

UPPLEVD RUMSKLANG

En studie i olika gestaltningar av rumsklang
med en visuell referens

PERCEIVED REVERBERATION

A study in the various portrayals of reverb with
a visual reference

Examensarbete inom huvudområdet Medier, estetik
och berättande
Grundnivå 30 högskolepoäng
Vårtermin 2014

Jonatan Sjöberg

Handledare: Lars Bröndum
Examinator: Toivo Burlin

Sammanfattning

Denna rapport undersöker om olika gestaltningar av rumsklang ger högre grad perceptuellt realism med en visuell referens. Undersökningen baseras på den forskning inom rumsklang och perception av dessa som skett tidigare. I dessa fall utan en visuell referens. I detta fall undersöks gestaltningen som ges av konvolutionsreverb, fördröjningsnätsreverb och en kombination av dessa. För att undersöka detta skapades tre videoklipp med samma bild och samma ljudeffekter med skillnad i den reverbteknik som användes. Undersökningen skedde med semi-strukturerade kvalitativa intervjuer som bestod av fyra frågor om informantens upplevelse. Det utfördes nio intervjuer med lika många informanter. Där varje informant endast fick se på ett av videoklippen. Resultatet av intervjuerna visade att ingen gestaltning gav högre grad perceptuell realism utan att så länge gestaltningen i sig låter bra spelar det mindre roll vilken teknik man använder. I en förlängning hade det varit intressant att utföra samma undersökning i en interaktiv kontext.

Nyckelord: Ljuddesign, Rumsklang, Reverb, Perceptuell realism

Innehållsförteckning

1	Introduktion	1
2	Bakgrund	2
2.1	Vad är rumsklang?	2
2.2	Närvaro som realism	3
2.3	Olika tekniker för att skapa rumsklang digitalt	3
2.4	Tidigare forskning kring rumsklang, ljud och bild	4
2.5	Diegesis och IEZA	5
3	Problemformulering	7
3.1	Metodbeskrivning	7
3.1.1	Undersökningsmetod	8
4	Genomförande	10
4.1	Artefaktbeskrivning	10
4.2	Arbetsprocessen	10
4.2.1	Inspelning	10
4.2.2	Ljudläggning och mixning	12
4.3	Pilotstudie	14
5	Utvärdering och analys	15
5.1	Artefakt 1 - Konvolution	15
5.2	Artefakt 2 – Fördröjningsnät	16
5.3	Artefakt 3 – Kombination	16
5.4	Slutsatser	17
6	Avslutande diskussion	19
6.1	Sammanfattning	19
6.2	Diskussion	19
6.3	Framtida arbete	20
	Referenser	21

1 Introduktion

Rumsklang är något alla människor upplever varje dag. Det är något som tas för givet och som man inte tänker på fören extremerna upplevs, till exempel i en gångtunnel eller i ett trångt utrymme. Rumsklangen ger mycket information om rummet man befinner sig i och om ljudkällan man hör. I skapandet av en virtuell värld är rumsklangen därför en viktig komponent. Rumsklangen utgörs av de ljud som studsar på alla ytor i ett rum och når örat efter de ljud som tagit sig till örat direkt. Att skapa rumsklang digitalt är därför en utmaning som kräver kraftfulla datorer. För att skapa rumsklang digitalt finns det tre olika kategorier av tekniker som används, fördröjningsnätverk, konvolution och beräknad akustik (Välimäki, Parker, Savioja, Smith & Abel, 2012, s. 1422). När man på svenska talar om återskapandet av rumsklang används ordet reverb. Reverb kallas tekniken, mjukvara eller hårdvara, för att simulera rumsklang. Rumsklang syftar då på fenomenet och reverb syftar på tekniker för att skapa rumsklang. I spel är det ofta reverb baserade på fördröjningsnätverk som används för att skapa rumsklang då dessa är effektiva och inte kräver mycket kraft att räknas ut dock ger de sämre resultat och konvolutionsreverb har därför börjar utvecklas för att vara mer effektiva i spelmotorer.

Studier visar att människor har god urskiljningsförmåga när det gäller att bedöma storleken på rumsklang och att rumsklang ger högre grad närvaro. Att skapa perceptuellt realistiska rumsklang kan därför anses vara viktigt för att skapa en perceptuellt realistiskt spelupplevelse. Genomgående för de studier som presenteras här är att de sker utan en visuell referens. Denna undersökning har därför i syfte att undersöka vad olika gestaltningar av rumsklang har för betydelse för den upplevda perceptuella realismen i en virtuell miljö med en visuell referens.

För att undersöka detta skapas tre ljudläggningar på ett videoklipp inspelat från spelet Arma 3 (Bohemia Interactive, 2013). Ljudläggningarna utgår från fördröjningsnätverk, konvolution och kombinationen av dessa. Informanter får titta på ett av dessa och sedan svara på frågor i en semistrukturerad intervju. Ljudläggningen av klippen utgår från effekt- och zonljud enligt IEZA-modellen (Van Tol & Huibers, 2008).

2 Bakgrund

Det sker mycket forskning kring virtuella miljöer och spelmotorerna blir allt mer avancerade (en spelmotor är system utformade för utvecklingen av spel med olika färdiga funktioner för rendering av grafik, fysik, ljud m.m.). En del i denna forskning berör specifikt ljud och hur människor kan urskilja ljud, ljudkällor och hur man kan simulera rummet i en virtuell kontext. Gemensamt för forskningen i detta område är att den ofta sker i en kontext utan visuell referens. I den forskning där man har en visuell kontext handlar ofta om en undersökning mellan bilden och ljudet. Ett exempel är en undersökning utförd av Skalski och Whitbred (2010) där man tittade på skillnaderna i upplevelsen mellan olika bildkvalitet och olika ljudåtergivningsformer, slutsatsen blev att bättre ljudåtergivning var viktigare än bra bildkvalitet för upplevd närvaro (närvaro handlar om till vilken grad man känner närvaro i en virtuell kontext (Lombard & Ditton, 2006). Närvaro är en översättning från det engelska ordet presence. I den forskning som presenteras här används ordet presence men denna text kommer att använda ordet närvaro). När en virtuell värld skapas är det därför inte bara viktigt att satsa på att återskapa det visuella utan också det auditiva och att skapa rummet i ljudet är en viktig del i detta.

Att återskapa rummet är en teknisk utmaning som ständigt är i utveckling. Denna undersökning kommer undersöka olika gestaltningar av rumsklang med en visuell referens. Under rubrikerna som följer kommer olika tekniker för att återskapar rumsklang digitalt att presenteras. Relevant forskning kring perceptionen av rumsklang kommer också att presenteras och slutligen kommer det att göras en viss presentation av hur denna undersökning kommer förhålla sig till ljud och rumsklang.

2.1 Vad är rumsklang?

Rumsklang är något människor och djur upplever varje dag. Rumsklang ger information om rummet man befinner sig i, om dess storlek, om det är möblerat eller ej och om ytorna är hårda eller mjuka. Rumsklang hjälper också hjärnan att placera en ljudkälla i rummet, om den är långt borta eller nära. Det finns tre kategorier av ljud som tas upp när det handlar om perceptionen av rummet. Det är de direkta ljudet, tidiga reflektioner och det som man kallar efterklang eller sena reflektioner (Huber & Runstein, 2010, s. 69). Det direkta ljudet är, som det låter, de ljud som når örat direkt, alltså de ljudvågor som tar sig från ljudkällan till örat obehindrat. Dessa ljud ger information om ljudkällans plats i rummet och ljudkällans storlek. Tidiga reflektioner är ljud som når örat efter att det studsats på de mest dominanta ytorna i rummet och kommer till örat kort efter det direkta ljudet, mellan 50 millisekunder och 80 millisekunder (Huber & Runstein, 2010, s. 70). De tidiga reflektionerna utgör en del av informationen om storleken på rummet. De ljudvågor som fortsätter att studsas runt i rummet och når örat efter 50-80 millisekunder kallas efterklang eller sena reflektioner. De sena reflektionerna kan ofta märkbart höras efter det att ljudkällan slutat låta. Dessa reflektioner ger ytterligare information om rummet och vad de innehåller. Ett rum med hårda, öppna ytor ger ofta fler och längre sena reflektioner och ger ett längre rumsklang (Huber & Runstein, 2010, s. 70-71). Eftersom dessa reflektioner studsats på flera ytor har flera frekvenser absorberats och ljudet har ofta en annorlunda klang än originalljudet. Man mäter rumsklangen från dess avklingningstid. Det är den tid det tar för rumsklangen att minska med 60 dB per sekund och detta bestäms av rummets absorptionsegenskaper (Huber & Runstein, 2010, s. 71) (Välimäki, m.fl, 2012, s. 1421). Värt att nämna är också att

det direkta ljudet och de tidiga reflektionerna förändras beroende vart man befinner sig gentemot ljudkällan och de sena reflektionerna ofta är konstanta eftersom spridningen av ljudet i rummet är den samma vart man än befinner sig i det.

