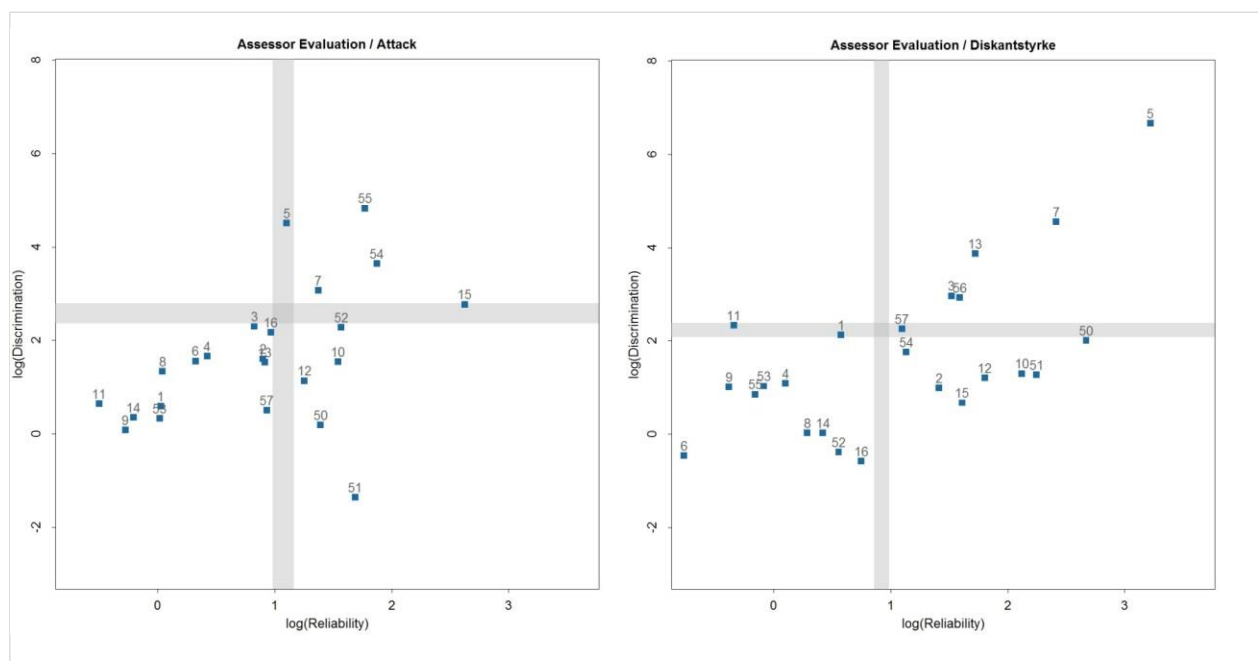
Disse egenskaber kan findes ud fra gentagne bedømmelser. Af tidsmæssige grunde blev der kun foretaget gentagelser af Jennifer Warnes nummeret.

SenseLabs eGauge metode⁶ er for hver attribut brugt til at kvantificere assessorernes evne til at diskriminere mellem stimuli og til at reproducere deres bedømmelser. Resultatet ses på Figur 4.

Helt overordnet må det siges, at mange assessorer har haft svært ved at bedømme forskellene mellem højtalerne med den valgte testmetode og på de valgte attributter. Normalt ser vi flere assessorer over støjgulvet, end det er tilfældet her.

⁶ eGauge - a measure of assessor expertise in audio quality evaluations. Gaëtan Lorho¹, Guillaume Le Ray², and Nick Zacharov². 1Nokia Corporation, Helsinki, 2DELTA SenseLab, Hørsholm, Denmark. AES 38'th International Conference, Piteå, Sweden, 2010.

Denne metode er på vej ind i den kommende ITU Standard ITU-R BS 1534-1: "Method for the subjective assessment of intermediate quality level of coding systems"



