

Brugbarhedsevaluering af Boss RC-20XL guitareffektapparat

- en case sammenligning af
heuristisk evaluering
og
brugertest med instant data analysis



Jesper Lumbye Andersen, Marianne Krosby og Poul Larsen

Master i IT, fagpakken Brugbarhed, faget Anvendt Brugbarhed
IT Vest, Aalborg Universitet, foråret 2010

Forord

Denne oppgaven er laget av tre studenter fra studiet "Master i IT" ved Aalborg Universitetet. Rapporten er en besvarelse på oppgaven som skal leveres i faget "Anvendt Brugbarhet" under fagpakken "Brugbarhet", forår 2010.

Vi har valgt å utføre en Brugbarhedsevaluering af Boss RC-20XL gitareffektapparat. Her vil vi sammenligner to metoder for brukbarhetsevaluering, nemlig heuristisk evaluering og brugertest analysert med analysemetoden Instant Data Analysis (IDA).

Vi tre studenter gjennomførte hver vår heuristiske evaluering av aparatet. Etterpå ble det gjennomført en brukertest med 4 test brukere, hvor det ble benyttet tænke højt teknikken. Det ble rett etter brukertesten gjennomført en brainstorming.

Videre vil vi også beskrive vår praktiske anvendelse av denne av IDA teknikken, da denne er utført delvis via Skype, p.g.a. prosjekt gruppens ulike geografiske plassering. Vi avslutter med en analyse av dataene etter sammenligning av de to metodene.

Tak til Tore L. Mogensen (TC Electronic) for idé til brugertestopgaver og med at rekrutere brugertestere. Tak til Mike Nickel (TC Electronic) for hjelp med at skaffe utstyr til brukertesten og for at deltage som observatør.

Referencer til kilder er angivet med firkant-parentes med en forkortelse og evt. med et side- eller figur-nummer. Forkortelsen kan findes i Kildelisten sidst i rapporten. Eksempel på kildereference: [CHI'92, s.397].

Indholdsfortegnelse

| | |
|--|-----------|
| Forord | 2 |
| Indholdsfortegnelse..... | 3 |
| 1. Indledning | 4 |
| 2. Systembeskrivelse | 5 |
| 2.1. RC-20XL Loop Station..... | 5 |
| 2.1.1. Viktige funksjoner | 5 |
| 2.2. Konceptuel Model..... | 6 |
| 3. Metodevalg | 8 |
| 3.1. Alternative Metoder | 8 |
| 3.1.1. Heuristisk Evaluering | 8 |
| 3.1.2. Tænke højt teknikken m/Instant Data Analysis | 9 |
| 3.1.3. Methods of Human Thinking | 9 |
| 3.1.4. Metodevalg | 10 |
| 3.2. Metodebeskrivelse..... | 11 |
| 3.2.1. Heuristisk evaluering | 11 |
| 3.2.2. Tænke højt teknikken | 14 |
| Video Data Analysis (VDA) | 15 |
| Instant Data Analysis (IDA) | 15 |
| 3.3. Brugbarhedsproblemers omfang og alvorlighed | 16 |
| 3.3.1. Lokale og globale brugbarhedsproblemer | 16 |
| 3.3.2. Brugbarhedsproblemers Alvorlighed | 16 |
| 3.4. Praktisk gennemførelse af brugertesten | 18 |
| 3.4.1. Afvikling af Brugertesten | 19 |
| 3.4.2. Analyse af data fra brugertesten | 20 |
| 4. Resultat af Evaluering | 22 |
| 4.1. Problemliste - Heuristisk evaluering | 22 |
| 4.2. Problemliste - Tænke højt test m/Instant Data Analysis..... | 25 |
| 4.2.1. Tidforbrug på de enkelte opgaver | 27 |
| 4.3. Feedback fra produktchef | 27 |
| 4.4. Sammenligning af resultater..... | 27 |
| Refleksion over forskellene mellem de to test resultater | 28 |
| 5. Konklusion | 30 |
| 5.1 Videre Arbejde..... | 30 |
| Kildeliste..... | 32 |
| Appendiks 1: For-spørgeskema, opgaver og | |
| evalueringsspørgeskema til testpersoner | 33 |
| A1.1. Spørgsmål før brugertesten | 33 |
| A1.2. Opgaver og spørgeskema efter brugertesten..... | 33 |
| A1.3. Sammenfatning på Spørgeskemaer | 34 |
| Bilag 1. Samtykkeerklæring | 36 |
| Bilag 2. Tekniske specifikationer for Boss RC-20XL..... | 37 |

1. Indledning

Når en virksomhed får en idé til et nyt produkt, er det naturligt, at undersøge tilsvarende produkter på markedet. For at få idéer til produktets interaktionsdesign, kan man udføre brugbarhedsevaluering på konkurrenternes produkter for at få en idé om, hvad man kan genbruge, og hvor man bør differentiere sit produkt, så brugerne med det nye produkt kommer til at opleve en bedre brugbarhed og brugeroplevelse. Når der arbejdes i idéfasen på et produktudviklingsprojekt, må der ikke investeres ret mange ressourcer i projektet. Det betyder, at det vil være interessant at se på, om man kan nøjes med en såkaldt discount metode for at evaluere brugbarheden af produkter.

Projektgruppen har i dette projekt valgt at kigge nærmere på to metoder, hvor der er fokus på at spare ressourcer, og som bl.a. adskiller sig ved om brugbarhedseksperter eller domæneeksperter, der evaluerer systemet. Det er inspektions-metoden heuristisk evaluering (HE), og den empiriske metode tænke-højt-test (TH) med efterfølgende Instant Data Analysis (IDA).

Den Heuristiskeevaluering er en ekspertvurdering av et produkt hvor en systematisk sjekker produktet opp mot en liste over heuristikker, dvs. viktige prinsipper for brukervennlighet. Metoden vi benytter her er introdusert av Jacob Nielsen og Rolf Molich i 1990, og senere revidert av Nielsen.

Den empiriske metoden tænke-højt-testen, går ut på at man tenker høyt mens en går igjennom testen. Det er i forbindelse med brukertesten blitt formulert en rekke oppgaver, som testpersonene skal gjennomføre. Oppgavene som vi benytter har er formulert av en produktsjef som har stort kjennskap til domenet og brukerne. Det sier seg selv at det for mange bedrifter vil være veldig resurskrevende å benytte en brukertest basert på tænke-højt metoden. Etter at testen var ferdig, ble det kjørt en IDA analyse. Denne teknikken går i utgangspunktet ut på om å få analysert og dokumentert i løpet av en dag. Dette gjør kanskje ikke denne testen til den beste, men det er en veldig fin metode for å finne de mest kritiske feilene.

Prosjektgruppen startet her med at vi hver for oss utførte en Heuristisk evaluering av produktet RC-20 Loop Station, som er en gitar effekt apparat. Det kan også kalles en phrase recorder eller gitar-effekt-pedal. Dette er et apparat som har et fysisk brukergrensesnitt, som tildels betjenes med fingrene og dels med bena.

Ut fra resultatene fra denne evalueringen ble det laget en felles problemliste, som også ble vektet etter feilens alvorlighetsgrad. Disse kategoriene er definert som kosmetisk, alvorligt eller kritisk. Vi anga også under hvilken heuristikk hvor feilen ble oppdaget. Vi utførte så en brukertest i et laboratorium, med 4 forskjellige brukere. I forbindelse med denne var kun to av prosjektgruppens deltagere til stede, og analysen i etterkant ble derfor avvirket noe anderledes enn beskrevet af Kjeldskov et al. Analysen ble utført i to omganger, det ble utført en IDA analyse rett etter avsluttet test, og en fortsettelse noe senere. Den tredje person i prosjektgruppen får videoopptaget av brukertesten til gjennomsyn og umiddelbart herefter gjennomfører prosjektgruppens deltagere en IDA analysedel vha. en audio- og chat-konferanse over Internettet. Desverre fikk vi her noen problemer med data linjen, slik at forholdene ikke ble helt optimale. Resultatene her ble også laget i en problemliste, og vektet på samme måten.

Resultatene fra de to evalueringer blev sammenlignet, og med heuristisk evaluering blev der fundet flere brugbarhedsproblemer end med tænke-højt-test m. IDA. Dette resultat er anderledes en det, som Karat et al. kom frem til og sammenholdt med det, som Kjeldskov et al. kom frem til. På baggrund af disse kilder havde vi forventet, at vi med tænke-højt-test m. IDA ville finde flere brugbarhedsproblemer end med heuristisk evaluering.

2. Systembeskrivelse

Vi skal her forklare litt nærmere hva et RC-20XL Loop Station gitareffektapparat er, si litt om viktige funksjoner. I tillegg forklare den konseptuelle modellen. Apparatets spesifikasjoner finnes i Bilag 2.

2.1. RC-20XL Loop Station

RC-20 Loop Station er en phrase recorder eller sampling-pedal. [RC-20XL guide]

Selve prosessen som benyttes her kalles overdubbing, og er en teknikk som ofte benyttes i musikkstudier. Dette går ut på at det legges ekstra lyd på tidligere innspilt musikk.

Det at man lager et spor eller track med rytmen i en sang, for så å overdubbe med soloinstrumenter, som gitar eller keyboard, og med en vokalist til slutt har vært en vanlig standardteknikk for innspilling av enkelte typer musikk siden 1960 tallet.



Det er flere grunner for at man lager en overdub, og en av de viktigste er at det gjør ting enklere. Skulle f.eks. bassisten i et band være midlertidig utilgjengelig, kan man spille inn alt annet, og legge til bassen senere. På samme måte vil man, hvis en sang trenger flere gitarstemmer, men kun har en gitarist, få den ene gitaristen til å spille inn begge stemmene og deretter legge stemmene sammen.

Du vil også kunne fremstå som at du er et band alene og det er veldig lett å kunne lage et "en persons band" i real time.

RC-20XL Loop Station er laget for utøvere som har behov for å lage og spille av kontinuerlige fraser på scenen. Det vil si det er svaret på mange utøvers drømmer, da det er en kompakt phrase recorder og den er utstyrt med en sampling-pedal som skal være forholdsvis lett å bruke.

RC-20XL kan benyttes på flere måter, og har en rekke forskjellige funksjoner, det er stort sett opp til din egen kreativitet og avgjøre hvordan den kan bli benyttet.

2.1.1. Viktige funksjoner

Loop Quantize-funksjonen gjør at man lett kan lage phrases med passende lengde.

Den har en samplingstid eller innspillingstid på opptil 16 minutter, som utgjør en total samplingstid på 5 minutter og 30 sekunder.

Man kan lagre inntil 11 looper på en gang, dvs. du kan lagre opptil 10 fraser som er loopet og 1 "one-shot" frase, selv når strømmen ikke er tilkoblet. Eget klikk og Loop-kvantisering gjør det raskt og enkelt å lage loops ved hjelp av fot pedaler.

Du kan gjøre endring av tempo i sanntid uten at tonehøyden eller tonearten endres. Dette kan også gjøres ved hands-free styring, dvs. du kan benytte kun pedaler ved opptak og avspilling.

Det finnes en overdub-funksjon for å legge lyd på lyd. Du kan også benytte Mikrofon-/ instrument- og Auxiliary-inngang for sampling av vokal eller musikk fra en CD. F.eks. så kan du ta opp phrases fra eksemel CD'en og overdubbe gitar og vokal oppe på den, eller ta opp en audio CD og bruke den for praktisering. Her kan både mikrofon- og instrumentinngangene brukes samtidig. Auto Start gjør at samplingen starter samtidig med lyden inn.