2.2 Närvaro som realism

Begreppet närvaro kan ha olika betydelse beroende på i vilken kontext man använder det (Lombard & Ditton, 2006). Närvaro är ett samlings begrepp för hur en individ känner närvaro i en virtuell kontext, till exempel film, tv eller spel. Det finns flera olika sätt som närvaro kan få en spelare kännas sig involverad i den virtuella världen. Ett exempel på detta är immersion där spelaren kan känna sig uppsluka i spelet av spelets handling och mekanik, ofta i kombination. Men immersion har lite att göra med realism utan mer om spelupplevelsen i sin helhet (Adams, 2010, s. 25-26). I detta fall kommer närvaro att används som ett mått på till vilken grad ett medium kan förmedla exakta representationer av objekt och händelser, så att dessa upplevs som verklighet och att spelaren på detta sätt känner en närvaro i den virtuella världen (Lombard & Ditton, 2006). Lombard och Ditton (2006) beskriver närvaro som realism utifrån två typer av realism, social realism och perceptuell realism. Social realism mäter graden av hur trolig eller sannolik ett scenario är, om det skulle kunna hända utanför mediet eller inte. Perceptuell realism behandlar graden av hur objekt, människor och ljud känns verkliga. Lombard och Ditton (2006) ger ett exempel på hur en scen från en science fiction film kan ha hög perceptuell realism men låg social realism då allt ser ut och låter som man skulle kunna förvänta sig men att händelserna inte är sannolika. Ett annat exempel som ges är att en tecknad serie kan ha hög social realism men låg perceptuell realism då scenariot är sannolikt men det inte representeras realistiskt i bild och ljud. Eftersom ljud spelar en roll i att uppnå högre grad närvaro i och med den perceptuella realismen så spelar alltså rumsklang och ljudets gestaltning en roll. För att mäta rumsklangens roll för upplevd realism i ett spel är därför närvaro lämpligt. Denna undersökning kommer att primärt utgå från den perceptuella realismen då det är denna del som behandlar just hur världen låter och ser ut.

2.3 Olika tekniker för att skapa rumsklang digitalt

För att skapa rumsklang i spel genereras den digitalt i spelmotorn. Det finns ett stort antal olika tekniker för att skapa rumsklang digitalt men alla är inte lämpliga att använda i en spelmotor då det antingen krävs för mycket datorkraft eller för att processen inte kan ske i realtid.

Reverbsalgoritmer kan delas upp i tre huvudkategorier, fördröjningsnät (delay networks), konvolution och beräknad akustik (computational acoustics) (Välimäki, m.fl, 2012, s. 1422). Inom dessa tre kategorier finns det ett antal olika tekniker som bygger på huvudprincipen men arbetar efter olika algoritmer. Fördröjningsnät går ut på att fördröja en signal (när ett ljud spelats in eller skapas digitalt kallar man det för signal), låta signalen gå igenom olika filter och föras tillbaka genom olika vägar enligt parametriserade efterklangsegenskaper. Konvolution innebär att man faltar en signal med en inspelad impulsrespons. Faltning är en matematisk operation på två funktioner för att skapa en tredje som då blir en modifierad version av någon av de ursprungliga funktionerna. Signalen man för in faltas med en impulsrespons för att skapa en utsignal med rumsklang. Impulsresponsen kan genereras på flera olika sätt, den kan antingen vara uträknad rent digitalt eller inspelad i ett rum. En impulsrespons av ett rum spelas in genom att spela upp antingen en sinusvepning eller en

transient, som ett startskott, och spelar in ljudet från den position som rumsklangen önskas fångas från (Välimäki, m.fl, 2012, s. 1434)(Nair, 2012). Den inspelade responsen konverteras sedan i ett program som utför faltningen. Konvolutionsreverb ger ofta bra representationer av rummet. Konvolutionsreverb är tunga för en dator att räkna ut, men ljudimplementationsverktyg och ljudmotorer för spel som till exempel Wwise (Audiokinetic Inc, 2014) har börjat ge stöd för konvolutionsreverb för användning i spel.

Med beräknad akustik låter man signalen driva en simulering av den akustiska energins utbredning i en modellerad geometri (Välimäki, m.fl, 2012, s. 1422). Man simulerar alltså ljudets reflektioner i ett modellerat rum. Den teknik som främst används i spelmotorer baseras på fördröjningsnät. Detta beror på att fördröjningsnät är effektiva och lätt drivna, på bekostnad av exakthet (Välimäki, m.fl, 2012, s. 1441). Men konvolutionsreverb börjar bli vanligare i spelmotorer, trots att dessa är tunga att driva. Vanligt är då att man använder en kombination av fördröjningsnät och konvolution. Utvecklingen kring reverb i virtuella miljöer går ständigt framåt och tekniker som ray tracing och förberäknade ljudvägssimuleringar testas för användning i spelmotorer (Chandak, 2012).

2.4 Tidigare forskning kring rumsklang, ljud och bild.

I undersökningar av Frissen, Katz och Guastavion (2009) utförde man fyra experiment där man testade vad olika ljudkällstimuli har för effekt på perceptionen av rumsklang. Det första var ett referensexperiment. I det andra experimentet genererade man nio impulsresponsor som hade en varierad avklingningstid mellan 1.48 och 2.12 sekunder i lika stora steg av 80 millisekunder. I det rumsklang som skapades spelade man upp en 170 millisekunder långt vitt brus och en mansröst som uttalar ett ord under 600 millisekunder. Deltagarna fick lyssna på två par stimuli där en av dessa, med rumsklangstid på 1.8 sekunder, användes som referens. Deltagarna fick efter att ha lyssnat på de två välja om de hade samma eller olika rumsklangstid. Resultatet visade på liten variation mellan brusparen och röstparen. Slutsatsen blev att typen av stimuli inte har någon effekt på förmågan att urskilja rumsklang. Detta visar dock inte om olika stimuli får rumsklangen att uppfattas olika. I experiment tre utförde man tester för att undersöka storleksbedömningen av rumsklang. Samma teknik användes här fast med tre olika ljudkällor, ett trumkomp, sång och tal, som spelades upp i tre olika längder vardera. Deltagarna fick lyssna på de olika ljuden och sedan sätta värden för att bedöma storleken på rumsklangen. Resultatet och slutsatsen blev att längden på ljudet inte hade någon betydlig effekt på den upplevda mängden rumsklang, storleken på rumsklangen, men att typen av ljud gjorde att man bedömde rumsklangen olika. Efter detta experiment valde man att utföra ett nytt då skillnaden i längd på de olika ljuden var för stor. I experiment fyra gjorde man därför samma test, denna gång med längden på ljuden bättre matchade. Experimentet gav resultat som var enhetliga med resultaten från experiment tre. Slutsatsen blev att typen av ljudstimuli inte påverkar urskiljningsförmågan av den upplevda rumsklangen men att den upplevda mängden rumsklang skiljde sig mellan olika ljudstimuli. Dessa experiment visar alltså att människor kan skilja på olika rumsklangstorlekar men eftersom experimenten utfördes utan någon visuell referens så visar dem inte hur rumsklangen upplevs i en visuell kontext.

I en annan undersökning utförd av Larsson, Västfjäll och Kleiner (2008) lade man fram hypoteserna att konsekvens mellan ljudinformationen och ljudets plats i rummet är en förutsättning för hög rumslig närvaro och att rumsklang ökar upplevde närvaro. Genom att simulera akustiken i ett rum med hjälp av Head Related Transfer Function, HRTFs, och

Binaural Room Impulse Response, BRIRs, kunde man låta deltagare lyssna på ljud som spelades upp i det simulerade rummet. Deltagarna fick lyssna på olika konstellationer av ljud bestående av ljud som rör sig, ljud som är stilla och artificiella ljud, i detta fall färgat brus. De fick lyssna på både utan rumsklang och med den simulerade rumsklängen. Deltagarna använde sig av hörlurar och bar bindel. Resultaten man fick var att tilläggnings av rumsakustik ökade närvaro och känslan av realism markant. Resultaten visade också att konsekvens mellan ljudinformationen och ljudets plats i rummet ökade känslan av närvaro. Denna undersökning är relevant eftersom den visar att rumsklang är en viktig del i att simulera verkligheten. Men eftersom den utfördes så att deltagarna inte hade någon visuell referens så visar den inget samband på hur närvaro och upplevd realism fungerar i en kontext där man har en visuell referens, till exempel i en film eller ett spel.

Undersökningarna visar alltså att människor kan urskilja storleken på ett reverb beroende på ljud som spelas i rummet och att reverbet ökar närvaro och den upplevda realismen i en simulerad virtuell kontext. Som tidigare konstaterat skedde båda dessa undersökningar utan en visuell referens. I en artikel gör Hulusic m.fl (2012) en sammanställning av en del av den forskning som skett inom just simuleringen av ljud och bild. Man tar upp att det finns en tydlig tvärmodal interaktion mellan synen och hörseln och att dessa påverkar varandra på olika sätt (Hulusic, m.fl, 2012, s. 118). Till exempel tar man upp ett fenomen där frekvensen av hur ett ljud fladdrar påverkar hur en person uppfattar och synkroniserar ljudet med ett ljus som flimmar, att ljusets flimmar korrigeras efter frekvensen av ljudets fladder. På liknande sätt beskrivs den så kallade buktalareffekten (Hulusic, m.fl, 2012, s. 119). Buktalareffekten innebär att ett ljud som spelas upp från en annan källa än representerat fortfarande kopplas till källan. Exemplet som ges är hur rösten från en person i ett tv-program kommer från högtalaren men fortfarande uppfattas komma från personen i bilden. Samma samband mellan syn och hörsel kan därför hävdas finnas i en rumslik kontext, alltså sambandet mellan hur ett rum ser ut och hur ett rum låter. Detta är något som tas upp i forskning och utvecklingen av nya tekniker för att generera rumsklang i virtuella miljöer och något som denna undersökning kommer att röra vid.