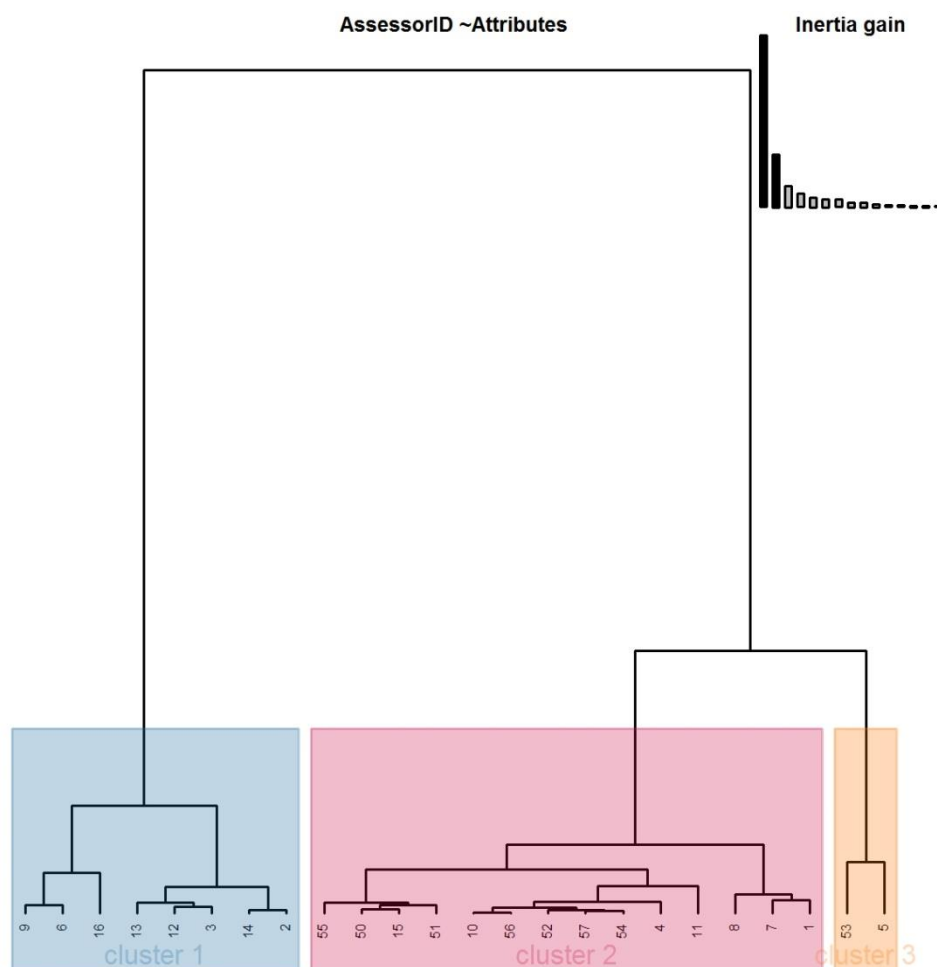
Figur 4

Reliability-discrimination grafer hvor assessorerne - de blå punkter- er identificeret ved numre. Eksemplerne viser resultaterne for attributterne attack (til venstre) og diskantstyrke (til højre). De grå områder viser det interval, hvor det statistiske støjguld med 95 % sandsynlighed befinder sig.

Ud fra resultaterne fra eGauge analysen er det besluttet i første omgang at udvælge de 14 assessorer, hvis resultater ligger på eller over støjguldvet for mindst to forskellige attributter.

Hvis man generelt ser på resultaterne fra disse 14 assessorer, finder man ud af, at ikke alle er enige om bedømmelserne, dvs. at der ikke enighed i det panel, der udgøres af de 14 udvalgte assessorer.

Der er derfor udført en klyngeanalyse (eng. Cluster analysis) af assessorernes resultater, se Figur 5. Det ses, at assessorerne falder i tre klynger. Det er valgt at bruge de assessorer fra klynge 2, dels fordi det er den største, dvs. der hvor der er flest, der har konsensus om bedømmelserne, og dels fordi den også indeholder nogen af de af SenseLabs lyttere, der erfaringsmæssigt er "skarpe" lyttere.

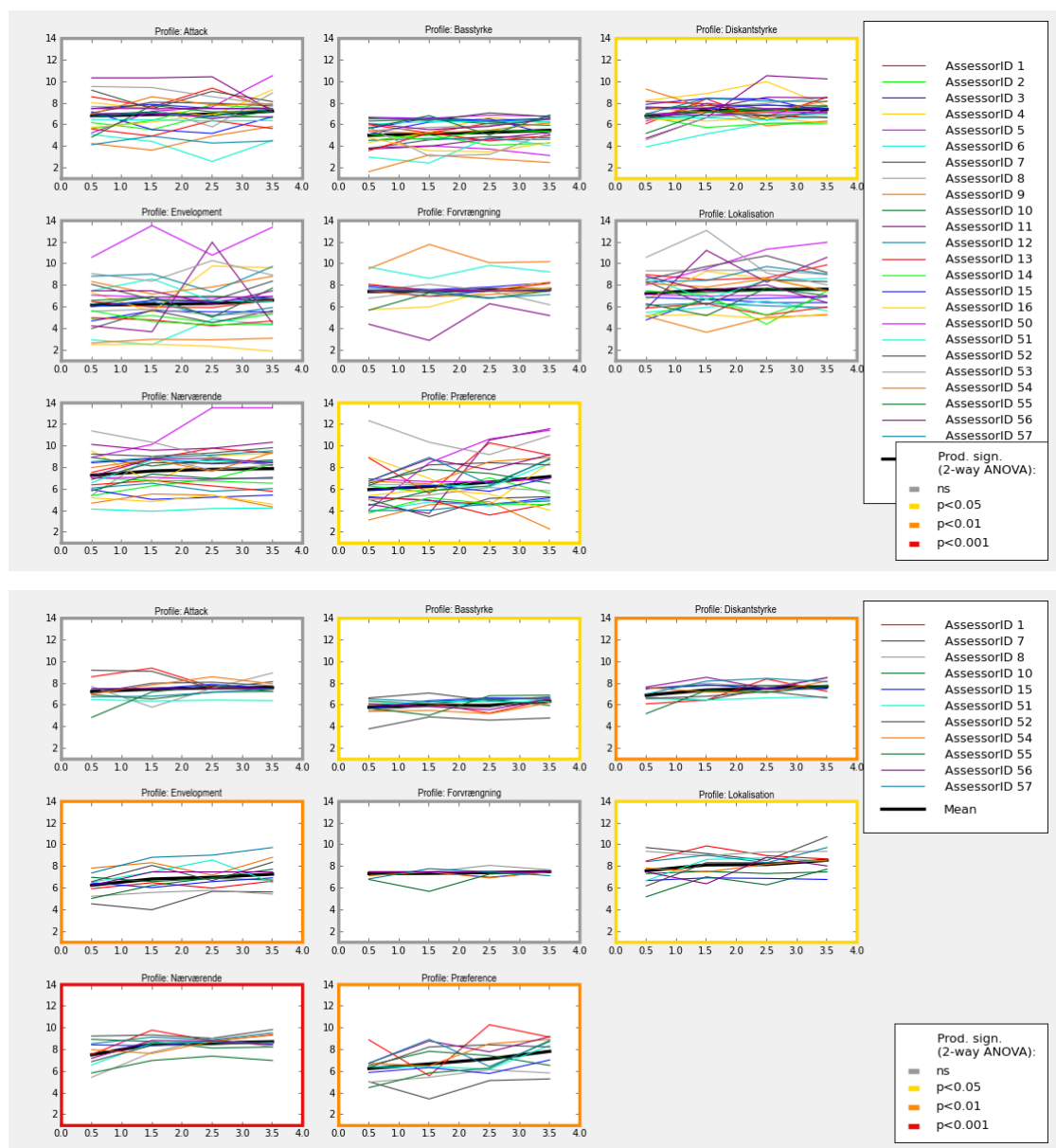


Figur 5

Klyngeanalyse af assessorernes resultater baseret på bedømmelser af alle stimuli og alle attributter. Det ses, at assessorerne falder i tre hovedgrupper.