Reverse-funksjon kan generere reverse sample playback uten ventetid, og man har også en angre funksjon. Angre funksjonen gjør at man lettere kan lage perfekte loops.



2.2. Konseptuel Model

Den nok viktigste del af den konseptuelle model for systemet er hvordan de optagede lydclip behandles og lagres. Der er to "arbejds-lydspor" og mulighed for at gemme op til 11 lydspor. De gemte lydspor kaldes "PHRASE" og for hhv. at gemme og slette dem benyttes betegnelserne "WRITE" og "DELETE". De to "arbejds-lydspor" er dér, hvor guitaristen arbejder med lyden. Det er dem, der høres og dem, der regieres i. Det primære af arbejdslydsporene er her, hvor optagelserne akkumuleres efterhånden som der laves nye optagelser. Det sekundære arbejdslydspor er her, der optages i. Lyde opbevares her midlertidigt, for at de kan slettes igen, hvis der er en fejl. Næste gang der optages oven i, akkumuleres lyden fra det sekundære lydspor i det primære lydspor, og kan nu ikke længere slettes - med mindre, man sletter alt, der er opbevaret i det primære lydspor. I forbindelse med arbejde på arbejdslydsporene benyttes betegnelserne "REC" for første optagelse i arbejdslydsporene (denne første optagelse bliver optaget direkte i det primære lydspor - det er også det, der sætter længden af optagelsen). "PLAY" betegner afspilning af arbejdslydsporene. "OVERDUB" betegner, at man optager i det sekundære arbejdslydspor. "UNDO" betegner, at man sletter det sekundære arbejdslydspor. "CLEAR" betegner, at man sletter både det primære og det sekundære arbejdslydspor. Og "STOP" betegner, at man stopper for afspilningen af arbejdslydsporene. Arbejdslydsporene har ikke en betegnelse på frontpanelet af apparatet, men den lyd, der er der i kaldes andre steder for "loop" og i nogle situationer for "one shot". Når man gemmer det, der er i arbejdslydsporene, i hukommelsen kaldes det for en "phrase" - gemmeoperationen akkumulere primær- og sekundær-lydsporene. Når man vælger en af de 11 phrases, læses det gemte lydspor fra hukommelsen over i det primære arbejdslydspor, samtidig med at det sekundære arbejdslydspor slettes.

Den formentlig næst vigtigste del af systemets konseptuelle model er styring af timing i optagelserne. Timing er ekstremt vigtigt i musik, og apparatet har derfor flere faciliteter til at hjælpe med at få styr på timingen. Der er dels en "AUTOSTART" funktion, som gør det muligt lade en kraftig lyd (første anslag) igangsætte en optagelse i det primære arbejdslydspor. Der er også en "LOOP QUANTIZE" funktionalitet, der gør det muligt automatisk at få justeret afslutningen af en optagelse, så den passer i timingen. Herudover er det muligt at slå takten an med enten en finger eller en fod ("TAP TEMPO") samt at lytte til ("GUIDE") og se takten ("TAP TEMPO") blinke i en indikatorlampe.

Det at brugeren forstår disse to dele af systemets konceptuelle model er af afgørende betydning for, om systemet er anvendelig i praksis for brugeren. Brugeren behøver dog formentlig ikke at forstå alle detaljerne, men skal have et grundlæggende begreb om det.

3. Metodevalg

Der findes i følge Nielsen og Molich [CHI'90, p.249] grundlæggende fire forskellige måder at evaluere brugbarhed på:

- Formelt med en analyseteknik.
- Automatisk med en computer-styret procedure.
- Empirisk med eksperimenter med test-brugere.
- Heuristisk ved på simpel vis at kigge på brugergrænsefladen og dømme den ud fra ens egen mening.

I følge Karat et al. [CHI'92, s.397] benyttes der dog i praksis to typer af metoder:

- Empirisk brugertest (eller brugbarhedstest), som kan foregå i et laboratorium eller i felten og med en eller anden form for efterfølgende analyse.
- Brugbarhedsgennemgang (eller inspektionsteknikker), som bl.a. omfatter kognitiv gennemgang, heuristisk evaluering og metaphors of human thinking evaluering.

Traditionelt bliver den empiriske test i den form, hvor man benytter tænke-højt-test og efterfølgende video-data-analyse til at lokalisere brugbarhedsproblemer ofte benyttet som en reference blandt alle metoderne. Det ses i flere artikler, hvor forskellige evalueringsmetoder sammenlignes med tænke-højt-test, bl.a. Karat et al. [CHI'92], eksperiment 3 hos Frøkjær & Hornbæk [MOT2008] og hos Kjeldskov et al. [IDA]. Det skyldes bl.a. at de brugbarhedsproblemer, som findes her er faktiske brugbarhedsproblemer, mens inspektionsmetoderne blot finder forventede brugbarhedsproblemer, og man undgår således problemer med "falske positive". Med faktiske brugbarhedsproblemer menes der, at en bruger (testperson) har oplevet en forsinkelse eller forhindring i at bruge systemet til at løse en opgave. Dette bygger på en antagelse om, at brugertestsituationen ikke afviger afgørende fra den rigtige brugssituation.

3.1. Alternative Metoder

I de følgende afsnit gennemgås en række relevante alternative brugbarhedsevalueringsmetoder, der er relevante i forhold til dette projekt. Metodernes fordele, ulemper og relevans diskuteres på baggrund forskelligt litteratur.

3.1.1. Heuristisk Evaluering

Grunden til at man ikke blot altid anvender empirisk brugertest er, at det er relativt resourcekrævende. Det kræver en del planlægningsarbejde at indkalde testpersoner og koordinere selve gennemførelsen af testen. Det kræver også en del mandskab at lede, observere og efterfølgende analysere testen. Ud fra et ønske om at reducere resourceforbruget ifm. brugbarhedsevalueringer er der opstået en række metoder, hvor det ikke er nødvendig at involvere "rigtige" brugere i evalueringen. Nielsen og Molich [CHI'90] introducerede 1990 metoden Heuristisk Evaluering. Denne metode er let at gå til, kræver meget få resourcer og kan finde de fleste brugbarhedsproblemer. Nielsen og Molich siger, at hver evaluator finder 20%-51% af brugbarhedsproblemerne [CHI'90, s.249], og det er ikke de samme, de finder. Så jo flere evaluators der, jo større bliver foreningsmængden af fundne brugbarhedsproblemer. Nielsen anbefaler 3-5 evaluators, som kan forventes at finde 50%-75% af brugbarhedsproblemerne [Nielsen HowTo, figur 2]. Nielsen skriver, at han i et eksempel har oplevet en faktor 48 mellem det beløb, der blev brugt på at gennemføre heuristisk evaluering og den forventede gevinst, som forbedringerne ville medføre. Dette bør naturligvis ses med en vis skepsis, men hvis blot

en del af den påståede gevindst kan opnås i andre situationer, kan det være værd at tage med. Nielsen og Molich fremhæver selv: "Hovedfordelene ved heuristisk evaluering er: Det er billigt. Det er intuitivt og det er let at motivere folk til at gøre det. Det kræver ikke forudgående planlægning. Det kan benyttes tidligt i udviklingsprocessen." [CHI'90, s.255]

Karat et al. har sammenlignet en række evalueringsmetoder, heriblandt tænke-højt-test og heuristisk evaluering og skriver: "Omkostnings-effektivitetsdata viser, at empirisk test kræver det samme eller mindre tid for at identificere hvert problem sammenlignet med inspektions-metoderne." [CHI'92, s.403]

Generelt kan det konkluderes, at empirisk brugertest finder mange brugbarhedsproblemer, og at det målt absolut er resourcekrævende, selv om det ikke er det per fundne brugbarhedsproblem. Heuristisk evaluering er let at gå til og kræver få ressourcer, men finder tilgængelig ikke så mange brugbarhedsproblemer. Hvis man benytter en iterativ udviklingsproces, hvor der udføres adskillige brugbarhedsevalueringer, kan man påstå, at en gentagen brug af heuristisk evaluering efterhånden vil få bugt med alle brugbarhedsproblemer. En kritik af heuristisk evaluering går også på, at det ofte er mindre væsentlige brugbarhedsproblemer, som metoden finder: "Et af de konsistente resultater af disse studier er, at heuristisk evaluering, når det udføres af eksperter, afslører mange flere lokale problemer end andre metoder. Brugbarhedstest (empirisk brugertest) gør det modsatte, det finder globale problemer." [D&R, s.82] Med lokale problemer menes, at omfanget er begrænset - ofte til f.eks. et enkelt skærmbillede. Mens der med globale problemer menes, at omfanget er mere omfattende som f.eks. dele af den konceptuelle model for interaktionsdesignet.

3.1.2. Tænke højt teknikken m/Instant Data Analysis

Gennem de senere år har der været tiltag dels i retning af at gøre empirisk brugertest mindre resourcekrævende og dels i retning af at forbedre brugbarhedsgennemgangsmetoderne. Et eksempel på det første er IDA-metoden, hvor analysedelen reduceres fra en minutiøs analyse af videooptagelserne med skrivning af log osv. til blot at være en brainstorm mellem testlederen og observatørerne som afslutning på en dag med tænke-højt-test. Kjeldskov et al. skriver, at i et konkret eksperiment blev 85% af brugbarhedsfejlene fundet på kun 10% af den tid, som en almindelig videoanalyse tager. [IDA, s.1]. IDA-metoden forudsætter, at der er flere (typisk tre) brugbarhedseksperter til stede under afviklingen af brugertesten. Efter brugertesten er der 1-1½ times brainstorm, hvorefter facilitatoren (som kan være en udenforstående) bruger ca. 1 time på at samle sammen og formulere en brugbarhedsproblemliste. Ved en brugertest med et halvt dusin testpersoner, kan det hele afvikles på én dag (man skal dog hertil lægge noget forberedelsestid).

3.1.3. Methods of Human Thinking

Som nævnt har et andet tiltag været at forbedre de metoder, som ikke kræver testpersoner. En ny metode, som blev introduceret af Frøkjær og Hornbæk i 00'erne er Metaphors of Human Thinking (MOT). Udførelsen af metoden minder om heuristisk evaluering, blot er heuristikkerne skiftet ud med en række metaforer for, hvordan mennesker tænker. Evaluatoren skal så vurdere brugbarheden i forhold til hver af dem. Frøkjær og Hornbæk synes, at eksisterende inspektionsteknikker lider under dels at de sjældent inddraget brugernes måde at tænke på eksplicit, og dels at anvisningerne (f.eks. heuristikker) kun er beregnet for et begrænset anvendelsesområde [MOT 2008, s.20:2]. De opstiller fem metaforer, der omhandler: Vaner, tankestrømme, bevidsthed, ufuldstændigheden i ytringer og viden. Frøkjær og Hornbæk har lavet en række eksperimenter, hvor de bla. sammenligner MOT-metoden med empirisk brugertest og

heuristisk evaluering (HE). Det konkluderes, at MOT finder brugbarhedsproblemer, der er mere alvorlige end dem, der findes med HE [MOT 2008, s.20:12], og MOT synes (i modsætning til HE) at finde brugbarhedsproblemer, der ikke forsvinder efterhånden som brugerne får mere erfaring [MOT 2008, s.20:13]. Hovedforskellen mellem MOT og HE er, at HE stræber mod at være simpel, hvilket har været et mål for metoden siden starten af dens udvikling [MOT 2008, s.20:24]. I forhold til tænke-højt-test (TA) vurderer nogle af deltagerne, at TA gav nogle resultater, der var lettere at fortolke, og at TA identificerer nogle brugbarhedsproblemer, som de ikke selv kunne have fundet ved en inspektion. Nogle deltagere bemærkede, at MOT var svær at forstå [MOT 2008, s.20:24]. Det er dog vigtigt at bemærke, at alle deltagere i sammenligningseksperimenterne var novicer og ikke brugbarhedseksperter. Dette har sandsynligvis en væsentlig betydning for, hvordan de forskellige metoder subjektivt er blevet vurderet i forhold til hinanden. Frøkjær og Hornbæk konkluderer, at deltagerne foretrækker tænke-højt-test i forhold til inspektionsmetoderne. Af inspektionsmetoderne foretrækkes *methods of human thinking* frem for heuristisk evaluering (og kognitiv gennemgang) [MOT 2008, s.20:31].

