

Kan ökad hörbarhet för TV-dialog erhållas genom användning av en centerhögtalare?

Nils Carlsson

Luleå tekniska universitet

C- uppsats
Ljudteknik

Institutionen för Musik och medier
Avdelningen för Medier och upplevelseproduktion



**KAN ÖKAD HÖRBARHET FÖR TV-DIALOG
ERHÅLLAS GENOM ANVÄNDNING AV EN
CENTERHÖGTALARE?**

Examensarbete ljudteknik – S0035F

av

Nils Carlsson – VT-10

Nils Carlsson, nilcar-7@student.ltu.se
Tel: 070-225 81 38

Institutionen för musik och medier
Luleå tekniska universitet/Luleå University of Technology
POSTADDRESS/POSTAL ADDRESS: SE-97187 Luleå, Sweden
BESÖKSADDRESS/VISITING ADDRESS: Universitetsområdet, Porsön, Luleå
TELEFON/PHONE: +46 (0)920 49 10 00 TELEFAX/FAX: +46 (0)920 49 13 99
HEMSIDA/WEBSITE: www.ltu.se

Abstrakt

TV-tekniken har gjort stora framsteg under 2000-talet. Digital TV-sändning har möjliggjort sändningar med fler ljudkanaler med högre upplösning. För att öka taltydligheten i TV-sammanhang används just flerkanalsljud då man implementerar en centerhögtalare istället för fantommitt. Denna uppsats undersöker vilka fördelar detta får för just taltydligheten. Ett lyssningstest genomfördes där subjekt dels fick bedöma uppfattbarheten i autentiska TV-reportage och dels fick sin egen förmåga att uppfatta tal satt på prov i ett Tal I Brus-test. Resultatet visar att användandet av en centerhögtalare bidrar positivt till taltydligheten i och med stabilare mittförankring och ger tydligare och bättre separation. Fördelarna med centerhögtalaren minskar dock markant om inspelningen är av dålig kvalitet, då även de stör ljud som finns på inspelningen ökar i tydlighet vid användandet av en centerhögtalare.

Innehållsförteckning

| | |
|--|----|
| 1. Inledning och bakgrund..... | 1 |
| 1.1 Reproduktion i två högtalare | 2 |
| 1.2 Hur örat hör riktningar | 2 |
| 1.2.1 Riktningar i horisontalplanet..... | 2 |
| 1.2.2 Riktningar i vertikalplanet | 3 |
| 1.2.3 Elevationseffekt..... | 3 |
| 1.2.4 Presedenseffekt | 4 |
| 1.3 Kamfiltereffekt | 4 |
| 1.5 Centerkanal och centerhögtalare | 5 |
| 2. Forskningsfråga | 6 |
| 2.1 Avgränsningar | 6 |
| 2.2 Definition av begrepp..... | 6 |
| 3. Metod..... | 7 |
| 3.1 Tal I Brus..... | 7 |
| 3.2 Ljudexempel | 7 |
| 3.2.1 Ljudexempel - SVT..... | 7 |
| 3.2.2 Ljudexempel - Tal I Brus | 8 |
| 3.3 Lokal | 8 |
| 3.4 Teknisk utrustning..... | 9 |
| 3.5 Subjekt..... | 9 |
| 3.6 Testets utformning..... | 9 |
| 3.6.1 Reportage från SVT | 9 |
| 3.6.2 Utformning av TIB-test..... | 10 |
| 3.6.3 Övrigt i testutformningen | 10 |
| 3.6.4 Andra faktorer i testmiljön..... | 11 |
| 4.Resultat | 12 |
| 4.1 Resultat för SVT-test | 12 |
| 4.2 Subjektens kommentarer till SVT-test | 13 |
| 4.2.1 Subjekt som föredrar 2.0-versionen av något inslag | 13 |
| 4.2.2 Subjekt som föredrar 3.0-versionen av något inslag | 13 |
| 4.3 Resultat för TIB-test | 14 |

| | |
|---|----|
| 5. Analys av resultat..... | 16 |
| 5.1 Analys av resultat i SVT-test..... | 16 |
| 5.2 Analys av TIB-test..... | 17 |
| 6. Diskussion..... | 17 |
| 6.1 Test med TV-reportage..... | 18 |
| 6.2 Tal med bakgrundsbrus..... | 19 |
| 6.3 Utveckling av undersökningen..... | 20 |
| 7. Slutsatser..... | 21 |
| Referenser..... | 23 |
| Bilagor..... | 24 |
| Bilaga 1 – Svarsformulär för lyssningstestet..... | 24 |

1. Inledning och bakgrund

På senare tid har stora tekniska framsteg gjorts på televisionsfronten. I och med digitaltelevisionens intåg har både bild-och-ljudkvalitet kunnat förbättras avsevärt. Ur ett ljudperspektiv är den största förändringen och förbättringen tillgången till flerkanalsljud vilket ger ljudteknikern verktyg för att producera mer avancerade ljudlandskap.

I Sverige har tekniken slagit igenom stort och Sveriges Television (SVT) producerar och sänder många av sina program i HD-kvalitet (High Definition) och med flerkanalsljud. Samtidigt har man påbörjat ett projekt för ökad hörbarhet för TV-dialog. [1] Detta kommer sig av att klagomål inkommit från tittare som anser sig ha svårt att uppfatta det som sägs i programmen. För att råda bot på detta har ett policydokument tagits fram där ett antal strategier och förhållningssätt för bättre hörbarhet föreslås, där flera direkt hänger ihop med möjligheten att sända just flerkanalsljud. [2]

Ett sätt att komma tillrätta med hörbarhetsproblematiken är att använda flerkanalsljud (minst 3.0 vilket innebär tre fronthögtalare med en centerhögtalare och vänster och höger högtalare 30° åt respektive håll sett från lyssnarpositionen) och i centerkanalen enbart placera dialog (alternativt monolog). Detta gör att konsumenter med en flerkanalsljudanläggning själva kan justera volymen på talet för att förbättra uppfattbarheten. Detta gör dock också att den tänkta ljudbilden förändras varför programmet inte längre motsvarar det som sändes ut. Det kräver också att konsumenten har rätt utrustning och framförallt kunskapen att använda den. Vidare måste dessa justeringar ändras i och med att man byter kanal eller att ett nytt program börjar för att behålla en kontinuerlig nivå och balans.