2.5 Diegesis och IEZA

Chion (1994) använder sig av termerna digetiska och icke-diegetiska ljud i sin diskussion om ljud i film. Digetiska ljud innefattar de ljud som finns i handlingen och kan anses vara realistiska i filmens kontext, ljud som finns i scenen. Motsatsen är då icke-diegetiska ljud och är de ljud som står utanför handlingen (Chion, 1994). Van Tol och Huiberts (2008) använde sig av dessa termer för att skapa IEZA (Van Tol & Huiberts, 2008). IEZA är ett ramverk för kategorisering av ljud i spel och består av fyra kategorier, Interface (gränssnitt), Effekt, Zon och Affekt. Effekt- och zonljud faller under diegetiska ljud medan gränssnitt- och affektljud faller under icke-diegetiska ljud (Van Tol & Huiberts, 2008, s. 3). Effektljud syftar till ljud som produceras eller tillhör ljudkällor i spelvärlden ofta interagerbara ljud, ljud som skapas av att spelaren interagerar med spelvärlden, som går att skapa eller skapas som effekt av integrering med spelvärlden. Det kan vara ljud som avatarens fotsteg, smällen från explosioner eller pip från knappar man kan trycka på. Zonljud refererar till diegetiska ljud som utgör miljöljud, man brukar kalla ljud i denna kategori ambientljud. Det är ofta ljud som inte direkt syns i bild men kan antas vara där och tillhöra spelvärlden, till exempel fågelkvitter eller muller i grottor (Van Tol & Huiberts, 2008, s. 3). Ljud kopplade till spelets gränssnitt är icke-diegetiska ljud eftersom dessa står utanför spelvärlden. Det är ofta ljud i ett spels menyer eller ljud som ska förmedla specifik information om spelet till spelaren, som

varningsljud eller bekräftande ljud kopplat till gränssnittet. Affektljud är också ljud som står utanför spelvärlden och är därför icke-diegetiska. Affektljud kan liknas med zonljud då dessa är till för att förmedla miljön i spelet. Affektljud består ofta av musik och kan användas för att förstärka en social, kulturell eller känslomässig referens i spelet (Van Tol & Huibers, 2008, s. 4).

Denna undersökning fokuserar på närvaro som realism i en spelkontext och kommer därför uteslutande att fokusera på effekt- och zonljud då dessa är en nödvändighet för att återskapa en representation av realism. Då gränssnittsljud och affektljud står utanför spelvärlden är dessa inte intressanta för undersökningen. Effektljud är nödvändiga då det krävs en ljudkälla för att skapa reverb och zonljud är nödvändiga för att bygga upp en perceptuellt realistisk ljudbild. Man kan också hävda att reverbet faller under både effekt och zon då reverbet är en effekt av en handling och är ljud som står utanför spelarens kontroll och bygger upp miljön.

3 Problemformulering

Utvecklingen inom spelmotorerna går konstant framåt och mer vikt läggs på ljuden, vilket man till exempel kan höra i spel som Battlefield 4 (Dice, 2013). Samtidigt kan reverben som fortfarande används inte helt återskapa exakta representationer av rummet i ett spel, även om ljudmotorer som till exempel Wwise (Audiokinetic Inc, 2014) har stöd för konvolutionsreverb. Ofta blandar man reverb baserade på fördröjningsnätalgoritmer med konvolutionsreverb där man kan för att få en lämplig blandning av direkta reflektioner, tidiga reflektioner och sena reflektioner. Den forskning som presenterats visar att människor har en god förmåga att urskilja storleken på rumsklang och att rumsklangen ger en förhöjd upplevd närvaro. Gemensamt för dessa undersökningar som presenterats är att de utförts utan en visuell referens. Denna undersökning studerar hur olika gestaltning av rumsklang påverkar den upplevda perceptuella realismen med en visuell referens. Frågeställningen lyder:

- Vilken gestaltning av rumsklang skapar högre grad perceptuell realism med en visuell referens, som gestaltas i realistisk stil?

Resultatet av denna undersökning kan fungera som ett hjälpmedel för ljudredigering och andra personer inblandade i skapandet av spel och film för hur man kan tänka kring användandet av reverb och gestaltning av rumsklang. Resultatet kan också vara intressant för framtida forskning inom områden som behandlar just sambanden mellan syn och hörsel. För hoppningsvis kan resultatet fungera som stöd och ge en inblick i hur den audiella perceptionen av rummet påverkas av den visuella referensen.

3.1 Metodbeskrivning

För att besvara frågeställningen användes en videinspelning av en spelsekvens från spelet Arma 3 (Bohemia Interactive, 2013), vilket har blivit godkänt av Bohemia Interactive. Ljudläggningen gjordes om och olika gestaltning av rumsklang användes, total tre ljudläggningar på samma sekvens skapades. Ljudläggningen utgick endast från effekt- och zonljud enligt IEZA modellen (Van Tol & Huibers, 2008). Rumsklangsgestaltningarna baserades på fördröjningsnät, konvolution och kombinationer av de båda. Till undersökningen fanns det alltså tre videoklipp med samma bild och ljud men med olika gestaltning av rumsklang. Valet att använda spelet Arma 3 (Bohemia Interactive, 2013) grundade sig i att spelet gestaltas i en realistisk stil, spelas i förstapersonsvy och är lätt att anpassa till olika situationer tack vare sina möjligheter till modifikation. Det var särskilt viktigt att sekvensen skedde i förstapersonsvy då undersökningen utgick från att lyssnaren hör från samma position som kameran befinner sig i, att använda annan vy skulle ändrat förutsättningarna för lyssnandet och den upplevda perceptuella realismen. Konvolutionsreverben baserades på impulsresponser från rum som var så lika de rum som gestaltas i videosekvenserna det var möjligt. Sekvenserna skedde i en inomhusmiljö med en rörelse genom flera olika rum. Fördröjningsnätverksreverb användes främst till att skapa de tidiga reflektionerna i rummet när det användes i kombination med konvolutionsreverb. Problem som kunde uppstå med skapandet av artefakten var att impulsresponsen aldrig helt skulle kunna spelas in i ett rum som ser exakt likadant ut som i videoklipppet. Det var därför extra viktigt att hitta ett rum med så lika dimensioner som möjligt. Detta var ett problem som också finns i skapandet av spel, om inte rummet i spelet är modellerat efter ett faktiskt rum.

3.1.1 Undersökningsmetod

Till undersökningen användes en kvalitativ metod för datainsamling. Data samlades in genom semistrukturerade intervjuer där informanterna fick titta och lyssna på ett av de tre videoklippen för att sedan svara på frågor angående deras upplevelse. Informanterna fick endast lyssna på ett av de tre klippen då de skulle göra en bedömning om ljudet i samma klipp. Risken som kunde uppstå om informanterna fick se och lyssna på alla tre för att sedan redogöra för sina upplevelser är att informanterna medvetet eller omedvetet gör jämförelser mellan dessa själva. Det kunde också uppstå problem med att informanterna inte kommer ihåg eller blandar ihop sina erfarenheter från klippen. Det var därför lämpligt att informanterna bara fick se och lyssna på ett av klippen. På det sättet kunde man enklare jämföra svaren från informanterna för att se skillnader i upplevelsen. Informanterna lyssnade genom ett par slutna hörlurar, detta för att minimera risken för utomstående ljud som kunde störa lyssnandet. Eftersom informanterna skulle använda sig av hörlurar och då videoklippen innehöll skottljud från ett vapen som är högre i volym än andra ljud var det noga att nämna att informanterna inte skulle öka volymen allt för mycket om det tyckte att det var lite för lågt. Detta för att undvika att informanterna skulle skada sin hörsel på något sätt. Här uppstod dock problem då det kunde vara svårt att avgöra vilken volym informanterna lyssnade på, vilket skulle kunna påverka deras upplevelse.