3.1.4. Metodevalg

I nærværende projekt evalueres brugbarheden for en guitar-effekt-pedal. Det er et apparat med en fysisk brugergrænseflade, som dels er fingerbetjent og dels er fodbetjent. Formålet er at få vurderet interaktionsdesignet mhp. brugbarhed som input til specifikationen for et nyt konkurrerende produkt. Det vil derfor være at foretrække, hvis evalueringsmetoden er god til at finde globale brugbarhedsproblemer.

I forbindelse med projektet er der en række omstændigheder, som det har været nødvendigt at tage hensyn til, og som derfor har en vis indflydelse på forsøgene. Medlemmerne i projektgruppen er spredt geografisk og bor i hhv. Århus, Oslo og Nørresundby. Det er ok i forbindelse med benyttelse af inspektionsmetoder, hvor der benyttes individuel inspektion. I forbindelse med brugertest er det u hensigtsmæssigt i forhold til afviklingen af tænke-højt-test, men i forbindelse med en efterfølgende evt. individuel dataanalyse er det ok.

For mange virksomheder kan omfanget af ressourcer, der skal benyttes til en brugertest med tænke-højt-test og efterfølgende videoanalyse virke overvældende og måske ligefrem forhindre, at der bliver lavet systematisk brugbarhedsevaluering. Det er derfor prioriteret i dette projekt at undersøge metoder, som har fokus på et lave resourceforbrug. Som tidligere beskrevet kan der ofte spares en del ressourcer på at benytte inspektionsmetoder frem for empiriske metoder, dog er resultatet ofte ikke så omfattende og så direkte anvendeligt. I dette projekt vil vi derfor sammenligne en empirisk metode med en inspektionsmetode. Som empirisk metode har vi valgt tænkt-højt-test med efterfølgende IDA primært for at holde resource forbruget til analysedelen nede. Som inspektionsmetode har vi valgt heuristisk evaluering, da det er den mest simple inspektions metode, der er let at anvende for novicer (som nærværende projektgruppes medlemmer).

I forbindelse med brugertest-afviklingen kan kun to af projektgruppens tre deltagere være til stæde, og IDA-metoden afvikles lidt anderledes end beskrevet af Kjeldskov et al. Vi har valgt at gennemføre analysen af to omgange, således at de to deltagere, der er med til afviklingen af brugertesten umiddelbart efter brugertesten gennemfører analysedelen. Den tredje person i projektgruppen får videooptagelser af brugertesten til gennemsyn og umiddelbart herefter gennemfører projektgruppens deltagere en IDA analysedel vha. en audio- og chat-konference over Internettet. Denne situation vil også kunne forekomme i praksis i virksomheder, som har flere mindre afdelinger eller søsterselskaber, der er geografisk spredt, og hvor der ikke er ressourcer til at have flere brugbarhedseksperter i alle afdelinger. Brugbarhedseksperterne i de forskellige afdelinger vil på lignende vis have mulighed for at benytte Internettet til at trække på

hinandens ekspertise. Det kan give en mere tung og langsommelig kommunikationsvej, og Instant i IDA, som betyder øjeblikkelig, er knap så øjeblikkelig, hvis der går længere tid fra brugertesten er gennemført og indtil alle brugbarhedseksperterne har fået set videoen af brugertesten. Jo længere tid, der går fra brugertesten er overstået og til data-analysen bliver gennemført, jo sværere er det at huske detaljerne for dem, der deltog i brugertesten som testleder og observatører.

Der er i forbindelse med brugertesten blevet formuleret en række opgaver, som testpersonerne skal forsøge at gennemføre. Opgaverne udspringer af en række typiske brugsscenarier opstillet af en produktchef, der har stort kendskab til domænet og brugerne.

Under afviklingen af den heuristiske evaluering har opgaverne været til rådighed for evaluatorene, som har kunnet benytte dem som inspiration. Dette giver en bedre mulighed for at sætte sig i brugerens sted og prøve at fokusere på den sammenhæng, som apparatet indgår i. Det kan muligvis øge chancen for at alvorligheds-kategorisere den bedre.

3.2. Metodebeskrivelse

I de følgende afsnit beskrives de to anvendte metoder i dette projektarbejde mere detaljeret.

3.2.1. Heuristisk evaluering

Heuristisk evaluering er en ekspertvurdering av et produkt hvor en systematisk sjekker produktet opp mot en liste over heuristikker, det vil si viktige prinsipper for brukervennlighet.[Heuristisk]

I utgangspunktet kommer ordet "to heuristic" fra gresk og betyr å finne eller oppdage. En heuristisk evaluering er brukt for raskt å komme frem til å løsning og er relatert til alminnelig strategier eller metoder for å løse problemer. "A heuristic" som et substantiv er en "tommel finger regel".

Bruken av denne metoden for evaluering av brukskvalitet for web ble introdusert for av Jacob Nielsen og Rolf Molich (1990), og senere revidert av Nielsen (1994). Metoden presenterte de 10 heuristikkene som brukes i dag.

Heuristiske evalueringer letter arbeidet i en evaluering, etter som man har visse retningslinjer å forholde seg til, og som peker på viktigheten i et system. Metoden er en teknisk evaluering som har til hensikt å avdekke både små og store problemer.

Testingen utføres som regel av en ekspert eller en med faglig tyngde. Selve evalueringsprosessen gjøres alene for å unngå påvirkning fra andre.

Heuristisk evaluering bør alltid utføres av flere, da en vil kunne avdekke flere feil. En alene vil ikke klare å oppdage alle feilene, og det anbefales å benytte fra 3 til 5 testere. Hvor mange som bør benyttes kommer an på hvor komplekst systemet er og kostnadsrammen på prosjektet.

Vanligvis tar en heuristisk evaluering fra en til to timer. Selvfølgelig vil det i noen tilfeller være nødvendig med en lengere evaluering, men det er da som regel bedre å splitte opp evalueringen i mindre sesjoner.



"Go ask your search engine."

Testen kan utføres på flere måter, men Jacob Nielsen anbefaler at en slik evaluering bør gå i flere steg.

- Ved at man først blir kjent med produktet/nettstedet
- Undersøker mer spesifikt med henblikk på de 10 heuristikkene

Den generelle anbefalingen er at man går igjennom det en skal evaluere minst to ganger. I første omgang for å få en total oversikt, og bli kjent med det som skal testes. I andre omgang går en igjennom og evaluerer med fokus og henblikk på de 10 heuristikker.

Resultatet fra en heuristisk evaluering lagres i et skjema hvor brukervennlighets (usability) problemer er listet opp med referanse til de ulike usability prinsippene (heuristikker) som ikke er oppfylt.

Dette gjøres ut fra hvordan testerene oppfatter problemet. Testerene kan ikke bare si at en ikke liker noe eller ikke er oppfylt, men dette må begrunnes. Begrunnelsen må gjøres ut fra heuristikken og bruk av design prinsippene.

Testerene må være så spesifikke som mulig, og hvert usability problem må listes separat og eventuelt relateres til et eller flere prinsipper. Er det et f.eks. flere problem i en dialog prosess skal alle problemene listes opp og forklares og med referanse til hver sin heuristikk.

Når man så er ferdig med å teste så leveres skjemaet inn og kommentarene på de ulike heuristikken blir sammenlignet mot skjemaene fra de andre testerne.

Sansynligheten er stor for at man finner de samme feilene, men med flere testere er det muligheter for å finne eventuelle uoppdagede feil. Vanligvis fanges det opp ca. 80% av problemene og det gjenstår ca. 20%. Og selvfølgelig kan årsaken til at brukbarheten til et nettsted ikke er bra, ligge her. Selvfølgelig er ikke heuristisk evaluering noen garanti for at et perfekt resultat, men det er en helt sikkert en god begynnelse å benytte Jacob Nielsens sine heuristikker og råd for å luke ut diverse problemer.

Etter en avsluttet evaluering kan man benytte en debriefing eller rundspørring blandt de testende. Dette kan foregå som et intervju med ekspertene etter evalueringen, besvaring av spørreskjema eller en åpen diskusjon. Her fokuserer en på de brukbarhetsproblemene som ble avdekket under evalueringen og andre generelle aspekter.

Til slutt så må man gradere problemene etter alvorlighetsgraden, slik at kritiske feil blir løst først.

De 10 heuristikken

I litteraturen finnes det beskrevet veldig mange mange forskjellige heuristikker, men vi har valgt å basere oss på Jakob Nielsens utvalg. [Heuristics]. Dette er 10 generelle prinsipper for design av bruker grensesnitt , og grunnen til at de er kalt "heuristics" er fordi de mer kan betraktes som "tommelfinger regler" enn som spesifikke design retningslinjer. I enkelte tilfeller kan det være behov for å legge til flere ved evaluering av spesielle type produkter.



1. **"Visibility of system status" – "Visning av systemets status"**
Systemet skal alltid gi tilbakemeldinger/status til bruker om hva som skjer.
2. **"Match between system and the real world" – "Systemet skal være tilpasset brukeren"**
Systemet skal snakke brukerens språk, dvs. benytte ord, fraser og konsepter som brukeren kjenner. Informasjonen bør opptre i en naturlig og logisk rekkefølge.
3. **"User control and freedom" – "Brukerkontroll, tydelig merket vei ut/tilbake"**
En bruker bør når som helst kunne komme seg ut av systemet, "emergency exit", uten å måtte gå igjennom hele dialogen, eller gå tilbake.
4. **"Consistency and standards"- "Konsistens og standarder"**
Brukeren skal aldri være i tvil, betydningen av ord, situasjoner og hendelser skal alltid være den samme. Følg platform konvensjoner og lag standarder..
5. **"Error prevention" – "Unngå feil i systemet"**
Selv om systemet har gode feilmeldinger, er det bedre med en god design som forbygger feilene i første rekke.
6. **"Recognition rather than recall" – "Reduser belastningen på brukerens hukommelse"**
Minimer brukerens hukommelses bruk. Dvs. gjør objekter, hendelser og valgmuligheter synlige. Brukeren skal slippe å huske informasjon fra en del av dialogen til en annen. Passende instruksjoner for hvordan systemet skal brukes, skal være synlig og lett å finne.
7. **"Flexibility and efficiency of use" – "Fleksibelt og effektivt å bruke"**
Tilby snarveier og hurtigtaster for ekspert brukerne, slik at systemet fungerer fint både for nybegynnere og eksperter.
8. **"Aesthetic and minimalist design" – "Estetisk og minimalistisk design"**
Dialogen skal ikke inneholde informasjon som er irrelevant eller sjeldent brukt. All unødvendig ekstra informasjon i dialogen konkurrerer med relevant informasjon og reduserer synligheten av denne.
9. **"Help users recognize, diagnose and recover from errors" – "Gode feil meldinger"**. Gode feilmeldinger skal hjelpe brukeren å gjenkjenne, finne årsaken til og rette opp feil"
Feilmeldinger skal være forklart i et klart språk, presist formulert og konstruktivt foreslå en løsning.
10. **"Help and documentation" – "Hjelp og dokumentasjon"**
Selv om det er best at systemet kan benyttes uten dokumentasjon, er det nødvendig å ha dokumentasjon og hjelp. Slik informasjon må være lett å søke i, den må fokusere på de oppgaver brukeren skal løse og må ikke være for stor.