1.1 Reproduktion i två högtalare

När ljud reproduceras i en 2.0-uppställning (vilket innebär 3.0 enligt ovan, dock utan centerhögtalaren) reproduceras allt ljud i ljudbilden av vänster respektive höger högtalare. Det finns alltså ingen riktig källa som reproducerar ljud i mitten av ljudbilden. Därför talar man ofta om en *fantomkälla* eller en *fantommitt*. Det är istället ett psykoakustiskt fenomen som gör att vi människor uppfattar ljudet som att det kommer från mitten. [3] Nedan följer en redogörelse dels för hur hjärnan tolkar olika hörselintryck så att vi kan höra något mellan två högtalare, och dels vilka negativa effekter detta får för det ljudmaterial som spelas upp.

1.2 Hur örat hör riktningar

För att uppfatta riktningar på ljud använder vi människor olika strategier beroende på varifrån ljudet kommer. Det handlar framförallt om följande två:

- Skillnader mellan höger och vänster öra vad gäller tid/fas samt intensitet/nivå.
- Hur ljudet studsar mot ytterörats olika vindlingar. Detta påverkar ljudet på olika sätt beroende på ljudets infallsvinkel.

Ovanstående går under begreppet Head Related Transfer Functions (HRTF eller "Huvudrelaterade överföringsfunktioner").

1.2.1 Riktningar i horisontalplanet

För frekvensområdet 200 Hz – 2000 Hz (alla frekvensangivelser är ungefärliga då de beror på hur stort/litet varje människas huvud och öra är) använder hörselsinnet skillnader i tid/fas för att avgöra vilken riktning ett ljud kommer ifrån. Ett ljud som har en infallsvinkel på till exempel 30° från vänster kommer att träffa vänster öra mindre än *en millisekund* före det träffar höger öra. Denna lilla skillnad använder hjärnan och hörselsinnet till att koda av vilken infallsvinkel ljudet har, i detta fall 30°. Då ljudet även träffar vänster öra först tolkar hjärnan det som att ljudet kommer 30° från vänster.

För frekvenser över 2000 Hz är våglängden så pass kort att det mänskliga huvudet blockerar vägen mellan öronen vilket leder till nivåskillnader mellan öronen på ca 8dB och uppåt. I detta område blir fasskillnaderna mycket små och ibland oanvändbara varför hörselsinnet använder skillnader i intensitet/nivå för att avgöra riktningen till ljudet [4].

1.2.2 Riktningar i vertikalplanet

När det kommer till att uppfatta riktningar i vertikalplanet spelar ytterörats form och utseende stor roll. Det handlar om den påverkan som ytterörats vindlingar ger ljudet och det förhållande som finns mellan direktljud och påverkat ljud som avgör ljudkällans placering i höjddled [4].

1.2.3 Elevationseffekt

Enligt ovanstående resonemang kommer alltså ett ljud som träffar både höger och vänster öra vid samma tid, med samma fas och samma intensitet och nivå att tolkas som att det kommer rakt framifrån. Det är helt enkelt detta som händer när man talar om fantommitt eller fantomkälla. Vänster och höger högtalare hjälps åt att skapa en ljudbild mellan högtalarna.

Problemet uppstår i och med ytterörats påverkan på ljudet. Ljudet som kommer från sidan träffar ytterörat med sådan vinkel att den färgning som ytterörat ger i stort sett stämmer överens med den påverkan som hade uppkommit vid en ljudkälla rakt framifrån, men ungefär 40° uppåt i vertikalled. Det gör att stereopanoramata mellan två högtalare ofta upplevs som en båge där mittenpartiet kommer från en punkt som befinner sig högre än högtalarnas faktiska placering. Detta kallas för *elevationseffekten* och är mycket individuellt beroende på ytterörats utseende [4].

Effekten är påtaglig i just TV-sammanhang där ljudet hamnar utanför bilden och korrelationen mellan dialog och den som talar mer eller mindre går förlorad.

1.2.4 Presedenseffekt

Tanken med en fantommitt är att den ska befinna sig i mitten av ljudbilden. Med hänvisning till avsnittet om elevationseffekt och hur det mänskliga örat uppfattar riktningar förstår man att det är en känslig process att återge en stabil fantommitt. Det enda som krävs är att lyssnaren befinner sig en aning närmare den ena högtalaren för att fantommitten ska kantra och luta åt den högtalare som man sitter närmast. Detta kallas för presedenseffekt. Även detta får extra stor betydelse när det handlar om ljud till bild, vilket är fallet i TV-sammanhang. Dialogen kommer då att hamna "vid sidan av" den som talar vilket kan upplevas som störande och göra det svårare att uppfatta talet. Effekten märks mer ju större avståndet mellan vänster och höger högtalare är. I fallet med en liten TV med inbyggda högtalare blir effekten marginell, men i och med större och bredare TV-apparater och att allt fler kopplar in externa högtalare (hemmabio, stereoanläggning och annat) ökar avståndet mellan högtalarna hastigt och effekten blir mycket påtaglig [3, 4, 5, 6].

1.3 Kamfiltereffekt

Kamfilter är ett fenomen som uppstår då ett ljud spelas upp från två källor på olika avstånd från lyssnaren/mottagaren (mottagaren kan vara en mikrofon i ett inspelningssammanhang, men principen är densamma). Då ljuden har olika lång väg att gå kommer de, precis som i resonemanget om elevationseffekt ovan, att anlända vid örat med olika fas. Ett mycket tydligt exempel är att skillnaden mellan de två ljuden är en halv våglängd, alltså att första ljudet har topp och det andra dal. I och med att plus och minus tar ut varandra kommer detta i teorin att resultera i total tystnad och utsläckning av ljudet. Detta kommer sedan att fortsätta uppåt i frekvensspektrat med regelbundna intervall där fasen ömsom blir positiv för bägge ljuden (alltså en hel våglängds fördröjning) och ömsom positiv för ena ljudet och negativ för andra ljudet. Detta kommer att leda till växelvisa förstärkningar och utsläckningar av olika frekvenser i ljudet som rent visuellt kommer att likna en kam, därav namnet (se bild 1) [5, 6].

Kamfiltret kommer att påverka frekvensgången hos programaterialet och i värsta fall märkbart försämra tydligheten hos tal om en utsläckning hamnar i de frekvensområden där taltydligheten finns.

