

DEN BINAURALA UPPLEVELSEN

Manipulering av ljud

THE BINAURAL EXPERIENCE

Manipulation of sound

Examensarbete inom huvudområdet Medier, estetik och berättande

Grundnivå 30 högskolepoäng

Vårtermin 2013

Victor Askling

Handledare: Peter Bryngelsson

Examinator: Per-Anders Nilsson

Sammanfattning

Detta examensarbete undersöker om man kan efterlikna binaurala ljud genom att manipulera monoinspelningar. Kan en monoinspelning med hjälp av en binaural *plugin* skapa samma effekt som en binaural inspelning har, där effekten som eftersträvas är lokaliseringsprecision och realism? Vad binauralt ljud innebär, hur det fungerar och dess för- och nackdelar presenteras. Texten tar dessutom upp viktiga aspekter om hur vårt lyssnade sker, vilket grundar sig på Michel Chion, Blesser och Salter teorier. Den frågeställning som ställs besvaras genom en kvantitativ och kvalitativ metodstudie. För att utföra studier har två ljudfiler skapats där den ena ljudfilen är en binaural inspelning, medan den andra är en manipulerad monoinspelning. Det slutgiltiga resultatet av undersökning var att det inte gick att simulera samma binaurala upplevelse och lokaliseringsprecision, som en riktig binaural inspelning ger.

Nyckelord: binaural, ljudmanipulering, Blesser, ljudlokalisering, realism, ljudupplevelse.

Innehållsförteckning

1.Introduktion.....	1
2.Bakgrund.....	2
2.1 Utveckling i tanke och teknologi.....	2
2.2 Sensationen, uppfattningen och betydelsen.....	3
2.3 Realism och ljudkvalitet.....	4
2.4 Binauralt ljud.....	5
2.4.1 Binauralt ljud i spel.....	7
3.Problemformulering.....	8
3.1 Metodbeskrivning.....	8
3.1.1 Urval.....	8
3.1.2 Datainsamling.....	8
3.1.3 Dataanalys.....	10
4.Projektbeskrivning.....	11
4.1 Arbetsprocess.....	11
4.1.1 Inspelning & Redigering.....	11
4.1.2 Mixning.....	12
4.1.3 Kvantitativt Pilottest.....	14
4.2 Potentiell förbättring.....	15
5.Utvärdering.....	16
5.1 Kvantitativt resultat.....	16
5.2 Kvalitativt resultat.....	20
6.Slutsats.....	23
6.1 Resultatsammanfattning.....	23
6.2 Diskussion.....	24
6.3 Framtida arbete.....	25
Referenser.....	26
Appendix A - Frågeformulär.....	28

1. Introduktion

I dagens samhälle har man en slags överenskommelse om hur ljud ska bearbetas eller förhålla sig i film- och spelproduktioner. Det finns även en uppfattning om vad som är realistiska ljud och hur man skapar realism enligt Michel Chion (1994), Barry Blesser och Linda-Ruth Salter (2007). Detta framförs med en koppling till den binaurala ljudvärlden, vilket har klassats som en mer realistisk representation av vårt hörande på grund av dess kapacitet att effektivt förtydliga ljudens placeringar (Blesser & Salter, 2007). Det diskuteras även hur binaurala ljud används i spelbranschen och hur de förhåller sig till spelarens immersion. Detta är en viktig del att undersöka eftersom man enligt Bob Bates (2004) uppskattar ljudets roll mer i spel än i film, då 3-D ljud blir en uppenbar tillgång för att intensifiera spelupplevelsen. Det kan även bidra till att filmupplevelsen ökar men Mavor och Durlach (1994) menar att detta inte är så populärt på grund av dess särskilda inspelnings- och manipuleringsprocedurer.

Metoden för att undersöka detta sker i både kvalitativ och kvantitativ form, kombinerat med intervjuer och test. Detta kommer belysa frågan om det finns en skillnad i ljudfilernas förmåga att visa ljudens position, samt om upplevelsen av ljuden förändras.

Det projekt som behövdes för att utföra denna undersökning består av två ljudfiler som spelats in på två olika sätt. Dessa sätt att spela in ljud beskrivs i kapitel fyra vilket är det kapitel som presenterar tillvägagångssättet bakom projektet, alltså vilka val som gjordes och varför.

Denna undersökning har som syfte att klargöra om den mer realistiska ljudvärldens uppfattning som upplevs av så kallade binaurala ljud, även existerar i ljud som har manipulerats med hjälp av en binaural *plug-in*. I sin bok *Spaces Speak, Are You Listening? : Experiencing Aural Architecture* gör Blesser och Salter (2007) viktiga slutsatser om hur vi lyssnar och uppfattar ljud. Dessa slutsatser är viktiga aspekter och inkluderas i sammanställningen av all data i kapitel sex som handlar om resultatsammanfattningen.

2. Bakgrund

2.1 Utveckling i tanke och teknologi

Tim. J Anderson skriver i sin bok *Making Easy Listening: Material Culture and Postwar American Recording* (2005) att ny teknologi hela tiden utvecklas för att göra filmproduktioner bättre, men att ljudet alltid har underskattats av konsumenterna. Han menar att konsumenten alltid har sett lyssnandet som en passiv del av upplevelsen, där man inte behöver vara aktiv i sitt lyssnande för att ta del av ljuden. Detta är enligt hans åsikt inkorrekt då han menar att ljud kan ge lyssnaren en annan nivå av förståelse.

The listening act is always an act of resolve, intent manifested at a particular level of understanding that may vary from subject to subject, time to time, space to space; it is never passive in the true sense of the word.

Anderson, Tim. J, 2005, s. 107

Anderson (2005) uttalar sig här om vad han anser är konsumenternas uppfattning och inte företagen. Enligt Clayton Crooks (2004) lägger företagen och bolagen en stor mängd tid och pengar på att förbättra filmens aspekter.

Over the past decade, there has been tremendous time and resources spent on improving these aspects of a movie, during which time we have seen the use of surround sound in both theatrical and home movie releases.

Crooks, 2004, s. 152

Det har alltså lagts en hel del resurser på att förbättra filmupplevelsen, men uppskattar konsumenterna deras ansträngning? I spelbranschen verkar konsumenterna ha en annorlunda syn på ljud enligt Bob Bates (2004), där han menar att det istället har blivit en central del av spelets design och mekanik. Ljud i spel har delvis syftet att skapa immersion hos spelaren så att han eller hon får mer inlevelse till spelet, men det har även syftet att påverka spelarens handlingar. I vissa spel har ljuden en sådan viktig roll att det kan handla om sin karaktärs överlevnad. Ett exempel på spel där detta är en nödvändighet är *Splinter Cell: Conviction* (2010), där man agerar som en agent ute på hemliga uppdrag. Dessa spel är designade för att spelaren ska smyga och förbli oupptäckt. Självklart är detta inget som spelaren är tvungen att göra, men om personen väljer att bli upptäckt kommer svårighetsgraden att öka, eftersom vakterna kallar på förstärkning och börjar attackera. Detta spelsätt har gjort det oerhört viktigt för spelaren att höra ljud som varnar eller tipsar att en vakt närmar sig runt hörnet.

Nu har man kommit så långt i utvecklingen att man kan återskapa en väldigt realistisk representation av vad vi egentligen hör. Den är så pass realistisk så att folk rycker till av obehag eller förundran över dess realistiska karaktär. Detta kallas för binauralt ljud eller 3-D ljud där man använder en speciell inspelningsteknik för att efterlikna hur det mänskliga örat uppfattar ljud. Denna teknik har visat sig vara exceptionell för att lokalisera ljud, eftersom den fångar upp den rumsliga akustiken så bra (Blessner & Salter, 2007).

Men hur upplever vi ljudet och finns det någon poäng med att förbättra den nuvarande teknologin? Kan vi fokusera på hur man uppfattar ljuden och fortsätta att försöka lura

lyssnaren som det tidigare har gjorts, eller har vi lyssnare blivit för kräsna? Detta reflekterar Mavor och Durlach över i *Virtual Reality : Scientific and Technological Challenges* (1994).

Specifically, in the auditory channel, the interface devices (earphones and loudspeakers) are essentially adequate right now. In other words, from the viewpoint of synthetic environment (SE) systems, there is no need for research and development on these devices and no need to consider the characteristics of the peripheral auditory system to which such devices must be matched.

Mavor & Durlach, 1994, s. 134

Mavor och Durlach menar att ljudteknologin för tillfället inte behöver mer forskning, utan den har kommit till en punkt där den anses vara tillräckligt bra. Ett exempel på vad som skulle kunna anses vara en tillräcklig teknologi är de binaurala *plug-ins* som används idag. Det finns *plug-ins* i redigeringsprogram och mjukvaror som kan simulera binauralt ljud genom att applicera en artificiell rumskaraktär hos ljuden. En vanlig sådan *plug-in* finns tillgänglig i *Logic Pro 9* (2010) enligt Kevin Anker och Orren Merton som har skrivit boken *Logic Pro 9 Power! : The Comprehensive Guide* (2010). Denna *plug-in* beskrivs som en bra resurs om man vill ge ljudet mer djup och närvaro. "The Binaural Panner is a very powerful, very flexible stereo imaging tool. It adds spatial directional clues to your audio, giving your source a depth or presence in a mix that a typical pan control can't achieve." (2010, s. 484). Detta är ett sätt att gå till väga för att försöka simulera binaurala ljud genom manipulation, om man anser att den binaurala inspelningstekniken känns begränsad eller opraktisk. Det är dock viktigt att ha i åtanke att även om denna metod kan förenkla det praktiska arbetet som ligger bakom skapandet av en binaural ljudbild, så är den fortfarande konstgjord. Den är fabricerad för att simulera binauralt ljud och frågan är om vi kan höra skillnaden.