Den samlede gruppe af lyttere, som både er med i klyngeanalysens gruppe 2 og som kvalificerer sig i henhold til eGauge analyserne, er nu på 11 personer. Resultaterne i afsnit 11 er, hvis ikke andet er nævnt, baseret på disse 11 personers bedømmelser.

Figur 6 viser bedømmelserne for hhv. alle assessorer og de 11 udvalgte assessorer. Ideelt skal hvert plot have streger (bedømmelser), der ligger nær hinanden (som ved forvrængning på nederste figur), men de behøver ikke ligge på en lige linje. Farven på plottenes ramme angiver signifikansen af resultatet; Grå er lig ingen signifikans og rød er lig højeste signifikans. Forskellene mellem det øverste og nederste plot i Figur 6 viser tydeligt, hvorledes den statistiske støj og spredningen på bedømmelserne mindskes ved at bruge data fra den udvalgte gruppe frem for alle assessorers data. Der ses stadig en del spredning på præferencebedømmelserne, men det er ”tilladt”, da det jo er en subjektiv sag, hvad man synes bedst om.



Figur 6

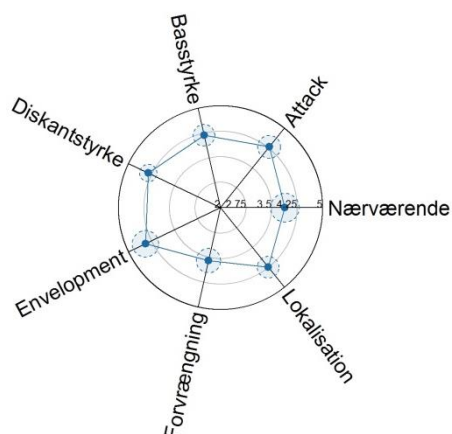
Bedømmelserne for de forskellige attributter (se overskrifterne) af de 4 højttalersystemer (0,5, 1,5, 2,5 og 3,5 på x-aksen ordnet efter stigende middelværdi) for Jennifer Warnes nummeret. Øverst er vist resultater for alle assessorer. Nederst er vist resultater for de 11 udvalgte assessorer.

11. Resultater

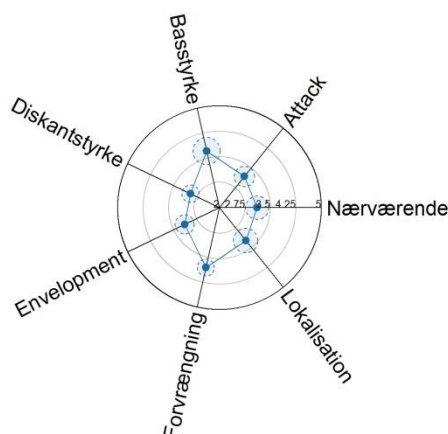
11.1 Profiler af lydkarakteristika

Profildigrammerne i Figur 7 bygger på data, der er normaliserede mht. middelværdi og spredning for hver attribut og for hver lytter, så akserne er et udtryk for relative forskelle. Punkterne angiver middelværdierne af bedømmelserne og cirklerne angiver 95 % konfidensintervaller.