Disse Heuristikker er beskrevet av Jacob Nielsen [Heuristics-list] og [CHI'90]

Alvorlighetsgradering av brukervennlighets problemer

Ved å gradere problemene etter alvorlighetsgraden, kan man få hjelp til å finne ut i hvilken rekkefølge problemene bør rettes. Slik at en løser kritiske feil først og kosmetiske til slutt.

Det viser seg nemlig at visse usability problemer kan ha en ødeleggende effekt på populariteten til et produkt selv om problemene i utgangspunktet er lette å overvinne. En slik feilrangering settes opp som en kombinasjon av hvor ofte problemene forekommer, og betydningen av hvorvidt det blir lett for brukeren å komme forbi dette problemet eller ikke. Det finnes flere forskjellige kategoriseringsmetoder. Den som er presentert her er laget av Nielsen og Molich. [Heuristics-rating]

Hvor alvorlig en feil er, er en kombinasjon av tre faktorer:

- Frekvensen av hvor ofte problemet oppstår, er det ofte eller sjeldent?
- Hvilken innvirkning har problemet for brukeren, er det lett eller vanskelig for brukeren å overvinne dem?
- Hvor ofte oppstår problemet, er det et engangsproblem som brukeren kan overvinne hvis de vet om det, eller vil brukeren gjentatt ganger bli forstyrret av problemet?



Ut fra disse faktorene kan vi lage en graderings skala fra 0 til 4 som kan benyttes for å gradere brukervennlighets problemer.

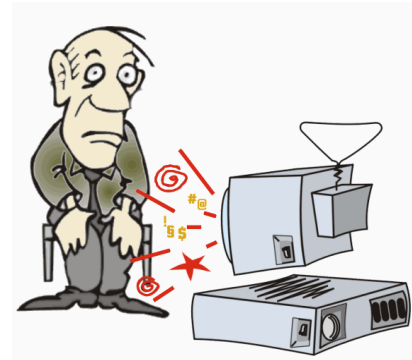
0. Jeg er ikke enig i at dette er et brukerproblem
1. Kosmetisk problem, som kan rettes opp dersom en har ekstra tid til overs
2. Mindre bruker problem, skal gis lav prioritet
3. Hovedproblem, viktig å rette opp skal gis høy prioritet
4. Kristisk problem, helt nødvendig å rette opp

3.2.2. Tenke høyt teknikken

Tenke høyt teknikken går ut på at man tenker høyt, mens en går igjennom testen. [D&R1999 s. 278]

I mange tester, så spør vi testdeltagerne om de kan tenke høyt, slik at vi kan høre hva de tenker når de tester. Vanligvis så tenker en ikke høyt når man arbeider. Selvfølgelig er det veldig forskjellig hva folk er i stand til å fortelle når de skal tenke høyt. De beste høyttankerne kommer med en sammenhengende ord flom, mens andre bare

mumler. Mesteparten er heldigvis midt i mellom. Det er lettere å forstå hvor en testdeltager støter på problemer ved høyt tenkning, og kommentarene kommer også med på et eventuelt video opptak. Å spørre noen om å tenke høyt når de utfører en oppgave kalles "taking a verbal protocol" og analyseringen av dette kalles "protocol analysis". Det har vært mange studier på om det å tenke høyt endrer måten å løse oppgaver på. Noen studier f.eks. Rhenius og Deffner (1990) viser at en bruker noe lengere tid, andre at en er mindre fleksibel med hensyn på endring av løsningsstrategier. Men, imidlertid strategien for å få løst problemet er den samme om man tenker høyt eller ei.



En annen interessant metode innen "tænke høyt" er at testdeltageren er stille under selve testen, derimot skal de kommentere i etterkant når de ser opptak fra testen. Dette var et interessant studie utført av Bowers og Snyder (1990) [D&R1999 s.279]. Det viser seg at disse forklarer bedre hvordan de tenkte, og hadde ofte et forslag til forbedring. Hvor ofte forstår en ikke en problemstilling bedre når en snakker høyt om den. Men selvfølgelig, det tar det mye lengere tid å først utføre testen for etterpå se på videoen og kommentere.

Det er viktig at test deltagerne lære metoden ved å tenke høyt, og en må forklare verdien av dette. En god ide er å kjøre en oppvarmingsoppgave. Noen vil selvfølgelig glemme seg under testen og ikke tenke høyt, da det er mye å tenke på når man tester. Da må test leder komme med forsiktige spørsmål slik at testdeltager kommer på gli igjen.

Analyse av resultatet fra testening utført med "tænke høyt" teknikken kan utføres på flere måter. Vi har bl.a. Video Data Analysis (VDA) og Instant Data Analysis (IDA).

Video Data Analysis (VDA)

Dette er den mest brukte meetode for analysering av brukervennlightes problemer. Vidoen fra usability testen analyseres og det produseres en detaljert log fil. [IDA] Usability problemene blir så beskrevet i detalj, og etterpå blir det rangert i forhold til alvorlighetsgrad, om det er kritisk, alvorlig eller kosmetisk. Til dette kan en benytte verktøyet "Væktøy til problemidentifikasjon", som er beskrevet før. Video Data Analysis er en veldig tidkrevende prosess.

Instant Data Analysis (IDA)

Denne teknikken går ut på å få analysert og dokumentert på en dag. Dette gjør at denne teknikken kanskje ikke er den beste, men at det er en fin metode til å finne de mest kristiske feilene. [IDA]

IDA teknikken er laget for å kunne kombineres med den velkjente "tænke høyt" teknikken. Denne teknikken utnytter det faktum at "tænke høyt" usability testing typisk involverer en test monitor og data logger med et høyt nivå av usability ekspertise. Under usability testen vil data logger og observatører merke seg hendelser og problemer. Testen er som regel utført med fra 4 til 6 test deltagere.



IDA teknikken utføres rett etter at usability testen er ferdig, allerede på samme dagen. Analyse sesjonen utføres som en brainstorming og varer i ca. en time. Denne ledes av en facilitator og utføres sammen med test monitor, observatører og data logger. Facilitator har ikke vært til stede under selve testingen, men spør spørsmål for å klargjøre. Det blir benyttet f.eks. whiteboard og flip-over for notering, screenshots, i tillegg kan det benyttes mind-map som notats teknikk. Mindmap er bl.a en

foretrukket metode ved Universitetet i Aalborg av Rune Thaarup Hoeegh. Selve metoden beskrevet i artikkelen [IDA] lister problemene i en listeform.

Etter at brainstormingen og analyse sesjonen er ferdig tar facilitator og bruker ca. 1 – 1½ t for seg selv og oppsummerer problemene i en liste. Listen består av problemet med en kort beskrivelse og klare referanser til systemet. Til slutt går test monitor, data logger og facilitator igjennom problem listen.

Formålet med sesjonen er at de produserer en liste med brukervennlighetsproblemer. Etterpå blir hvert problem rangert etter alvorlighet, fra kritisk, alvorlig eller kosmetisk. Poenget er ikke å identifisere så mange usability problemer som mulig, men å finne de mest kritiske.

Det viser seg at en som regel finner 85% av de kritiske og 68% av de alvorlige feilene, og i tillegg at det tar ca. 10% av tiden det tar å benytte vanlig videoanalyse.

3.3. Brugbarhedsproblemers omfang og alvorlighed

Dumas og Redish [D&R1999, s. 322-326] beskriver, hvordan man kan kategorisere brugbarhedsproblemer ud fra omfanget af problemet set i forhold til systemet (scope) og ud fra alvorligheden (severity).

3.3.1. Lokale og globale brugbarhedsproblemer

Omfanget af et brugbarhedsproblem kan karakteriseres som lokalt eller globalt. Lokale brugbarhedsproblemer er karakteriseret ved at være begrænsede til et enkelt sted eller en enkel funktionalitet. Det kan f.eks. være en uklar formulering, der gør at brugeren kan komme i tvivl et sted, hvor han skal foretage et valg. Lokale brugbarhedsproblemer er ofte relativt lette at rette op.

Globale brugbarhedsproblemer er karakteriseret ved at være mere gennemgribende. Det har ofte en konceptuel karakter, som gør, at der er noget grundlæggende i interaktionsdesignet, der ikke fungerer. Globale brugbarhedsproblemer kan være svære at gennemskue, da der i brugbarhedsevalueringen kan være rapporteret flere tilsyneladende forskellige lokale problemer, men som set i et større perspektiv skyldes det samme.

Dumas og Radish anbefaler, at man er omhyggelig med at undersøge alle brugbarhedsproblemer for, om de er en del af et globalt problem, så man ikke blot laver symptombehandling, hvis man retter problemerne i et lokalt perspektiv.

3.3.2. Brugbarhedsproblemers Alvorlighed

Dumas og Redish foreslår at inndeile brugbarhedsproblemer i fire niveauer [D&R1999, s. 324]:

1. Afskærer for gennemførsel af opgaven.
2. Skaber betydelig forsinkelse og frustration.
3. Mindre indflydelse på brugbarheden.
4. Ubetydelig - evt. en mulig forbedring i fremtidig version.

I et tiltag for at gøre sådan en kategorisering meget operationel, har Stage og Skov [S&S2010] foreslået et skema, der kaldes Værktøj til Problemidentifikation (se figur 3.1). Her er der en inddeling i tre niveauer:

- Kritisk
- Alvorligt
- Kosmetisk

Værktøjet er især tænkt anvendt i forbindelse med analyse af videooptagelser fra tænke-højt-brugertest. Ifølge Stage [Stage] skælnes der ikke mellem lokale og globale brugbarhedsproblemer.

| | Forsinkelse i at løse opgaven (relativt til testpersonens normale arbejdstempo) | Forståelse | Frustration | Testleder |
|------------------|---|--|--|--|
| Kritisk | Forhindres i at løse opgaven. | Forstår ikke, hvordan informationer i systemet kan bruges i løsningen af en opgave. Gentager de samme informationer forskellige steder i systemet. | Bliver meget frustreret over, at hun ikke kan finde ud af det, og opgiver helt at løse opgaven. | Får betydelig hjælp af testlederen (kunne ikke have løst opgaven uden). |
| Alvorligt | Forsinkes i adskillige sekunder. | Forstår ikke, hvordan en bestemt funktionalitet fungerer eller aktiveres. Kan ikke forklare systemets virkemåde. | Bliver tydeligt irriteret over noget man ikke kan gøre eller huske eller noget ulogisk man skal gøre. Tror at hun har ødelagt noget. | Bliver ledt på vej af testlederen. |
| Kosmetisk | Forsinkes i nogle få sekunder. | Gør ting som ikke kan forklares (det skal man bare). | Bliver lidt irriteret over måden noget skal gøres på. | Testlederen stiller et spørgsmål, som får testpersonen til selv at komme på løsningen. |

Figur 3.1. Værktøj til Problemidentifikation [S&S2010].

Formålet med at inddele brugbarhedsproblemerne i alvorlighedsniveauer er at kunne prioritere vigtigheden af at få dem løst. Kritiske brugbarhedsproblemer vil formentlig holde produktet tilbage fra frigivelse til produktion, mens kosmetiske brugbarhedsproblemer ofte vil kunne accepteres i det endelige produkt - især hvis det let kan opdateres ude hos brugerne gennem en software-opdatering hentet via Internettet.