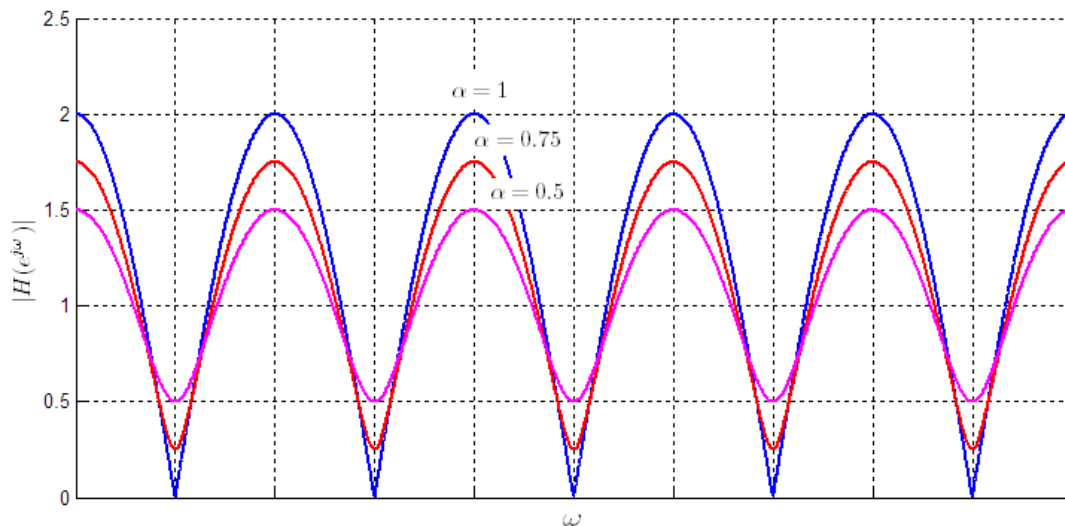


Bild 1 -frekvensgång för kamfilter
http://upload.wikimedia.org/wikipedia/commons/f/fe/Comb_filter_response_ff_pos.png

1.5 Centerkanal och centerhögtalare

Ovanstående fysiska fenomen påverkar alla ljudet negativt. Frekvensgång och riktningssupplevelser förändras vilket i vissa fall kan försvåra hörbarheten avsevärt om vissa kritiska frekvenser dämpas.

Alla problem härstämmer som sagt från att det inte finns någon riktig källa som återger ljudet i mitten utan endast en fantomkälla. Således kan man komma ifrån ovanstående problematik genom att koppla in en centerhögtalare som matas från en speciell centerkanal. En centerhögtalare kan enkelt kopplas in utan större tankemässiga eller ekonomiska investeringar från TV-konsumenten. Dessutom kommer en hel del nya TV-apparater av widescreentyp redan med en inbyggd centerhögtalare som är klar att användas [7].

Frågan är då hur mycket *endast* användandet av en centerhögtalare förbättrar hörbarheten utan att justera inbördes volymer mellan tal och bakgrundsljud.

2. Forskningsfråga

Med hänvisning till ovanstående blir forskningsfrågan som följer: *kan ökad tydlighet och hörbarhet för TV-dialog uppnås genom att lägga till en centerhögtalare för dialogen, utan att i övrigt påverka programmaterialet?*

2.1 Avgränsningar

Forskningsfrågan kommer i denna uppsats att endast gälla lyssning i sweetspot (den teoretiskt bästa lyssningspositionen). Vidare kommer undersökningen att gälla en akustiskt behandlad lokal med en mellanstor surroundanläggning för professionellt bruk. Slutligen kommer undersökningen att göras utan bild för att till fullo kunna fokusera på ljudet.

2.2 Definition av begrepp

I uppsatsen förekommer begreppen (tal-)tydlighet och (tal-)uppfattbarhet. Med tydlighet menas möjligheten för lyssnaren att kunna isolera och fokusera på talet i en större ljudbild. Med uppfattbarhet menas hur enkelt det är för lyssnaren att utan ansträngning ta till sig det som sägs. Båda begreppen är mycket tätt sammankopplade och det ena leder direkt till det andra. Både god tydlighet och god uppfattbarhet krävs för att kunna tillgodogöra sig dialog på ett bra sätt.

3. Metod

För att besvara forskningsfrågan utfördes ett lyssningstest. Testet gick ut på att lyssna på olika ljudexempel innehållande dialog och olika former av bakgrundsljud i dels en 2.0-uppställning och dels en 3.0-uppställning. Subjekten fick därefter bedöma uppfattbarheten i de olika ljudexemplen.

Subjekten fick också göra ett Tal I Brus-test i de båda högtalaruppställningarna för att objektivt kunna bedöma om någon ökning i taltydlighet hade skett.

För benämning av högtalare används härnäst följande termer: Left/L (vänster högtalare fram), Right/R, (höger högtalare fram), Center/C, (högtalare i mitten fram), Left surround/Ls och Right surround/Rs (vänster respektive höger högtalare bak).

3.1 Tal I Brus

Tal I Brus (TIB) är ett standardiserat test som används inom hörselvården för att undersöka hörselnedsättningar och utvärdera vilken effekt till exempel en ny hörapparat ger för patienten. Tal i Brus används som ett komplement till audiogrammet då de mäter olika saker. För att mäta hur man uppfattar tal i en tyst miljö fungerar audiogrammet bra, då uppfattning av tal i tyst miljö till största delen beror på förmågan att registrera alla frekvenser. När det gäller tal i miljöer med bakgrundsljud är det mer förmågan att *skilja frekvenser åt* som spelar störst roll för hur god uppfattbarhet som uppnås [8].

3.2 Ljudexempel

Ljudexemplen som spelades upp var av två olika typer; dels två faktiska reportage/inslag från SVT i Umeå, dels ett TIB-test

3.2.1 Ljudexempel - SVT

Exemplen från SVT var i originalutförande mixade i 5.0, så en nedmix till 3.0 och 2.0 gjordes, vilket innebär att signalen till Ls och Rs mixades till L och R

med oförändrad nivå. I fallet med 2.0 togs även C bort och dess signal mixades till L och R för att erhålla fantommitt. För att erhålla samma nivå på dialogen i C som i fantommitten användes dels en decibelmätare och dels togs tre pilotsubjekt in som fick ställa in den nivå de uppfattade som lika. Dessa värden hamnade väldigt nära det teoretiska värdet som säger att dubbelt antal ljudkällor ger 3dB ökning i ljudtryck. Således sänktes nivån på dialogen med 3dB när den mixades till L/R. Exempelen från SVT var följande:

- "Himmelska hundar" från programmet "Fråga Doktorn" som sändes 2009-10-19. Inslagstid 6:08. Intervjuer med olika hundägare på välbesökt kafé.
- "Buttericks" från programmet "Go' kväll" som sändes 2009-09-30. Inslagstid: 7:59. Inslag om Butterickshuset på Drottninggatan i Stockholm. En del intervjuer ute på gatan och en del inne i husets trappuppgång av sten.