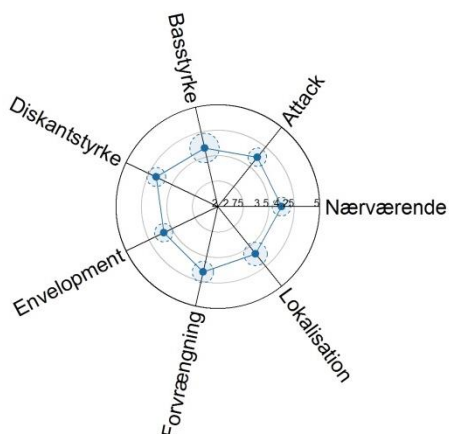
SystemLabel : Q113_Kompakt



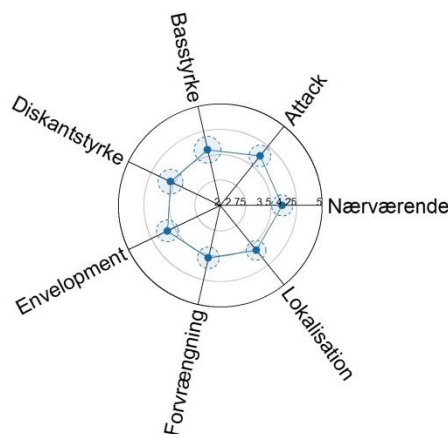
SystemLabel : Q113_Væg



SystemLabel : SA2K



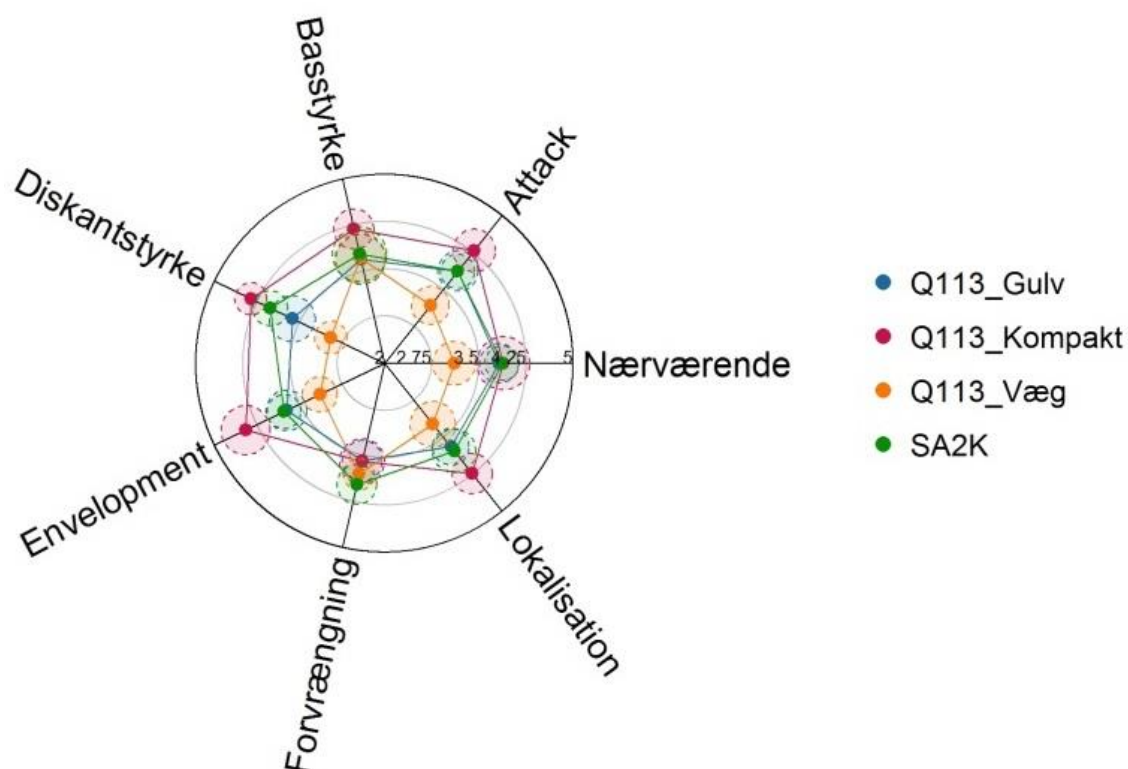
SystemLabel : Q113_Gulv



Figur 7

Profildigrammer som karakteriserer lyden fra de fire højttalere i testen. Diagrammerne er baseret på normaliserede data for de 11 udvalgte assessorers bedømmelser af alle lydeksempler. Skalaerne er relative og cirklerne angiver 95 % konfidensintervaller.

I Figur 8 er profildiagrammerne sammenstillet for at lette sammenligningen. Som en tommelfingerregel kan man antage, at hvis disse konfidensintervallerne (cirklerne) overlapper, kan der ikke påvises signifikante forskelle mellem middelværdierne.

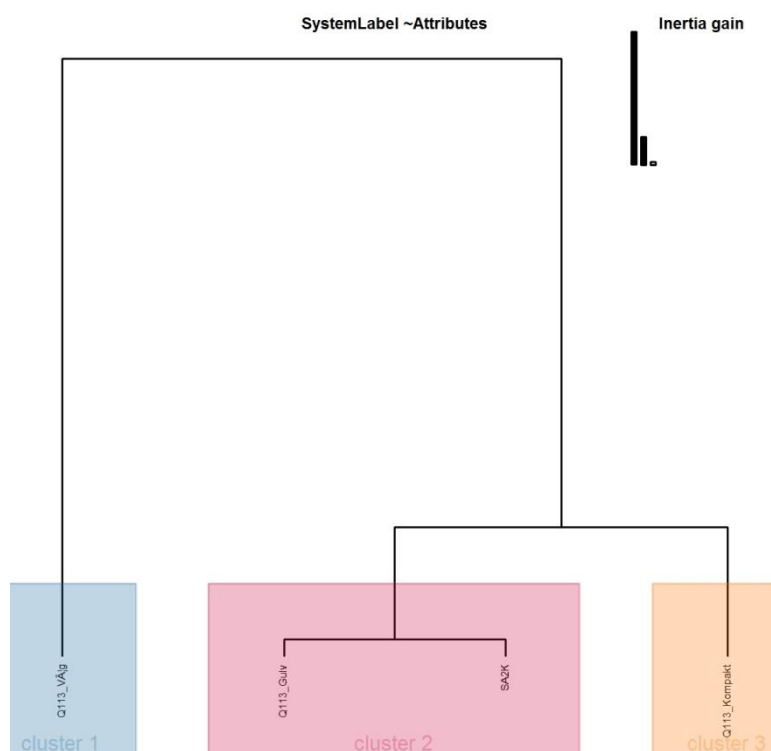


Figur 8

Lydkarakteristikprofiler baseret på normaliserede data for de 11 udvalgte assessorers bedømmelser af alle lydeksempler. Skalaerne er relative og cirklerne angiver 95 % konfidensintervallerne.

Det ses, at der er signifikante forskelle mellem kompakt- og væghøjtaler på mange attributter, medens gulvhøjtaleren og SA2K ser ret ens ud. Dette bekræftes også af en klyngeanalyse, se Figur 9, hvor det ses, at gulvhøjtaleren og SA2K falder i samme gruppe.

Det skal bemærkes, at alle attributter for hver assessor er bedømt i samme omgang for hver højtaler. Det kan medføre en afsmittende virkning mellem bedømmelserne af attributterne.