Informanterna som användes fick inte besitta en allt för stor kunskap om ljuddesign eller produktion för att undvika överanalys av enskilda ljuds gestaltning i artefakterna. Informanterna skulle också ha någon form av erfarenhet av spel eller virtuella världar. Eftersom undersökningen behandlar detta område var det lämpligt att informanterna hade någon form av erfarenhet av detta. På det sätt undveks oklarheter och oförståelser som kunnat uppstå, till exempel att en informant skulle uppmärksamma saker i videon som en informant med spelarefarenhet inte skulle. Saker som är rent speltekniska så som spelets gränssnitt. Något urval eller restriktion av ålder eller kön gjordes inte då det ansågs att inga av dessa faktorer spelade någon större roll för undersökningen. Antalet informanter som användes var totalt nio, tre per videoklipp. Vilket gav tillräckligt god generaliserbarhet och precision. De frågor som informanterna svarade på började med hur de generellt upplevde klippet de sett och lyssnat på för att sedan gå mer specifikt in på ljudet. Frågorna lyder:

- Beskriv kort vad som skedde i klippet.
- Var det något i klippet som du tyckte stack ut?
 - Om "Ja", vad?
- Tycker du att ljuden lät som du hade förväntat dig om du hade hört dem i verkligheten?
 - Om "Nej", vad vart det som inte stämde?
- Matchar ljudets representation rummets storlek?
 - Om "Nej", på vilket sätt?

Eftersom frågeställningen behandlar en jämförelse av upplevelsen av de olika gestaltningarna av rumsklang var det mer lämpligt med denna form av semistrukturerad intervju, då det handlar om att bli informerad av informanten snarare än ren datainsamling (Østbye, Knapsko, Helland & Larsen, 2008, s. 101-102). Intervjun utgår från frågorna som presenterats ovan men det tillkomma följdfrågor som formades efter situationen. Anledningen till att rumsklngen inte nämns i frågorna är för att undvika att informanten leddes till ett förväntat svar. Om informanten inte reflekterade över rumsklngen eller upplevde att det var något som inte passade så är det i sig relevant för frågeställningen.

Risken med att använda denna form av semistrukturerad intervju var att informanternas svar kunde bli svårtolkade och att forskarens eget syfte kunde påverka. Detta kan vara problematiskt för undersökningens reliabilitet. Vad som också var viktigt var att påpeka för informanten att det inte fanns något rätt eller fel svar, detta för att undvika att informanten ger svar som den tror förväntas och på det sättet påverka reliabiliteten. Det var också viktigt att informera informanterna om att de fick avbryta när som helst under sitt deltagande. Detta var extra viktigt då det finns en sekvens i videoklippet där vapen avfyras, vilket eventuellt kunde vara obehaglig för någon av informanterna. I skapandet av videoklippet görs det också klart att vapen kommer att avfyras för att förbereda informanten på detta. Informanterna informerades också om att deras deltagande kunde vara helt anonymt om de önskade detta. Det finns många som inte vill ha sitt namn eller identitet lagrat utanför deras kontroll. Alternativet till anonymitet skulle också göra deltagandet mer avslappnat och för att undvika att informanterna skulle känna att deras namn skulle användas i oklara syften, speciellt då syftet med undersökningen inte avslöjades för informanterna innan deltagandet. Syftet med undersökningen avslöjades inte för informanterna då detta skulle kunna påverka resultatet. Informanterna skulle då tänka på rumsklanen automatiskt vilket ville undvikas.

4 Genomförande

4.1 Artefaktbeskrivning

Artefakten består av ett videoklipp inspelat från spelet Arma 3 (Bohemia Interactive, 2013). Det ska klargöras att tillstånd att använda denna videoinspelning för icke kommersiellt syfte har givits av Bohemia Interactives publishing director, Ota Vrtátko, i en mailkonversation. Tre versioner av videoklippets skapades där varje version har samma ljudläggning men använder olika tekniker för att återskapa rumsklngen till ljudläggningen. Konvolution, fördröjningsnät och en kombination av dessa.

Videoklipppet spelades in i förstapersonsvy från spelarkarakters perspektiv. I videoklipppet rör sig karaktären upp till ett hus, öppnar dörren och rör sig in i huset. Inne i huset rör sig karaktären fram till mitten av rummet och stannar upp. Här kommer ytterligare en karaktär ut från ett anslutande rum, genom en dörr, och går fram till spelarkaraktären och säger att den ska röra sig upp till husets andra våning och skjuta på en bil som står parkerad utanför. Spelarkaraktären rör sig därefter upp för en trappa som finns i rummet och ställer sig vid ett fönster, tar fram ett vapen och avfyra vapnet mot bilen som står utanför. Videoklipppet har en tydlig progression med utrymme för effektljud som kan få resonera fritt i rummet och skapa rumsklang. Valet av rum gav också bra förutsättningar för att hitta ett liknande rum att spela in impulsrespons i. Ljudläggningen skedde i programmet Pro Tools 10 (Avid Technology, 2014). Pro Tools 10 (Avid Technology, 2010) valdes då det har goda förutsättningar för inspelning av ljud, ljudläggning av film och då det var detta program som fanns tillgängligt. Ljudläggningen mixades i stereo med syfte att spelas upp ur hörlurar.

4.2 Arbetsprocessen

Arbetsprocessen skedde i flera olika steg och kan delas in i tre huvudsteg: Inspelning, ljudläggning och mixning. Arbetssättet har utgått från de förutsättningar som finns i produktionen av spel och film. Allt arbete har alltså förhållit sig till de begränsningar och designval som skulle göras i en riktig produktion. Det kan handla om ljud som används inte är inspelningar av den exakta ljudkällan som syns i bild. Det kan också innebära att impulsresponsen inte spelades in i ett rum som var exakt likt det som syns i bild. Processen har i sin helhet följt denna progression med få undantag där ljud också spelats in, antingen på nytt eller som tillägg, under ljudläggningen. Detta är vanligt inom ljudläggning då det är här man kan veta om ljuden som spelats in är lämpliga eller om det behöver rättas till. Valet att dela upp processen på detta sätt grundade sig i att det skulle vara lättare att analysera och beskriva. Det är också en vanlig progression för ljudarbete som detta. Efter arbetsprocessen skedde också en mindre pilotstudie för att ytterligare analysera arbetet och för att synliggöra eventuella brister i produktionen.

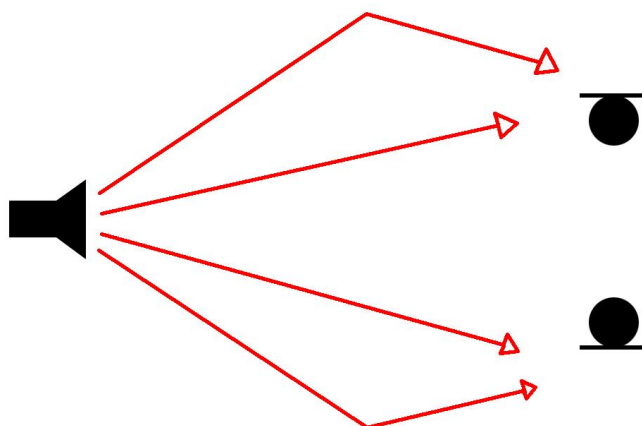
4.2.1 Inspelning

Arbetet började med inspelning av videoklipppet. Det började här då det var videoklipppet som bestämde vilka ljud och hur impulsresponsen skulle spelas in. Inspelningen skedde med programmet Fraps (Beepa Pty Ltd, 2014), ett program som spelar in vad som sker på skärmen. Först spelades det in ett scenario där spelaren, tittaren, rör sig genom ett hus och går upp till andra våning på samma hus. Denna inspelning valdes bort då det saknades ljudkällor som kunde resonera i rummet. Det spelades då in ytterligare ett scenario där

spelaren rör sig i rummet och möter ytterligare en spelarkaraktär som talar till spelaren. Denna inspelning valdes då den hade tillräckligt med ljudkällor för att kunna få en tydlig resonans i rummet.

När videoklippet var bestämt kunde det ljud som skulle användas identifieras. Ett rum för impulsresponsen kunde också identifieras. Alla ljud, förutom vapnet, spelades in specifikt för ljudläggningen av videoklippet. Inspelningen av vapnet hämtades från en tidigare inspelning. Nästa steg var att spela in den impulsrespons som skulle användas. För att hitta ett lämpligt rum mättes rummet i videoklippet genom att hitta ett vanlig mått på dörrar och använda en av dörrarna i spelet för att räkna rummets volym. Ett rum som stämde överens med rummets volym och tänkt ytmaterial men hade vissa skillnader i geometri hittades och fick bli det rum som användes. Rummet hade samma takhöjd, djup och material på väggar och golv. Rummet hade också samma totala volym. Att hitta ett rum som exakt stämde överens med det rum som avbildas i videoklippet kunde inte frambringas. Rummet som användes fick då baseras på det data som samlats in om rummet i videoklippet. Detta är ett problem som kan anses tillhöra de begränsningar och designval som görs i denna form av produktion, eftersom rummet är fiktivt och alltså inte finns i verkligheten.

För inspelning av impulsresponsen finns det ett antal olika metoder och uppsättningar av uppspelning- och inspelningsutrustning som går att använda. Metoderna innebär olika uppsättningar utrustning och ger varianterna full mono, mono till stereo, full stereo och olika varianter av surround upptagning. Valet föll på en mono till stereo variant, där man spelar upp impulsen ur en högtalare och spelar in den i stereo med två mikrofoner. Denna variant valdes då det är en enklare uppsättning att arbeta med och då ljudläggningen inte innehåller några ljudkällor som behöver panoreras kraftigt är något håll. Den optimala metoden är full stereo då den ger möjligheten att ge en exaktare representation av rummet när man vill kunna panorera kraftigt (Helbing, 2013). Mikrofonerna sattes upp på en ställning och riktades i motsatta riktningar för att återskapa en relation mellan två öron, konsthuvudstereo. På engelska kallas detta "binaural" och innebär att man simulerar förhållandet mellan öronen på ett huvud, ofta med mikrofoner i en modell av ett huvud (Pike, 2013). Uppsättningen kan ses i illustrationen som visas i figur 1. Mikrofonerna som användes var ett par Røde NT1-A.



Figur 1 Bilden ovan visar hur mono till stereo uppsättningen som användes ser ut. En högtalare spelar upp impulsen medan två mikrofoner fångar upp den. Bilden är egenskapad för rapporten.

4.2.2 Ljudläggning och mixning

Ljudläggningen handlade om att klippa, rensa och placera ljuden som spelats in och val av reverb. Ljuden delades upp i det två kategorierna effekt- och zonljud enligt IEZA-modellen. Zonljuden byggdes upp av två olika ljud, ett för utsidan av huset och en för insidan. Effektljuden består av fotsteg, dörrar, dialogen och vapnet. Ljuden valdes för att skapa en så trovärdig miljö som möjligt. Alla ljud förutom vapnet kunde spelas in i ett behandlat rum för att uppnå minsta möjliga rumsklang i inspelningarna. Eftersom vapnet måste spelas in på en säker plats gick det inte ställa höga krav på akustiken. Följde blev att inspelningen fick en mycket lång och hörbar efterklang av området. I ljudläggningen löstes detta genom att noga klippa ut den transient som vapenljudet består av. På detta sätt kunde transienten skickas till reverbet och efterklngen från inspelningen fick agera efterklngen utanför byggnaden.

Det var först efter att ljudläggningen var klar som reverben kunde läggas på. Det reverb som valdes var Waves IR-L och Waves TruVerb. Det finns en stor mängd olika reverb baserade på de båda teknikerna. Vilket man använder grundar sig ofta i personligpreferens eller budget. Waves IR-L valdes till det konvolutionsreverb som användes. Det valdes då det har en enkel process för behandlingen av impulsresponsen och ansågs vara bekvämast att arbeta i. Waves TrueVerb valdes till det fördröjningsnätsreverb som användes. Det valdes då det har bra kontrollmöjligheter för att ställa in rummets volym. Det har även separata kontroller för direktljud, tidiga reflektioner och sena reflektioner. Detta gjorde att man mer precist kunde kontrollera mängden tidiga reflektioner i mixningsfasen. Det skapades en huvudmix som alla de tre varianterna skulle utgå från. Varje ljudlager mixades för volym och frekvens. Alla ljud som skulle gå genom reverben fick en "send" till varje reverb (en "send" är en utgång som gör att man kan skicka en signal till en extern källa innan signalen går ut ur huvudutgången). Totalt skapades det fyra AUX-spår med ett reverb på vardera. Konvolutionsreverbet mixades först och det var inte mer än att lägga på reverbet och sedan styra hur mycket signal som skulle skickas från varje spår till reverbet för att påverka hur mycket reverbet skulle ha effekt på ljudet. Ljud som är närmare lyssnaren skickar mindre signal till reverbet och tvärt om. Inställningarna som används kan ses i figur 2 nedan.



Figur 2 Bilden ovan visar de inställningar som användes för konvolutionsreverbet. Skärmdump av Waves IR-L.

Fördröjningsnättsreverbet mixades sedan och tanken var den samma, att lägga reverbet och skicka så mycket signal som behövdes. Men eftersom TrueVerb inte fungerade på samma sätt som IR-L (IR-L utgår från den inspelade impulsen och kan därför behöva en annan styrka på signalen) var alla spår tvungna att skicka lika mycket signal. Här blev det då problem att styra avståndet i rummet. Eftersom TrueVerb har en inställning för avståndet mellan ljudkälla och lyssnare valdes det att göra ytterligare ett spår med ett TrueVerb. Den fick samma inställningar med skillnad i inställningen "Distance", som ställer in avståndet mellan ljudkälla och lyssnare. Det ljud som inte skapades av spelkaraktären fick gå genom detta reverb. Inställningarna för fördröjningsnättsreverben kan ses i figur 3 nedan. En sista varianten är kombinationen av konvolution och fördröjningsnät. I denna variant användes samma inställningar som användes för konvolutionsvarianten med tillägget av ett TrueVerb som var inställt på att bara skicka tidiga reflektioner.



Figur 3 Bilden ovan visar inställningarna som användes för fördröjningsnättsreverben. Skärmdump av Waves TrueVerb.



Figur 4 Bilden ovan visar de inställningar som användes för kombinationen av konvolutionsreverbet och fördröjningsnättsreverbet. Skärmdump av Waves IR-L och TrueVerb.

Videoklippen delades upp i artefakt 1 som använde konvolution, artefakt 2 som använder fördröjningsnät och artefakt 3 som är kombinationen. Eftersom undersökningen ska undersöka vilken gestaltning av de olika teknikerna som ger högst grad perceptuell realism mixades varje variant för optimal ljudbild oberoende av varandra. Detta är ett krav för att behålla god validitet då resultatet inte ska påverkas av en dålig mix eller låg kvalitet på inspelade ljud. Länkar till artefakterna står att finna i appendix D.

4.3 Pilotstudie

Efter slutförandet av artefakten utfördes en pilotstudie. Syftet med pilotstudien var främst att se hur artefakterna fungerade, om det fanns mixningstekniska problem som kunde gör att artefakterna inte gick att använda. Den utfördes även för att se om metoden var en lämplig metod för att undersöka problemformuleringen. Pilotstudien utfördes på en mindre skala än huvudundersökningen var tänk. Totalt deltog tre personer i pilotstudien, en person per videoklipp. Studien utfördes över Skype (Skype Technologies S.A, 2013). Informanterna instruerades att använda hörlurar och titta på videoklippen på fullskärmsläge. Efter att informanten tittat på klippet ställdes frågorna från det tänka formuläret. Informanterna fick innan undersökningen veta att deras svar kunde vara anonyma och att det inte finns något rätt eller fel svar på frågorna. Intervjuerna kan läsas i appendix A, B och C.

I den första intervjun, appendix A, hade informanten titta och lyssnat på klippet med konvolutionsreverbet. På frågan om något stack ut kom det fram att röstskådespeleriet kändes underligt och att informanten uppfattade rumsklngen av vapnet som plåt från när kulan träffar bilen. På de följande frågorna tyckte informanten att ljuden både lät som den förväntade sig om den hört dem i verkligheten och att ljudens representation matchar rummet storlekt. Här kom det inte fram någon ytterligare information som var relevant för eventuell justering av artefakten. Att informanten uppfattade rumsklngen när vapnet sköt som att kulan träffar bilen var intressant. Detta kan härledas till att informanten inte har erfarenhet av hur ett vapen som avfyras i den miljön låter. Detta är inget som ändrades i artefakten då ljudet är ett resultat av reverbet och svaret ger en fingervisning i hur en lyssnare kan uppfatta detta.

I den andra intervjun, appendix B, fick informanten titta och lyssna på klippet med fördröjningsreverbet. I denna intervju svarade informanten att fotstegen var för höga i volym, att de inte passade i förhållande till de andra ljuden och att rösten lät som den var i ett större rum än vad som förväntats. Utöver detta var det inget som var relevant för justering av artefakten eller frågorna. I den tredje intervjun, appendix C, fick informanten titta och lyssna på klippet med kombinationen av konvolution och fördröjningsnät. I denna intervju svarade informanten att rösten lät som den var i ett mindre rum, att volymen på fotstegen var för höga och här nämndes också rumsklngen av vapnet som plåten från bilen som skjuts på, informanten diskuterade dock att det kunde vara rumsklngen. Informanten tyckte in att ljudets representation matchade rummets storlekt och menade att rummet lät för litet, på fotsteg och röst.

Resultatet av pilotstudien visade att frågorna fungerar för syftet, att fotstegen behövde mixas om alternativt bytas ut helt och att man uppfattade vapnets rumsklang som plåten från bilen då kulan träffar den. I de fall där rumsklngen nämndes, som i intervju två och tre, behövs viss justering göras i volym och mängd.