3.2.2 Ljudexempel - Tal I Brus

Testet består i att 50 ord sägs av en röst samtidigt som brus spelas upp. Talets nivå låg i detta test 3dB starkare än bakgrundsbruset. Testet gjordes två gånger med varje subjekt; en gång där tal och brus spelades upp av L och R vilket skapade en fantommitt, och en gång där bruset spelades upp av L och R medan talet spelades upp av C. Testet kommer från en skiva [9] där brus och tal i originalutförande ligger i mono. Bruset mixades till L/R med en nivå-sänkning på 3dB enligt förfarandet och resonemanget i 3.2.1. I 2.0 mixades även talet till L/R med en sänkning på 3dB. Liknande test har gjorts i andra undersökningar för att undersöka taltydlighet. [10]

3.3 Lokal

Testet utfördes i Kontrollrum 5 vid Luleå Tekniska Universitets institution för Musik och Medier beläget i Piteå. Det är ett rektangulärt rum på 5 x 10 x 3 (BxLxH i meter) som är akustiskt behandlat för att fungera som just studiokontrollrum. Lokalen används i undervisningen som kontrollrum för TV/broadcastmixning. Kontrollrum 5 valdes dels för att akustiken i rummet

skulle vara så jämn och neutral som möjligt för att inte påverka resultatet, samt att rummet redan är bestyckat med en professionellt surroundsystem uppställt enligt vedertagen standard (ITU-R BS.775). [11]

3.4 Teknisk utrustning

Alla ljudstimuli spelades upp med ett Pro-Tools HD-system genom Digidesgin 192-interface. Signalen tappades ut via TDIF till ett Studer Vista 8 mixerbord för att göra de två exemplen enkelt tillgängliga för subjekten. Från Vistabordet gick signalen ut till en Genelec 7071A-sub som distribuerade signalen vidare till de fem Genelec 1032A-högtalare som utgjorde surrounduppställningen.

3.5 Subjekt

Testgruppen utgjordes av 20 stycken studerande vid Institutionen för Musik och Medier, varav 12 män och 8 kvinnor. Åldern på subjekten var i spannet 19-28 år. Subjekten kontaktades via ett allmänt mejl till deras studentmejladresser. Varje subjekt genomförde testet i ett svep utan pauser eller andra avbrott. Testet tog cirka 20 minuter att genomföra.

3.6 Testets utformning

3.6.1 Reportage från SVT

Subjekten fick lyssna på varje ljudexempel i två versioner; en 2.0 (A) och en 3.0 (B). Det stod subjekten fritt att byta mellan A och B med hjälp av en knapptryckning. Subjekten ombads bedöma uppfattbarheten i inslaget. Uppfattbarhet definierades som ”hur enkelt det är att höra och uppfatta innehållet”. När subjekten kände sig nöjda lämnades ett svar innan subjektet fortsatte till nästa ljudexempel. Svar lämnades på följande sjugradiga skala;

- A är mycket bättre än B
- A är bättre än B
- A är lite bättre än B
- A och B är lika

- B är lite bättre än A
- B är bättre än A
- B är mycket bättre än A

Vilken version som var A respektive B slumpades i alla tester för att sprida ut eventuella felfaktorer jämnt över all data så att bra resultat skulle kunna erhållas.

3.6.2 Utformning av TIB-test

I TIB-test var det subjektens uppgift att efter bästa förmåga upprepa de ord som sades. Rätt ord gav en (1) poäng och fel ord gav noll (0) poäng. Då testet gjordes två gånger med varje subjekt (ett i 2.0 och ett i 3.0) var de 50 orden olika för de olika gångerna för att undvika att subjekten skulle kunna lära sig orden. Författaren satt bredvid som testledare och antecknade svaren direkt i Microsoft Excel.

3.6.3 Övrigt i testutformningen

Längst ner på svarsformuläret fanns ett utrymme för övriga kommentarer som att till exempel lämna en kommentar eller motivering till sitt svar. Testledaren instruerade även subjekten kortfattat i början av varje delmoment hur testet skulle gå till. I övrigt var testledaren passiv och svarade inte på några frågor under testet.

Ordningen på SVT-delen och TIB-delen slumpades även den för att sprida ut eventuella felkällor jämt över alla data. Ingen test/träningsrunda genomfördes med subjekten. Detta för att det i SVT-testet fanns gott om tid – från 6 till 8 minuter – att skaffa sig en uppfattning om materialet. TIB-testet är utformat för att genomföras utan träningsomgång. En träningsomgång hade flyttat upp medel- och maxvärde på skalan men de inbördes tendenserna hade funnits kvar.

3.6.4 Andra faktorer i testmiljön

Alla subjekt befann sig mitt i lyssningpositionen, i sweetspot. Detta är inte så vanligt förkommande vid TV-tittande då folk rör sig och sitter utspridda på olika platser i rummet. Att ha subjekten sittandes i sweetspot valdes för att minska variablerna i datan och göra datamängderna lättare att båda hantera rent praktiskt och att jämföra vid senare analys.

Vidare täcktes högtalarna i rummet för med ett akustiskt transparent tyg. Detta gjordes för att undvika att de visuella intrycken skulle störa lyssningen. Subjekten skulle inte med hjälp av ögonen kunna identifiera den punkt som ljudet kom ifrån. Testen spelades upp på en konstant ljudnivå; 65dB för testet med SVT-reportagen och 67dB för TIB-testet. Dessa värden var de som testledaren lyssnat på för att kontrollera testet. Subjekten erbjöds att justera volymen i början av testet, men ingen valde att göra så.



Bild 2 - testmiljön i Kontrollrum 5.

4.Resultat

4.1 Resultat för SVT-test

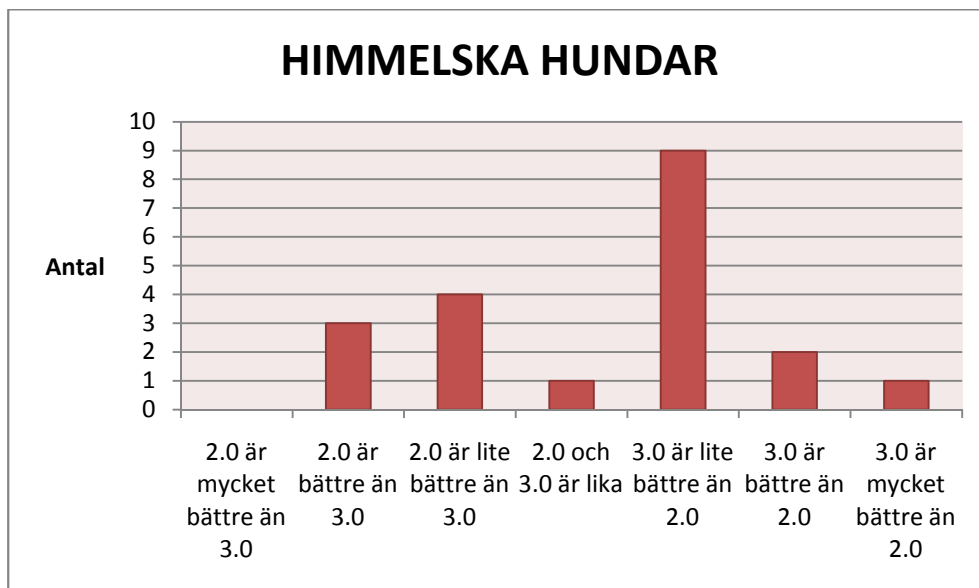


Diagram 1 - fördelning av svar för inslaget "Himmelska hundar"