For at kunne sammenligne de forskellige brugbarhedsevalueringsmetoder, er det vigtigt at benytte en ensartet kategorisering for at kunne sammenligne resultaterne. I dette projekt benyttes værktøjet i figur 3.1.

3.4. Praktisk gennemførelse af brugertesten

Dette afsnit beskriver den praktiske gennemførelse af brugertesten. Først selve afviklingen af brugertesten, og herefter den praktiske gennemførelse af data-analysen.

3.4.1. Afvikling af Brugertesten

Opgaverne til brugertesten blev til på baggrund af et udkast fra produkt-/forretningschefen. Han hjalp også med at skaffe testpersoner. Der endte med at være i alt fire testpersoner. Det var alle meget erfarne guitarister. To af dem var ikke professionelle, mens de to andre i gennemsnit optræder et par gange om ugen.

Der var afsat én time til hver tester, og hele brugertesten blev afviklet i løbet af én dag. Dagen startede med en pilottest, hvor produktchefen var testperson - han havde ikke selv tidligere prøvet apparatet. Denne pilottest blev benyttet til at tjekke, at alt udstyr virkede, og at opgaveark og spørgeskema var ok.

Når hver af testpersonerne ankom blev de vist hen til et forrum til testlaboratoriet. Her blev der udleveret en samtyggeerklæring (se bilag 1), som blot skulle læses igennem, og det første spørgeskema med spørgsmål til deres baggrund. De blev herefter vist ind i testlaboratoriet af testlederen, hvor de hilste på de to observatører. Figur 3.1. viser et billede af testlaboratoriet, hvor testpersonen sad på den røde stol og testlederen på den sorte stol. I baggrunden ses vinduet ind til observatørrummet og stativet med videokameraet. De fik udleveret arkene med opgaverne - der var én opgave per ark. Testlederen bad nu testpersonen om at være opmærksom på at "tænke højt" mens han løste opgaverne. Efter opgaverne var løst, fik testeren udleveret et andet spørgeskema, hvor der for hver af opgaverne blev spurgt, om den havde været let at løse og om, hvor vigtigt det var, om den var let at løse. Begge disse spørgsmål skulle besvares på en fem-punkt Likert-skala. Desuden var der et kommentarfelt. Endelig var der nogle generelle spørgsmål bl.a. om oplevelsen med at benytte en looper. Appendiks 1 i denne rapport indeholder spørgeskemaspørgsmålene med svar og opgave formuleringerne.

Det varierede meget, hvor hurtige testpersonerne var til at gennemføre opgaverne. Den hurtigste tog cirka en halv time, og skulle ikke have hjælp undervejs, mens den der tog længst tid tog knap fem kvarter, og skulle hjælpes flere gange undervejs for at kunne komme videre. Alle opgaverne blev dog løst af alle testpersoner.

Under brugertesten tog de to observatører noter, og den ene observatør styrede også optagelse af video. Testlederen tog ikke notater, men koncentrerede sig udelukkende om at guide testpersonerne gennem testen som beskrevet af Kjeldskov et al. [IDA, s.4].



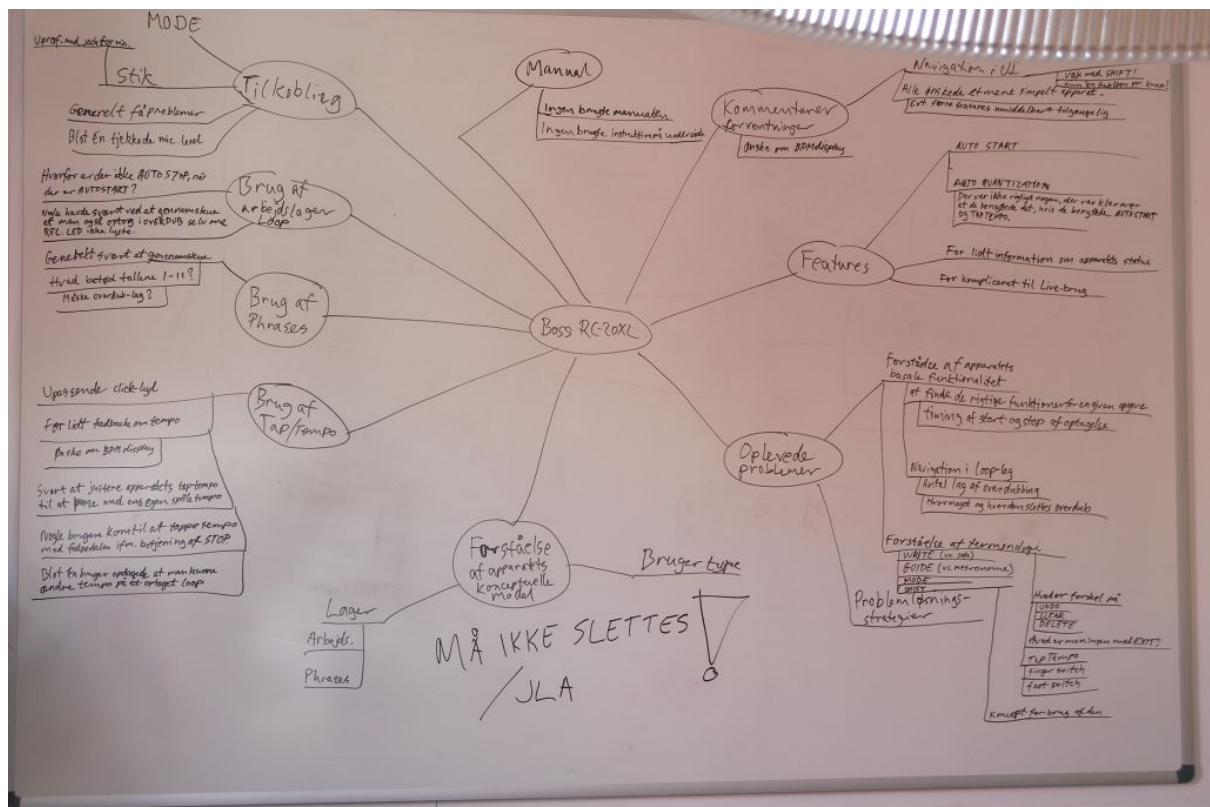
Figur 3.2: Brugertest-faciliteter. Ved siden af Boss RC-20XL på gulvet ligger opgaverne og manualen til apparatet - ingen af testpersonerne kiggede i manualen! Den røde stol blev benyttet af testpersonen, den sorte stol af testlederen. Bag den mørke røde sad den ene observatør, der også styrede videooptagelsen. Videokameraet står på en trefod fjernest i billedet. Den anden observatør stod ca. der, hvor dette billede er taget fra.

3.4.2. Analyse af data fra brugertesten

Projektgruppens oprindelige tanke var at lave Instant Data Analysis (IDA) som beskrevet af Kjeldskov et al. [IDA] umiddelbart efter brugertestene var afviklet. Da én af projektgruppens medlemmer ikke havde mulighed for at være til stede under afviklingen af brugertesten, valgte vi af praktiske årsager at dele analysen i to dele: Første del var en brainstorm, hvor testlederen og de to observatører deltog. Anden del var en Skype-konference, hvor det tre projektgruppemedlemmer deltog.

Første del af data-analysen foregik i et almindeligt mødelokale, hvor der blev lavet en mindmap (se figur 3.3). I modsætning til Kjeldskov et al. fremgangsmåde, blev der ikke endnu lavet en fælles, prioriteret problemliste. De følgende dage fik det sidste projektgruppemedlem gennemset videooptagelserne af brugertesten, og tog observatørnotater. Af praktiske årsager gik der ca. 1½ uge fra afviklingen af brugertesten indtil der blev forsøgt afholdt Skype-konference, hvor de tre projektgruppemedlemmer på baggrund af observatørnoter og mindmap skulle have lavet en fælles problemliste med alvorligheds-kategorisering i kitisk, alvorligt og kosmetisk.

Desværre var der så alvorlige tekniske problemer, så til trods for at det havde være testet af i forvejen, lykkedes det ikke at for afviklet en Skype-konference. Som nødløsning blev i stedet noterne benyttet, og der blev lavet en fælles liste på baggrund af i alt to sæt observatørnoter fra brugertesten, en mindmap fra brainstorm umiddelbart efter brugertester samt et sæt observatørnoter fra gennemsyn af videooptagelser. Problemlisten kan ses i afsnit 4.2.



Figur 3.3. Mindmap fra brainstorm umiddelbart efter brugertesten.

4. Resultat af Evaluering

Dette kapitel præsenterer brugbarhedsproblemlisterne fra hhv. heuristisk evaluering og tænke-højt-testen m. IDA. Brugbarhedsproblemerne er kategoriseret ens på baggrund af Værktøj til Problemidentifikation, som er vist i figur 3.1. Kapitlet slutter med en sammenligning af de to lister.

4.1. Problemliste - Heuristisk evaluering

Alle tre medlemmer af projektet lavede heuristiske evaluering af Boss RC-20XL. Dette er det samlede resultat af disse evalueringer. Er fejlen identificeret af flere brugere vil det fremgå af kategorien. For overskuelighedens skyld er fejlene samlet på heuristikker. Inden for en heuristik er fejlene ordnet med de mest alvorlige først. I kategori-kolonnen er med almindelig type angivet projektgruppens kategorisering, mens der med fed, kursiv er angivet produktchefens kategorisering.

| Id | Beskrivelse | Heuristik | Kategori |
|-----------|---|---------------------------------|---|
| 1 | Der benyttes de samme indikatorlamper til at vise Mode og Stop Mode, hvilket betyder at Stop Mode valget er skjult for brugeren det meste af tiden. | 1 (Visibility of system status) | Kritisk Kritisk |
| 2 | Det er svært at gennemskue om Phrase Select har to niveauer (eller banker) ud fra Loop Phrase og One Shot Phrase indikatorlamperne, eller om disse lamper blot reflekterer den systemtilstand, som vælges med Loop / One Shot under Auto Start knappen. | 1 (Visibility of system status) | Kritisk Alvorlig |
| 3 | Hvis man har flere end en overdub buffer: Hvordan ser man det på apparatet? | 1 (Visibility of system status) | Kritisk Kritisk |
| 4 | Man kan ikke se, om der er optaget noget i den midlertidige hukommelse. | 1 (Visibility of system status) | Kritisk, Alvorlig (Set af 2 evaluatore) Kritisk |
| 5 | Nederst mellem pedalene var det "Rec" (rød) når jeg tar opp, "Play" (grønn) når jeg spiller av og "Overdub" (gul), var noe vanskelig å se hvis man benyttet pedalene. | 1 (Visibility of system status) | Alvorlig Alvorlig |
| 6 | Lysniveauet i trykknapperne (som f.eks.) Reverse er væsentligt svagere end i de forskellige indikator LED. De er svære at se i kraftig lys. | 1 (Visibility of system status) | Kosmetisk (Set af 2 evaluatore) Kosmetisk |
| 7 | Når man holder Stop/Clear fodkontakten nede i nogle sekunder blinker Shift Delete knapperne på en måde, så man kunne tro, at apparatet er i stykker - hurtig og med svag lysstyrke. | 1 (Visibility of system status) | Kosmetisk Kosmetisk |
| 8 | Røde lysdioder til venstre er for svage og matcher dårligt til den røde plade. | 1 (Visibility of system status) | Kosmetisk Kosmetisk |

| | | | |
|----|--|---|-------------------------------|
| 9 | Systemet var ikke alltid helt tilpasset meg, da det var en del ord og fraser jeg ikke kjente. Men, sansynligvis vil dette være enklere for en som er kjent med konseptet i utgangspunktet. | 2 (Match between the system and the real world) | Kosmetisk Alvorlig |
| 10 | Når jeg skulle slette en "Phrase" skulle jeg holde "Shift(EXIT)" og "Delete" nede samtidig. Begge knapper begynte å lyse rødt, noe som ikke sluttet før jeg trykker ned "Exit" igjen. | 3 (User control and freedom) | Alvorlig Alvorlig |
| 11 | Når jeg skulle lagre en "Phrase": Da jeg trykket "Write" begynte den å blinke rødt og sluttet ikke før jeg trykket "Exit". | 3 (User control and freedom) | Alvorlig Alvorlig |
| 12 | Systemet hadde en "Exit" knapp mitt på, denne burde muligens vært i en annen farge | 3 (Usercontrol and freedom) | Kosmetisk Kosmetisk |
| 13 | Exit buttons primære funktion er som "cancel" ved viktige operation, som skal bekræftes. Ellers brukes den kun som "shift". den burde have et annet navn og man burde kunne lave et bedre knapp design | 4 (Consistency and standards) | Kritisk Alvorlig |
| 14 | Når det gjaldt knappene så hadde de to verdier alt etter som "Shift(EXIT)" var brukt dvs. "EXIT" knappen fungerte som "Shift" sammen med andre knapper. ("REVERSE" – "Reverse", "TAP TEMPO"- Opptaks tempo, "WRITE" – Lagre eller save, "EXIT" – Exit, "AUTO START" – Auto start. "Shift(DELETE)" – sletter phrase, "Shift(LOOP/ONE SHOT)" - switzer mellom å stille av en phrase en gang fra start til slutt, eller å spille den om og om igjen i loop. Shift(STOP MODE)" – denne stopper også avspillingen, men du kan velge å bruke denne etter fadeout.) | 4 (Consistency and standards) | Alvorlig Kritisk |
| 15 | Knappen "MODE" viste seg å være opptaks metode for lyden. Slik som "Normal", og "Center cancel" var greit, men "Flat amp simulate" sa meg ingen ting. Disse stod grupert under "Aux in". Men, det viste seg at en kunne velge "Inst" og "Mic" også. Her var ikke heller mappingen helt tydelig. | 4 (Consistency and standards) | Alvorlig Alvorlig |
| 16 | Ved bruk av "Shift(GUIDE SELECT)" - endring av guide tone. Denne må trykkes ned flere ganger, da guide tone endrer seg 8 typer. Dette er ikke logisk, knapp burde vært markert anderledes. | 4 (Consistency and standards) | Alvorlig Alvorlig |
| 17 | Loop Phase LED og One Shot Phrase kan hver betyde tre forskjellige ting og sammen betyde to ting, hvilket i alt giver otte forskjellige informationer. | 4 (Consistency and standards) | Alvorlig Kritisk |
| 18 | Loop Quantize LED er placeret ved siden af Guide drejeknappen, med de har ikke umiddelbart noget med hinanden at gøre. Bryder med gestaltlovene. | 4 (Consistency and standards) | Alvorlig Alvorlig |

| | | | |
|----|---|---|-------------------------------|
| 19 | Ellers så manglet Phrase recorderen en "Av/På" knapp. | 4 (Consistency and standards) | Kosmetisk Kosmetisk |
| 20 | Mangel på konsistens på en del av lede tekstene. F.eks. "Level" - Av spillings volum, play volume, "Guide" - volume på guide tonen, "Input level" - input volume, "Mic" - input volume mikrofon, "Inst"- input volume instrument o.s.v. | 4 (Consistency and standards) | Kosmetisk Kosmetisk |
| 21 | Baksiden var ganske godt merket, med hvilke kabler som skulle inn hvor. Det eneste jeg ble i tvil om var "Output" og en pil med "Power on" | 4 (Consistency and standards) | Kosmetisk Kosmetisk |
| 22 | Knapper til at skruer op og ned for lyden mangler startmarkering og retning | 4 (Consistency and standards) | Kosmetisk Kosmetisk |
| 23 | I højre side af apparatet er der noget rod med hvad tekstene binder sig til. visuelt læner "Input level" sig mere til tingene nedenunder, fremfor de to dreje knapper ovenfor. | 4 (Consistency and standards) | Kosmetisk Kosmetisk |
| 24 | Den hang seg hvis det stod på uten at den ble brukt. Måtte da slå strømmen av og på. | 5 (Error prevention) | Alvorlig Kritisk |
| 25 | Når man slukker apparatet: det er en feil at man kan miste alle data, hvis den står i en forkert state (se advarsel side 11 i manualen) | 5 (Error prevention) | Alvorlig Kritisk |
| 26 | Hadde litt problemer med å huske hva de forskjellige knappene betydde, da det ikke var helt entydig. I tillegg var det en del innebygde funksjoner som fungerte ved å trykke på en knapp flere ganger f.eks. "TAP TEMPO". | 6 (Recognition rather than recall) | Alvorlig Kosmetisk |
| 27 | Når man velger stopmode: Man får ingen visuell beskrivelse av hva de forskjellige knapper betyr, man skal lese det i manualen og huske det | 6 (Recognition rather than recall) | Alvorlig Kosmetisk |
| 28 | Man mangler også informasjon hva der er gemt. Et muligt forslag kunne være en indikation af hvor lange nogle optagelser der er gemt i bufferen/phrases | 6 (Recognition rather than recall) | Kosmetisk Kosmetisk |
| 29 | Det var litt vanskelig for meg å bedømme om det var funksjoner her som var irrelevante, da jeg ikke har helt oversikt over hvilke funksjoner som er mest brukt. | 8 (Aesthetic and minimalist design) | Kosmetisk |
| 30 | Hvad betyder det, at teksten ved nogle knapper er skrevet med inverteret skrifttype. | 8 (Aesthetic and minimalist design) | Kosmetisk Kosmetisk |
| 31 | Tap Tempo knappen har en skjult funtionalitet kaldet Time Signature. | 9 (Help users recognize, diagnose, and recover from errors) | Kritisk Kritisk |
| 32 | Ingen mulighet for feilmeldinger, fantes ikke noe display | 9 (Help users recognize, diagnose, and | Alvorlig Alvorlig |

| | | | |
|----|--|-----------------------------|----------------------------|
| | Det fantes en rekke status lamper, men tror ikke disse hadde noen funksjon ved eventuelle feil. | recover from errors) | |
| 33 | Side 17 i manual: Hva skjer der når man overdubber mere end een gang? sletter man det der ligger i bufferen, ligges det oveni eller ligger der to midlertidige buffere? Har man to buffere mangler jeg indikation af hvor mange buffere man har. (Havde jeg lyd på kunne jeg have eksperimenteret mig frem til det rigtige). Jeg forventer man kun har een temporær buffer, men så forvirrer det der der står om fortryd og fortryd-fortryd mig på side 17 | 10 (Help and documentation) | Alvorlig Alvorlig |
| 34 | Manglet norsk dokumentasjon. Denne var kun på engelsk og tysk. Quick start guide var kun på tysk. Fant pluselig også noe informasjon under, som ikke var så klart formulert. | 10 (Help and documentation) | Kosmetisk Kosmetisk |
| 35 | Manual side 13: "press the MODE button to get the Inst & Mic to light". hvorfor? Hva har det med stopmodes at gøre? (det fandt jeg ud af senere, men jeg var forvirret, derfor fejl). | 10 (Help and documentation) | Kosmetisk Kosmetisk |

Det er slående at kun to fejl ud af 37 blev fanget af mere end én evaluator. Dette indikerer at man med fordel kunne lade flere evaluere apparatet.

4.2. Problemliste - Tænke højt test m/Instant Data Analysis

Brukertesten ble utført i et laboratorie med 4 test brukere og to av prosjekt medlemmene til stede. Den ene som test leder og den andre som observatør. Den tredje prosjekt deltager observerte ved hjelp av videooptaket fra testen. Her er det samlede resultatet fra denne brukertesten.

Den første oppgave bestod i at brukeren skulle gætte hvad de forskjellige knapper skulle brukes til. de ting man fandt under denne oppgave gemmes i en separat tabel. Brukeren kom også med mange kommentarer under testen. Vi vil lave en liste med de mest interessante af disse.

| | Usability problem | | | | | |
|----|--|------------------|---|---|---|---|
| Id | Beskrivelse | Kategori | 1 | 2 | 3 | 4 |
| 1 | Vælger den forkerte input mode ved tilkobling af apparatet | AlvorligAlvorlig | | x | | |
| 2 | Opfatter "Tap lyden" som den optagede lyd | Alvorlig | x | | x | |
| 3 | Ved optagelse af lyd bliver brugeren frustreret over at man ikke kan optage (Få "rec" lampen til at lyse). Problemet var at der allerede var noget i bufferen, derfor kunne man kun overdubbe og afspille. | Kritisk | x | x | | |
| 4 | Kunne ikke styre tempo på Tab lyd | Alvorlig | x | | | |

| | | | | | | |
|----|--|-----------|---|---|---|---|
| 5 | Kunne ikke bruge autostart funktion, anammede ikke at man skulle trykke på venstre pedal, for at looperen var klar til at starte. | Kosmetisk | x | | | |
| 6 | Havde dårlig forståelse for pedalernes funktion. | Kosmetisk | | | | x |
| 7 | forventede at "start" og "stop" lå på samme pedal | Kosmetisk | | | x | x |
| 8 | Manglede noget feedback ved "Clear" (var uklar på om det var gjort eller ej) | Kosmetisk | | | | x |
| 9 | Ville slette en phrase, men gemte i stedet (forventede dobbelttryk på "write" ville aktivere den sekundære funktion "delete") | Kritisk | | | x | x |
| 10 | Brugeren kunne ikke forstå forskellen mellem "record" og "overdub" | Kosmetisk | | | x | |
| 11 | Det var uklart hvad men skulle gøre for at starte med at optage. | Kosmetisk | | | | x |
| 12 | Fejlede med at slette, idet man ikke indså hvordan man gjorde. | Alvorlig | | x | x | |
| 13 | Brugeren ville have Looperen i "Overdub mode" for at slette sidste overdub. derved kom brugeren til at ligge et nyt overdub ind, som han slettede. Resultat: Der skete tilsyneladende ingenting ved at slette det der lå i bufferen. | Kosmetisk | | | | x |
| 14 | Brugeren kom til at skifte tempo ved at trykke adskillige gange på højre pedal. | Alvorlig | | | | x |
| 15 | Brugeren forsøger at slette det sidste ved at vælge overdub og derefter aktivere "Delete" (han skulle have holdt højre pedal nede i 2 sekunder i stedet). | Alvorlig | | | | x |
| 16 | Ventede for længe med at spille ved overdub, der var forvirring over hvad der reelt blev overdubbet. | Kosmetisk | | | | x |
| 17 | Kom til at slette indhold i buffer ved at skifte phrase | Kritisk | x | | | x |
| 18 | Ved at gemme bufferens indhold i phrase:Trykker "write", ser "write" blinke og trykker på "exit" for at få den til at holde op. (derved fortryder man gem, den rigtige adfærd er at trykke på "write" igen for at bekræfte at man vil gemme) | Kosmetisk | | | | x |
| 19 | Brugeren havde en forventning om at alt hvad der blev overdubbet lå som separate lag, var frustreret over at han ikke kunne administrere disse lag. | Kosmetisk | | | | x |
| 20 | Med det formål at gemme indholdet i bufferen trykkede brugeren een gang på "write" derefter skiftede han phrase, men trykkede ikke på write igen. Han skiftede phrase igen, hvorved bufferen blev slettet. | Alvorlig | | | | x |

4.2.1. Tidforbrug på de enkelte opgaver

| | 1 | 2 | 3 | 4 |
|----------|------|------|-------------------------------|--|
| Opgave 1 | 0:45 | 1:50 | 2:25 | 3:15 |
| Opgave 2 | 8:05 | 2:35 | 23:05 (Test leder griber ind) | 14:50 |
| Opgave 3 | 0:20 | 1:00 | 1:00 | 1:00 |
| Opgave 4 | 0:30 | 0:35 | 0:30 | 15:30 |
| Opgave 5 | 0:35 | 0:45 | 0:10 | 0:40 |
| Opgave 6 | 1:50 | 1:00 | 3:00 | 11:30 |
| Opgave 7 | 2:20 | 1.00 | 12:00 | 4:30 (Testleder stillede et ledende spørgsmål) |

Der var en slående forskel på de forskellige testbrugere. De to første var ret kendt med at bruge pedaler, mens det var relativt nyt land for de to sidste brugere. Det store tidforbrug nogle steder dækker over at der gik "legestue" i den, hvor brugerne ikke var strengt fikseret på opgaven, men i stedet legede med looperen. Dette gjaldt i særdeleshed for de sidste to brugere. Man skal derfor være varsom med at overfortolke disse tidsrum.