5 Utvärdering och analys

Undersökningen som genomförts innefattade nio informanter som tittade på videoklippen och svarade på de fyra frågorna. Intervjuerna skedde semi-strukturerat via Skype eller facebook's chat, med tal eller text. De nio informanterna delades upp mellan de tre videoklippen så det var tre informanter per videoklipp. Informanterna informerades om att använda hörlurar och att titta på videoklippen i fullskärm, vilket alla informanter gjorde. Efter informanten tittat på videoklippen fick informanten besvara fyra frågor som behandlade deras upplevelse av videoklippen. I den första frågan tillfrågades informanten att kort beskriva videoklippen. Denna fråga var en kontrollfråga och hade till syfte att kontrollera om informanten titta på videoklippen korrekt. De resterande frågorna behandlar informantens upplevelse av videoklippen. Dessa frågor hade till syfte att se om informanten på något sätt upplevde att rumsklangen var korrekt eller ej. Till varje fråga fanns det utrymme för följdfrågor. Denna metod var lämplig då det var lättare för informanten att beskriva sin upplevelse och då möjligheten till följdfrågor gjorde att det var lättare att utröna oklarheter i informantens beskrivningar. Den var också lämplig då den gav möjlighet att gå djupare i specifika uppfattningar. Svagheten med metoden är att stor vikt läggs på kvalitén av transkriberingarna av intervjuerna. I de fall där intervjuen skett med text var detta inte ett problem.

Den intressanta informationen från varje intervju var huruvida informanten upplevde att rumsklangen inte stämde med bilden som visades. Detta skulle då påvisa huruvida olika gestaltningar av rumsklang påverkar den upplevda perceptuella realismen. För detta krävdes en analys av de transkriberingar som gjorts. För att lättare analysera kondenserades transkriberingarna till sin kärna. Något Graneheim och Lundman (2004) kallar att kondensera meningen till meningsskoder som sammanfattar innehållet i ett långt svar. Vad man då får är kortare, mer hanterbara meningar. Dessa var enklare att använda vid analysen. För att hålla hög reliabilitet med denna metod krävdes det också hög kvalitet på transkriberingarna. Att man beaktar tonfall, pauser och om informanten är osäker eller säker i sitt svar. Detta för att kunna analysera vikten i det informanten säger. De kondenserade intervjuerna kan läsas i appendix E, F och G. Nedan presenteras och analyseras resultaten av intervjuerna. I analysen ställs först informanterna som titta på samma videoklipp mot varandra för att sedan jämföra resultaten från varje videoklipp.

5.1 Artefakt 1 - Konvolution

I intervjuerna för det första videoklippen, som baserades på konvolution, var det ingen informant som på första frågan, om den tyckte något stack ut, nämnde rumsklangen. Informant ett uppmärksammar dock att huset var tomt. Intressant här är att informant tre nämnde att kulan som träffade bilen lät burkigt, något som också nämndes i pilotstudien. Detta är intressant då det inte var lagt något sådant ljud. Det som informanten syftar på är en effekt av konvolutionsreverbet och den höga transitenen från vapenljudet. På frågan om de tyckte att ljuden lät som de förväntat sig om de hört dem i verkligheten var det inte heller någon av informanterna som nämnde rumsklangen. Här nämnde informant tre kulan som träffade bilen igen. Den sista frågan behandlade specifik rumsklangen och om informanterna tyckte att ljudens representation matchade rummet storlekt. Här var det bara en av de tre informanterna som ansåg att rumsklangen stämde överens. De två andra informanterna menade båda att det kunde varit mer rumsklang. Informant 1 nämnde geväret och menade

att det skulle varit mer rumsklang. Informant tre menade att det lät bra men att fotstegen och rösten borde resonerat mer i rummet. Antalet som nämner rumsklngen eller ej kan ses i tabell 1.

Det var alltså ingen informant som direkt uppmärksammade rumsklngen fören de direkt blev tillfrågade. Informant ett hänvisade till att rummen var tomma och därför förväntade sig mer rumsklang. Informanten nämnde också själv att den inte hade någon erfarenhet av att höra vapen i verkligheten.

5.2 Artefakt 2 – Fördröjningsnät

I intervjuerna för videoklippet baserat på fördröjningsnät var det en av de tre informanterna som hänvisade till rumsklngen på fråga om de tyckte något stack ut. Informant fyra nämnde avsaknaden av ambiens (ljud som faller inom zonljud) och felpositionering av rösten. Informant fem nämnde att skotten från geväret lät som de befann sig utomhus. Den information som dessa två informanter gav kan relateras till rumsklngen. Både felpositioneringen av rösten och att skotten lät som de är utomhus kan vara effekter av för lite rumsklang. Dock nämnde ingen av informanterna rumsklngen specifikt som något som stack ut utan valde att nämna specifika ljud på frågan. Informant sex nämnde att skotten stack ut tack vare volymen i relation till de övriga ljuden. På frågan om informanterna tyckte att ljuden lät som de hade förväntat sig om de hade hört dem i verkligheten var det bara en som direkt nämnde rumsklngen. Informant fyra nämnde att det var för lite reverb och att det lät tort. Här nämnde informant fem att skotten lät som de sköts utomhus igen. Informant sex ansåg att det lät bra. Informant fyra utvecklade på det den ansåg stack ut och nämnde att anledningen var avsaknaden av rumsklang och att rösten var felpositionerad tack vare detta. På den sista frågan om informanterna tyckte att ljudets representation matchar rummets storlekt svarade två av de tre informanterna att det inte tyckte det. Informant fyra ansåg att det var för lite reverb, alltså för lite rumsklang, och informant fem hänvisar igen till att skotten låter som de avfyras utomhus. Informant sex menade att det lät bra.

Till videoklippet som baserats på fördröjningsnät var det alltså två av tre informanter som direkt nämnde saker som är produkter av rumsklngen, de nämnde inte specifikt att rumsklngen skulle vara anledningen fören andra frågan då informant ett nämner avsaknaden av reverb. Informant fem nämnde att skotten lät som de avfyrades utomhus i varje fråga, en effekt som är ett resultat av reverbet. Ett antagande skulle kunna göras om att informant fem redan i början av intervjun menade att det är för lite rumsklang.

5.3 Artefakt 3 – Kombination

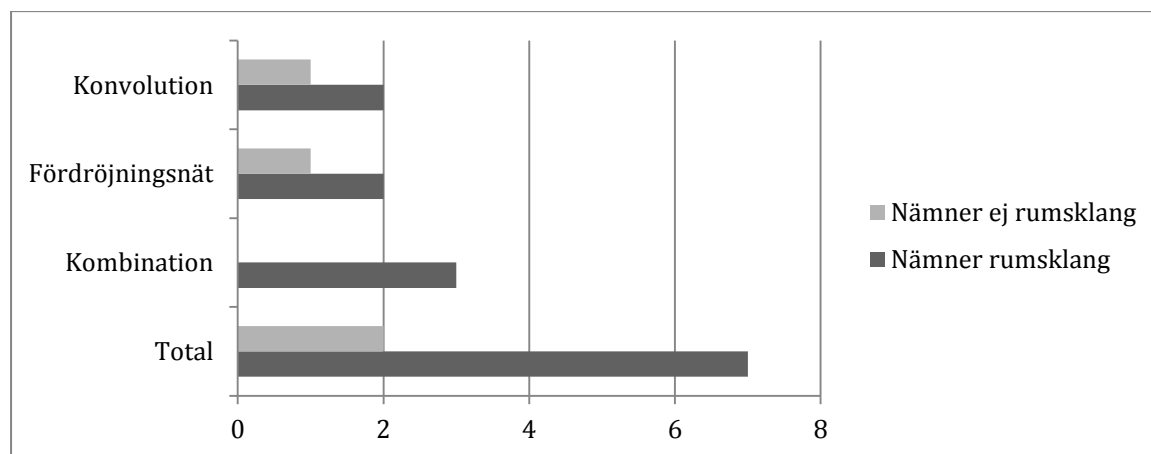
Svaren från informanterna till videoklippet baserat på kombinationen av konvolution och fördröjningsnät var mer blandade. På den första frågan om informanterna tyckte att något stack ut svarade informant sju direkt att det var för lite eko och hänvisade till att huset var tomt och att rösten inte var placerad bra i rummet. Informant åtta svarade att skottet från geväret lät burkigt och plåtigt, likt informant tre svarade på videoklippet baserat på konvolution. Dock gjordes ingen referens till rumsklngen eller att det skulle varit kulan som träffar bilen. Informant nio tyckte att ljudeffekterna lät bra men att ljuset såg ofärdigt ut. På frågan om informanterna tyckte att ljuden lät som de förväntat sig om de hade hört dem i verkligheten nämnde informant sju att ljuden lät bra men att reverbet inte stämde på rösten.

Vilket kan vara sammankopplat med att informanten tyckte att det var för lite rumsklang. Informant åtta tog upp att geväret lät burkigt, likt första frågan. Här poängterade informanten att den hade dålig erfarenhet av hur vapen låter, vilket kan spela roll i hur informanten uppfattade ljudet. Informant nio ansåg på denna fråga att det borde varit mer reverb, till skillnad från alla andra informanter. På frågan om informanterna tyckte att ljudets representation matchar med rummet storlekt förväntade sig informant nio att rummet skulle vara större och att rumsklangen därför var för liten. Informant sju ansåg att rumsklangen var för kontrollerad, informanten menade att rumsklangen skulle vara mer okontrollerat med tanke på att rummet var tomt. Informant åtta tyckte att rummet lät mindre än de såg ut.

Informanternas svar till videoklipplet som baserades på kombinationen var alltså mer blandade. Utöver att informant nio ansåg att rumsklangen var för stort, vilket informanten var ensam om i hela undersökningen, så svarade de övriga två informanterna att rumsklangen var litet eller okontrollerat.