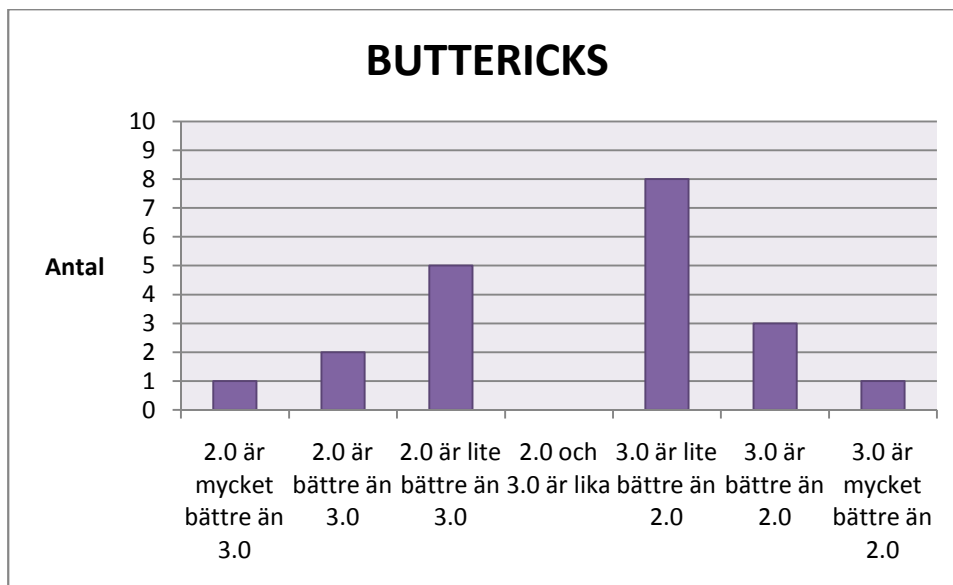


Diagram 2 - fördelning av svar för inslaget "Buttericks"

4.2 Subjektens kommentarer till SVT-test

Exempel A respektive B byts här för tydlighetens skull ut till att benämnas med 2.0 och 3.0.

4.2.1 Subjekt som föredrar 2.0-versionen av något inslag

” Ljudkvaliteten kändes högre på 3.0, men man stördes mer av bakgrundsljudet och klangen än i 2.0”

” på exempel 1 tyckte jag att svar nr.5 passade bäst in på min upplevelse, inte i första hand för att det var lättare att uppfatta innehållet, utan mer för att det var ett mer lyssnarvänligt ljud som gjorde att man inte satt och tänkte på t.ex. dom skarpa S ljuden utan tankte mer på innehållet”

” Det kändes som det var lättare att uppfatta taltydligheten vid det brusiga exemplet i ex:2 just för att det var mer brus”

” I exempel 2 var talet högre men på något sätt lite störande”

”Nu när det inte är något ljus brus blir A tydligare för jag hör mindre bakgrundsljud och mindre diskantigt”

” När brus och tal låg centerhögtalaren så blev det otydligare.”

4.2.2 Subjekt som föredrar 3.0-versionen av något inslag

” 3.0 känns mycket tydligare i taluppfattbarheten. Det känns dock mer naturligt att lyssna på exempel 2.0. 2.0 har en mer omslutande känsl men är sämre i taluppfattbarheten. Ansträngningen var mindre vid lyssning på exempel 3.0. Det hade dock gärna fått vara lite mer ambiens”

” 3.0 kändes tydligare än 2.0 av samma anledningar som i exempel 1. Allt skrivet ovan om exempel 1 för A tycker jag i exempel 2 om B. ”

” Elevationseffekten blev väldigt tydlig.”

” Separeras talet bättre från bakgrundssorlet”

” I exempel ett kändes 3.0 som att den använde sig av centerkanal och därför hade talet som en äkta monokälla. Överlag var taltydligheten god i både exemplen, men något klarare och lättare att lokalisera i 3.0. I Exempel två var upplevelsen likartad men omvänd. Annan frekvensgång på talet ide båda vinnande exemplen; olika högtalare eller fantomcenter vs diskret centerhögtalare?”

” Jag upplever att totalnivån är starkare i 2.0 än 3.0. Däremot är ”signal-brusförhållandet” bättre i 3.0. Därför tar 3.0 en knapp seger”

” Talet låg ”längre fram” ”

” Bakgrundsljud stökar mindre med taket i 3.0 än i 2.0”

” 3.0 kändes lite tydligare även om det inte på något vis var svårt att tyda 2.0. Talet i 3.0 kändes starkare o med en annan tinal balans med mer bas så 3.0 var lite mer genomträngande resten av ljudbilden.”

” Talet blir tydligare i 3.0 även fast jag hör mer bakgrundsljud, bakgrundsljudet i 2.0 är lägre men då är det ett ”ljusare” brus som är mer störande.”

4.3 Resultat för TIB-test

Tabell 1 - antal rätt i TIB-test

| ANTAL RÄTT I TIB-TEST (AV MAX 50) | | |
|--|----------------|----------------|
| | TIB 2.0 | TIB 3.0 |
| | 24 | 28 |
| | 14 | 21 |
| | 17 | 28 |
| | 23 | 26 |
| | 22 | 29 |
| | 21 | 26 |
| | 16 | 26 |
| | 20 | 28 |
| | 29 | 25 |
| | 23 | 29 |
| | 25 | 29 |
| | 21 | 21 |
| | 21 | 25 |
| | 22 | 28 |
| | 16 | 22 |
| | 20 | 28 |
| | 22 | 26 |
| | 17 | 29 |
| | 17 | 23 |
| | 27 | 26 |
| SUMMA: | 417 | 523 |
| MEDEL: | 20,85 | 26,15 |