Figur 9

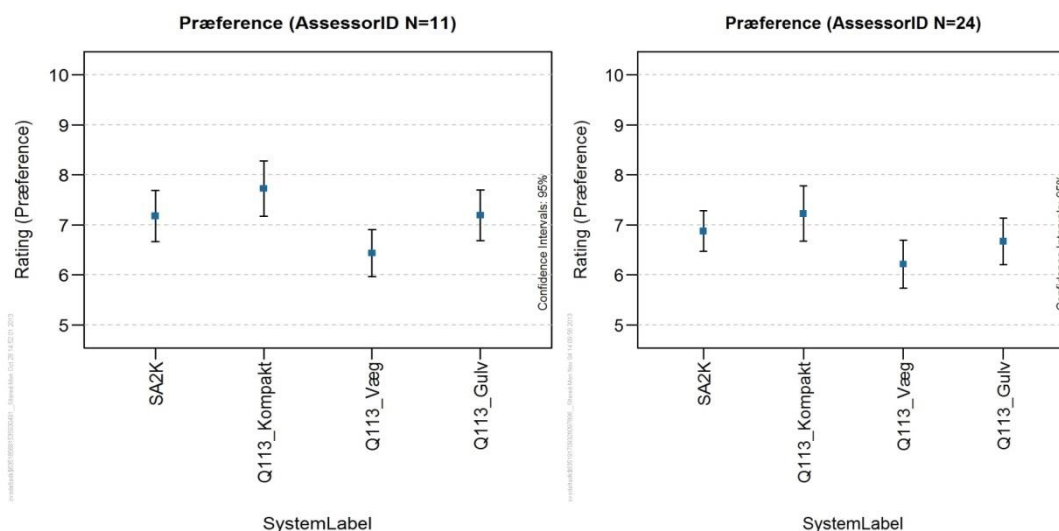
Klyngeanalyse af bedømmelserne på de fire højttalersystemer baseret på data for de 11 udvalgte assessors bedømmelser af alle lydeksempler og alle attributter.

11.2 Præference

Medens der tilstræbes konsensus omkring bedømmelserne af attributterne (der jo betragtes som en objektiv karakteristik af højttaleren) er det ikke nødvendigt med konsensus på præferencemålingen, da præference jo er en subjektiv og personlig sag. Som Figur 10 viser, synes der dog at være en vis grad af enighed.

Det eneste signifikante er, at kompakhøjttaleren foretrækkes frem for væghøjttaleren. Selvom middelværdien af kompakhøjttaleren er den højeste, er den ikke signifikant bedre end SA2K og gulvhøjttaleren, ligesom væghøjttaleren ikke er signifikant mindre foretrukket end de to sidstnævnte, selv om den har den laveste middelværdi.

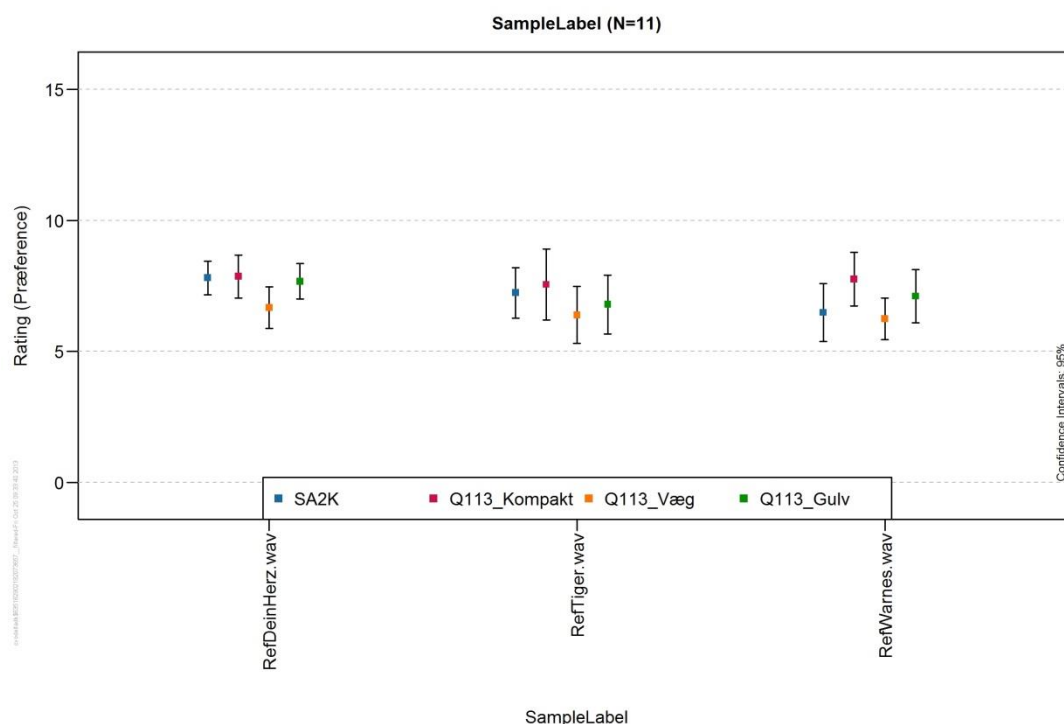
For at se, om de 11 udvalgte assessorer har præferencer, der afviger fra hele den samlede gruppe på 25 lyttere, kan man sammenligne venstre og højre graf i Figur 10. Det ses, at der ikke er signifikante afvigelser i præferencerne for de to grupper. At konfidensintervallerne i de to grafer (med hhv. 11 og alle lyttere) er næsten ens er tegn på, at spredningen er mindre for de udvalgte 11 lyttere.



Figur 10

Præferencebedømmelser, 11 udvalgte assessorer (tv.) og alle assessorer (th.). Bedømmelserne er middelværdier over alle lydeksempler. De lodret markerede intervaller er 95 % konfidensintervaller.

Endelig kan det jo være interessant, om bedømmelserne afhænger af musikken. Af Figur 11 ses, at der ikke er nogen signifikant afhængighed. Hvis man ser bort fra signifikansen (hvilket er farligt!), er der en tendens til, at kompakthøjttaleren ligger stabilt højt og at gulvhøjttaleren ligger lavt. SA2K er den højttaler, der viser den største afhængighed af musiknummeret, uden at det dog er signifikant.



Figur 11
Præferencebedømmelser for de tre musikeksempler med de 11 udvalgte assessorer.

12. Konklusion

DELTA bød ind på opgaven som led i en metodeudvikling. Testen blev gennemført med lyttere, hvis kvalifikationer ikke var dokumenterede, og som ikke var trænet i brugen af de anvendte attributter. Attributterne var udvalgt på baggrund af et kvalificeret skøn, men det var uvist, om de med den anvendte metode med en referencehøjtaler kunne afsløre forskelle mellem testemnerne.

Resultaterne baseret på alle lytteres bedømmelser gav kun signifikante resultater på én attribut og præferencebedømmelsen, så med henblik på at udvælge de mest kvalificerede lyttere blev bedømmelserne underlagt en nærmere undersøgelse.

Indledningsvis blev der foretaget en analyse af lytternes resultater med SenseLabs eGauge-værktøj. Dette kræver, at der er lavet en test med gentagelser. Resultaterne fra Jennifer Warnes musikeksemplet blev benyttet til dette, da det af tidsmæssige grunde var det eneste, der blev bedømt med gentagelser. Ud fra resultaterne blev 14 lyttere, som kunne diskriminere mellem højtalerne samtidig med at de kunne gentage egne bedømmelser på mindst to attributter, udvalgt.

Herefter blev der foretaget en klyngeanalyse (eng.: Cluster analysis) af lytterne. Klyngeanalysen grupperer lytterne på en sådan måde, at de der har mest til fælles, dvs. er mest enige i deres bedømmelser, kommer i samme gruppe. Det er her vigtigt at bemærke, at der er tale om bedømmelse af objektive attributter, så der er altså ikke tale om subjektive egenskaber som præference, der blev målt i en separat test.

Klyngeanalysen opdelte lytterne i tre grupper. Den største gruppe var på 14 personer og indeholdt desuden flere af SenseLabs skarpeste lyttere, så det blev besluttet at gå videre med denne gruppe.

Fællesmængden af de to kvalifikationstests udgjordes af 11 lyttere. Baseret på deres bedømmelser af alle tre musikeksempler blev højttalerne karakteriseret i form af profildia-grammer, så deres egenskaber let kunne sammenlignes. Lydkarakteristikken for hhv. kompakt-højttaleren og væghøjttaleren skilte sig ud, medens SA2K og gulvhøjttaleren ikke blev vurderet signifikant forskellige. En supplerende klyngeanalyse bekræfter dette.

Hvilken højttaler lytterne foretrak, blev målt i en separat præferencetest. Resultatet var, at kompakt-højttaleren blev foretrukket frem for væghøjttaleren. Selv om middelværdien af kompakt-højttaleren er den højeste, blev den ikke fundet signifikant bedre end SA2K og gulvhøjttaleren. Ligeledes er væghøjttaleren ikke signifikant mindre foretrukket end de to sidstnævnte, selv om den har den laveste middelværdi.

Det viste sig også, at de 11 udvalgte lytteres præferencer ikke afviger fra hele den samlede gruppe af lyttere.

Der var ikke signifikante forskelle på præferencerne for højttalerne for forskellige musik-eksempler. SA2K var den højttaler, der viste den største afhængighed af musiknummeret, uden at dette dog var signifikant.

Alt i alt må det konkluderes, at selv om vi her har afprøvet en ny metodik med indirekte sammenligning - som formentlig er mindre følsom end en direkte sammenligning - har vi påvist signifikante forskelle mellem højttalerne. Når resultaterne er opnået ved en blindtest som denne, kan man stole på, at det virkelig er lydkarakter og lydkvalitet, som har gjort udslaget uden visuelle bias-effekter og rummets indflydelse. Den systematiske forsøgsteknik har også givet mulighed for at udvælge de mest pålidelige assessorer og give information om, hvilke forskelle mellem testemnerne der reelt er signifikante.

13. Anneks A - Faciliteter og udstyr

13.1 Lytterum

SenseLabs lytterum med indre mål på L: 7,8 m, B: 4,7 m og H: 2,6 m er opført efter kasse-i-kasse princippet for at give god lydisolation. Det inderste rum er vibrationsisoleret fra det yderste. Rummet opfylder standarderne EBU Tech 3276 og ITU-R BS.1116-1 til multikanalets højttaler lyttetest. Det betyder bl.a. meget lav baggrundsstøj og en frekvensuafhængig efterklangstid (så der ikke sker en farvning af lyden) på ca. 0,2 sekunder over 160 Hz og svagt stigende mod lavere frekvenser.