5.4 Slutsatser

När man tittar på den information som kommit fram från intervjuerna och sammanställningarna av informanternas svar kan man se att det inte finns ett klart svar på vilken variant eller gestaltning som gett högre grad perceptuell realism. Svaren har varit olika på varje variant. Det var få informanter som direkt uppmärksammade rumsklangen. Men några som hade synpunkter på ljud där rumsklangen kan varit orsaken. Fyra av de nio informanterna menade att rumsklangen kunde varit större. I videon med konvolutionsreverb svarade två av de tre att rumsklangen kunde varit större, medan en tyckte att rumsklangen kunde varit mindre. För fördröjningsnätet var det också två av de tre informanterna som hade synpunkter på rumsklangen och på videon med kombinationen var det en informant som menade att rumsklangen var för liten. En av de tre menade att rumsklangen var för stor. Det totala antalet informanter som nämnde rumsklangen direkt eller indirekt kan ses i diagrammet i figur 5 nedan.



Figur 5 Diagrammet visar antalet som i intervjuerna nämner rumsklangen.

Generellt anmärkte informanterna på enskilda ljud innan de blev direkt tillfrågade om de tyckte att rumsklangen stämde. Sista frågan som informanterna fick svara på rörde rummets storlekt. Utöver storleken var det bara två informanter som nämnde rumsklangens karaktär. Det var då informant tre och åtta nämnde burkigheten och plåtljudet. Båda dessa

informanter tittade och lyssnade på videor med konvolutionsreverb. Informant åtta tittade på kombinationen. Slutsatsen man kan dra av detta är att ingen av de olika teknikerna ger en gestaltning som direkt ger högre perceptuell realism åt lyssnaren, även om konvolutionsreverb vid en djupare analys stämmer bättre överens med verkligheten. Så länge man återskapar en storlek på rummet som stämmer är det få som kommer uppmärksamma det. Vad som dock är intressant är att transienten av vapnet resonerar mycket olika i konvolutionsreverb och fördröjningsnätsreverb, därav burkigheten. Här spelar erfarenheten hos lyssnaren in, om den hört ett vapen innan eller ej och om den kan föreställa sig hur den skulle låta i ett mindre rum.

6 Avslutande diskussion

6.1 Sammanfattning

Arbetet som presenteras i denna rapport har haft till syfte att undersöka om olika gestaltningar av rumsklang, på grund av olika tekniker, ger olika grad av perceptuell realism med en visuell referens. För att undersöka detta skapades tre videoklipp, inspelade från spelet Arma 3 (Bohemia Interactive, 2013), som använder sig av olika tekniker för att skapa rumsklang. De olika teknikerna var konvolution, fördröjningsnät och en kombination av dessa. Varje videoklipp hade samma ljudläggning men med skillnad i rumsklangen. Varje videoklipp mixades också för bästa möjliga ljudbild för att undvika validitetsproblem. Själva undersökningen skedde genom semi-strukturerade kvalitativa intervjuer med nio informanter. Informanterna fick bara se ett av videoklippen.

Resultatet av undersökningen har varit blandat. Slutsatsen var att ingen av de olika gestaltningarna direkt gav någon högre grad perceptuell realism än den andra. Så länge storleken på rumsklangen stämmer överens med den visuella referensen kommer det vara enskilda ljudeffekter som uppmärksammas först, om något uppmärksammas alls. Den mesta kritiken som kommit fram i frågorna har varit direkta mixningstekniska saker. Så som relationen mellan volymer i olika ljud och kvalitén på olika ljudeffekter. Det har varit både undersökningens styrka och svaghet. Samtidigt som ljudläggningen och mixningen gjorts så bra som möjligt så har den också baserats på de begränsningar som en vanlig produktion har. Frågorna var också generella över informanternas upplevelse vilket kan ha gjort att informanterna inte reflekterat över rumsklangen, även om det kan varit något de tänkt på, utan gjort att de uppmärksammade andra saker först. Det ligger alltså stora mängder subjektivitet bakom resultatet och informanternas svar. Informanternas egen erfarenhet av ljuden och miljöerna i fråga kan också spelat roll i de svar som givits.

6.2 Diskussion

I den forskning som presenterats i bakgrunden, Frissen, Katz, Guastavion (2009) och Larsson, Västfjäll och Kleiner (2008), har man visat att människor kan skilja på storleken mellan två olika rumsklang och att rumsklangen ökar den upplevda realismen i en virtuell kontext. Dessa undersökningar har som bekant skett utan en visuell referens. Denna undersökning behandlar samma område men med en visuell referens. Resultatet av denna undersökning visar att genom att lägga till en visuell referens blir resultatet inte lika klart som det varit i de två presenterade undersökningarna. När informanten får ett visuellt media att förhålla sig till utöver ljudet börjar man göra jämförelser mellan det man ser och det man hör. I slutändan spelar den teknik man använder för att återskapa rumsklang, i ett spel eller film, mindre roll utan det är slutresultatet som räknas.

I en undersökning som undersöker personers egna uppfattningar om deras upplevelse av ljud finns det många faktorer som kan påverka resultatet. Eftersom svaren är rent subjektiva spelar informanternas tidigare erfarenheter och bakgrund roll i sammanhanget. Det kan vara erfarenheter som om de har erfarenhet av att höra ljuden i sammanhanget, till exempel vapnet, om de har erfarenhet av ljudbehandling eller om de spelar eller tittar mycket på film. Även informanternas personliga preferenser spelar roll, om de tycker att enskilda ljud låter bra eller inte. Detta är saker som kan påverka resultatet av undersökningen. Det har därför varit viktigt att få informanter med olika bakgrunder och som inte har allt för stor erfarenhet

av ljuddesign eller ljudläggning. I detta fall har undersökningen lyckats bra med att inte använda en allt för homogen grupp informanter. Informanterna har olika bakgrund till ljud och kommer från olika delar av världen.

6.3 Framtida arbete

Arbetet som utförts här har utgått från de begränsningar som finns i en produktion av spel eller film och den begränsade utrustningen som fanns tillhanda. Det hade varit intressant att mer exakt undersöka samma frågeställning i en mer kontrollerad miljö. Det hade också varit intressant att använda sig av en faktisk spelsession där informanten själv fått styra karaktären i en virtuell miljö. I den virtuella miljön hade man då kunna modelera ett rum exakt efter ett verkligt rum där man spelat in en impulsrespons och på så sätt fått en mer exakt representation av rumsklngen i den virtuella miljön. På det sättet hade man kunnat få en bättre uppfattning om hur olika gestaltningar av rumsklngen påverkar den perceptuella realismen.

I en förlängning eller om det hade funnits mer tid hade det varit intressant att använda en större grupp informanter från flera olika bakgrunder. Det hade också varit intressant att komplettera metoden med en kvantitativ metod, till exempel ett frågeformulär. På det sättet hade man kunnat täcka en större grupp människor och på så sätt fått en bättre grund att kunna generalisera från. I en förlängning skulle man även kunna tillverka flera olika miljöer för att också där få en större bredd i undersökning och för att kunna se om det görs någon skillnad i olika miljöer. Man skulle också kunna försöka innefatta de mest använda spelmotorerna för att använda den teknik som används i branschen just nu. Hade undersökningen skett på ett företag som har större resurser och tillverkar sin egen spelmotor hade man kunna implementera de olika teknikerna och även undersöka tekniker som ännu inte används. Så som hörselfömedlare och riktningsvarierande efterklang (Aural Proxies and Directionally-Varying Reverberation) eller förberäknad vågsimulering för realtids utbredning (Precomputed Wave Simulation for Real-Time Sound Propagation) (Menhorn, 2012).

Referenser

- Adams, E. (2010) *Fundamentals of Game Design* (2:a upplagan). Berkley: New Riders.
- Audiokinetic Inc (2014) *Wwise* (Version: 2013.2) [Datorprogram]. Audiokinetic Inc. Tillgänglig på Internet: <https://www.audiokinetic.com/>.
- Avid Technology (2014) *Pro Tools 10* (Version: 10.3.8) [Datorprogram]. Avid Technology. Tillgänglig på Internet: <http://www.avid.com/US/>.
- Beepa Pty Ltd (2014) *Fraps* (Version: 3.5.99) [Datorprogram]. Beepa Pty Ltd. Tillgänglig på Internet: <http://www.fraps.com/>.
- Bohemia Interactive (2013) *Arma 3* (Version: 1.10.114700) [Datorprogram]. Bohemia Interactive. Tillgänglig på Internet: <http://www.arma3.com/>.
- Chandak, A. (2012) *Reverb: The Science and the State-of-the-Art*. Tillgänglig på Internet: <http://designingsound.org/2012/12/reverb-the-science-and-the-state-of-the-art/> [Hämtad Februari 12, 2014].
- Chion, M. (1990/1994) *Audio-Vision: Sound on Screen*. New York: Colombia University Press. [Ursprungligtitel: *L'Audio-Vision: Son et image au cinéma*]
- Electronic Arts (2013) *Battlefield 4* (Version: 1.0) [Datorprogram]. Electronic Arts. Tillgänglig på Internet: <http://www.battlefield.com/battlefield-4>.
- Frissen, I., Brian, F.G. & Gustavino, C. (2010) Effect of Sound Source Stimuli on the Perception of Reverberation in Large Volumes. *Auditory display: 6th international symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009: revised papers.*, Köpenhamn, Danmark, Springer. s. 358–376.
- Graneheim, U.H., Lundman, D. (2004) Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*. 24, s. 105-112.
- Helbling, G. (2013) *Convolution for mono and stereo*. Tillgänglig på Internet: <https://www.soundeffects.ch/en/tutorials/create-impulse-responses/convolution-reverb-for-mono-and-stereo.php>. [Hämtad April 9, 2014].
- Huber, D. & Runstein, R. (2010) *Modern Recording Techniques*. Burlington: Focal Press.
- Hulusic, V. Harvey, C. Debattista, K. Tsingos, N. Walker, S. Howard, D. & Chalmers, A. (2012) Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction. *COMPUTER GRAPHICS forum*. 31(1) s. 102-131.
- Larsson, P., Västfjäll, D. & Kleiner, M. (2008) Effects of auditory information consistency and room acoustic cues on presence in virtual environments. *Acoustical Science and Technology*. 29(2), s. 191–194.
- Lombard, M., Ditton, T. (1997) At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*. 3(2). Tillgänglig på Internet:

- <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1997.tb00072.x/full> [Hämtad Februari 5, 2014].
- Nair, V. (2012) *Recording Impulse Responses*. Tillgänglig på Internet: <http://designingsound.org/2012/12/recording-impulse-responses/>. [Hämtad Februari 17, 2014].
- Menhorn, J. (2012) *Reverb: The Science And The State-of-the-Art*. Tillgänglig på Internet: <http://designingsound.org/2012/12/reverb-the-science-and-the-state-of-the-art/>. [Hämtad Maj 17, 2014].
- Pike, C. (2013) *Listen Up! Binaural Sound*. Tillgänglig på Internet: <http://www.bbc.co.uk/blogs/legacy/researchanddevelopment/2013/03/listen-up-binaural-sound.shtml>. [Hämtad April 10, 2014].
- Skype Technologies S.A (2014) *Skype* (Version: 6.3.59.105) [Datorprogram]. Skype Technologies S.A. Tillgänglig på Internet: <http://www.skype.com/>.
- Välimäki, V., Parker, J., Savioja, L., Smith, J. & Abel, J. (2012) Fifty Years of Artificial Reverberation. *IEEE Transactions on Audio, Speech, and Language Processing*. 20(5), s. 1421-1448.
- Van Tol, R. & Huiberts, S. (2008) *IEZA: A Framework For Game Audio*. Tillgänglig på Internet: http://www.gamasutra.com/view/feature/131915/ieza_a_framework_for_game_audio.php [Hämtad Februari 6, 2014].
- Østbye, H. Knapskog, K. Helland, K. & Larsen O.L. (2007/2008) *Metodbok för medievetenskap*. Malmö: Liber. [Ursprungligtitel: *Metodebok for medievitenskap*]

Appendix A - Pilotintervju 1

- Beskriv kost vad som skedde i klippet.
Jag såg ur en persons synvinkel som gick in i en byggnad. Där mötte han en person med ett gevär och så sa han till en att gå upp på andra våningen och skjuta på bilen som står parkerad. Så han gick upp på andra våningen, sköt ett par gånger och kollade sig om och så slutade videon.
- Var det något i klippet som du tycker stack ut?
Äh, det allra första jag tänkte på var la att röstskådespelet var lite off men. Och sen också tänkte jag medan han sköt på bilen att eh, det lät nästan som att jag vet inte. Men eh, metalljuden som kom från bilen där lät nästan som om de var precis framför honom. Liksom precis framför honom. Annars tycker jag att gevärsljudet var riktigt häftigt, väldigt bra gjort.
- Tycker du att ljudet lär som du hade förväntat dig om du hade hört dem i verkligheten?
Ah, ah faktiskt. Förutom själva ljuden när de träffade bilen då, de lär lite för nära på ett sätt.
- Matchar ljudets representation rummets storlekt?
Definitivt, faktiskt. Det anmärkte jag också när jag kollade på videon att lokaliseringen av ljudet var bra när den andra personen kom in i bilden. Det lät liksom som rummet var tomt vilket det var.

Appendix B - Pilotintervju 2

- Beskriv kort vad som skedde i klippet.
Man fick följa en spelare som gick in i en byggnad. Därefter dök en man upp och bad spelaren gå upp en trappa och skjuta på en bil från ett fönster.
- Var det något i klippet som du tyckte stack ut?
Ja. Spelarens fotsteg hördes väldigt mycket i ljudbilden. Mannens röst lät som om han befann sig i ett större rum än vad som representerades i bilden.
- Tycker du att ljuden lät som du hade förväntat dig om du hade hört dem i verkligheten?
Både ja och nej. Fotstegen lät väldigt mycket högre än vad jag hade förväntat mig. Vapnet lät dock som jag hade förväntat mig.
- Matchar ljudets representation rummets storlek?
Både ja och nej. Vapnet och rumsambiansen kändes bra representerat. Fotstegen kändes out of place.

Appendix C - Pilotintervju 3

- Beskriv kort vad som skedde i klippet.
Vi gick in i ett hus å träffa en snubbe som sa att vi skulle upp å panga på en bil.
- Var det något i klippet som du tyckte stack ut?
Ja. Rösten stack ut. Lät inspelad i ett litet rum. Bör varit mer eko alternativt lägre volym. Skottljuden låter feta. Men är plåtljuden från bilen man träffar eller av ekot i rummet? Om det är bara är eko i rummet så gillade jag det första 2 skotten mest.
- Tycker du att ljuden lät som du hade förväntat dig om du hade hört dem i verkligheten?
Ja, men volymen är lite väl hög på fotstegen tycker jag dock.
- Matchar ljudets representation rummets storlek?
Nej. Fotstegen låter som man är i ett klart mindre rum. Samma sak gäller rösten. Trappan låter bra och även skotten. När man tar fram vapnet låter även det rätt fett.

Appendix D - Artefakter

Artefakt 1 som använder sig av konvolution:

<https://www.youtube.com/watch?v=mwZcuotbl1Q>

Artefakt 2 som använder sig av fördröjningsnät:

https://www.youtube.com/watch?v=AxOQo_EBWbI

Artefakt 3 som använder sig av en kombination av konvolution och fördröjningsnät:

<https://www.youtube.com/watch?v=-hQhbThiDM>

Appendix E - Kondenserade intervjuer, artefakt 1

Frågor	Informant 1	Informant 2	Informant 3
Var det något i klippet som du tyckte stack ut?	Geväret hade hög volym. Huset var tomt.	Personen rör inte läpparna. Inget ljud för skotten på bilen. Att man tittar runt i slutet.	Personen talar drygt, ointresserad. När kulan träffar bilen lät burkigt.
Tycker du att ljuden lät som du hade förväntat dig om du hade hört dem i verkligheten?	Aldrig hört riktigt vapen. Hyffsat verkligt. Fotstegen på golvet lät bra om det var på hårt golv.	Ja.	Ungefär. Kulan som träffar bilen.
Matchar ljudets representation rummets storlek?	Matchade inte. Mer eko med tanke på rummets tomhet. Märks på geväret och talet.	Ja med tanke på rummets kalhet.	Lite mer rumsklang. Låter bra. Helheten. Mer på fotsteg och röst.

Appendix F - Kondenserade intervjuer, artefakt 2

Frågor	Informant 4	Informant 5	Informant 6
Var det något i klippet som du tyckte stack ut?	Lite ambiens. Övergången i fotstegen. Fel position på röst. Igen kula träffar bilen.	Fotstegen var lite tajta. Skotten lät utomhus.	Shooting drew attention. The sound peeked. Higher than other sounds.
Tycker du att ljuden lät som du hade förväntat dig om du hade hört dem i verkligheten?	Dörren snabbare än ljudet. Rösten fel positionerad. Skottljuden är högre i verkligheten. Lite reverb. Lät tort.	Fotstegen lät högt. Skotten utomhus.	Sounded real.
Matchar ljudets representation rummets storlek?	Tänkte på det. För lite reverb.	Nej. Skotten utomhus. Fotstegen för höga.	I would say so.

Appendix G - Kondenserade intervjuer, artefakt 3

Frågor	Informant 7	Informant 8	Informant 9
Var det något i klippet som du tyckte stack ut?	Lite eko. Huset var tomt. Rösten ej placerad.	Geväret, lät burkigt/plåtigt.	Sound effects were good. Lighting looks unfinished.
Tycker du att ljuden lät som du hade förväntat dig om du hade hört dem i verkligheten?	Ja. Reverber var lite off på rösten.	Geväret lät burkigt. Dålig erfarenhet av vapen.	The room was empty. More reverb.
Matchar ljudets representation rummets storlek?	Ja och nej. Det var för kontrollerat.	Rummet lät mindre.	No. Expected the room to be bigger.