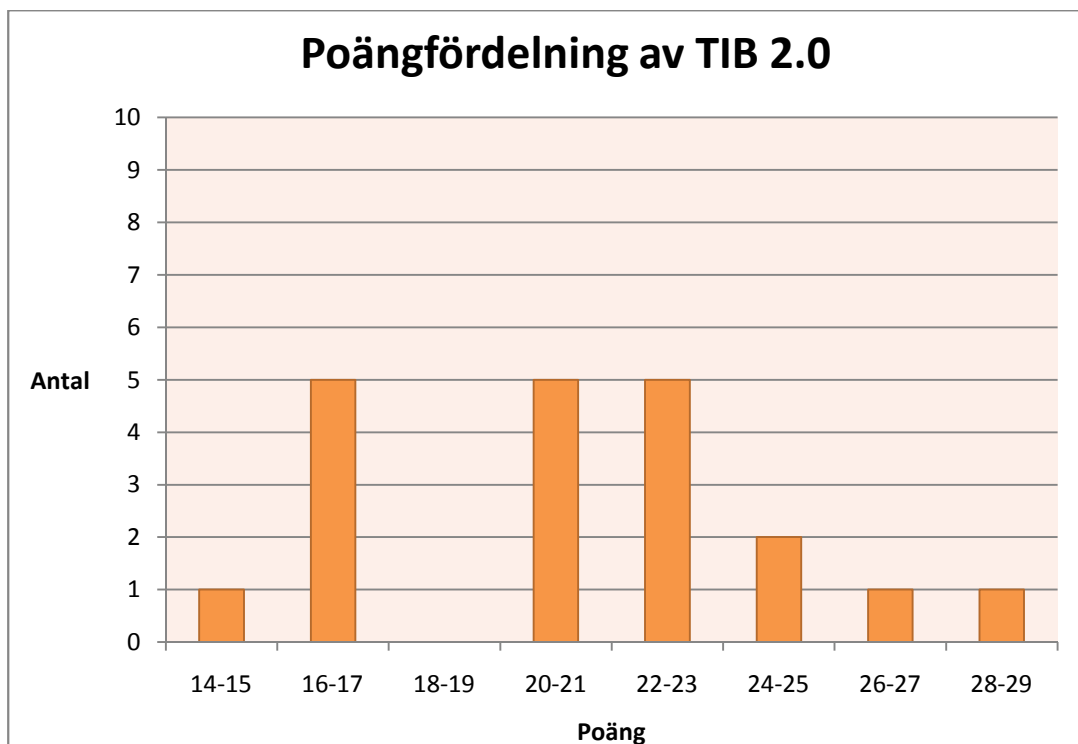


Diagram 3 - fördelning av poäng i TIB 2.0

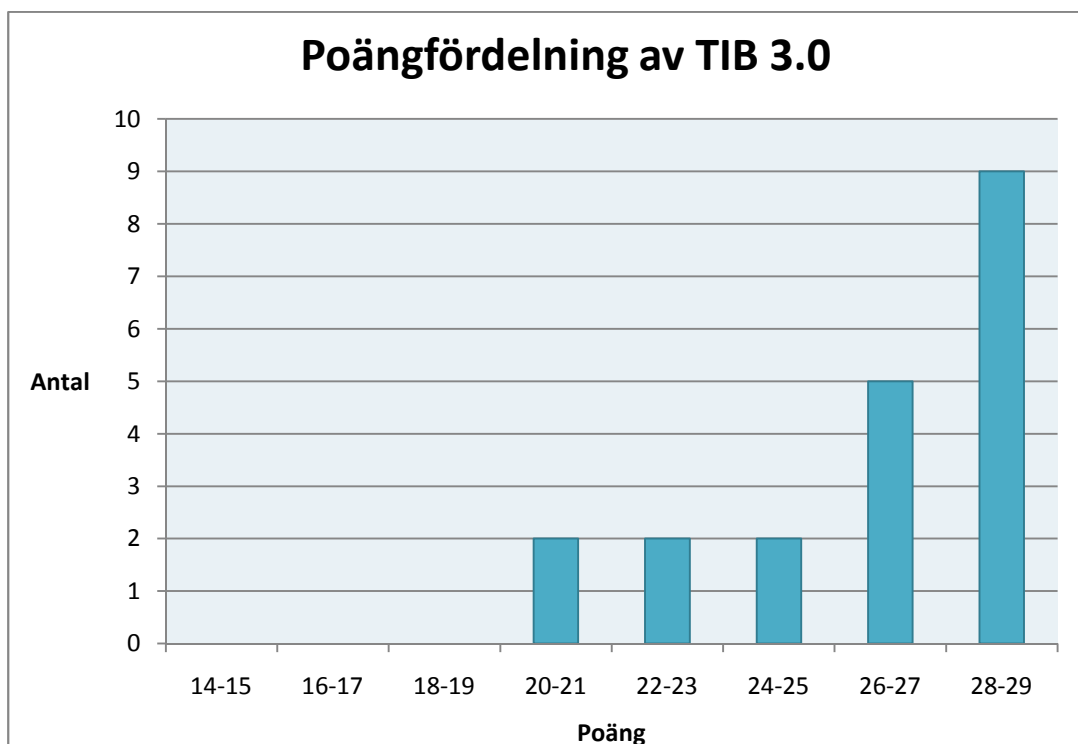


Diagram 4 - fördelning av poäng i TIB 3.0

5. Analys av resultat

Resultaten i avsnitt 4. visade på att en det fanns en skillnad mellan 2.0 och 3.0. En statistisk analys gjordes för att undersöka resultatet och se om den skillnad som fanns var signifikant.

5.1 Analys av resultat i SVT-test

För att möjliggöra analys resultaten i SVT-testet omsattes svaren till poäng enligt följande:

- 2.0 är mycket bättre än 3.0 – gav 3 poäng
- 2.0 är bättre än 3.0 – gav 2 poäng
- 2.0 är lite bättre än 3.0 – gav 1 poäng
- 2.0 och 3.0 är lika – gav 0 poäng
- 3.0 är lite bättre än 2.0 - gav - 1 poäng
- 3.0 är bättre än 2.0 – gav - 2 poäng
- 3.0 är mycket bättre än 2.0 – gav - 3 poäng

Detta gav följande medelvärde:

Tabell 2 - medelvärde för SVT-test.

| MEDELVÄRDE FÖR SVT-TEST | |
|-------------------------|-------|
| Himmelska hundar: | -0,3 |
| Buttericks: | -0,25 |

Ett T-test genomfördes i Microsoft Excel. Testet var ett så kallat "one sample T-test" där ett uppmätt resultat jämförs mot ett förväntat. I detta fall valdes det förväntade resultatet till "0", alltså att det inte fanns någon skillnad mellan 2.0 och 3.0.

Detta gav ett p-värde på 0,3679 för "Himmelska hundar" och ett p-värde på 0,4980 för "Buttericks". Detta fick klassas som otillräcklig signifikans. Testet visar att de skillnader som finns mellan 2.0 och 3.0 med 36,79% respektive 49,8% chans beror på slumpen.