2.2 Sensationen, uppfattningen och betydelsen

Tidigare forskning om ljud och rumsuppfattning (Blessner & Salter, 2007 & 2009) har undersökt hur vi människor uppfattar och bearbetar ljud. Under detta kapitel så presenteras teorier om olika moment av uppfattning som människans hörande består av. Det första momentet är *raw sensation* som innebär vår fysiska förmågan att registrera ljud. Detta moment behandlar inte ljudets mening eller betydelse utan är bara den biologiska egenskapen att kunna registrera ljuden.

Det andra momentet är *preception* vilket kopplar de registrerade ljuden till lyssnarens individuella erfarenheter. Här får ljudet en betydelse och kan lokaliseras tack vare att man fått lärdom genom åren om hur ljud fungerar i ens kultur och samhälle. Språk är ett exempel på hur erfarenhet och lärdom har gett ordernas ljud en betydelse som vi sedan sammansätter till en mening. Här menar Blessner och Salter (2007 & 2009) att lyssnande påverkas av den kultur och samhälle man lever i, att det är detta moment av lyssnande som formar vad det är man känner igen och kan förstå beroende på vad man har exponerats för tidigare.

Det sista momentet som sker i vårt lyssnande är *emotionally engaged listening* som kopplas till våra känslor eller sinnesstämning. I denna process får ljud en djupare betydelse till den individ som lyssnar och kan innefatta olika betydelser för olika människor. Denna process menar Blessner och Salter (2007 & 2009) är svår att undersöka eftersom ljuden blir väldigt

subjektiva och personen som blir undersökt inte nödvändigtvis har förmågan eller vetskapen som behövs för att uttrycka sin sinnesstämning.

Moreover, a listener may not have the linguistic skills to describe affective reactions, and a researcher may not have an objective means for observing neurological responses corresponding to emotions.

Blessner & Salter, 2007 & 2009, s.13

Michel Chion (1994) är en annan författare som har skrivit om hur vårt lyssnande sker och hur vi uppfattar olika ljud annorlunda beroende på vilken lyssningsmetod som används. Han har gått djupare in i momentet *perception* som Blessner och Salter (2007) skriver om, där han går igenom tre olika sätt att lyssna. Dessa består av *casual listening*, *semantic listening* och *reduced listening*. Han förklarar att *casual listening* är det vanligaste sättet att lyssna där vi samlar in information från ljudkällan, att man skaffar sig en uppfattning om vad det är som låter. Detta sätt att lyssna blandar sig även med annan information som minnen, känslor och bilder men mera. Detta gör att den även är det mest opålitliga sättet att lyssna eftersom en bild kan påverka vad man tror det är som låter, vilket man i film utnyttjar mycket (Chion, 1994).

Semantic listening behandlar istället ljud som koder där man tolkar och försöker ta reda på ljudets underliggande betydelse. Det vanligaste exempel där detta lyssnings sätt används är i språk som egentligen består av olika varianter av ljud som förändras beroende på den kultur och samhälle man lever i (Chion, 1994). Dessa olika varianter av ljud formar konsonanter och vokaler som i sig bildar ord med en viss betydelse som har olika betydelser beroende på vem som lyssnar.

Det tredje sättet att lyssna är *reduced listening* och beskrivs av Chion (1994, s. 29) som "...the listening mode that focuses on the traits of the sound itself, independent of its cause and of its meaning". Här lyssnar man efter hur ljudet låter i sig självt istället för vad, varför eller om det betyder något. Blessner och Salter (2007, s. 183) förklarar även att man behöver vara en erfaren lyssnare som har tränat upp en förmåga att koppla bort de andra två sätten att lyssna. De menar att det redan är så djupt involverade i lyssningsmomentet *perception*, där vårt lyssnande sker automatiskt. Detta betyder att man måste bryta sig ur den naturliga instinkten som försöker anknyta ljud till ett igenkännande genom att bara lyssna på hur ljudet låter i sig självt.

With this type of listening, a sound reduces neither to a sonic event, such as a dog bark, nor to a semantic code, such a musical melody, but remains a pure sound for which there is no linguistic or experiential representation.

Blessner & Salter, 2007, s. 183

Ljudet får varken koppling till semantiska värden eller information som den lyssnande kan analysera, utan är helt naket utan mening.

2.3 Realism och ljudkvalitet

Har binaurala ljudfiler kapacitet till att visa mer realism än stereo- eller monoformat? Var ifrån grundar vi våra åsikter om vad som är realistiskt? Är ett realistiskt ljud en perfekt återspeglning av hur ljudet låter i verkligheten där den rumsliga kontexten och akustiken

formar ljudet, eller påverkar vår *perception* vad vi anser är realistiskt? Chion (1994) skriver att vi egentligen inte vet sanningen om hur ljud låter. Filmbranschen har utvecklat en slags överenskommelse som handlar mer om återgivning av ljud som i sig inte behöver vara realistiska i dess sanna mening. Denna överenskommelse har på något sätt övertagit vad allmänheten uppfattar som realistiskt, där publiken gör en kopplingen till ett minne om hur realistiska ljud låter. Chion (1994, s. 108) menar dock att detta minne har influerats av film, vilket gör att publiken inte räknar med den rumsliga akustiken som tillhör ljudets sanna form.

For another thing, when the spectator hears a so-called realistic sound, he is not in a position to compare it with the real sound he might hear if he were standing in that actual place. Rather, in order to judge its "truth," the spectator refers to his memory of this type of sound, a memory resynthesized from data that are not solely acoustical, and that is itself influenced by films.

Chion, 1994, s. 108

Han hävdar även att de ljud som brukar tolkas som realistiska är sådana ljud som är av sämre kvalitet (1994, s. 108). Ett exempel är att man hör brus i bakgrunden eller att ljudbilden i allmänhet känns relativt obehaglig på grund av avvikelser som inte är planerade eller kontrollerade. Chion (1994, s. 108) förklarar genom att skriva att "In much the same way for sound, the impression of realism is often tied to a feeling of discomfort, of an uneven signal, of interference and microphone noise, etc."

Blessner och Salter (2007) antyder att samhället under den industriella revolutionen började en grupp människor definiera vad kvalitativa ljud innebar. Denna grupp skapade en tidskrift som hette *Journal of Acoustical Society of America* år 1929, där kontroll över ljud utsågs vara det nya sättet att hantera ljud på (Blessner & Salter, 2007). Man började forma en uppfattning om att kvalitét av ljud innebär rent och tydligt ljud utan eko, reverb eller rumsklang. Här ser man ett tydligt sammanhang där ljud som upplevs rena utan tillhörande akustiska aspekter anses vara kvalitativa, medan ljud som innehåller brus och oregelbundna missljud anses vara mer realistiska. Igen så behöver detta inte vara så som vår sanna uppfattning av ljud är, men i film- och spelbranschen har detta använts när produktioner där ljud är tillhörande komponent.

2.4 Binauralt ljud

De som har kunskap om hur en binaural inspelning går till, vet att det inte är alltför olik en stereospelning. Båda använder två kanaler för att spela in ljuden för att få en bredare ljudbild, där ljuden får en riktning vilket inte representeras om man lyssnar på en monospelning. Skillnaden mellan stereo och binaural är att den binaurala inspelningen oftast använder en så kallad dummy head, där man sätter in två mikrofoner i öronen på ett artificiellt människohuvud för att efterlikna dess avstånd och riktning. Detta gör man för att återskapa samma förutsättningar som existerar i processen bakom vårt mänskliga hörande. Alltså att använda samma separerad längd, riktning och hinder mellan ljudingångarna (mikrofonerna). Problemet som kan uppstå med denna metod är att man fortfarande har svårigheter att höra när ett ljud kommer upp-, ner- bak- eller framifrån. Detta är fallet om det inte finns öron på det artificiella huvudet, vilket man undviker om man använder en dummy head med öron.

Det har visat sig att öronen är en viktig del för att förhindra dessa svårigheter och det kommer beskrivas tydligare lite senare i detta kapitel. Blesser och Salter (2007) förklarar att Pauli Minnaars och hans arbetskamrater istället satte in mikrofonerna i öronen på en riktigt människa för att få fram en mer korrekt representation av ljudet. Dock är det även viktigt att ha i åtanke att människors öron är olika från person till person, vilket innebär att risken finns att ljudkällornas position uppfattas olika (Blesser & Salter, 2007).

Det som anses vara hindret i en binaural inspelning är själva huvudet eftersom det blockerar ljudet för ett av öronen beroende på vart ljudkällan är lokaliserad. Till exempel om ljudet kommer från vänster så kommer huvudet hindra ljudet att färdas direkt till det högra örat vilket gör att ljudets karaktär förändras. Ljudet blir filtrerat och förlorar styrka i frekvenser. Detta fenomen kallas för den akustiska skuggan. Jan Schnupp, Israel Nelkan och Andrew King (2010) beskriver detta fenomen som en viktig del för att uppleva den binaurala ljudbilden.

Depending on their wavelength, incident sounds may be reflected by the head and torso and diffracted to the ear on the opposite side, which lies within an “acoustic shadow” cast by the head.

Schnupp, Jan, Nelken Israel & King, Andrew, 2010, s.178

Gary Gottlieb (2007, s. 53) skriver om att det finns tre komponenter som är fundamentala förutsättningar för att det ska bli en binaural ljudupplevelse. Den första komponenten som nämns kallas för *interaural intensity* och avgör hur långt bort ljudkällan är i relation till lyssnaren beroende på ljudets intensitet. Den andra komponenten är *interaural arrival time* vilket är ankomsttiden från ljudens uppkomst tills det når öronen. Denna funktion har att göra med hur hjärnan avgör vilken riktning ljuden kommer ifrån. Ifall ett ljud anländer tidigare hos det högra örat än det vänstra så innebär det att ljudkällan finns någonstans på höger sida. De små marginaler som kan skilja i ljudens ankomsttid behandlar vår hjärna för att hitta den mer exakta riktningen som ljuden kommer ifrån. Dock finns det vissa riktningar som inte kan urskiljas ifrån varandra med denna funktion till exempel om ankomsttiden är samma hos båda öronen. Detta innebär att ljudkällan antingen är rakt framför, över, bakom eller under oss vilket den tredje komponenten *physiology* behandlar. I detta fall handlar det om fysiologin bakom människans öron och hur de är formade. Denna form och olika densiteter hos örat möjliggör vår förmåga att höra små skillnader i ljudets karaktär, beroende på vart ifrån det kommer. Dessa små skillnader som skapas är skillnader i ljudets frekvensomfång. Ljud är vibrationer i luften som mäts av svängningar per sekund och desto större svängningar som sker, desto mer bas innehåller ljudet. Oftast så innehåller naturliga ljud som till exempel någon som talar inte bara en sorts svängning, utan den innehåller flera olika svängningar. Detta gör att ljudet blandar olika bas- och diskantfrekvenser vilket skapar ett frekvensomfång. I detta omfång kan till exempel den lägsta frekvensen (svängning per sekund) vara 80 Hz och den högsta frekvensen vara 9000 Hz. Det är här örat filtrerar ljudet för att skapa små skillnader som hjärnan sedan kan analysera. Ett ljud som kommer framifrån kan exempelvis innehålla en frekvenstopp på 7000 Hz, men istället ha en frekvenstopp till 6300 Hz när det kommer bakifrån. Denna skillnad gör att örat kan urskilja på om ljud kommer bak- eller framifrån, över- eller underifrån (Gottlieb, 2007).

Dessa tre komponenter avgör i vilken riktning ljuden kommer ifrån eller hur långt borta de är i relation till lyssnaren. Dock är dessa tre bara de grundprinciper som finns för binaural

hörsel. Det finns även andra komponenter som påverkar ljudens karaktär som vår hjärna sedan behandlar för att räkna ut vart källan kommer ifrån (Moller, 2006). Ett exempel är den akustiska skuggan som är nämnt innan vilket skapas av att en ljudvåg blir böjd när den reflekteras emot människans kropp eller huvud innan den träffar öronen (Schnupp, Nelkan & King, 2010).

2.4.1 Binauralt ljud i spel

Även om spelföretagen har lagt ner mycket tid och pengar på att utveckla de ljudteknologiska aspekterna som är nödvändiga för höja upplevelsen, så menar Michael Duggan (2007) att det fortfarande är begränsat. Han antyder att begränsningen till att efterlikna verkligheten i en fiktiv värld leder till att musiken istället ska väcka den inlevelse som inte teknologin kunde framföra. Andy Farnell som har skrivit boken *Designing Sound* (2010, s. 318) berättar att ”The limitation that immediately presents itself is that sampled sound is fixed in time.”. Han antyder att oavsett hur man försöker redigera om ett ljud så är den fortfarande bara en enstaka version av ett ljud som kan bete sig på flera olika sätt. Detta menar han har påverkat hur ljud används i den traditionella spelbranschen där ljud är bundna till speciella händelser som aktiveras av spelarens handlingar. Dessa ljud kan sedan modifieras av olika realtidsfunktioner som förändrar ljudens karaktär beroende på spelarens agerande, som till exempel panoreringar, rumsakustik och eko tillägg.

Begränsningen av binauralt ljud beror på att spelvärlden är interaktiv där spelaren påverkar vad som inträffar i spelets värld. Detta betyder att ljuden inte är förutbestämda i styrka, panorering eller karaktär utan de förändras beroende på spelkaraktärens position och riktning till ljuden. Till exempel om ett ljud från ett tåg dyker upp på högra sidan av spelkaraktären, men spelaren väljer att snurra på karaktären så att ljudet istället är framför oss. Då panorerar ljudet om sig och dess akustik förändras. En binaural inspelning har redan en förutbestämd panorering, styrka och akustik vilket gör att om man vill implementera detta i spelsammanhang så riskerar man att ta bort det som gör att det känns som en tredimensionell upplevelse.

Denna realtidsfunktion är orsaken till varför det finns svårigheter att använda binaurala inspelningar, eftersom de redan har förutbestämda panoreringar och volymer som ger dem en tredimensionella effekt. Tas detta bort, och realtidsfunktionen används istället, kommer den fiktiva världen att kännas mer sammanhängande med ljuden. Problemet är att en mjukvara modifierar artificiella förhållanden för det binaurala ljudet, vilket betyder att dess ursprungliga tredimensionella effekt kan försvinna.

Den metod som används idag i produktioner av en fiktiv ljudvärld baseras på det manipuleringsätt som nämnts innan, där ljudesignern använder sig av stereo- och monofomat. Dessa ljudfiler manipuleras i efterhand av ljudesignern tills han eller hon känner att ljuden kan implementeras in i spelet. Detta betyder att ljuden inte alls behöver låta så som de gjorde när man spelade in dem, utan en vattenkokare kan till exempel bli ett ljud för ett passerande tåg i spelet. Denna förändring som sker vid manipulation är viktigt att ta hänsyn till, speciellt om man vill försöka efterlikna en binaural upplevelse med hjälp av manipulation. Anledning till detta är att binaurala ljudfiler har förutbestämda förhållanden som bidrar till dess realistiska karaktär.

3. Problemformulering

FRÅGESTÄLLNING: Är den binaurala upplevelsen och uppfattningen av ljud möjlig att simulera genom manipulation av ljudfiler i monoformat?

I tidigare forskning har den binaurala inspelningstekniken bevisats vara bättre för att lokalisera ljud och för att uppleva en mer naturlig återspeglning av vårt hörande. Varför är det fortfarande en mindre okänd metod? Ett problem som ofta dyker upp i samband med binaurala inspelningar är att det finns vissa svårigheter med att manipulera ljudet i efterhand. Till exempel om vi har spelat in en person som går ifrån det bakre högra hörnet i ett rum till det bakre vänstra hörnet med binaural inspelningsteknik. Lyssnar man på denna ljudfil så hör man tydligt att den redan känns panorerad och bestämd i styrka för att visa avstånd och riktning. Den förändrar även sitt frekvensinnehåll när den upplevs att vara rakt bakom än gentemot hur den låter i det bakre högra hörnet och det är dessa förutbestämda förhållanden som gör metoden begränsad. Dock kan man undvika denna begränsning genom att spela in ljuden i mono och använda en *plug-in* som simulerar den binaurala effekten, som enligt Kevin Anker (2010) ska fungera utmärkt. Frågan är dock om denna metod kan jämföras med en riktig binaural inspelning, förändras inte uppfattningen eller upplevelsen av ljuden?

3.1 Metodbeskrivning

Den undersökningsmetod som används för att besvara denna frågeställning är en kombination mellan kvantitativa och kvalitativa studier. De båda används eftersom det är två frågor som behöver besvaras vilket är om det finns en skillnad i lokalisering och uppfattning, men dessutom om det finns en skillnad i känsla och upplevelse. Därför behöver olika undersökningsmetoder tillämpas.

Den kvantitativa metoden används för att samla in data som kan ge någon form av verifierande resultat, som i sig kan anses vara en orsak till det kvalitativa resultatet. Henny Olsson och Stefan Sörensen (2011) förklarar att man använder kvantitativ metod för att bekräfta något om sin frågeställningen, där den antingen kan verifieras eller förkastas. Den kvalitativa metoden används istället för att undersöka en frågeställning på ett djupare plan där man som forskare har ett inifrånperspektiv. Här vill man ha en mer subjektiv version av data ifrån testpersonerna vilket oftast betyder att man fokuserar på en mindre skala av personer, men har mer djupgående frågor (Olsson & Sörensen, 2011).

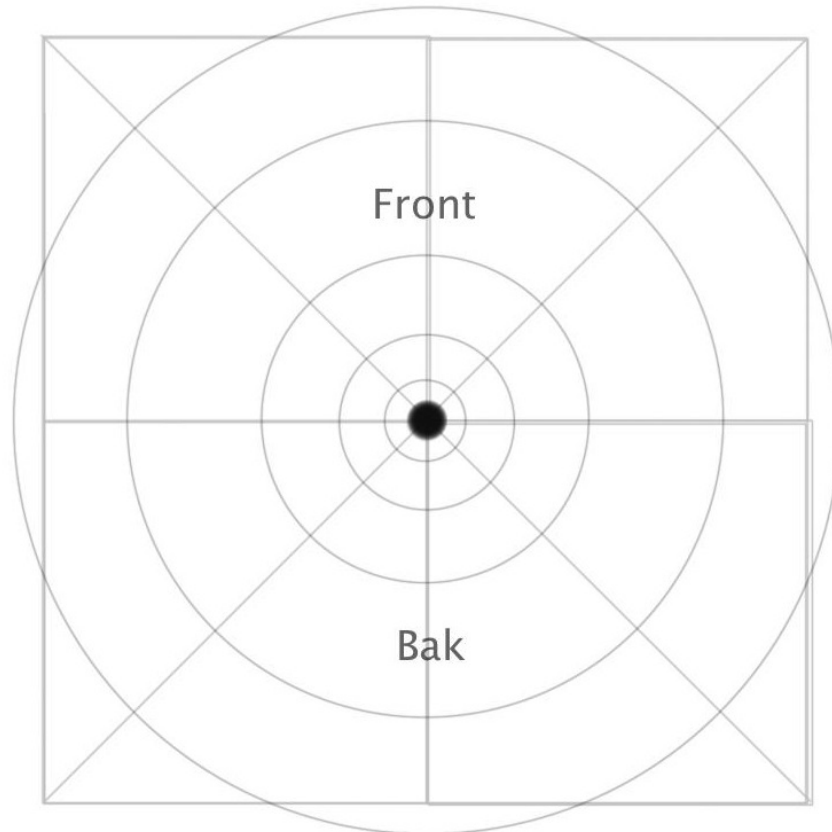
3.1.1 Urval

Urvalet av testpersoner består av sjutton stycken studenter som går på högskolan i Skövde. De är i åldern mellan tjugo till trettio och har ingen tidigare erfarenhet av inspelning, redigering eller mixning av ljud eller musik. Målet var att ha tjugo stycken testpersoner så att en jämn fördelning skulle vara möjlig, dock så var det bara sjutton stycken som deltog.

3.1.2 Datainsamling

Det den kvantitativa metoden undersökte var om lokaliseringsförmågan skiljer sig åt i de olika ljudfilerna, och i så fall hur mycket. Insamlingen av data utfördes genom att spela upp den icke-manipulerade ljudfilen för grupp 1 där de fick som uppgift att i en mall rita de

ljudkällor som de hör och dess position. Grupp 2 lyssnar istället på den manipulerade versionen som utgår efter Blesser och Salters (2007) tolkning på hur ljud ska bearbetas, vilket är en monofil som ska efterlikna en binaural ljudfil. Båda grupperna använder samma mall för att ett resultat skulle vara möjligt att jämföras och mätas, se figur 1.



Figur 1. Lokaliseringsmall för den kvantitativa metoden.

Ljudfilen spelades upp två gånger så att den som lyssnade fick en uppfattning om ljudens karaktär i relation till varandra. Det kan vara svårt att på första försöket rita ut placeringen av ett ljud som upplevs nära, om ljudet därefter upplevs ännu närmre. Mallen behöver därför få en koppling till ljuden så att man som testperson kan avgöra ljudens position i relation till varandra. Till exempel att det 3:e ljudet upplevs längre bort än det 2:a ljudet, men det känns fortfarande närmre än det 1:a ljudet. Detta betyder att det 1:a ljudet är längst bort, det 3:e är i mitten och det 2:a är närmast. Det är detta tankesätt som den dubbla uppspelningen ska sätta igång, så att det blir mer tydligt för testpersonerna att förstå mallens storlek i relation till ljuden.

Det är sjutton stycken personer som testas i den kvantitativa undersökningen där de delades in i två grupper. *Grupp 1* bestod av nio stycken personer och lyssnade på den binaurala ljudfilen som inte har manipulerats, medan *Grupp 2* lyssnade på monofilen. De instruerades att fokusera på att lokalisera ljuden genom att rita dess placering på den förbereda mallen. Denna mall fungerar som ett mätverktyg för hela den kvantitativa undersökningen så att varje individuellt resultat kan jämföras, vilket förtydligar om det finns skillnader i förmågan att lokalisera ljud och i så fall hur mycket. Det blev en ojämn fördelning mellan grupperna eftersom det bara var sjutton stycken testpersoner som deltog. Efterfrågan av testpersoner

från början var att tjugo stycken skulle delta för att en jämnare fördelning skulle bli möjlig. Olyckligtvis så blev detta inte fallet, vilket har inkluderats som en påverkande variabel i diskussion om den insamlade datan.

Det kvalitativa testet består av strukturerade intervjuer, där sju stycken av personerna som gjorde den kvantitativa undersökning utfrågades. Denna undersökningsmetod har som syfte att ta reda på hur deltagarna upplevde den binaurala effekten, i den ljudfil som de lyssnade på.

3.1.3 Dataanalys

Den kvantitativa datan analyserades genom att applicera värden till det resultat som testgrupperna bidragit med. Dessa värden kommer senare i texten att förklaras i kapitel fem, där en utvärdering av resultaten framförs. Dessa värden jämfördes med ett facit som representerar ljudens korrekta värden i form position och förflyttning, vilket medförde att ett resultat kunde sammanfattas och diskuteras.

Den kvalitativa datan analyseras istället genom en latent innehållsanalys som enligt Olsson och Sörensen (2011) är en metod där forskaren ska undersöka den underliggande meningen bakom vad som egentligen sägs i intervjun. Denna metod används eftersom det är den generella känslan om upplevelsen som har betydelse, alltså om man kan se att det finns ett underliggande mönster bakom folks tankar och åsikter.

4. Projektbeskrivning

4.1 Arbetsprocess

Det som utfördes under denna iteration av examensarbetet var att skapa de två ljudfiler som senare användes i undersökningen. Skapandet av dessa filer hade en arbetsprocess bestående av tre moment. Första momentet var inspelningen av ljud samt redigering av dem. Det andra var mixningen, där effekter applicerades. Det tredje momentet bestod av ett pilottest för att se om testpersonen förstod vad som skulle göras och även för att se om ändringar behövde utföras.

Den praktiska delen i detta examensarbete var att skapa de olika binaurala ljudklippen som undersökning baserar sitt resultat ifrån. Två ljudfiler har skapats, den ena filen är inspelad i mono medan den andra i stereo. Den som är i stereo har spelats in med binaural utrustning och har helt *torra ljud*, vilket i detta sammanhang betyder ljud som inte har manipulerats i form av effekter eller *plug-ins*. Den andra ljudfilen har istället *blöta ljud*, vilket betyder att de har manipulerats. *Torra* och *blöta* ljud är svenska översättningar på de engelska termerna *dry* och *wet sounds*, vilket är vanliga termer inom ljudsammanhang. För enkelhetens skull kommer denna text att vara konsekvent och använder därför bara de engelska termerna i fortsättningen. Jeffrey P. Fisher (2011, s. 156) förklarar skillnaderna hos dessa termer genom med orden ”Generally, you choose what you want to change, select an audio effect, adjust its parameters as needed, adjust the mix of dry (original sound) and wet (effected sound), and preview your work.”. Med andra ord så menar han att ljud som inte har någon applicerad effekt, anses vara *original sounds* vilket är ljud i dess ursprungliga form utan påverkan av manipulation.

Denna förklaring finns med för att förtydliga skillnaderna mellan de två ljudfilerna. Utan den kan otydligheter förekomma, om vad som skiljer ljudfilerna åt och behöver därför uppmärksammas.

Det andra ljudklippet är som sagt monoversionen. Den innehåller inspelningar som grundar sig på den inspelnings- och redigeringsmetod som Blesser och Salter (2007) menar att film- och mediabranscherna använder sig av. Detta har i ett tidigare stycke redan förklarats där kontroll, enligt Blesser och Salter (2007), innebär kvalitet. Det man strävar efter i denna metod är att ha en så tydlig och ren ljudkälla som möjligt, med så lite oregelbundna missljud som möjligt. Man vill åt ljudkällans egna läte och inte hur det påverkats av sin miljö som exempelvis reverb, rumsklang och eko.

4.1.1 Inspelning & Redigering

Inspelningen utfördes genom att en kollega tog på sig ett par binaurala mikrofonsnäckor och satte sig på stolen, där ljuden sedan uppkom runt omkring honom. Dessa snäckor ser ut som vanliga hörselsnäckor som man använder för att lyssna på musik när man ska ut och jogga, men är lite annorlunda. Dessa har ett litet membran i varje snäcka som används för att spela in ljud. Dessutom så har de två sladdar kontakter där ena är en sladd som kopplas in om man vill lyssna med snäckorna och den andra om man vill spela in. Har man möjlighet att koppla in båda ska kan man alltså lyssna på den binaurala inspelningen samtidigt som man spelar in. Kollegans huvud var hela tiden riktat rakt fram och var stilla under inspelningarna,

så att ljudkällornas position inte förändrades. Poängen med denna inspelningsmetod var att fånga upp rummets akustik, samt den akustiska skugga som gör små förändringar i ljudens karaktär.

Inspelning av ljuden gjordes i en studio på Högskolan i Skövde där en stol sattes i mitten av rummet så att testpersonen fick uppleva en relativt jämlig akustik hos ljuden, oavsett vilken riktning de kommer ifrån. Efter detta var det bara att börja att spela in. De flesta ljud som användes var statiska i sin position och riktning, som till exempel en pistol som laddas höger om testpersonens huvud. De ljud som inte var statiska förflyttar sig och fanns med för att se om även dessa ljud påverkades av manipulation. Ett exempel på ett sådant ljud som används är en person som förflyttade sig i rummet, där fotstegens position blev lite annorlunda efter varje steg.

Den manipulerade ljudfilen spelades istället in med en riktad shotgun-mikrofon. Detta gjordes för att få en så tydlig ljudkälla som möjligt, där akustiken inte påverkar ljudet. Som tidigare nämnt så antyder Blesser och Salter (2007) att en inspelning med kvalitet är en inspelning med lite brus och den valda ljudkällan är så tydlig som möjligt. Detta var målet med dessa icke-binaurala inspelningar. För att fånga upp de ljud som förflyttades sig i rummet assisterade samma kamrat som agerade som en dummy head i denna process, genom att följa ljudkällan medan den förflyttades. Varför även dessa ljud förflyttar sig i monofilen beror på att det kan vara svårt att få ett ljud att låta naturligt om man ska spela in det på en och samma position. Detta handlar förstås inte om alla ljud, men till exempel personen som går kan låta konstigt och fel om han eller hon inte förflyttar sig. Varför det kan kännas som detta beror på att det blir en annorlunda rörelse om en person står stilla och går på stället, än om han eller hon förflyttar sig i rummet.

4.1.2 Mixning

Enligt Gottlieb (2007) behöver ljud vara kopplat till den kontext som den slutgiltiga produktens miljö förmedlar. Till exempel om genren för en film är action så kommer tittarens antagande av hur ljuden låter i dessa filmer göra att de anser att de mer extrema versionerna av ljud passar till filmens miljö. En pistol som skjuter i en actionfilm bör låta stort och maffigt medan en pistol i en drama- eller dokumentärfilm bör låta mer som det gör i verkligheten, det vill säga mindre pompös.

Sounds need to be understandable within the context of their given environment. Context is determined by the content of the material, which comes with an assumption of the parameters regarding the final product.

Gottlieb, 2007, s. 29

Detta var en fråga som i början av projektet kom till ytan, där det svåra var att bestämma hur ljuden i den manipulerade versionen skulle manipuleras. Det behövdes riktlinjer om vad manipulationen skulle bestå av, så att alla ljud fortfarande har en samhörighet i den kontextuella miljön. Ifall de binaurala ljuden har en realistisk miljö bör också de icke-binaurala ljuden omfatta samma miljö, annars kan den kvalitativa undersökningen påverkas av de olika miljöerna där ljuden berättar olika upplevelser. Poängen av undersökningen var att ta reda på om binaurala och binauralt simulerade ljud kan förmedla samma information och upplevelse, inte om man kan påverka dess kontextuella miljöer annorlunda. Detta löste

sig genom att fokusera på realism i båda ljudfilerna så att den kvalitativa undersökningen om upplevelse inte påverkas av detta.

Detta skapade dock ytterligare ett problem. Eftersom Gottlieb (2007) anser att ljud bör vara kopplade till den kontextuella miljö man vill förmedla, så valdes realism då detta är något som de binaurala ljuden ligger i framkant med. Främsta anledningen till detta är för att monofiler är sådana filer som spelar in ljud på en kanal istället för två. Detta betyder att information som förmedlar ljudets riktning utsluts och ljudkällan upplevs lika mycket i höger och vänster öra. Stereo spelar, som tidigare nämnts, in ljud på två kanaler. Detta gör även binaurala inspelningar vilket gör att en riktning redan är bestämd och ljudet upplevs mer realistiskt. För att skapa denna realism i en monofil så behöver ljuddesignern i ett redigeringsprogram för panorera och mixa ljuden till målet är uppnått. Att redigera och mixa en binaural ljudfil på samma sätt gör att man riskerar att förlora de viktiga aspekterna som tillhör binauralt ljud. Ljud är som sagt vibrationer i luften som studsar på väggar och objekt. Dessa kollisioner med objekt kan reflektera tillbaka ljudet till personen som lyssnar, men från en annan riktning än den riktningen som ljudet ursprungligen kommer ifrån. Ljudet omger alltså den som lyssnar men har olika volym- och klarhetsskillnader som informerar lyssnaren om ljudkällans riktning. Denna information kan försvinna om man manipulerar en binaural eller stereo ljudfil, samt är det svårt att återge i en monofil.

Problemet med detta är ljuden ska förmedla realism men samtidigt utföras på en professionell nivå där ljuden ska anses ha ett kvalitativt värde. För att förtydliga varför detta kan anses vara motsägelsefullt så påminner vi oss om Chions (1994) och Blesser och Salters (2007) förklaring till realism och kvalitet. Som nämnts innan så skriver Chion att ljud som innehåller till exempel brus, anses vara en form av dålig kvalitet av ljud, vilket åskådaren upplever som realism. Blesser och Salter (2007) skriver också att det mest accepterade sättet att behandla ljud i branschen, om kvaliteten är viktigt, är att ha kontroll. Ska man dra en slutsats om vad som egentligen menas här, är att göra en inspelning med brus och missljud om man vill fånga upp en känsla på realism i ljudklippet.

För att lösa denna paradoxala situation applicerades resonemanget att efterlikna den binaurala inspelningens karaktär av ljud och realism. Alltså att anpassa och utnyttja den kontroll man har över monoinspelningen, för att få den att låta som mer den binaurala. För att göra detta användes effekten *equalizer* (EQ) och binaural *plug-in* på monoinspelningarna. *EQ:n* användes för att både ta bort och öka styrkan hos vissa frekvenser för att efterlikna den binaurala inspelningen så mycket som möjligt. Till exempel dörren som öppnas. Den omfattade starkare frekvenser i basregistret i monofilen, än vad den gjorde i den binaurala ljudfilen. För att efterlikna den binaurala versionen utnyttjades *EQ:n* här för att sänka basregistret så att de lät mer lika.

För att tydligare visa hur den binaurala effekten applicerades på monoinspelningen visas nedanför en bild på den inställningen i den *plug-in* som användes. Denna inställning tillhör ljudsekvens med en person som knackar, trycker en kod, öppnar och stänger dörren. Som man kan se är riktningen inte till väst, men nästen mittemellan nordväst och väst. Detta val gjordes för representera hur inspelningen av samma ljud utfördes i den binaurala inspelningen. Radien på denna cirkel är fem meter vilket inte är exakt som det var i den binaurala inspelningen. Här har radien förstörats eftersom akustiken som detta gav efterliknade den binaurala inspelningen mer än om en mindre radie att satts ut.



Figur 2: Binaural inställning på dörrljud-sekvensen.

När ett ljud skulle ändra sin position automatiserades ljudklippet. Detta betyder att man som ljuddesigner använder sig av linjer i redigeringsprogrammet för att ändra ljudets karaktär, beroende på vart i ljudet man vill att det ska bli en förändring. Det som visas på bilden nedan är ljudklippet med fotstegen. Det gröna sträcket här representerar vilken riktning som ljudet ska komma ifrån och hur den ska förändras över tid. I början är den inställd att komma direkt bakifrån, men dras mer till höger med tiden. Till slut stannar riktninglinjen på bilden och förblir samma riktning under resterande tid av ljudklippet. Under denna sekvens är ljudet helt panorerat till höger.



Figur 3: Binaural inställning för automatisering av fotstegen.

4.1.3 Kvantitativt Pilottest

Pilottestet skedde på högskolan i Skövde med två studenter som inte har någon erfarenhet

inom redigering, mixning eller inspelning av ljud eller musik. Detta test handlade främst om hur man som ansvarig för undersökning ska förhålla sig till testpersonen, där man förklarar det han eller hon behöver veta utan att påverka resultatet. Innan varje test fick testperson kort information om vad testet handlade om, vilket var om lokalisering av ljud. De fick dock inte reda på vilken typ av ljudfil de skulle lyssna på, eftersom det inte var väsentligt. Anledningen till varför detta stycke har benämningen "kvantitativ" i rubriken beror på att de inte utförde något pilottest av de strukturerade frågorna, som den kvalitativa undersökningen består av.

Det första testet som utfördes använde den manipulerade versionen. Det enda som sades under detta test var att personen skulle lyssna och lokalisera var ljuden kommer ifrån. Efter att han lyssnat en gång fick han lyssna en gång till, men efter denna gång skulle han markera på mallen vart han upplevde att ljuden kom ifrån. Poängen med detta test var inte att få svar på undersökningens frågeställning, utan poängen var att se om testpersonen förstod sin uppgift. Dessutom var det viktigt att se om ljudfilen var ändamålsenlig när det kommer till volym och tidslängd. Testet visade sig vara lyckat. Personen förstod och det fanns inga klagomål på ljudfilens kvalitet.

Det andra testet utfördes istället med den icke-manipulerade versionen och eftersom det fungerade enligt planerna under det första testet användes samma instruktioner denna gång. Testpersonen förstod och hade inga frågor om vad ska skulle göras, däremot råkade han avbryta sitt lyssnande en kort sekund eftersom han faktiskt trodde att ett ljud inträffade i verkligheten. Redan från början var detta något som anades skulle inträffa och tidigt in i projektet så var det planerat att testpersonen skulle blivit tillsagd att sitta still, oavsett vad han eller hon tror händer i den riktiga världen. Detta nämndes aldrig under detta test på grund av glömska och kom i den riktiga undersökningen att lyftas fram mer tydligt för testpersonen.

4.2 Potentiell förbättring

Denna rubrik finns med om potentiellt intresse, efter detta examensarbete, finns för att undersöka frågeställningen vidare.

En förbättring av projektets utförande kan troligtvis ha varit ett införskaffande av en så kallad dummy head, så att min kollega inte hela tiden behövde gå ifrån sina egna förpliktelser för detta projektets skull. En dummy head är inte en stor utgift men problemet är som beskrivit innan i detta arbete så inkluderar den inte öron som filtrerar ljuden. Detta hade lett till ett införskaffande av ett par öron med form och material som ska efterlikna fysiologin hos människans öron så mycket som möjligt. Denna kvalitet på öronen behöver nödvändigtvis inte ha en större betydelse men det vore tråkigt om fallet var sådant att detta var viktigt, då studien möjligtvis skulle fallera tack vare dess dåliga kvalitet. Detta stycke är, som sagt, en ändring som skulle ha inträffat om kostnader inte hade varit ett hinder.

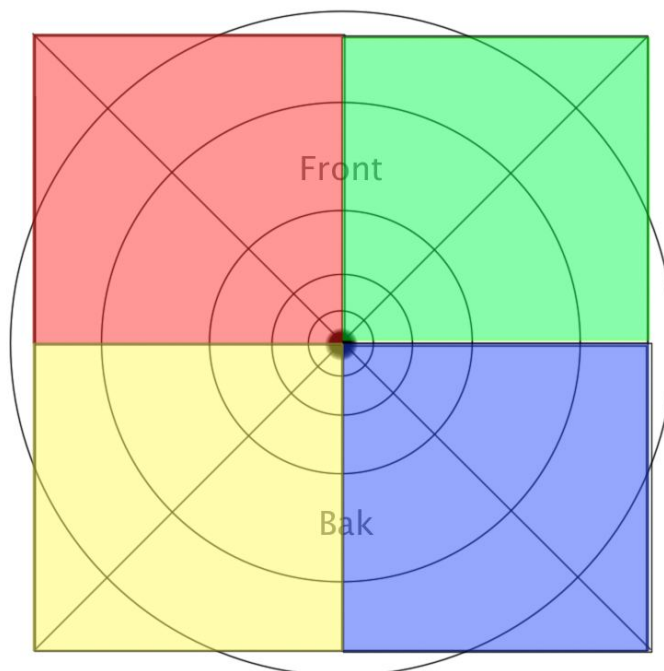
5. Utvärdering

Både den kvantitativa och kvalitativa metoden utfördes på Högskolan i Skövde på studenter, vilket nämnts innan, mellan åldern 20 till 30. Eftersom pilottestet visade att förmågan att utföra ett test kan fallera om man som ansvarig inte är konsekvent i sitt utförande, så har ett dokument skapats för att påminna hur intervjuaren ska förhålla sig till testpersonen. På detta dokument står frågorna som ställdes under intervjuerna, men också riktlinjer om hur strukturerade intervjuer ska utföras. Detta var ett sätt att påminna sig själv om att lyssna och inte försöka påverka den utfrågade med ord eller följdfrågor.

5.1 Kvantitativt resultat

De kvantitativa testen utfördes först eftersom de ansågs skulle gå snabbast att utföra. Det fanns inte så mycket svårighet att hitta personer som vill ställa upp, vilket kan bero på att de fick en kexchoklad om de deltog. Dock blev det svårare på slutet att få folk att ställa upp, speciellt när även den kvalitativa metoden skulle tillämpas. Detta medförde att bara sju personer av tjugo gjorde den kvantitativa studien, där sju av dem även var kvalitativa.

Under jämförelsen av de kvantitativa testen uppstod ett problem. Detta problem handlade om hur testen skulle mätas korrekt för att få ett mätbart resultat. Först utfördes en variant av mätverktyg där *Lokaliseringsmallen* delades upp i fyra hörn, där varje hörn hade olika färger. Idén bakom detta var att om 8 av 10 testpersoner, för den binaurala ljudfilen, uppfattade ett ljud i det blåa området medan 6 av 10 uppfattade detta i ljudfilen i mono, så kan man generellt säga att den binaurala ljudfilen uppfattas bättre. Om man förutsätter att det blåa området är korrekt från början.

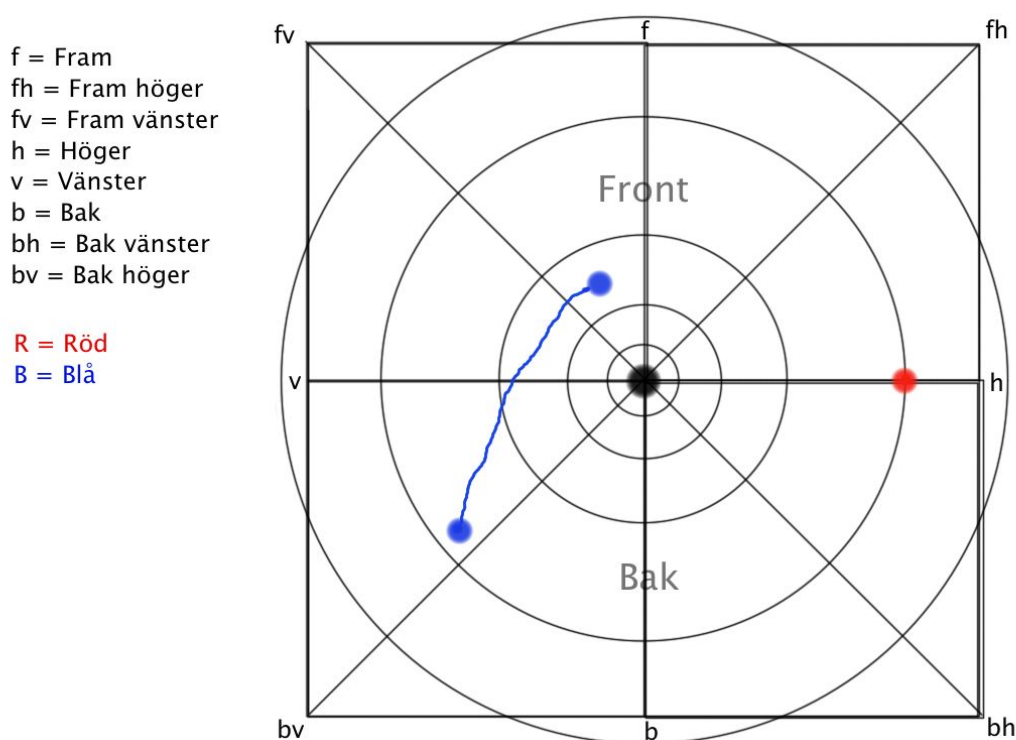


Figur 4: Kvantitativt mätverktyg. Version 1

Dock var denna generalisering alldeles för generell för att göra någon vetenskaplig slutsats

eftersom det fanns en sådan liten skillnad i hur de olika resultaten såg ut. Eftersom detta inte kändes rätt så började ett nytt mätverktyg struktureras och ta form, vilket är det verktyg som slutligen har använts.

I detta verktyg handlar det istället om att man namnger de olika sträckan på mallen, som visar vilken riktning ljuden förhåller sig till mittpunkten av rummet (fyrkanten). Syftet bakom detta var att tydliggöra testernas resultat i form av generaliserbar riktning, istället för generellt område som färgmätverktyget visade. Detta passade bättre för att jämföra resultaten eftersom det fokuserar mer på personen i mitten och hans eller hennes uppfattningen om riktning. Hur själva benämningen av riktning avgjordes, var att förkorta ordens riktning ned till en bokstav och sedan använda cirklarna som avstånd i form av nummer.



Figur 5: Kvantitativt mätverktyg. Version 2.

Det man ser i *Figur 3* är att *Ljud R (röd)* är placerad på sträcket med betäckningen *h* med avståndet *4*. *Ljud R* får därför värdet *h4* och när dessa punkter och sträck får värden så kan de börja jämföras, för att se om det finns likheter eller skillnader. Dock så skapar även detta mätverktyg problem när det kommer till generalisering. För det vi kan se på *Ljud B (blå)* är att ljudet inte är exakt placerat på linjerna, utan är placerade mer i mitten. I dessa fall får man göra tilläggslinjer som tydliggör om ett ljud upplevs mer åt det ena hållet än det andra, vilket simplifierar det ursprungliga resultatet till ett värde som kan jämföras med resterande värden.

Binaural ljudfil	Ljud 1	Ljud 2	Ljud 3	Ljud 4	Ljud 5	Ljud 6
Facit	b4-bh4-h4-h2	fv5	bh4-b3-bv4	h1	f3	v2
Person 1	bh4-h4-h3	v4	v4-bv5	fh4	f3	v5
Person 2	b4-bh4-h4-fh4-h4	bv4	fh4-f4-fv4-v4	fh2	f3	v4
Person 3	bh5-h5-bh5	fv5	bh3-bv4	h2	f1	v5
Person 4	h4-bh4	bv5	fh4-f4-fv4	h1	f3	bv3
Person 5	bh5-h3	v5	h4	h1	f3	fv4
Person 6	b4-bh4-h4-bh4	v4	bh4-b4-bv4-v4-fv4	fh2	b1	fv3
Person 7	b5-bh5-h5-bh5-b5	b5	bh4-b4	bh4	bv4	bh5
Person 8	h3-h4-h2	v4	h3-0-v5	fh1	v3	bv2
Person 9	h4	v5	h3-bh2-b2-bv2-v3	bv4	f3	bv5

Tabell 1: Kvantitativt mätverktyg för jämförelse av värden för den binaurala ljudfilen.

Med denna tabell kan man jämföra de olika värdena som den tidigare mallen har bidragit med. Det som visas i denna tabell är hur många personer och hur deras resultat ser ut, i jämförelse med det facit som innan har dokumenterats.

Det resultat som denna tabell visar är att åtta stycken av alla femtiofyra ljud har placerats korrekt enligt facit, medan arton stycken nästan är korrekta. Dessa ljud som nästan är korrekta har närliggande värden baserat på facit, där skillnaden kan bestå av ett för högt eller för lågt värde i avstånd. Alltså, om värdet i facit är 3 så är ett närliggande värde 2 eller 4. Detta gäller även för riktningen där värdena istället består av bokstäver, till exempel så är ett närliggande fall hos värdet f lika med fh eller fv . Ett annat exempel är det tidigare visade *Kvantitativa mätverktyget*, där *Ljud R* har det sammansatta värdet $h4$. I detta fall kan närliggande värden vara $h3$, $h4$, $fh3$, $fh4$, $fh5$, $bh3$, $bh4$ och $bh5$. De ljud som återstår är de som har inkorrekta värden, vilket inte har värden som är tillräckligt nära för att anses vara

närliggande.

Vissa ljud består av sektioner av värden som är sammansatta med bindesträck, som till exempel *h3-bh4-b5*. Dessa sektioner representerar de ljud som förändrar sin position på inspelningarna. I detta fall representerar värdena ett ljud som finns på höger sida av lyssnaren och förflyttar sig till nästa punkt, vilket är snett bakåt med ett lite längre avstånd. Ljudet fortsätter sedan att röra på sig och är nu bakom lyssnaren, där avstånden har ökat ytterligare. Dessa ljud har även varit svårast att bedöma om de bör anses vara närliggande eller inkorrekta svar, eftersom en testperson kan ha närliggande värden i första delen av sekvensen men inkorrekt i sista delen. I dessa fall har den majoriteten avgjort om sekvensen ska anses var närliggande eller inkorrekt. Det finns även värdet *o* vilket betyder att ljudets position upplevdes i, över eller under testpersonen.

Mono ljudfil	Ljud 1	Ljud 2	Ljud 3	Ljud 4	Ljud 5	Ljud 6
Facit	b4-bh4-h4-h2	fv5	bh4-b3-bv4	h1	f3	v2
Person 1	bh3-h3-fh3	v4	bh3-b2-bv3	h3	f2	bh4
Person 2	fh3-h3-bh3	bv3	b3-b4	bh2	f1	v3
Person 3	bh3-bh5-h2	v5	b3-bv4	h2	bv2	bv4
Person 4	f4-fh4-h2	v4	h2-o-v2	bv1	f1	v3-v4
Person 5	f4-fh4-h4-h2	v4	fv1-h3-h5	h4	f3	bv4
Person 6	h5-h3-h1	v5	bh1-b1-bv1-v3	bh2	bv2	bv4-v4
Person 7	b3-bh3-h3-h2	v3	bh3-b2-bv3	h2	f2	v3
Person 8	b5-bh4-h4-h3	v3	b3-bv4	h2	f2	v4

Tabell 2: Kvantitativt mätverktyg för jämförelse av värden för monoljudfilen.

Resultatet av testet för ljudfilen som är i mono ser annorlunda ut. Här är det bara ett ljud som har korrekt värde enligt facit, medan tjugoen stycken ljud är närliggande och tjugosex stycken är inkorrekta. I detta fall består inte testen av femtiofyra värden som det föregående, eftersom efterfrågan av testpersoner i detta test inte uppfylldes. Här består testen istället av

fyrtyoåttan värden.

5.2 Kvalitativt resultat

I denna del har den kvalitativa datan utvärderats och sammanfattats. Detta utfördes genom förkorta testpersonernas svar och förtydliga deras ståndpunkter och attityder, genom att göra en latent innehållsanalys. Deras svar som visas i tabellerna nedan är alltså inte deras ursprungliga svar, utan en sammanfattning och tolkning av dem.

Frågor för binaurala ljudfilen.	Var det svårt att lokalisera?	Var något svårare att lokalisera?	Var något lättare att lokalisera?	Vad var din reaktion?	Finns ljud som dessa i spel och film?	Vilket sorts spel skulle de passa till?	Kvantitativa värden
Person 1	Nej, det var ganska tydligt.	När de började röra på sig.	Pistolen, den var väldigt tydlig.	Att det var ballt att höra ljuden röra på sig.	Det vet jag inte.	Skräck, för att bygga upp stämning av hur ljud rör sig.	1 korrekt. 2 närliggande. 3 inkorrekta.
Person 2	Nja, lite grann kanske. [Helt min tolkning om personens attityd genom hela intervjun]	Rakt framifrån, det var svårt att avgöra avstånd.	Pistolen, den var distinkt, kompakt och nära.	Ja, verklighetstroget. Annars kan det låta väldigt platt. Höger och vänster typ.	Dessa ljud är mer verklighetstroga. Det kändes som ljuden fanns i miljön omkring än.	Äventyrsspelet 3D, som Skyrim. Även FPS är en given kandidat.	1 korrekt. 2 närliggande. 3 inkorrekta.
Person 3	Egentligen inte. Det svåra var att översätta det på pappret.	Tidningsgrejen. Den kändes väldigt nära och svårt att veta vilket håll den kom ifrån.	Pistolen var tydligast. Men även bollen och dörren kändes tydliga.	Jag trodde faktiskt att det var någon som kom in i rummet, så jag blev förvånad	De kändes mer naturliga, i spel och film så är de mer överdrivna.	Realistiska spel. Första personen men inte nödvändigtvis aktion, utan mer undersök	1 korrekt. 2 närliggande. 3 inkorrekta.

				när dörren visade sig vara ett ljud i testet.		a saker.	
Person 4	Det var inte så problematiskt.	Dörren var svår att förstå att det var ett ljud, kanske för att det finns en dörr här i rummet.	Fyran var väldigt lätt, alltså mantelrörelsen. Men de flesta var lätta men jag vet ju inte om det är rätt.	”Ah, nu fasen kommer någon och stör när jag gör testet” Var reaktion av ljudet från dörren.	De spel jag spelar använder mycket audiolokalisering. Som till exempel Battlefield 3.	FPS eftersom man lever första person hela livet. Därför passar dessa ljud in där.	2 korrekta. 1 närliggande. 3 inkorrekta.

Tabell 3: Kvalitativt mätverktyg för att jämföra åsikter om den binaurala ljudfilen.

Frågor för ljudfilen i mono.	Var det svårt att lokalisera?	Var något svårare att lokalisera?	Var något lättare att lokalisera?	Vad var din reaktion?	Finns dessa ljud i spel och film?	Vilket sorts spel skulle de passa till?	Kvantitativa värden.
Person 1	Njae, vissa var svårare.	Pappret kändes på mig. Han som går var svår att veta om han gick framåt eller bakåt.	Fyran och tvåan. Alltså pistolen och dörren.	Det var udda, man brukar inte höra så mycket fram och bakåt i ljuden. Häpnades även av dörrpipet.	Det känns mest som att man inte fokuserar på bak eller framifrån. Utan ifall något inte syns så är det bakifrån.	Skräckspel, då kan man riktning på fotstegen. Det kan även var viktigt i FPS, fast kanske inte fram och bak som det är med	4 närliggande. 2 inkorrekt a.

						höger och vänster.	
Person 2	Jo att "pin-pointa" exakt. Många ljud kändes bakom mig.	Pappret, kanske pistolen och när han gick.	Bollen var lättast. En studs här, sen en där och en där liksom. Den konstanta förflyttningen förenklade det.	Nej, inte egentligen. Inget speciellt.	Nja, alltså det finns inte så många spel som man måste fokusera så här "Okej, vart hörde jag dörren öppnas."	I skräckspel är ljud en stor grej, så där hade det passat. Men även "stealth-action" spel, och även "shooters."	2 närliggande. 4 inkorrekt a.
Person 3	Vissa var tråliga.	De som var längst bort.	Bollen var tydlig, den hörde man gick från höger till vänster.	Nej, inte egentligen.	Jo det gör det väl, tror jag.	Pusselspel.	3 närliggande. 3 inkorrekt a.

Tabell 4: Kvalitativt mätverktyg för att jämföra åsikter om ljudilen i mono.

6. Slutsats

I detta kapitel sammanställs alla resultat för att få ett enhetligt resultat som denna studie har lett till. Det presenteras även viktiga variabler som är värda att inkluderas i undersökning för att man ska få mer förståelse om vad detta resultat egentligen innebär.

6.1 Resultatsammanfattning

Efter att ha analyserat den kvantitativa datan ser man att de olika ljudfilerna skiljer sig åt, när det handlar om hur bra ljudlokaliseringen uppfattas. Det första man ser var att mängden ljud som korrekt placerats, enligt facit, är betydligt mer i den binaurala ljudfilen än den i mono. Detta var inget chockerande resultat eftersom att det finns mer information i binaurala ljud än vad det finns i mono, vilket skapar dess realistiska karaktär. Denna kännedom har förklarats innan i denna text, under rubriken bakgrund. Exempel på en sådan information, är den akustiska skuggan som presenteras av Schnupp, Nelken och King (2010). Dock så fanns det mer närliggande värden i monofilen än vad det gjorde i den binaurala, vilket antyder att den binaurala *plug-in* som användes fungerade bättre när det kommer till att visa en generell riktning och position. Denna observation ger anledningen till att anta att den binaurala ljudfilen representerar vissa ljud bättre än andra. Alltså att när ett ljud registreras så uppfattas det antingen helt korrekt eller helt fel. Binaurala ljud ger ingen bra generell uppfattning av ljudens riktning och position, utan ger en korrekt eller felaktig position.

Detta är en väldigt grov generalisering av denna observation men är fortfarande intressant att iaktta om någon sanning ligger i det. Fokuset i detta examensarbete ligger på att undersöka om det överhuvudtaget finns skillnader i en simulation av binaurala ljud och riktiga binauralinspelningar. Denna skillnad gäller både i uppfattning och i upplevelse av de olika ljudfilerna, vilket det har visat sig att det gör i uppfattning om ljudens position.

Det viktigast att ta med sig från den kvalitativa studien är hur testpersonen upplevde ljuden, eftersom det är detta som den kvalitativa metoden ska svara på. För att få ett rakt svar på detta kan man jämföra testpersonernas svar på frågan "Vad var din reaktion?". Det raka svar man kommer fram till är att en skillnad i upplevelse även existerar hos de olika ljudfilerna, alltså inte bara uppfattningen av ljudfilerna. Fyra av fyra som hade lyssnat på den binaurala ljudfilen ansåg att ljuden upplevdes naturliga eller verklighetstroga medan två av tre som lyssnat på ljudfilen i mono inte märkte något speciellt. Baserar man en slutsats på denna information så finns det en skillnad i upplevelsen av de olika ljudfilerna. Den binaurala ljudfilen är, i så fall, bättre lämpad för att förmedla en mer realistisk upplevelse.

Frågeställningen för denna undersökning är om den binaurala upplevelsen och uppfattningen av ljud är möjlig att simulera genom manipulation av ljudfiler i monoformat? Det slutgiltiga svaret på detta är inte är möjligt att se. Undersökningen visar att en ljudfil i mono som har blivit manipulerad inte kan ge samma information och upplevelse som en binauralt inspelad ljudfil kan. Både de kvantitativa och kvalitativa studierna i denna undersökning stödjer detta påstående.

6.2 Diskussion

I denna underrubrik kommer de viktiga variablerna att presenteras och diskuteras, vilket kan påverka resultatets trovärdighet. Poängen här är att ge en bättre förståelse om vad de olika studiernas resultat innebär.

Det resultat som den kvantitativa studien gav i denna undersökning visade, som sagt, tydligt att det finns en skillnad mellan de olika ljudfilerna. Dock är det värt att nämna att den metod som användes för att skapa jämförelsebara värden, överdriver skillnaden som finns mellan dem. Denna skillnad som framförs i de olika testerna hade varit betydligt mindre om en utvärdering med mer precisa mätverktyg använts. Detta hade lett till en mer rättvis analys och utvärdering av resultaten, vilket bara kan förbättra en undersökning. Detta var dock något som blev påtänkt alldeles för sent i arbetsprocessen och tiden fanns inte för denna optimering av mätverktyget.

Eftersom efterfrågan av den mängd testpersoner som ursprungligen skulle göra testerna inte uppnåddes blev omständigheterna problematiska, speciellt för de kvalitativa testerna. Dessa tester skulle ursprungligen bestå av fem mot fem, alltså totalt tio tester för både den binaurala ljudfilen och monoljudfilen. Istället blev det sju stycken, fyra mot tre vilket ger det slutgiltiga resultatet mindre trovärdighet. Dels för att det blir ojämn fördelning mellan de olika ljudfilerna, men också för att en person får en procentuellt större chans att påverka resultatet vilket inte är fördelaktigt i en undersökning.

De genusrelaterade frågorna i denna undersökning har också en roll som inte bör undvikas. Som tidigare nämnt så skriver Blesser och Salter (2007) att människors öron och huvudform är annorlunda från person till person. Detta är viktigt ha i åtanke hela tiden när man analyserar de individuella resultaten, eftersom observationer av sådana frågor kan förklara varför vissa resultat skiljer sig från andra. I denna undersökning var det dock bara fyra av sjutton av det kvinnliga könet och baserat på resultaten från de olika metoderna, så fanns det inget mönster att uppmärksamma.

När det kommer till vilka ljud som används i ljudfilerna finns det många aspekter som kan påverka uppfattningen av ett ljud. Chions (1994) term *casual listening* kan beskriva det lyssnandesätt som använts när man utfört dessa test. *Casual listening* har förklarats i bakgrunden men för att uppdatera minnet om vad det är så förklaras det igen. Detta lyssningsätt är det vanligaste och händer hela tiden. Det är det lyssningsätt som samlar information från ljudkällor för att sedan tolkas av hjärnan för att förstå vad det är som låter. Dock nämner även Chion (1994) att det är det minst pålitliga sättet att lyssna, eftersom den kan påverkas av annan information som till exempel personliga minnen. Detta är något som möjligtvis har påverkat det resultat som de kvalitativa intervjuerna gav. *Ljud 2* är en ljudsekvens som består av två ljud. Det ena är knapptryckningar av en kod för att öppna en dörr, det andra är dörren som öppnas och stängs. Dessa ljud är inspelade på Högskolan i Skövde och eftersom testpersonerna var studenter på denna skolan så är det logiskt att anta att denna ljudsekvens har en annan betydelse än de andra ljuden. Det finns alltså skäl till att tro att deras upplevelse om vad som majoriteten ansåg var mer verklighetstroget, var det ljud som de trodde inträffade på riktigt. Det är en ljudsekvens som de hör hela tiden i deras vardagliga miljö och kan därför påverka vad de anser är mer verkligt.

6.3 Framtida arbete

Det resultat som denna undersökning bidragit med öppnar möjligheten för andra frågor och undersökningar. Exempelvis vore det intressant att se hur mycket man kan manipulera en binaural ljudfil innan realismen och lokaliseringsprecisionen förminskas. En annan frågeställning är om det finns andra sätt utnyttja monoinspelningar för att sedan omvandla de till binaurala inspelningar. Detta skulle kunna utföras genom att spela upp manipulerade monoinspelningar i högtalare som sedan spelas in av binaural utrustning. Även här är det intressant att se om den binaurala känslan och upplevelsen fortfarande finns kvar.

Några svar från intervjuerna ansåg att vissa ljud vara lättare att lokalisera, vilket de flesta hade en ganska gemensam åsikt om. Däremot var åsikterna om vad som var svårt att lokalisera helt annorlunda. Detta kan bero på det som nämndes i diskussionen där vissa ljud kan påverkas av personliga minnen, känslor eller bilder. Även detta är värt att undersöka eftersom man kanske kan få en tydligare bild av vad för sorts ljud man ska undvika i binauralt format, eftersom de möjligtvis förvirrar mer än vad de förtydligar.

För framtida arbeten som handlar om samma eller liknande frågeställning som denna undersökning gör, så finns det många förbättringsaspekter att tänka på. Många har nämnts innan i denna text, men det som skulle tillföra mest trovärdighet till undersökning vore att göra testen i en större skala. Dessutom så skulle ett införskaffande av en dummy head förenkla inspelningen av binaurala ljud, så att man kan sköta det mesta själv.

Referenser

- Anderson, Tim J. (2005) *Making Easy Listening : Material Culture and Postwar American Recording*. Minneapolis, MN, USA: University of Minnesota Press.
- Anker, Kevin; Merton, Orren. (2010) *Logic Pro 9 Power! : The Comprehensive Guide*. Boston, MA, USA: Course Technology / Cengage Learning.
- Bates, Bob. (2004) *Game Design, Second Edition*. Boston, MA, USA: Course Technology / Cengage Learning.
- Blessner, Barry; Salter, Linda-Ruth. (2007/2009) *Spaces Speak, Are You Listening? : Experiencing Aural Architecture*. Cambridge, MA, USA: MIT Press.
- Chion, M. (1990/1994) *Audio-vision*. New York Chichester: Columbia University Press. [Ursprunglig titel: *L'Audio-vision*].[Elektronisk version]. Tillgänglig på Internet: http://soundenvironments.files.wordpress.com/2011/11/michelchion_audio-vision.pdf. [Hämtat 12.12.20].
- Crooks, Clayton, II. (2004) *Awesome 3D Game Development : No Programming Required*. Hingham, MA, US: Charles River Media / Cengage Learning.
- Duggan, Michael. (2007) *Official Guide to 3D GameStudio*. Boston, MA, USA: Course Technology / Cengage Learning.
- Farnell, Andy. (2010) *Designing Sound*. Cambridge, MA, USA: MIT Press.
- Fisher, Jeffrey P.. (2011) *Soundtrack Success : A Digital Storyteller's Guide to Audio-Post Production*. Boston, MA, USA: Course Technology / Cengage Learning.
- Gottlieb, Gary. (2007) *Shaping Sound in the Studio and Beyond : Audio Aesthetics and Technology*. Boston, MA, USA: Course Technology.
- Logic Pro 9. (2009) [Datorprogram] Apple Inc. Tillgänglig på Internet: <http://www.apple.com/se/logicpro/>
- Moller, Aage R.. (2006) *Hearing : Anatomy, Physiology and Disorders of the Auditory System*. Burlington, MA, USA: Academic Press.
- National Research Council Staff; Mavor, Anne S. (Editor); Durlach, Nathaniel I. (Editor). (1994) *Virtual Reality : Scientific and Technological Challenges*. Washington, DC, USA: National Academies Press.
- Olsson, H., Sörensen, S. (2011) *Forskningsprocessen - Kvalitativa och kvantitativa perspektiv (3:e uppl.)*. Stockholm: Liber
- Schnupp, Jan; Nelken, Israel; King, Andrew. (2010) *Auditory Neuroscience : Making Sense of Sound*. Cambridge, MA, USA: MIT Press.
- Ubisoft Montreal. (2010) *Tom Clancy's Splinter Cell: Conviction*. [Tv-spelsprogram].

Ubisoft. Tillgänglig på Internet: <http://www.ubi.com/SE/Games/Info.aspx?pId=7757>.

Appendix A - Frågeformulär

Dessa intervjuer kommer att spelas in och transkriberas (skrivas ner i textform). Du har full anonymitet men jag kommer att skriva ner ditt namn för att jag själv ska komma ihåg intervjuerna. Dessa intervjuer är till för att ta reda på en mer kvalitativ och djupgående information om hur du upplever de ljud du hör.

1. Var det svårt att lokalisera ljuden, alltså att höra var de kom ifrån? Motivera.
2. Fanns det några/något ljud som var svårare att lokalisera? Vilket/vilka?
3. Fanns det några/något ljud som var lättare att lokalisera? Vilket/vilka?
4. Upplevde du något speciellt eller annorlunda med ljuden när du lyssnade? Hur var din reaktion?
5. Finns det någon skillnad på dessa ljud och de som redan finns i spel och film? Vilken?
6. Om dessa ljud skulle finnas i spel, vilken typ av spel skulle de passa bäst till? Varför?

SAKER ATT TÄNKA PÅ

Strukturerad intervju

Samma frågor ställs till alla och alla svar kategoriseras i någon specifik svarskategori. Samma frågor kan ställas i ett frågeformulär.

Allmänna riktlinjer vid intervju

- Motivera informanten att berätta
- Skapa tillit och ge bekräftelse
- Poängtera att man vill lära sig
- Lyssna
- Var konkret

De kategoriseras

...genom att göra en latent innehållsanalys. Tolkning av uttalandets innebörd.