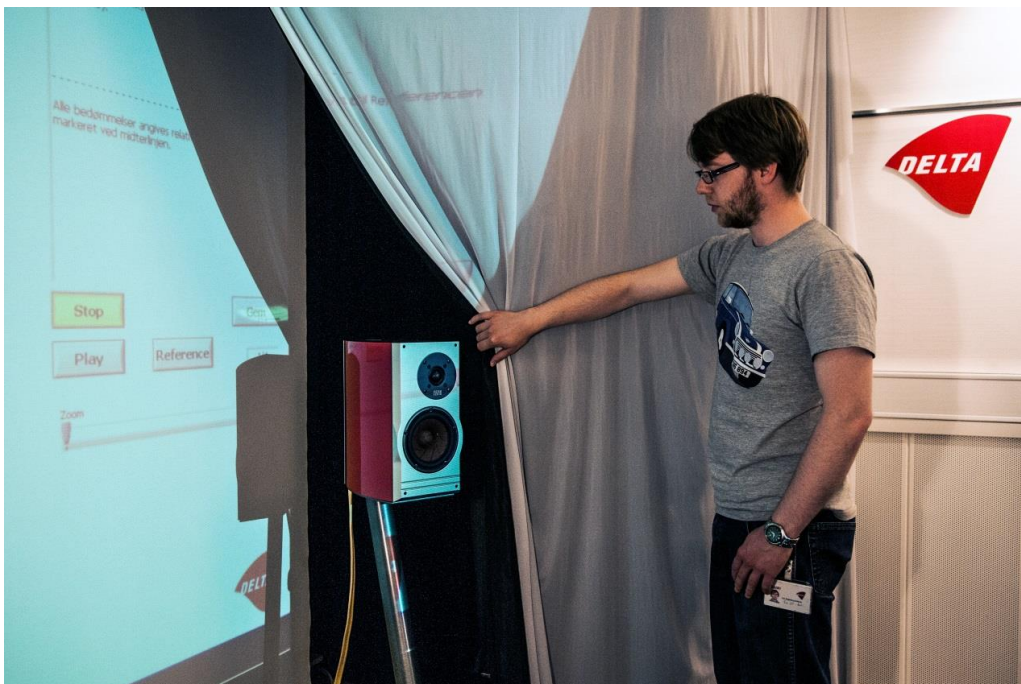
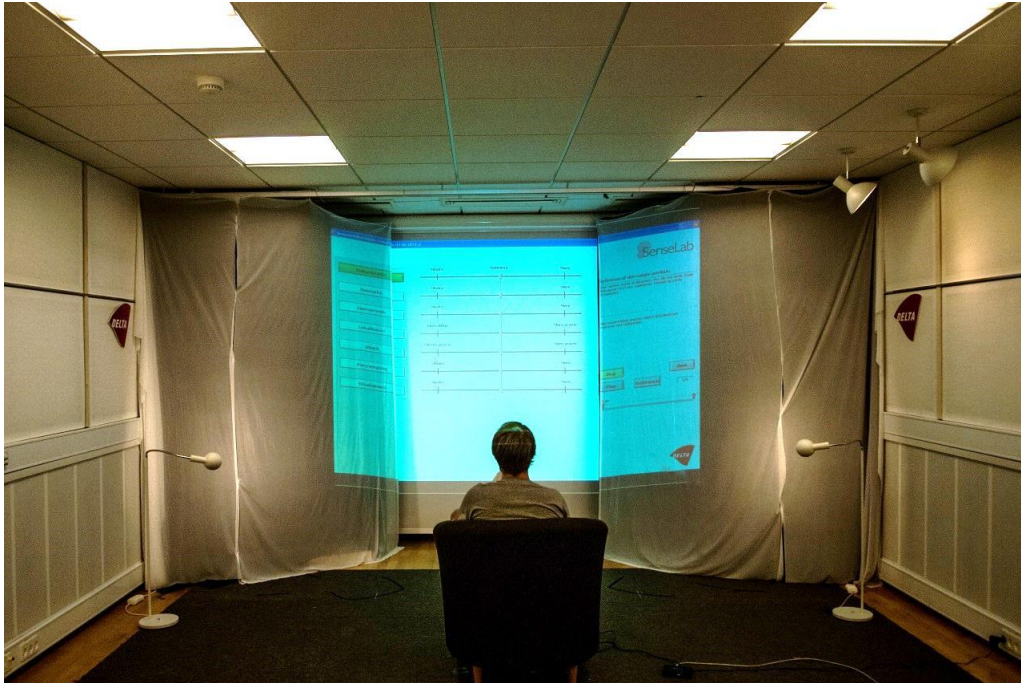
13.2 Udstyr

Apparat	Fabrikat	Type
PC	Lenovo	Think Center
Lydkort	RME	Fireface 800
Kalibrator	Brüel & Kjær	4230
Lydtrykmåler	Brüel & Kjær	2270
Forstærker	Primare	I32 (2 x 120 W)
Højttalerkabler	Van der Hul	CS-122 hybrid, 5 m, 3 mm ²

Tabel 3

Udstyr der blev brugt til lyttetest.

13.3 Opstilling



Figur 12
Testopstillingen i DELTA SenseLabs lytterum.

Højttalernes placering i rummet er vist på Figur 2.



14. Anneks B - Instruktion til lytterne

Kære panelmedlem

I dette projekt skal du bedømme lyden fra 4 forskellige højttalere.

Testtiden er estimeret til 2 timer i alt.

De 4 højttalere skal bedømmes i forhold til de 7 definerede attributter (skalaer) for 3 forskellige musikstykker. Musikstykkerne og lyden fra højttalerene vil du få dybere kendskab til i den indledende træningsrunde.

Alle bedømmelser i denne test udføres relativt til en Reference-højttaler der er til din rådighed. Hvis testemnet ikke adskiller sig fra Referencen sættes bedømmelsen udfor reference-linjen midt på skalaen (se skærbillede på næste side).

Du skal i denne test bedømme 'Præference' samt 7 attributter:

- Diskantstyrke
- Basstyrke
- Nærværende
- Lokalisation
- Attack
- Forvrængning
- Envelopment

Definitionen af hver attribut bliver angivet på skærmen under testen.

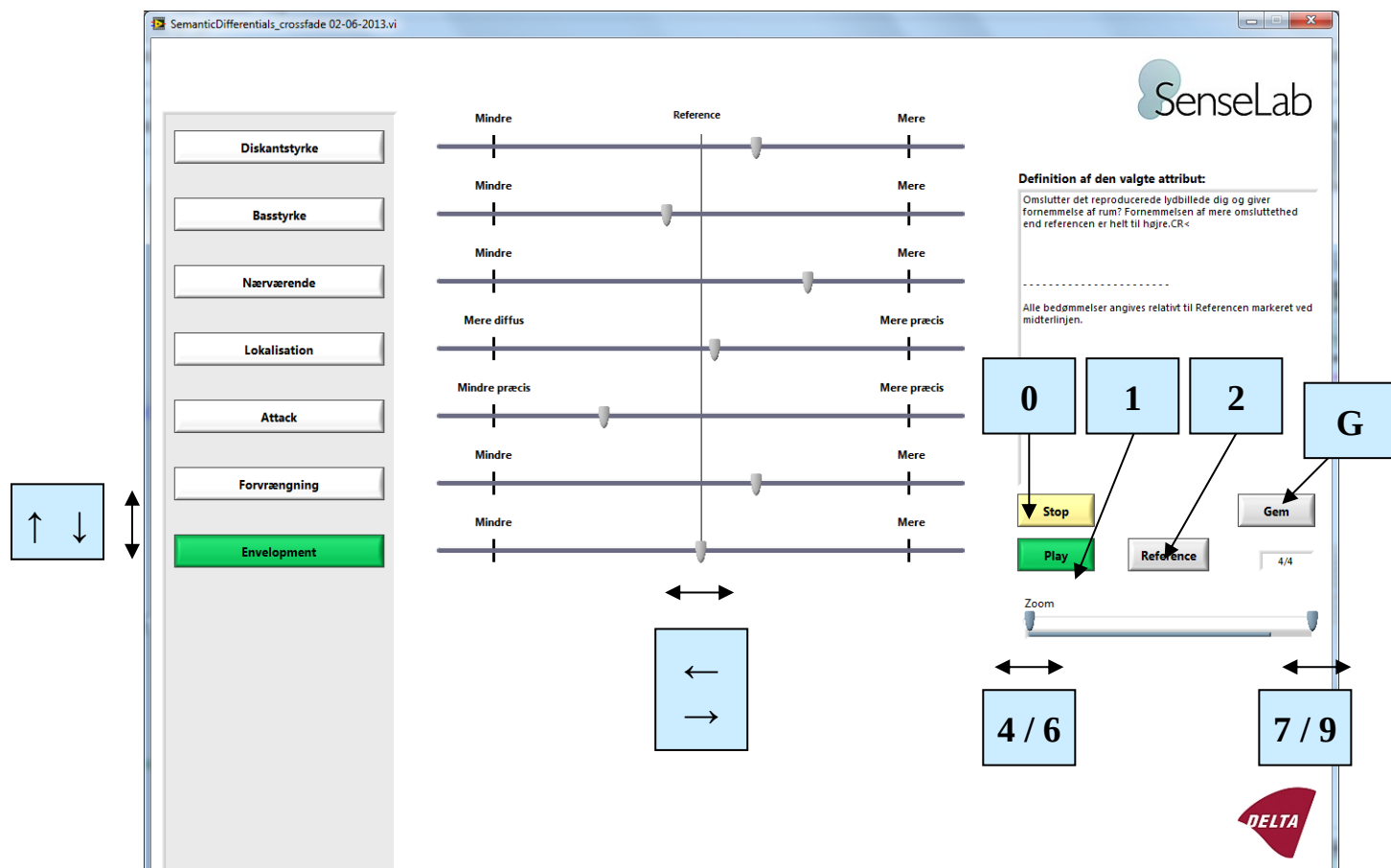
Du har mulighed for at zoome ind på et relevant område i musikstykket hvis du ønsker, men undlad at zoome for snævert ind på enkelte begivenheder. Du vil modtage en gennemgang af softwarens betjening på næste side.

Hvis du har spørgsmål undervejs, er i tvivl om opgaven eller driller teknikken er det vigtigt, at du kontakter forsøgslederen med det samme.

Sluk venligst din mobiltelefon under testen og hold øje med tiden og husk at holde pause ca. hver 20. minut!

Betjening af softwaren:

Skærbilledet vil se således ud:



Du betjener funktionerne med tastaturet. Se hvilke genvejstaster der styrer hvad.

Genvejene er lagt på det numeriske tastatur primært, i håb om at forenkle betjeningen.

Hvis du ønsker at nulstille zoom'et tast 5 eller 8.

Du vil starte testen med en træningsrunde for at sikre at du er indforstået med betjeningen af både software og forståelse af attributterne.

God fornøjelse!

Mvh SenseLab



Attributbeskrivelser

Præference

Din subjektive vurdering af lydkvaliteten.

Diskantstyrke

Den relative styrke af diskanten, dvs. de lyse toner (høje frekvenser) fra f.eks. bække-
ner, triangel og s-lyde (hvislelyde).

Basstyrke

Den relative styrke af bassen, dvs. de dybe toner (frekvenser) f.eks. mandestemmer,
bas, stortromme. Må ikke forveksles med basområde, som angiver basens udstrækning.

Nærværende

Indtrykket af at et instrument eller en vokal træder frem fra lydbilledet og er tilstede i
rummet.

Lokalisation

Kan instrumenter og stemmer placeres tydeligt og adskilt i det rumlige i lydbillede?
Hvor præcist står de enkelte lydkilder i rummet?

Attack

Hvor god er højttalerens evne til at gengive transienter/impulslyde? Gengives instru-
mentanslag fra stortromme og bas præcist og distinkt?
Er bassen stram og kontant eller er den blød og buldrende?
Mindre præcis høres ved at anslaget breder sig ud i tid og at klimakset i anslaget udvi-
skes.

Forvrængning

Er der uren lyd eller mislyde som f.eks. hvislen, skratten, klipning eller forvrængning
fra højttaleren?
Hvis ingen forvrængning er hørbar placeres svaret på midterlinjen svarende til Refe-
rencen.

Envelopment

Omslutter det reproducerede lydbillede dig og giver fornemmelse af rum omkring dig?
Fornemmelsen af mere Envelopment (omsluttethed) end Referencen bedømmes til høj-
re for midterlinjen.