Den låga signifikansen beror på att många subjekt svarat i mitten av skalan; "2.0 är lite bättre än 3.0" eller "3.0 är lite bättre än 2.0". Vidare är skillnaden i antalet subjekt som föredragit 2.0 respektive 3.0 för liten för att något signifikant resultat ska kunna fastställas.

5.2 Analys av TIB-test

En liknande analys genomfördes för resultatet i TIB-testet. Ett T-test gjordes i Microsoft Excel. Testet var ett så kallat "paired T-test", då samma grupp subjekt lyssnat på två olika stimuli.

Resultatet blev ett p-värde på $7,487 * 10^{-6}$ vilket får klassas som mycket god signifikans. Testet visar att de skillnader som finns mellan de bägge TIB-omgångarna endast med 0,007487‰ chans beror på slumpen. Detta betraktas som helt försumbart. Skillnaderna mellan TIB-resultaten beror inte på slumpen utan är en effekt av ökad tydlighet.

6. Diskussion

Syftet med denna uppsats var att undersöka om det gick att förbättra uppfattbarheten för TV-dialog genom att implementera en centerhögtalare. Ett tvådelat lyssningstest genomfördes där subjekt dels fick bedöma uppfattbarheten i autentiska TV-reportage samt fick sin förmåga att uppfatta tal i kraftigt bakgrundsbrus satt på prov. Nedan följer diskussioner kring varje del i testet för att sedan mynna ut i en slutsats.

6.1 Test med TV-reportage

Enligt de resultat som kom ur detta test finns vid första anblicken en skillnad. Efter att resultatet analyserats visar det sig att skillnaderna inte kan klassas som signifikanta och resultatet har låg validitet. En trend börjar dock utkristallisera sig. 3.0-versionerna har fått fler röster överlag och även fått något högre frekvens på de högre betygen. Ett sätt att förbättra signifikansen i den här undersökningen skulle vara att utöka testet med fler subjekt. Detta skulle göra att de tendenser som nu verkar finnas skulle ha chansen att bli tydligare. På så vis skulle klarare och säkrare slutsatser kring resultatet kunna dras.

Många subjekt utnyttjade utrymmet för ”övrigt” på svarsformuläret till att lämna en motivering till sitt svar (se 4.2). Bland de subjekt som föredragit 3.0-versionen av något reportage förekommer ofta kommentarer som kan sammanfattas med ”tydligare klang och/eller separation på talet” och ”stadigare mitt”. Detta kan direkt härledas till teoridelen och de olika bieffekter som reproduktion i två högtalare ger ljudet. ”Tydligare klang” kommer sig av frånvaron av *kamfiltereffekt*. [5, 6]. Vid återgivning i en centerhögtalare kommer man helt bort från de växelvisa utsläckningar/förstärkningar som sker vid återgivning i fantommitt. Detta gör i sin tur att talet låter naturligare med ett korrekt och oförändrat frekvensinnehåll. Alla vokalljud och konsonantljud finns kvar och förhållandet mellan dessa återges så som de en gång spelades in. Denna naturligare och rakare klang uppfattas av många som tydligare och gör det enligt subjekten lättare att fokusera på talet och det går att uppfatta det som sägs med mindre ansträngning.

På samma sätt kan kommentaren om ”stadigare mitt” kopplas till framför allt teorin om *presedenseffekten* [3, 4, 6] men också till viss del till *elevationseffekten* [4]. Några subjekt har rent konkret nämnt båda effekterna i sina kommentarer.

Vid återgivning i centerhögtalare finns inget presedenseffektsfenomen. Detta beror som tidigare sagts på att det finns en riktig ljudkälla som faktiskt reproducerar ljud i mitten av ljudbilden. Fantommitten å andra sidan lider av presedenseffekten varför mitten av ljudbilden kommer att luta åt antingen

vänster eller höger beroende på var lyssnaren är placerad eller hur denna rör sitt huvud under lyssningen. Detta gör det svårt att tydligt identifiera en punkt som talet kommer ifrån varför det blir svårare att fokusera och följdaktligen svårare att uppfatta vad som sägs.

Detsamma gäller i viss mån för elevationseffekten. Dock kommer inte denna att ge upphov till samma förflyttningar som presedenseffekten. Elevationseffekten placerar talet en bit ovanför högtalarna på en plats som upplevs som direkt onaturlig. Denna effekt är i högsta grad individuell men kan vid stora vertikala förflyttningar vara mycket störande och helt ta bort fokus från programinnehållet.

En annan sida av samma mynt är brus och andra störljud som läckt in i den talmikrofon som använts vid reportaget. Bland de subjekt som föredragit 2.0-versionen av något inslag var kommentarer i stil med ”bättre ljudkvalitet i 3.0 men bruset stör för mycket” vanliga. Detta gällde bägge inslagen men var något vanligare i Buttricksreportaget. Då reportagen är inspelade ute i ”verkligheten”, på kaféer och andra offentliga miljöer, kommer större eller mindre problem med störljud att uppstå. Ska materialet sedan återges med centerhögtalare bör man vara extra noga med just denna problematik. Man bör hålla i huvudet i fallet med en talkanal med mycket störljud och brus får ju även *bruset* ta del av den förbättrade tydlighet och stadigare återgivning som en centerhögtalare erbjuder. Vid kraftigt brus kan detta bli mycket störande för lyssnaren varför tydligheten helt går förlorad. Anledningen till att reportaget ”Buttricks” oftare fått denna kommentar beror förmodligen helt enkelt på att den faktiskt *är* brusigare. ”Buttricks” var enligt SVT inspelat med enklare reportageutrustning varför materialet inte var det bästa.

6.2 Tal med bakgrundsbrus

Det andra delmomentet – TIB-testet – var med som ett sätt att konkret testa hur pass mycket bättre uppfattbarheten blev i och med centerhögtalaren. Resultaten här visar att det finns en tydlig, signifikant skillnad mellan de båda testomgångarna, där subjekten nästan genomgående är bättre i 3.0-versionen än i 2.0. Endast tre subjekt avviker från detta. Två subjekt är bättre i 2.0 än 3.0

– ett med mariginella 27-26 och ett med 29-25. Det tredje subjektet får samma poäng i bägge omgångarna – 21-21. Gemensamt för dessa tre subjekt är att de alla har börjat med 3.0-versionen av sitt TIB-test. Detta gör att man inför 2.0-versionen hade en klarare bild av hur testet skulle gå till, samt att man (medvetet eller undermedvetet) tränat upp och utvecklat en lyssningsstrategi för att lösa uppgiften och därigenom presterat bättre trots de i teorin sämre förutsättningarna. I övrigt visar resultaten av TIB-testet på att man helt klart erhåller bättre uppfattbarhet och tydlighet för tal i och med att man använder en centerhögtalare.