Det man kan sige er at det store tidsforbrug ved opgave 2 hos alle brugere dækker over at de til at starte med har lidt svært ved at bruge looperen. Opgave 2 er den første opgave hvor de reelt skal bruge looperen til noget.

Gennemgående kan man sige at testbrugerne typiske tilgang til at bruge apparatet var at man legede og gættede sig til hvordan man skulle bruge det. Det er f. eks. slående at alle uden undtagelse komplet ignorerede manualen, som lå lige ved siden af apparatet. Dette bør man tage hensyn til når man designer udstyr til denne målgruppe.

4.3. Feedback fra produktchef

Inden færdiggørelsen af denne rapport, var det muligt at få produktchefens feedback på problemlisten fra den heuristiske evaluering - det kan ses afsnit 4.1. Problemlisten for den heuristiske evaluering indeholder i alt 35 brugbarhedsproblemer. Af disse var produktchefen enig i alvorligheds-kategoriseringen af de 26, hvilket svarer til ca. 74%. Der, hvor der var uenighed var der i alle tilfælde blot ét trin til forskel. Der var med andre ord ikke brugbarhedsproblemer, der blev vurderet kosmetisk af den ene part og kritisk af den anden part.

4.4. Sammenligning af resultater

I denne figuren har vi sammenlignet resultatene fra de to testene Heuristisk evaluering og Tænke højt teknikken m/Instant Data Analysis. Hver kolonne repræsenterer et usability problem. En sort firkant identificerer at problemet var identificeret ved den specifikke teknikken, og en hvid firkant at problemet ikke var identificeret med denne teknikken, men funnet ved hjælp af den andre teknikken.

Heuristisk evaluering



Kritisk



Alvorlig



Kosmetisk

Tænke højt m/IDA



Kritisk



Alvorlig



Kosmetisk

Refleksion over forskellene mellem de to test resultater

I alle tre kategorier (kritisk, alvorlig og kosmetisk) finder heuristisk evaluering flere brugbarhedsproblemer end tænke-højt-test m.IDA. Umiddelbart overrasker dette resultat, idet det adskiller sig væsentligt den konklusion, som Karat et al. kommer frem til: "Den empiriske test identificerede det største antal problemer, og identificerede et afgørende antal relativt alvorlige problemer, som blev overset af inspektionsmetoderne." [CHI'92, s.402] Denne forskel kan næppe alene tilskrives, at Karat et al. har benyttet en mere grundig analyse end IDA ifm. den empiriske test.

At sammenligne de to resultater og finde ud af hvor man havde fundet den samme fejl var langt fra trivielt. Ved den heuristiske inspektion fokuserede man på detaljer i brugerinterfacet. Ved IDA testen fokuserede man på hvad brugeren lavede af fejl, eller på hvad det var han ikke kunne gøre. Ved begge tests blev der fundet mange fejl som ikke blev fundet ved den anden metode. Man kan spørge sig om hvorfor.

Kigger man på de fejl som kun blev fundet ved den heuristiske inspektion så er der mange detaljer ved brugerinterfacet, som slet ikke kom i spil under testen. Eller også var af en natur, som ikke nødvendigvis vil medføre problemer for brugeren.

Kigger man på fejl som blev fundet under IDA testen kan se at de også var af nogle bestemte typer:

- Fejl, hvor problemet for brugeren var at gennemskue hvordan en handling skulle udføres. (eks: "Hvordan sletter jeg en phrase?")
- Fejl, hvor en operation krævede at brugeren udførte mere end een handling på brugerinterfacet. (eks: Hvis man gemmer en phrase skal man bekræfte ved at trykke på write igen).
- Fejl hvor brugeren ved fejl kom til at udføre operationer, han ikke havde intensioner om. (eks: brugeren kommer til at slette bufferens indhold når han skifter til en anden phrase).

Man kan spørge sig selv hvad man kan gøre for at finde disse fejl under heuristisk inspektion. Der er flere bud:

- Man kan bygge nogle scenarioer, som de opgaver man bruger under brugertesten. Derved kan man nemmere se om der er nogen basale operationer der ikke er selvindlysende i brugereinterfacet.
- Man kan stille sig nogle andre spørgsmål under inspektionen f. eks. Metaphors of human thinking.

Endelig skal man ikke glemme at der var en stor forskel på eksperterne der udførte den heuristiske inspektion og brugerne under IDA testen: Den første gruppe havde læst manualen, og sat sig ind i funktionen på den måde. Brugerne, alle som een, prøvede sig frem.

5. Konklusion

Alle tre medlemmene av prosjektgruppen gjennomførte en Heuristisk evaluering av Boss RC-20XL. Dette førte til at vi samlet fant 35 feil og derav 6 kritiske, 14 alvorlige og 15 kosmetiske. Det er kanskje slående at kun to feil av de 35 ble oppdaget av mer en en testperson. Dette viser da at vi kunne vært flere testere. For oversiktens skyld er feilene listet opp etter heuristikker.

I kolonnen med kategori er den med almindelig type angitt fra prosjektgruppens kategorisering, mens der med fet, kursiv er angitt av produktsjefen.

Ved tænke-højt testen, som ble analysert ved Instant Data Analysis (IDA) ble det oppdaget 20 feil, derav 3 kritiske, 7 alvorlige og 10 kosmetiske.

Denne brukertesten ble utført i et labratorie med 4 test brukere og to av prosjekt medlemmene til stede. Den ene som test leder og den andre som observatør. Den tredje prosjekt deltager observerte ved hjelp av videooptaket fra testen.

Meningen var at vi skulle utføre IDA ved hjelp av Skype, men desverre hadde vi alvorlige tekniske problemer, til trods for at vi hadde testet det i forveien. Som nødløsning ble det i stedet laget en felles liste på baggrund av i alt to sett observatørnoter fra brukertesten, en mindmap fra brainstorm umiddelbart etter brukertester, samt et set observatørnoter fra gennemsyn af videooptaket.

Det var ikke så lett å sammenligne resultatet fra de to testene for å finne ut hvor vi hadde funnet samme feilen.

Ved den heuristiske inspeksjonen fokuserte man jo mest i detaljer angående brukergrensesnittet, mens man ved tænke-højt testen fokuserte man på hva brukeren gjorde av feil, eller hva han ikke fikk til å utføre. Det viste seg at vi fant flest feil ved den Heuristiske inspeksjonen, men det ble ved begge testene funnet feil som ikke ble funnet ved den andre metoden. Av 35 feil funnet ved heuristisk evaluering ble 10 funnet ved tænke-højt, og av 20 tænke-højt ble 8 funnet ved heuristisk metode.

I afsnit 2.2 blev der beskrevet to særligt vigtige dele af den konceptuelle model for systemet. Det var dels hvordan lydclip bliver gemt på forskellige niveauer og dels hvordan timingen af lydklippene kan styres. Det er begge to afgørende faktorer for, om apparatet er anvendeligt for en guitarist. Brugbarhedsproblemer omkring intuitiv forståelse af disse fundamentale dele af den konceptuelle model blev meget tydeligt afsløret i den empiriske brugertest, mens det slet ikke kom klart til udtryk i den heuristiske evaluering.

Dette understøtter at heuristisk evaluering i højere grad finder lokale end globale brugbarhedsproblemer, som beskrevet sidst i afsnit 3.1.1. - og som Nielsen og Molich skriver om heuristisk evaluering: "Metoden er biased med evaluatorernes aktuelle tankesæt og skaber normalt ikke gennembrud i det evaluerede design." [CHI'90, s.255]

5.1 Videre Arbejde

Dette projekt har fokus på selve evalueringen af brugbarheden. Et naturligt næste skridt vil være at se på, hvordan resultatet af brugbarhedsevalueringen skal anvendes. I nærværende sammenhæng, hvordan listerne med de fundne brugbarhedsproblemer i praksis kan bruges i forbindelse med at få specificeret et nyt produkt, som har et bedre interaktionsdesign end Boss RC-20XL. Noget af det ligger formentlig i den måde, som brugbarhedsproblemerne er formuleret på. Når man har lavet en brugbarhedsevaluering - enten en inspektion eller en brugertest - er ens fokus særlig rettet på konkrete detaljer, og det kommer formuleringerne af brugbarhedsproblemerne let til at være præget af. For at de for alvor kan være en hjælp i enten arbejdet med at specificere et

interaktionsdesign, eller med at implementere det i et apparat, er det vigtigt, at brugbarhedsproblemerne er formuleret med henblik på det formål. Man kan også forestille sig, at det i nogle situationer kan være en hjælp, hvis brugbarhedseksperterne med baggrund i deres fagligt indsigt kommer med idéer løsninger.

Kildeliste

[RC-20XL guide]

Loop Station RC-20 XL Phrase Recorder Owner's Manual, Boss Corporation

<http://no.wikipedia.org/wiki/Overdubbing>

<http://www.bosscorp.co.jp/products/en/RC-20XL/brochures.html>

http://www.sound1.com/portal/asp/produkt detalj1.asp?Produkt_ID=677

[Nielsen HowTo] Jakob Nielsen, How to Conduct a Heuristic Evaluation, useit.com

http://www.useit.com/papers/heuristic/heuristic_evaluation.html (besøgt 2010-05-22)

[D&R1999] Joseph S. Dumas & Janice C. Redish, *A Practical Guide to Usability Testing*, Revised Edition, Interlect Ltd., 1999, ISBN 1-84150-020-8.

[S&S2010] Jan Stage og Mikael Skov, *Værktøj til Problemidentifikation*, 2010, udleveret som undervisningsmateriale ifm. kurset Brugertest.

[CHI'90] Jakob Nielsen & Rolf Molich, *Heuristic Evaluation of User Interfaces*, p. 249-256, CHI '90 Proceedings, ACM, 1990.

[CHI'92] Clare-Marie Karat, Robert Campbell, and Tarra Fiegel, *Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation*, p. 397-404, CHI '92, ACM, May 3 -7, 1992.

[MOT2008] Frøkjær, E. and Hornbæk, K. 2008. *Metaphors of human thinking for usability inspection and design*. ACM Trans. Comput.-Hum. Interact. 14, 4, Article 20 (January 2008)

[IDA] Jesper Kjeldskov, Mikael Skov og Jan Stage, *Instant Data Analysis: conducting usability evaluations in a day*.

<http://www.cs.aau.dk/~jesper/pdf/papers/NordiCHI04-final.pdf>

[Stage] Informationer givet af Jan Stage i forbindelse med seminarerne i faget Usabilitytest, Aalborg Universitet, foråret 2010.

[Heuristics]

<http://en.wiktionary.org/wiki/heuristic>

<http://www.useit.com/papers/heuristic/>

http://www.useit.com/papers/heuristic/heuristic_evaluation.html

<http://www.data.no/Index.aspx?PageID=36&Language=no-NO&PublicationView=Show,102,3>

[Heuristics-list]

http://www.useit.com/papers/heuristic/heuristic_list.html

[Heuristics-rating]

<http://www.useit.com/papers/heuristic/severityrating.html>

Appendiks 1: For-spørgeskema, opgaver og evalueringsspørgeskema til testpersoner

Dette bilag indeholder spørgeskemaerne og svarene i koncentreret form. Opgaverne blev udleveret til testpersonerne i form af 8 sammenhæftede ark, med én opgave på hvert ark. Efter hver opgave stod der på siden: "Fortæl testlederen når du føler, du er færdig med opgaven."

A1.1. Spørgsmål før brugertesten

1. Indtast venligst dit navn: A, B, C, D
2. Hvad er din alder? 38, 36, 28, 21
3. Hvor mange år har du spillet guitar? 25, 25, 18, 10
4. Hvor mange bands spiller du i i øjeblikket? 5, 2, 1, 1
5. Hvor mange gange har du spillet live det sidste års tid? 150, 3, 15, 100
6. Tjener du penge på dit guitar spil? Nej:1, Ja:3
7. Hvor mange guitarer har du? 8, 4, 12, 10
8. Hvilken combo eller amp./cab. har du? ...
9. Hvilke guitar-effekt-pedaler har du? ...
10. Hvor meget erfaring har du med at benytte en looper-effekt?
Har aldrig benyttet en looper-effekt - 1 3 0 0 0 - Benytter en looper-effekt som en fast del af min optræden.
11. Hvor meget kender du allerede til Boss RC-20/RC-20XL?

2 - Kender den ikke.
2 - Kender til den, men har ikke prøvet den.
0 - Har engang prøvet den.
0 - Har selv haft én.
0 - Har selv én, men benytter den ikke så ofte.
0 - Har selv én og benytter den ofte.
0 - Andet.

A1.2. Opgaver og spørgeskema efter brugertesten

Opgave 0: Tag apparatet op og fortæl, hvad du tror man kan gøre med de forskellige knapper, og hvad lamperne kan fortælle.

Opgave 1: Forbind apparatet med kabler, så der kan komme lyd igennem fra guitar og fra mikrofon til forstærker/højttaler.

Meget let - 2 0 2 0 0 - Meget svært
Ikke vigtigt - 0 0 1 1 2 - Meget vigtigt

Opgave 2: Lav et loop i 4/4. Optagelsen af takten skal passe i timing, så når den loopes lyder det rigtigt. De 4 slag i takten kan du evt. lave ved at tromme 1, 2, 3, 4 på guitaren.

Meget let - 1 0 1 2 0 - Meget svært
Ikke vigtigt - 0 0 0 2 2 - Meget vigtigt

Kommentarer: "Jack til mikrofon...!! >-(

Opgave 3: Lav et overdub med guitaren (dvs. optag oven i).

Meget let - 1 1 1 1 0 - Meget svært
Ikke vigtigt - 0 0 0 3 1 - Meget vigtigt

Opgave 4: Slet det du har optaget i overdub (i opgave 3) uden at slette det, du optog først (i opgave 2).

Meget let - 0 1 0 1 2 - Meget svært
Ikke vigtigt - 0 0 1 2 1 - Meget vigtigt

Kommentarer: "Ulogisk, som at lave et @ på en PC!"

Opgave 5: Lav et overdub med mikrofonen.

Meget let - 1 2 1 0 0 - Meget svært
Ikke vigtigt - 0 0 2 1 1 - Meget vigtigt

Opgave 6: Gem det du har optaget i hukommelsen på apparatet. Tjek at det er gemt korrekt ved at optage et nyt (tilfældigt) loop, og derefter at genkalde det, du har gemt i hukommelsen.

Meget let - 1 1 1 0 1 - Meget svært
Ikke vigtigt - 0 1 0 1 2 - Meget vigtigt

Opgave 7: Slet det igangværende loop og optag et nyt (tilfældigt) loop.

Meget let - 1 0 1 1 1 - Meget svært
Ikke vigtigt - 0 1 0 1 2 - Meget vigtigt

Supplerende spørgsmål

Er apparatet indbydende og let at gå til, eller skal du bruge manualen?

Meget let - 0 0 1 1 2 - Meget svært
Ikke vigtigt - 0 0 1 2 1 - Meget vigtigt

Kommentarer: "Igen ulogisk.", "Afhænger af forkendskab til tilsvarende produkter."

Er det sjovt at bruge looperen, bliver du grebet af det?

Meget let - 1 0 2 1 0 - Meget svært
Ikke vigtigt - 0 0 1 1 2 - Meget vigtigt

Kommentarer: "Havde været nemmere at more sig, hvis maskinen havde gjort det nemmere at more sig."

Er alle features på looperen nødvendige, eller kunne man nøjes med mindre?

For få - 0 0 0 3 0 - For mange
Ikke vigtigt - 0 2 1 0 0 - Meget vigtigt

Er det let nok at indstille lydniveauerne (levels) på apparatet?

Meget let - 0 3 0 0 0 - Meget svært
Ikke vigtigt - 0 0 1 2 0 - Meget vigtigt

Andre kommentarer: "Der kunne godt være en demo-mode.", "Lidt svær at komme ind i, men når man først er inde i systemet blir det klart lettere, men en svær effekt at starte ud med."

A1.3. Sammenfatning på Spørgeskemaer

De fire testpersoner er mellem godt 20 og knap 40 år. De har alle spillet guitar siden deres tidlige teenageår, de spiller i et eller flere band og optræder med det. To af testpersonerne er professionelle og optræder live i gennemsnit flere gange om ugen. Ingen af testpersonerne benytter looper-effekt i deres musikudøvelse - en enkelt har enddog aldrig prøvet en looper-effekt før. Halvdelen af testpersonerne har hørt om Boss RC-20XL, men ingen af testpersonerne har prøvet den før.

Én af testerne havde lidt problemer med at koble apparatet til, mens en anden tester ikke brød sig om den anvendte stiktype. Men de to andre testere oplevede det som meget let at koble apparatet til og alle syntes, det var let at indstille lydniveauerne. Generelt opfattes det som vigtigt at kunne tilslutte og indstille apparatet.

At optage et loop (opgave 2) fandt to at testerne var svært, samtidig med at to syntes, at det var vigtigt, at denne opgave var let. Det at optage et loop er den mest grundlæggende og forventelig anvendte funktion på et looper apparat, og man kunne derfor forvente, at det burde være den opgave, der var lettest og mest intuitiv at udføre.

At overdubbe (opgave 3) og efterfølgende at slette det igen (opgave 4) blev oplevet lidt forskelligt, idet det generelt oplevedes som svært at slette det. Begge dele vurderes som at det generelt er vigtigt at det er let at gøre. At optage et overdub med mikrofon (opgave 5 og næsten en gentagelse af opgave 3) blev generelt oplevet som let, hvilket også var forventet.

At gemme et loop på en plads i hukommelsen (opgave 6) blev oplevet meget forskelligt - en enkelt syntes endda, at det var meget svært, og det at slette det igen (opgave 7) blev også oplevet forskelligt, men generelt lidt sværre. Generelt blev det vurderet som vigtigt til meget vigtigt, at det var let at gøre. Dog var der én af testpersonerne, der vurderede, at det ikke var vigtigt.

Generelt oplevede testerne, at apparatet var svært til meget svært at gå til, og de vurderede generelt ikke, at det var specielt sjovt at bruge looperen. De vurderede, at det generelt var vigtigt, at det var let og sjovt. Det var svært at komme i gang med den, og én foreslog, at der kunne være en demo-funktionalitet til at hjælpe én igang.

Bilag 1. Samtykkeerklæring

TC Electronic A/S
Sindalsvej 34
PO Box 2206
DK-8240 Risskov
Denmark
CVR no: 12 90 07 32
phone: +45 8742 7000
fax: +45 8742 7010
tcelectronic.com
tc@tcelectronic.com

tc electronic



AALBORG UNIVERSITET

Samtykkeerklæring ved deltagelse i brugertest hos TC Electronic

Formål

En gruppe studerende ved Aalborg Universitet og TC Electronic beder dig deltage i en brugerundersøgelse af apparater til musikere. Ved at deltage i denne undersøgelse hjælper du os med at forstå, hvordan brugere oplever betjening af apparater til musikere og dermed at gøre vores apparater lettere at benytte og at lære at betjene.

Omgivelser

Brugerundersøgelsen finder sted i TC Electronics teststudie, hvor du bliver observeret, mens du bruger produktet.

Indsamlet Information

Vi vil opsamle information om, hvordan du bruger apparatet. Vi vil bede dig om at udfylde spørgeskemaer og interviewe dig. Vi vil bruge den information, du giver os, sammen med information indsamlet fra andre testere, til at anbefale forbedringer af apparater og udvikle nye apparater.

Videoptagelse

Alt eller noget af dit arbejde med apparatet og interview med dig vil blive videofilmet. Ved at give dit samtykke her, giver du tilladelse til, at de studerende og TC Electronic må benytte din stemme, dine udtalelser og videoklip med dig, men ikke dit navn, med det formål at evaluere apparatet og vise resultatet af vores brugertest.

Fortrolighedserklæring

Denne brugerundersøgelse er primært en del i forbindelse med et uddannelsesforløb ved Aalborg Universitet, og der vil ikke umiddelbart være noget, der kræver at det bliver holdt fortroligt.

Bekvemmelighed

Ved længerevarende brugertest, har vi planlagt pauser, men du må gerne holde en pause, hvis du har behov for det. Du skal blot sige til testlederen, at du ønsker en pause.

Frihed til at afbryde brugertesten

Du må afbryde og trække dig tilbage fra brugertesten til enhver tid.

Frihed til at stille spørgsmål

Hvis du har nogle spørgsmål må du spørge testlederen nu, eller på et hvilket som helst tidspunkt senere under brugertesten.

Bilag 2. Tekniske specifikationer for Boss RC-20XL

- Nominal Input Level Inst: -20 dBu (variable)
- Mic: -40 dBu (variable)
- Aux In: -10 dBu
- Input Impedance Inst: 1 M ohms
- Mic: 1 k ohms
- Aux In: 47 k ohms
- Nominal Output Level Output: -20 dBu
- Output Impedance 1 k ohms
- Recommended Load Impedance 10 k ohms or greater
- Internal Memory Recording time: 16 min.(max.)
- Maximum number of saved phrases 11 phrases (max.)
- Connectors Inst jack (1/4 inch phone type), Mic jack (1/4 inch phone type), Aux In jack (Stereo miniature phone type), Phrase Shift jack (1/4 inch TRS phone type), Reverse jack (1/4 inch phone type), Output jack (1/4 inch phone type), AC Adaptor jack
- Power Supply DC 9V: Dry battery (R6/LR6 (AA) type) x 6, AC Adaptor (PSA-series)
- Current Draw 120 mA (9 V max.)
- Accessories Owner's Manual, Dry battery (AA type) x 6, Sound Library for RC-20XL (Sample Phrase CD)
- Options AC Adaptor (PSA-series), Foot Switch (FS-5U)
- Bredde 173 mm
- Dybde 158 mm
- Høyde 57 mm
- Vekt 1.2 kg