En av anledningarna till det tydliga resultatet är att TIB-testet är helt befriat från ”tycke och smak”. Alla subjekt var musik/mediastudenter och har i och med sin utbildning stor vana av att konsumera ljud (TV, radio och/eller musik) i just traditionell 2.0-uppställning. Således känner man sig hemma med fantommitten och vet hur den låter varför man i SVT-delen av test kanske föredrog en 2.0-version. TIB-test är då mer svart och vitt. Antingen lyckas man som subjekt uppfatta rätt ord eller så gör man det inte. Det gör att TIB-testet ger en mer objektiv bild av hur stora förbättringar man kan erhålla för taltydlighet genom att implementera en centerhögtalare.

6.3 Utveckling av undersökningen

Som tidigare nämnts bör undersökningen göras om med ett större antal subjekt för att se om det på så vis går att dra några slutsatser ur resultatet för SVT-testet. Vissa tendenser finns i resultatet redan nu, men de behöver bli mycket tydligare för att det ska gå att erhålla några signifikanta resultat och få fram ett resultat som är generellt.

Då jag som suttit med som testledare under de två dagar som lyssningstestet genomförts samt ytterligare en dag för intrimning och förberedelser har jag hört de olika ljudexemplen ett stort antal gånger. Jag är också den enda som suttit vid sidan om sweetspot under testets gång (se stol och datorskärm till höger i bild 2). Vid så kallad off-axislyssning är fördelarna med en centerhögtalare mycket påtagliga. När ett subjekt lyssnat på 2.0-versionen har ljudbilden kanträt ordentligt åt höger och resultatet har blivit något som

påminner om en monofonisk återgivning, men en där man uppfattar fasfel när man rör på huvudet. Även ett av mina pilotsubjekt kommenterade detta då denne vid ett tillfälle satt bakom lyssningspositionen vid intrimineringen av testet.

För att undersöka detta bör testet även utökas till att inkludera lyssning utanför sweetspot. Detta för så långt det är möjligt efterlikna en äkta TV-situation och verkligen se vilka effekter som implementationen av en centerhögtalare kan bidra med i sådant sammanhang.

7. Slutsatser

I och med den digitalsända televisionens intåg har nya möjligheter öppnat sig för ljudtekniker. Fler ljudkanaler med högre upplösning ger fler verktyg att göra intressanta och bra TV-program. En aspekt av TV-ljudet som diskuterats flitigt är taltydligheten. Ett sätt att förbättra denna är att använda just flerkanalsljud och implementera en centerhögtalare för dialogen i TV-programmet. Denna uppsats syftade till att undersöka vilka förbättringar användandet av en centerhögtalare gav med fokus på just dialogens tydlighet. Ett lyssningstest visar att en centerhögtalare helt klart bidrar positivt till taltydligheten. Centerhögtalaren erbjuder en fast förankrad mittpunkt i ljudbilden som inte lider av fasfel/kamfilter och inte heller kantrar åt något håll när lyssnaren rör på sig.

Dock är det viktigt att hålla i huvudet att den gamla devisen ”skräp in, skräp ut” fortfarande gäller i allra högsta grad. Centerhögtalaren är inget magiskt redskap som direkt löser problem med taltydlighet. En dåligt gjord inspelning med mycket störljud i talkanalen gör att även störljuden får del av den ökade separation och tydlighet som centerhögtalaren ger, vilket direkt motverkar just talets tydlighet. Däremot kan centerhögtalaren användas för att öka tydligheten på en redan bra inspelning ytterligare ett par snäpp.

Det är fortfarande upp till oss ljudtekniker att se till att de inspelningar och ljudupptagningar som görs håller så hög kvalitet som möjligt för att på så vis erhålla god taltydlighet i TV-programmen. Centerhögtalaren är ett bra tillskott

i verktygslådan för ljudtekniker och är bra att ta fram i arbetet med att producera bra och välljudande TV-program för konsumenterna där man inte ska behöva undra vad det var som egentligen sades.

Referenser

[1]http://svt.se/content/1/c8/01/04/83/61/svt_horbarbetsprojektet2007.pdf
läst 2010-03-19, kl 09.40

[2]<http://svt.se/content/1/c8/01/04/83/61/Broschyr%20GOD%20H%D6RB%20ARHET%20I%20SVT.%20Vers.%202009.pdf>, läst 2010-03-19 kl 10.00

[3] Vickers, Earl “*Fixing the phantom center: diffusing acoustical crosstalk*”. Presenterat vid 127th AES Convention i New York, USA, 2009. Paper nummer: 7916.

[4] Evers, Hans ”*Stereoteknik*” 4:de upplagan. Kompendium för ljudingenjörer. Utgivet av Sveriges Television och Sveriges Radio 2003.

[5] Alton Everest, F ”*Master handbook of acoustics*” 4:de upplagan. Utgiven av McGraw-Hill 2001. ISBN:0-07-136097-2

[6] Faller, Christof “*Multiple loudspeaker playback of stereo signals*”. Publicerat i J. Audio Eng. Soc., Vol. 54, No. 11, november 2006.

[7] Vickers, Earl “*Frequency domain two-to-three channel upmix for center channel derivation and speech enhancement*”. Presenterat vid 127th AES Convention i New York, USA, 2009. Paper nummer: 7917.

[8] Taylor, Bryan “*Speech-in-noise-tests: how and why to include them in your basic test battery*”. Publicerat i The Hearing Journal vol 56, nr 1, januari 2003.

[9] CD-skiva “*Svensk talaudiometri*”. Utgiven av C A Tegnér AB (<http://www.categner.se/ex%20shop/CDskivor.html>). Artikelnummer: 10-210.

[10] Shirley, Ben, Kendrick, Paul och Churchill, Claire “*The effect of stereo crosstalk on intelligibility: comparison of phantom center image and a central loudspeaker source*”. Publicerat i J. Audio Eng. Soc., vol. 55, nr. 10, oktober 2007

[11] Rumsey, Francis och McCormick, Tim “*Sound and recording – an introduction*” 5:e upplagan. Utgiven av Focal Press 2006. ISBN:978-0-240-51996-8.

Bilagor

Bilaga 1 – Svarsformulär för lyssningstestet

Sessionsnummer: 1

Bedöm uppfattbarheten i följande exempel. Med uppfattbarhet menas ”hur enkelt det är att höra och uppfatta innehållet”.

Följande är skalan:

1. A är mycket bättre än B
2. A är bättre än B
3. A är lite bättre än B
4. A och B är lika
5. B är lite bättre än A
6. B är bättre än A
7. B är mycket bättre än A

Exempel 1

Svar:

Exempel 2

Svar:

Övrigt, till exempel kommentarer/motiveringar till svaret/svaren ovan, andra tankar om testet etc.

Svar: