

Musik ur tomma luften
– en musikleksak baserad på Kinect

EMMA FRID
och ANDREAS ALMQVIST



**KTH Datavetenskap
och kommunikation**

Musik ur tomma luften – en musikleksak baserad på Kinect

E M M A F R I D
o c h A N D R E A S A L M Q V I S T

DM229X, Examensarbete i medieteknik om 15 högskolepoäng
vid Programmet för medieteknik 300 högskolepoäng
Kungliga Tekniska Högskolan år 2012
Handledare på CSC var Kjetil Falkenberg Hansen
Examinator var Sten Ternström

URL: www.csc.kth.se/utbildning/kandidatexjobb/medieteknik/2012/almqvist_andreas_OCH_frid_emma_K12066.pdf

Kungliga tekniska högskolan
Skolan för datavetenskap och kommunikation

KTH CSC
100 44 Stockholm

URL: www.kth.se/csc

Abstract

This paper investigates how the Microsoft Kinect camera can be used in playful music applications. The aim of the study was to gain knowledge on how the Kinect can be used in a motion capture based musical toy. A focus group in combination with prototype testing of a simple beat machine controlled by a Kinect was used, in order to examine the limitations and possibilities of a Kinect interface. In addition, the commercial interest for a musical toy based on Kinect was investigated. The subjects in the study consisted of youths of age 21-28 years with variable musical interest. It was found that there was a considerable interest for a product based on the musical toy prototype among the subjects. However a majority felt that this system would not be well suited for professional music applications. The biggest problem identified during the prototype testing was the latency inherent within the system. This, and the fact that motion capture data retrieved from a Kinect is somewhat imprecise, makes the camera more suitable for entertainment applications than for instruments with high precision requirements. After receiving feedback from the prototype tests, a final improved prototype was developed. No further user tests were carried out for this prototype.

Sammanfattning

Denna artikel undersöker möjliga användningsområden för en Microsoft Kinect-kamera i lekbetonade musikapplikationer. En fokusgrupp samt prototyptester av en enkel musikleksak, vars effekter kontrollerades av en Kinect-kamera, användes för att undersöka vilka möjligheter och begränsningar som Kinect som gränssnitt medför. Syftet med studien var att undersöka hur en musikapplikation baserad på Kinect bör och kan utformas. Vidare undersöktes huruvida en sådan musikapplikation skulle kunna vara av kommersiellt intresse. Vissa slutsatser om en potentiell målgrupp kunde dras efter ett antal utförda användartester. Testanvändarna utgjordes av ungdomar i åldern 21-28 år, och ett intresse för en produkt liknande prototypen kunde påvisas hos dessa testanvändare. Däremot ansåg en majoritet att detta system inte skulle lämpa sig för professionellt musikanvändande. Den största brist som upptäcktes hos prototypen var fördröjningen i systemet. Fördröjningen samt det faktum att Kinect-kameran inte erbjuder så exakt positioneringsdata jämfört med andra motion capture-system medför att Kinect-kameran är bättre lämpad för att användas inom lekbaserade applikationer än inom avancerade musikinstrument och produkter avsedda för professionella ändamål. Med resultaten från prototyptesterna i åtanke utvecklades slutligen en förbättrad prototyp. Denna prototyp genomgick inga användartester.

Innehåll

1	Inledning	3
1.1	Bakgrund	3
1.2	Syfte och problemformulering	3
1.3	Frågeställningar	3
1.4	Avgränsningar	4
1.5	Målgrupp	4
2	Teori	5
2.1	Motion capture	5
2.2	Kopplingen mellan musik och kroppsrörelser	6
2.3	Att designa instrument	8
2.4	Mappning i motion capture-baserade system	11
2.5	Air Instruments	12
2.6	Att generera musik i 2D	14
2.7	Kinect-kameran	15
2.8	Kinect och musik - tidigare forskning & projekt	17
3	Metod	20
3.1	Angreppssätt	20
3.2	Förstudie	20
3.3	Fokusgrupp och brainstorming	20
3.4	Prototyputveckling	21
3.5	Prototyptest	21
3.6	Metodkritik	21
	3.6.1 Urval	22
	3.6.2 Validitet	22
	3.6.3 Reliabilitet	22
4	Resultat	24
4.1	Prototypen	24
4.2	Fokusgrupp och brainstorming	28
	4.2.1 Frågor som ställdes vid fokusgruppen	29
4.3	Prototyptester	31
	4.3.1 Frågor som ställdes vid prototyptesterna	32
4.4	Slutgiltig Prototyp	34

5	Diskussion	36
5.1	Analys av mappning	38
5.2	Slutsats	40
5.3	Framtida forskning och tillämpningsområden	41

1 Inledning

I detta kapitel presenteras bakgrunden till projektet, studiens syfte samt dess frågeställningar. Här redogörs även för projektets avgränsningar samt målgrupp.

1.1 Bakgrund

Alla ljud härstammar från någon typ av rörelse. Den musikaliska kommunikationskedjan kan med fördel betraktas utifrån ett rörelsebaserat perspektiv: människan kontrollerar musikinstrument med hjälp av kroppsrörelser och musikinstrument genererar i sin tur rörelse hos luftpartiklar. Dessa partikelrörelser förmedlas därefter till en lyssnares hörselsystem, varefter hörselmembranet sätts i rörelse. Det råder således ett intimt samband mellan begreppet musik och begreppet rörelse. Musik besitter en förvånansvärd förmåga att förmedla en känsla av rörelse och att uppmåna till dans eller gester. Betänk bara hur ljudet av musik får de flesta av oss att vilja stampa fötterna i golvet, dansa eller att nicka med huvudet i takt (Godøy & Leman 2009). Följande studie avser att undersöka hur denna koppling mellan rörelse och musik kan utnyttjas vid utvecklingen av en musikleksak som styrs med en Kinect-kamera.

1.2 Syfte och problemformulering

Syftet är att undersöka hur en Kinect-kamera kan användas i en *motion capture-baserad* musikapplikation. Med motion capture menas en process där ett objekts position och rörelser i ett rum detekteras, mäts och spelas in (Dyer et al. 1995). I vårt fall utgörs objektet i fråga av den mänskliga kroppen och rörelsen består av kroppsrörelser. Vidare syftar denna studie till att klargöra vilka möjligheter och begränsningar som Kinect-kameran införlivar i ett sådant rörelsebaserat system för ljudproduktion.

1.3 Frågeställningar

De frågeställningar vi kommer utgå ifrån är:

- Hur ser marknaden för motion capture-baserade musikapplikationer ut idag?
- Finns det ett intresse för musikapplikationer som styrs med kroppsrörelser?

- Hur bör en musikapplikation som styrs med kroppsrörelser utformas? Hur parar man ihop kroppsliga rörelser med ljudeffekter på ett sätt som känns naturligt för användaren?
- Hur skulle en musikapplikation baserad på motion capture-teknik från en Kinect-kamera kunna låta, fungera och styras? Vilka möjligheter och begränsningar finns?
- Kan denna musikapplikation fungera som ett instrument, eller är den bäst lämpad som leksak?

1.4 Avgränsningar

Utifrån studiens förutbestämda omfattning samt rådande resurstillgångar valde vi att definiera en rad begränsningar för detta arbete. Studien är därför främst baserad på resultat från test av en egenutvecklad prototyp. Denna prototyp genererar uteslutande ljud från förinspelade ljudklipp, och syntiserar inga nya ljud. Någon djupgående analys av den utvecklade prototypens gränssnitt kommer inte genomföras, även om användbarheten till viss mån kommer behandlas och diskuteras. Ingen jämförande studie mellan vanliga gränssnitt för modifikation av musik (så som exempelvis gränssnitt för MIDI-kontrollers) och rörelseigenkännande motion-capture system kommer att genomföras. Andra motion capture-system kommer enbart att behandlas ytligt i undersökningen. Således kommer enbart begränsade slutsatser om rörelsebaserade gränssnitt för ljud i allmänhet att kunna dras.

1.5 Målgrupp

Vi föreställer oss att det främst är unga personer med teknik- och musikintresse som kan komma att uppskatta vår produkt. Av praktiska skäl har vi valt att fokusera på KTH-studenter, vilka kan anses vara både teknikvana och relativt unga. Det skulle kunna finnas anledning att bredda målgruppen även till yngre personer, samt att närmare undersöka skillnaden mellan de som har mer eller mindre teknik- och musikintresse. Vi har emellertid valt att begränsa målgruppen till unga personer i åldern 20-30 år med teknikvana och varierande grad av musikalisk erfarenhet.

2 Teori

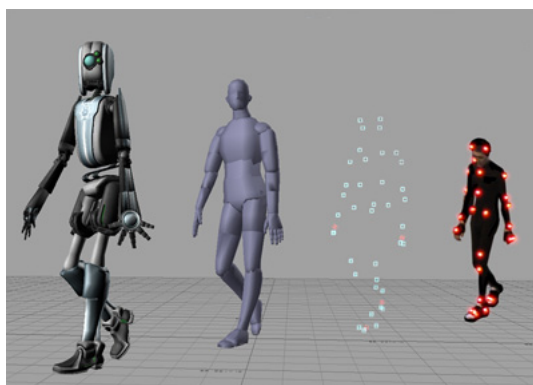
I detta kapitel presenterar vi information som på ett sätt eller annat är relevant för vårt arbete.

2.1 Motion capture

Motion capture är en teknik som går ut på att mäta ett objekts position eller orientering i ett rum (Dyer et al. 1995). Positioner eller rörelser i rummet detekteras, mäts och spelas in. Den inspelade informationen kan sedan spelas upp i realtid, alternativt sparas och analyseras i efterhand. Motion capture används oftast för att dokumentera kroppsliga rörelser eller ansiktsuttryck, men tekniken har även en rad andra användningsområden. Efter att positioneringsdata har fångats upp av ett motion capture-system så konverteras denna oftast till ett för datorn lämpligt format, varefter den inspelade informationen kan bearbetas utifrån tillämpningsområde och situation. Vissa motion capture-system arbetar i realtid och genererar positioneringsdata användes då interaktivt för att ge direkt återkoppling tillbaka till användaren. Detta kan till exempel betyda att data som erhållits från en handrörelse återges i direkt form på en datorskärm samtidigt som rörelsen utförs. Med andra ord kan rörelsen spelas upp direkt, ungefär som en live-streamad videosekvens. Det finns även motion capture-system som inte arbetar i realtid. För dessa gäller att uppmätt data behöver bearbetas efter att ha spelats in, oftast för att kunna användas inom animering eller datorbaserad grafik.

Det finns en rad olika typer av motion capture-system. I denna uppsats behandlas främst motion capture med Kinect-kamera. För att få en uppfattning om vilka andra lösningar som finns, samt för att kunna jämföra Kinect-kamerans prestanda mot andra lösningar, är det relevant att nämna några andra etablerade motion capture-system. De tre vanligaste idag är mekanisk motion capture, optisk motion capture samt elektromagnetisk motion capture. Vid mekanisk motion capture används ett antal metallbitar som sätts ihop till ett konstgjort skelett. Detta skelett fästs sedan på användaren och när denna rör på sig kommer metall-skelettet att förflyttas på samma sätt som användarens kropp förflyttar sig. Sensorer som sitter utplacerade mellan de olika metallbitarna används för att generera data som representerar kroppsdelarnas olika positioner och rörelser. I optisk motion capture används istället ett antal reflekterande prickar tillsammans med ett antal kameror för att registrera data om positioner och rörelser. Informatio-

nen om var prickarna befinner sig beräknas med hjälp av triangulering. Ett exempel på optisk motion capture kan ses i figur 1. Här används triangulering för att beräkna positioner för olika kroppsdelar och en datoranimerad robotfigur kan på så sätt genereras utifrån erhållen positioneringsdata. En fördel med optisk motion capture är att denna genererar väldigt detaljerad data som är fri från brus. Den tredje typen av motion capture som är värd att nämna är elektromagnetisk motion capture. Denna baserar sig på att användaren bär elektromagnetiska mottagare, vars positioner beräknas i förhållande till en statisk elektromagnetisk sändare. En fördel med elektromagnetisk motion capture är att metoden kan användas i realtid (Furniss 1999).



Figur 1: Exempel på hur motion capture-teknik kan användas. Här beräknas positionen för olika kroppsdelar med hjälp av optisk motion capture-teknik, varefter en animerad bild kan genereras utifrån insamlad data (Wikipedia 2012a).

2.2 Kopplingen mellan musik och kroppsrörelser

Jensenius deklarerar i sin avhandling *Action-Sound – Developing Methods and Tools to Study Music-Related Body Movement* (2007) att kroppsrörelser spelar en viktig roll både för musikalisk perception och upplevelse vid ett musikaliskt framträdande. Han söker vetenskapliga belegg för uttalandet att musik är rörelse och att människan tänker på musik just som en rörelse. En poäng är att människan skapar musik genom att röra sig och att vi dessutom uppfattar musik medan vi rör på oss. Jensenius menar att hela kroppen spelar en viktig roll vid vår uppfattning och upplevelse av musik. Västerländsk konstmusik är en förkroppsligad aktivitet och bör således utvärderas just

som en sådan. Vidare kan man, i de metaforer som används i notskrift för att beskriva hur musiken skall spelas upp, observera ett flertal olika referenser till just kroppslig rörelse. De flesta italienska tempoangivelserna är associerade till mänskliga rörelser. Exempel på detta är termen *andante* (“lugnt, gående”) samt *corrente* (“springande”). Ett annat exempel är *corrente*, springande. Jensenius diskuterar hur olika typer av musikrelaterade rörelser har olika stark koppling till ljud. Han betonar att direkta ljudgenererande rörelser så som att spela ett musikinstrument har en nära och intim koppling till begreppet ljud, medan dansrörelser är mer löst kopplade till ljudbegreppet.

I boken *Musical Gestures - Sound, Movement, and meaning* diskuterar Godøy & Leman (2009) begreppet gest och dess koppling till musik. Författarna definierar termen gest som en rörelse utförd av någon kroppsdel, till exempel en hand eller ett huvud. En gest är en rörelse som syftar till att förmedla en mening eller en idé. I ett kapitel i boken diskuterar Leman kopplingen mellan musik och gester, samt hur budskap kan förkroppsligas med hjälp av gester i musik. Leman argumenterar för att musik till viss utsträckning är samma sak som gester. Han konstaterar att gester utgör ett kroppsligt led i vår musikaliska perception samt att musik både innehåller gester och i sig kan ses som gester. Musik kan beskrivas som en speglade process, där den kroppsliga gestalten formar mening i musiken. Vidare gäller att strukturella delar i musik, så som korta musikmotiv inom ett större stycke, kan imiteras med kroppsliga gester. Sådana musikaliska mönster kallas ofta för “gestures” på engelska, just för att dessa kan imiteras med den mänskliga kroppen.

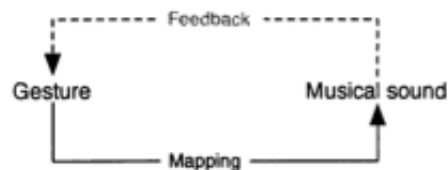
I publikationen *Music via Motion: A distributed framework for interactive multimedia performance* (Leonard & Ng 2011) nämns hur musikaliska uppträdanden är nära förknippade med kroppsrörelser på många olika nivåer, både för instrumentalisten och lyssnaren. Författarna hävdar att det råder en stark koppling mellan kroppsliga rörelser och musik samt perception av musik. Leonard och Ng menar vidare att de motion capture-system som arbetar i realtid erbjuder ett mycket uttrycksfullt gränssnitt. Sådana system minimerar dessutom de fysiska begränsningar som vanligtvis karakteriserar analoga instrument.

Från det ovanstående drar vi slutsatsen att det råder ett samband mellan be-

greppet musik och kroppslig rörelse. Denna tankegång utgör ett viktigt fundament i vår studie och styrker dessutom idén om att ett kroppsrörelsebaserat gränssnitt kan tillföra en mängd fördelar vid musikskapande.

2.3 Att designa instrument

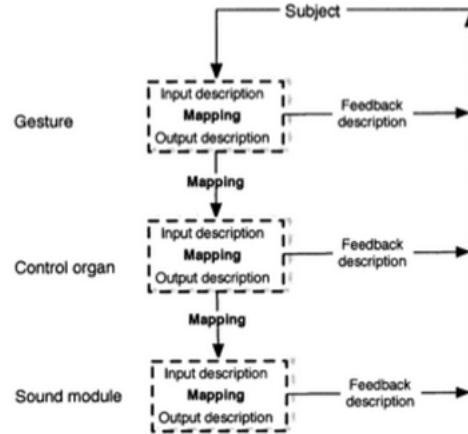
När man vill beskriva eller designa en produkt har man stor nytta av att förstå vilka parametrar som står till förfogande för manipulation, alltså vilka egenskaper som kan förändras hos det som skall designas. En modell som beskriver hur en typ av instrument fungerar ger den som ska beskriva, designa eller utvärdera liknande instrument goda förutsättningar i sitt arbete (Kvifte & Jensenius 2006). Ett musikinstrument kan förstås utifrån en modell som baseras på en processkedja med både instrumentet och instrumentalisten inkluderade. En sådan processkedja illustreras i figur 2. I den enklaste modellen över ett instrument så leder en gest utförd av instrumentalisten till att ett ljud produceras. Däremellan sker så kallad *mappning* (“mapping”). Med mappning menas att ett moment i processkedjan pekar mot nästa moment som en kausal följd av det föregående momentet. Mappning är ett begrepp som beskriver matchningen mellan ett objekts tänkta och dess verkliga funktion. Det sista steget i ett instruments processkedja utgörs av återkoppling till instrumentalisten. Detta sätt att förstå instrumentalistens användande av ett instrument synliggör flera olika delar av interaktionen med instrumentet.



Figur 2: Processkedjan från gest till ljud. En användare av ett instrument utför en gest som leder till att ljud produceras, gesten mappas sedan till ett ljud, varefter ljudet ges som feedback tillbaka till användaren (från Kvifte & Jensenius 2006).

Den enkla modellen som beskrivs ovan kan utvidgas, vilket illustreras i figur

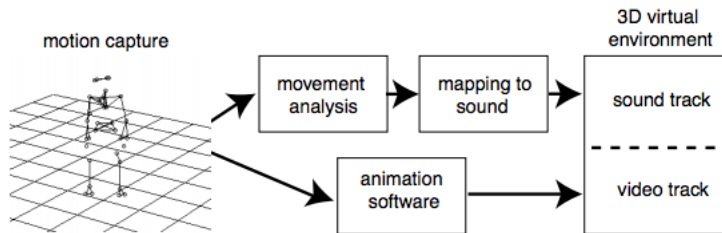
3. I den utvecklade modellen tas instrumentalisten genom en serie av multi-modala händelser. När en gest utförs får instrumentalisten först återkoppling från kroppen genom att denne känner och kanske också hör eller ser hur gesten genomförs. Om gesten innebär att man tar i ett fysiskt instrument så kommer även instrumentet ge återkoppling med exempelvis hur det känns att ta i. Slutligen ger instrumentet återkoppling genom det ljud som produceras. Varje delsteg kan analyseras utifrån mappningsbegreppet. Mappning kan här ses som översättningen mellan en typ av beskrivning till en annan (Kvifte & Jensenius 2006). Ett ingående meddelande mappar till ett utgående i varje delsteg. På detta sätt förändras rörelsen i en kropp, en gest, till en rörelse i ett instrument, till exempel vibrationer i en sträng.



Figur 3: Detaljerad modell för processkedjan från gest till ljud. Ett instrument beskrivs i tre delar. Användaren initierar en kedja av händelser. Instrumentets syfte, att producera ljud, är en del av den utgående återkopplingen från kedjans sista del (från Kvifte & Jensenius 2006).

Ett instrument där interaktionen sker med hjälp av motion capture skiljer sig från den modell som presenteras i figur 3. För ett motion capture-system existerar det nämligen ingen naturlig fysisk återkoppling direkt från instrumentet till användaren. Återkoppling kan då tillföras genom att introducera ett grafiskt gränssnitt som återger hur instrumentet uppfattar de rörelser som instrumentalisten utfört. Detta kan exempelvis ske med hjälp av videosekvenser som återger rörelserna. I figur 4 beskrivs ett förlopp från motion capture till ljud- och videoåterkoppling. En annan möjlighet för mo-

tion capture-baserade system är att ge så kallad haptisk återkoppling, vilket innebär att systemet ger återkoppling till känselsinnet genom exempelvis vibration eller motstånd.



Figur 4: En motion capture-baserad musikprodukt. Figuren visar händelseförloppet för musikalisk kontroll hos en motion capture-baserad produkt. Rörelser analyseras och resulterande data mappas till ett ljud. Samtidigt används mjukvara för att generera en animerad video med visuell återkoppling till användaren (Bevilacqua 2002).

Ett problem som man ställs inför när ett elektroniskt musikinstrument ska designas är fördröjningen mellan den ljudgenererande handlingen och det genererade ljudet. Det är önskvärt och i många fall nödvändigt att den återkoppling som man får från ett instrument inte kommer långt efter det att man utfört en kontrollerande handling. Det är svårt att spela musik om systemet introducerar för mycket fördröjning. När det gäller system som använder sig av motion capture kan det, beroende på vilket system som används, uppstå viss fördröjning på grund av att systemet måste analysera rörelserna från instrumentalisten. För system med motion capture utan haptisk återkoppling konstaterades det i en studie utförd vid Helsingfors Tekniska Universitet (Mäki-Patola & Hämmäläinen 2004) att den kortaste tidsfördröjning som en användare kan uppfatta (JND) ligger runt 30 ms. Denna slutsats drogs efter tester som utfördes på en Theremin. Av testandvänderna uppfattade 35 % att en fördröjning på 100 ms var störande vid spelande av instrumentet. I studien deltog 16 personer och resultatet beräknades ha en riskmarginal på 2.6 %.

2.4 Mappning i motion capture-baserade system

Vid all form av produktutveckling där slutprodukten inbegriper någon form av mänsklig interaktion bör man reflektera över hur produktens gränssnitt skall utformas. För ett rörelsebaserat gränssnitt är det relevant att fundera över hur de kroppsliga rörelserna skall kopplas ihop med ljudeffekter. Hur kan mänskliga rörelser och gester tolkas musikaliskt i ett gränssnitt? För utvecklingen av en Kinectbaserad musikapplikation handlar mappningen om att konstruera en naturlig koppling mellan de rörelser som användare utför, och de ljudeffekter som dessa genererar.

Jensenius (2007) diskuterar i sin avhandling att människan kan förstå ljud genom att gissa hur dessa är producerade. Han diskuterar sambanden mellan händelser och ljud där det finns en mekanisk och akustisk koppling mellan själva händelsen och ljudet (“action-sound couplings”). Den mänskliga förmågan att föreställa sig och förutspå ljud tycks bero på vår kunskap om objekts akustiska egenskaper samt de mekaniska egenskaperna hos händelserna. Vår mentala bild av ett ljud baseras på händelsen som producerar ljudet, ljudet i sig samt kunskap om kopplingen mellan dessa två faktorer. Denna kunskap är djupt rotad i vårt kognitiva system och med hjälp av denna kunskap kan vi tänka oss in i vilket ljud som produceras av en specifik handling.

I samma studie behandlas också förhållanden mellan handlingar och ljud där kopplingen mellan de båda inte är lika uppenbar. Dessa lösa kopplingar beskrivs som förhållanden (“action-sound relationships”) snarare än direkta samband. Sådana förhållanden existerar exempelvis i elektroniska instrument. För sådana instrument råder det oftast inte något intuitivt samband mellan ljudet och den ljudproducerande handlingen. Jensenius berättar att det således finns två olika strategier vid design av elektroniska enheter; praktisk design och kreativ design. Vid praktisk design försöker man återskapa direkta kopplingar mellan ljud och handling. Sambandet mellan handling och ljud blir således naturlig och stark. Vid kreativ design skapar man istället nya förhållanden mellan handling och ljud, och dessa måste inte nödvändigtvis vara intuitiva. Jensenius påpekar vidare att det är viktigt att fokusera på egenskaper hos interaktionen som sker mellan användare och enhet. En impulsiv handling bör exempelvis orsaka ett impulsivt ljud, medan en återhållsam handling istället bör generera ett återhållsamt ljud. Klangfärgens egenskaper och styrkan på ljuden skall också motsvara den ljudgenererande handlingen. Om man följer dessa kriterier så får man auto-

matiskt ett naturligt och intuitivt gränssnitt (Jensenius 2007).

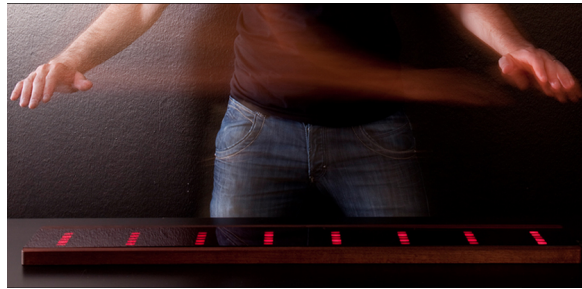
För en musikprodukt baserad på motion capture i realtid gäller att en mängd kontinuerliga funktioner skall mappas till olika kroppsrörelser med minimal fördröjning. Wanderley (2000) framhäver att man vid design av nya digitala och geststyrda instrument bör studera karaktären hos gester som producerats av erfarna instrumentalister. Han menar att man bör analysera och identifiera karaktärsdrag hos gesterna i dess speciella kontext och på så sätt få värdefull kunskap som hjälper till vid designprocessen. Wanderley poängterar vidare att förhållandet mellan kroppsliga gesters olika variabler och indata i ljudsyntes i ett digitalt instrument är långt ifrån uppenbar. Det råder ingen konsensus om vilka mappningsstrategier eller designmodeller som skall användas vid utvecklingen av gestbaserade digitala musikinstrument (Wanderley 2000).

Henze et al. (2010) diskuterar i arbetet *Free-Hand Gestures for Music Playback: Deriving Gestures with a User-Centred Process* hur man väljer vilka gester som ska styra specifika funktioner i ett system för uppspelning av musik. Författarna föreslår en utvärderingsmetod för mänskliga rörelser där ingående intervjuer utförs med användarna. Resultatet av en sådan utvärdering antas vara något förfinade gester med starkare koppling till det som av användaren uppfattas som intuitivt. Henze et al visar i sitt arbete att det inte finns något enkelt sätt att ta reda på vilka gester som lämpar sig väl i en applikation, om man inte först frågar många användare. Även Jensenius (2007) nämner att det är svårt att dra generella slutsatser om kroppsliga rörelser och hur dessa skall uppfattas ur ett ljudperspektiv. Sammanfattningsvis drar vi slutsatsen att det finns ett antal olika studier som behandlar utvärdering av mänskliga gester i ett antal specifika sammanhang, men att resultaten från dessa studier inte kan generaliseras i någon vidare mån. För att åstadkomma god mappning och på så sätt generera användarvänliga gestbaserade gränssnitt behöver man med andra ord utföra en rad olika produktspecifika användartester.

2.5 Air Instruments

Air Instruments, eller luftinstrument, är virtuella musikinstrument som genererar ljud genom att instrumentalisten härmar en ljudproducerande gest i luften. Ett luftinstrument producerar ljud utan att musikanten har någon direkt fysisk kontakt med instrumentet i sig. Exempel på luftinstrument är Virtual Slide Guitar från Helsingfors Tekniska universitet (Pakarinen et al.

2008) och Air Piano (Yosha 2012), se figur 5 . För båda dessa instrument gäller att ljud genereras genom att användaren rör sina händer i luften, utan direkt kontakt med vare sig en gitarr eller en keyboard.



Figur 5: Air Piano. En innovativ MIDI och OSC-kontroller som styrs av handrörelser (från Yosha 2012).

Den mentala bilden av ett musikinstrument kan till viss del åsynliggöras vid studier av hur användare interagerar med ett luftinstrument. Att undersöka hur etablerade luftinstrument fungerar kan således vara av intresse för vår studie, eftersom stora delar av prototyparbetet går ut på att se till att den mentala bilden av hur ljud produceras överensstämmer med den verkliga bilden. I en studie utförd av Godøy et al. (2005) observerades testpersoner som utförde luftspelning (“air playing”) till förinspelad musik. Med andra ord studerades hur användarna härmade musikgenererande gester samt hur användarna låstsades framföra musikstycken i luften. Likheterna mellan de gester som krävs för att generera ljud på ett fysiskt instrument och de faktiska härmande gesterna jämfördes. Resultaten visade att en större förståelse för kopplingen mellan ljud och rörelser kan uppnås just genom att observera hur människor härmar ljud Godøy et al. menar att studier av mänsklig interaktion med ett luftinstrument kan ge information om vilka aspekter av en rörelse som användare upplever som relevanta i gesten. På motsvarande sätt kan man också få information om kroppsrörelser som inte är relevanta och således kan förbises vid implementation av ett nytt instrument.

En av de svårigheter som uppstår vid utformning av både luftinstrument och gestbaserade musikprodukter är att dessa system inte introducerar någon naturlig taktill återkoppling till användaren (Godøy et al. 2005). Den återkoppling som presenteras är istället främst återkoppling som kommer från kroppen då gesten utförs samt hörbar återkoppling från den musik som produceras. I kapitlet ovan presenterades en modell för hur ett instrument

kan beskrivas i tre delsteg. För ett motion capture-baserat gränssnitt gäller att det andra delsteget i figur 3 (“control organ”) inte automatiskt bidrar med någon återkoppling till användaren. För att förbättra upplevelsen av kontroll över instrumentet kan det därför vara bra att introducera någon form av “artificiell” återkoppling, exempelvis i form av grafik som visar när användaren aktiverar någon del av instrumentet. Detta enligt principen att återkoppling alltid ska ges från systemet för varje handling utförd av användaren (Preece et al. 2002).

2.6 Att generera musik i 2D

Att studera hur tvådimensionella system är utformade kan vara värdefullt även vid utveckling av musikprodukter som baserar sig på en tredimensionell parameterrymd. Genom att observera hur man i 2D valt att koppla ihop rörelser till olika ljudeffekter kan man få kunskap om hur motsvarande rörelser i 3D skulle kunna mappas till olika ljudeffekter. Det existerar idag en uppsjö av produkter på marknaden där ljud genereras med hjälp av rörelser över en tvådimensionell rymd, exempelvis på pekskärm. Två exempel är KORG Electronics Kaoss Pad KP3¹ och Kaossilator KO-1². En Kaoss Pad är en kombinerad MIDI kontroller, effektprocessor och sampler med touchskärm där ljudet kan manipuleras i realtid. Touch-skärmen kan användas för att kontrollera effekter så som filter, tremolo, flanger och reverb. Vidare kan Kaoss Pad:en användas som en MIDI-kontroller, där fingrets position på x- och y-axeln anger värden som används för att generera två kontinuerliga MIDI strömmar (KORG 2012a).



Figur 6: KORG Electronics Kaoss Pad KP 3 (från KORG 2012a).

¹<http://www.youtube.com/watch?v=1hdhCSSWn-s&feature=relmfu>

²<http://www.youtube.com/watch?v=zkyLvRhOUPM&feature=relmfu>

Kaossilatorn KO-1 beskrivs på KORGS hemsida (KORG Electronics 2012b) som en användarvänlig touch-pad som tillåter användaren att skapa musik och rytmiska fraser med hjälp av 100 olika synthprogram. Synthprogrammen innehåller allt från instrument till beat-mönster, och man kan välja mellan 31 olika skalor samt 50 olika arpeggios. Genom att röra fingret över touch-skärmen kan man generera en stor mängd olika ljud i realtid. X-koordinaten på skärmen styr en parameter och y-koordinaten styr en annan. De genererade ljuden kan sedan spelas in. Med hjälp av en Kaossilator kan man loopa ett flertal olika ljud samtidigt, och på så sätt skapa en ljudbild i flera lager.



Figur 7: KORG Electronics Kaossilator KO-1 (från KORG Electronics 2012b).

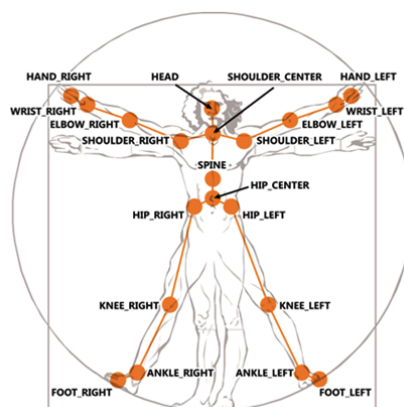
2.7 Kinect-kameran

Kinect är en användarkontroller som producerats av Microsoft med syftet att användas tillsammans med spelkonsolen Xbox-360. I de Xbox-spel som utvecklats för Kinect-kamera används motion capture för att registrera en spelares position i rummet. På detta sätt kan eventuella styrdon ersättas med kroppen. Kinect-kameran, vilken kan ses i figur 8, innehåller en infraröd kamera, två vanliga kameror samt en rad med mikrofoner. Med hjälp av dessa och mjukvara för bildigenkänning kan positioner för olika kroppsdelar beräknas på en användare som står placerad framför Kinect-kameran. Kinect ger möjlighet till registrering av ett tjugotal olika leder i kroppen. Se figur 9. Med hjälp av dessa registreringar kan kameran till exempel avgöra om någon har böjd eller rak arm, står på ett ben, hukar sig eller liknande. Kinect har även stereo-seende, och med andra ord kan kameran användas för att detektera positioner inte bara i horisontell och vertikal ledd utan också i djup. Positioneringen av kroppens alla större leder kallas ofta för “skeleton

tracking” och med hjälp av denna kan man generera en skelettliknande grafisk avatar av användaren. Microsoft tillåter utveckling av både kommersiella och icke-kommersiella applikationer för Kinect och tillhandahåller för detta ett programutvecklingspaket, Kinect for Windows Software Development Kit (SDK) (Microsoft 2012) (Wikipedia 2012b).



Figur 8: Kinect-kameran. Microsofts Kinect-kamera utvecklad för Xbox (från Microsoft 2012).



Figur 9: Skeleton Tracking. Kinect-kameran kan registrera 20 olika punkter på en kropp. Dessa kan sedan sättas ihop till ett virtuellt skelett (från Embodied Sensor Technology 2011).

2.8 Kinect och musik - tidigare forskning & projekt

Många forskare och utvecklare har sedan Kinect-kamerans lansering i 2010 arbetat med att utveckla olika typer av applikationer som kan styras med hjälp av kroppslig rörelse. Följaktigen finns det idag en mängd olika applikationer där en eller flera Kinect-kameror används för att reglera ljud, ljudeffekter och musik. Många av dessa applikationer har släppts som öppen källkod. Det finns också några olika forskningsprojekt som behandlar ämnesområdet. De flesta forskningsprojekt som ämnar förena motion capture-data från Kinect med musik utgår från ungefär samma tekniska upplägg: Kinect-kameran används för att generera positioneringsdata, denna data överförs och översätts sedan till lämpligt format, varefter musik kan syntetiseras med hjälp av Max/MSP eller Pure Data alternativt annan mjukvara för ljud så som till exempel Ableton Live. Ett projekt som är värt att nämna som bra exempel är "Creating Musical Expression using Kinect" vid Yonsei University i Sydkorea (Yoo et al. 2011). I detta projekt användes skelett-data från mjukvaran Skeleton Tracker. Denna positionerings-data översattes sedan till MIDI-data, varefter en mängd olika ljud kunde genereras. Därefter skapades musik med hjälp av ett program som skrivits i programmeringsspråket Max/MSP.

Ett annat projektexempel är *Interactive Music: Human Motion Initiated Music Generation Using Skeletal Tracking By Kinect*³ utfört av Berg (2012) med kollegor. Detta projekt syftade till att omvandla mänskliga rörelser till musik, och att sedan omvandla denna musik till notskrift. För att kunna använda Kinect-kameran tillsammans med en dator används ramverket OpenNI och NITE Middleware. För att koppla ihop motion capture-mjukvara med en musikgenerator användes sedan Open Sound Control Library (OSC). Ett antal parametrar som definierar mänskliga rörelser beräknas och skickas sedan vidare som OSC-meddelanden till en musikgenerator. Den musikgenererande applikationen utvecklades i MAX/MSP och musiken kunde även dokumenteras i notskrift.

Det finns en del intressant mjukvara som syftar till att fungera som en brygga mellan Kinect-kamerans rörelseigenkänning och produktion av musik. Ett exempel är Kinectar Performance Platform, ett MAX/SMP-baserat gränssnitt som är utvecklat för att användas tillsammans med Microsofts Xbox Kinect (Vik 2012). En skärmdump från Kinectar kan ses i figur 11. Med hjälp av Kinectar kan Kinectkameran användas som en avancerad

³<http://www.youtube.com/watch?v=7DnKPMZaJ5I>

MIDI-kontroller. Mjukvaran är tänkt att passa både nybörjare såväl som mer avancerade musikproducenter. Med hjälp av Kinectar kan data från handrörelser förmedlas till externa MIDI-kontroller eller MIDI-mjukvara via MIDI-protokoll eller OSC-meddelanden. Kinectar kan användas för att kontrollera enkla parametrar så som exempelvis ett lågpasfilter, eller programmeras för att passa betydligt mer avancerade ljudkonfigurationer.



Figur 10: Kinectar. Mjukvaran Kinectar används för att koppla samman mänskliga gester med MIDI-kontrollers eller MIDI-mjukvara (från Vik 2012).

Det finns en rad olika exempel på projekt där en Kinect-kamera använts för att skapa och generera ljud i konstnärliga tillämpningar och ljudinstallationer. Ett exempel är *LoopJam - A Collaborative Musical Map on the Dance Floor*, ett projekt som utfördes år 2011 vid Université de Mons i Belgien (Frisson et al. 2011). LoopJam är en interaktiv ljudinstallation som syftar till att skapa musik utifrån rörelser hos en hel grupp människor. Olika ljud och loopar har placerats ut över en yta, en "ljudkarta". Ljud aktiveras och spelas upp när deltagarna rör sig över ljudkartan och gestikulerar på olika sätt. Kombinationen av alla deltagares gester genererar en mängd olika ljud som synkas och på så sätt skapas musik. LoopJam använder sig av en Kinect-kamera och mjukvaran OpenNI för att få positioneringsdata. Denna data används sedan för att spela ljud från ett ljudbibliotek.

Kinect-kameran kan även användas för att konstruera nya elektroniska instrument. Exempel på ett sådant instrument är Therenect⁴ (Kaltenbrunner 2010). Detta instrument utvecklades vid Interface Culture Lab (University of Art and Industrial Design in Linz, Österrike). Therenect är en virtuell Theremin för Kinect. Therenecten definierar två antennpunkter, och utifrån handens position i förhållande till dessa punkter kan tonhöjd och volym hos

⁴<http://vimeo.com/17494162>

en enkel oscillator kontrolleras. Avståndet till kontrollpunkterna kan ändras genom att röra handen i en fri rymd. Rörelserna som genererar ljud i There-necten är med andra ord tänkta att vara relativt lika de rörelser som används för att spela på en riktig Theremin.

Idag finns det en mängd musikrelaterade tv-spel ute på marknaden, så som exempelvis GuitarHero och SingStar. Det existerar dock enbart ett fåtal spel som utvecklats för Xbox och Kinect. Två exempel på sådana musikrelaterade spel är Air Band samt Musical Feet (Kinect Aku 2012b). Airband är ett spel för en till två spelare där man på ett lekfullt sätt kan uppleva att man spelar i ett band. Det finns fem olika musikstilar att välja mellan och tre olika instrument. Instrumenten sätts igång genom att utföra olika gester. Användaren har inte så stor möjlighet att påverka musiken som spelas, men styr vilka instrument som skall ljuda samt när i låten som dessa ska spela (Kinect Aku 2012a). Musical Feet är ett annat spel, där man först väljer ett specifikt ljud (till exempel ett musikinstrument eller ett djurläte) varefter man sedan kan dansa över en virtuell keyboard. Musiken kan sedan spelas in och sparas.

3 Metod

3.1 Angreppssätt

Arbetet inleddes med förstudier och litteratursökning. Litteratursökningen följdes av brainstorming i fokusgrupp, produktutveckling samt prototyptester och utvärdering. Vid undersökningen användes främst kvalitativa metoder. Den främsta orsaken till detta var att vi hade begränsat med tid att tillgå. För att kunna genomföra bra kvantitativa undersökningar krävs det i regel betydligt mer tid och resurser än vad vi hade utrymme för. De kvalitativa metoder som vi använde oss utav var brainstorming i fokusgrupp samt prototyptest i form av observationsintervju. Syftet med att utföra dessa undersökningar i grupp var att skapa en god gruppdynamik där användarna kunde inspirera varandra att tänka kreativt, och att på så sätt skapa en bra diskussion om vår idé och produkt.

3.2 Förstudie

För att få en inblick i ämnet inleddes arbetet med en litteratursökning. Syftet med denna initiala litteratursökning var främst att kartlägga likartade projekt och forskning inom ämnesområdet. Litteratursökningen inriktade sig mestadels på att finna vetenskapliga artiklar samt rapporter och vi genomförde därför sökningen i de digitala databaser som tillhandahålls av KTH. Vi sökte också bland tillgängliga publikationer från "Sound and Music Computing Network" (SMC), "International Conference on New Interfaces for Musical Expression" (NIME) samt "International Computer Music Conference" (ICMC).

3.3 Fokusgrupp och brainstorming

Innan prototyputvecklingens start genomfördes brainstorming i fokusgrupp. Att använda en fokusgrupp är en effektiv metod där mycket data och information kan samlas in samtidigt. Metoden är också relativt enkel och okomplicerad att genomföra (Robson 2002). Brainstormingen genomfördes i typen semistrukturerad intervju. Vid en semistrukturerad intervju av denna typ har man förberett ett antal frågor, men de som intervjuas uppmanas att svara fritt och intervjun kan mer liknas vid ett samtal (Gulliksen & Göransson 2002). Fokusgruppen bestod av sex olika deltagare från Medieteknikprogrammet på KTH. Gruppdiskussionen utgick från ett antal öppna frågor och syftade till att skapa en kreativ diskussion om vår produktidé.

Målet var att undersöka om det finns ett intresse för den här typen av produkt, samt att i så fall kartlägga vilka egenskaper och funktioner som är önskvärda hos produkten.

3.4 Prototyputveckling

Första steget i prototyputvecklingen utgjordes av studier av källkod från ett liknande projekt som utfördes av Hansen, al Moubayed och Bresin vid institutionen för Tal Musik och Hörsel på KTH och presenterades på ForskarFredag 2011, en temadag anordnad av föreningen Vetenskap och Allmänhet. Prototyputvecklingen skedde därefter i en iterativ process där programmering kombinerades med testning, utvärdering och reflektion samt fortsatt programmering. Vid prototyputvecklingen användes programmeringspråken C# och Pure Data ³.

3.5 Prototyptest

När prototypen var färdigutvecklad planerades ett antal prototyptester med tio användare från Medieteknikprogrammet på KTH. Av de tio användarna så hade fem även deltagit i den fokusgrupp som beskrivits tidigare. Prototyptesten genomfördes i grupper om två användare i taget och var av typen observationsintervju. Vid en observationsintervju kombineras observationer av hur användare arbetar med en viss uppgift med intervjuer (Gulliksen & Göransson 2002). Tanken med att utföra dessa intervjuer i par var att skapa en diskussion om prototypen och på så sätt få mer kvalitativ data från testerna. Syftet med prototyptesterna var att undersöka hur väl vår produkt mottogs av målgruppen. Prototyptesterna syftade också till att kartlägga eventuella brister hos produkten samt att tydliggöra vilka krav som ställs på den här typen av produkt för att den ska fungera väl och uppfattas som enkel och rolig att använda. Vidare användes prototyptesterna för att undersöka hur mappningen mellan funktion och ljud skulle utformas för ett rörelsebaserat gränssnitt.

3.6 Metodkritik

Alla vetenskapliga undersökningsmetoder har brister, vilket i värsta fall kan leda till att felaktiga slutsatser kan dras. Vi har genom hela arbetsprocessen försökt ta detta i beaktande för att minimera de felaktigheter som kan uppkomma vid en undersökningsprocess. Vi har valt att använda ett flertal

³<http://puredata.info/>

metoder för att svara på de frågeställningar som presenterades i inledningen. Genom att kombinera resultat och information från alla dessa olika metoder hoppas vi kunna dra slutsatser med relativt hög tillförlitlighet.

3.6.1 Urval

De resultat som erhöles från våra undersökningar är baserade på en relativt homogen och liten grupp testanvändare. Merparten av dessa användare hade ett stort intresse för musik och spelade också instrument. Det är möjligt att dessa testanvändare inte är helt representativa för gruppen “unga teknikintresserade”, vilket är definitionen på den målgrupp som vi riktat in oss på. Det är med andra ord svårt att generalisera de erhållna resultaten till unga teknikintresserade i allmänhet. Eftersom alla testanvändare passade in bra i den avsedda målgruppen så drar vi emellertid slutsatsen att resultaten från våra undersökningar är relevanta för vår frågeställning, trots att urvalet av testanvändare varit tämligen riktat och snävt.

3.6.2 Validitet

Validitet är ett begrepp som beskriver hur väl en undersökningsmetod mäter det som denna metod egentligen har för avsikt att mäta (Bell 2000). Denna undersökning har främst baserats på relativt fria undersökningsmetoder så som semistrukturerade intervjuer och observationsintervjuer. Användarna uppmanades att komma med många egna synpunkter, snarare än att svara på direkta koncisa frågor. Det kan vara svårt att uppnå hög validitet med undersökningsmetoder som har denna lösa karaktär och resultaten skall därför beaktas med viss försiktighet. Vidare kan vi som intervjuare ha mer eller mindre stor inverkan på undersökningarnas resultat, vilket man också bör bära i åtanke.

3.6.3 Reliabilitet

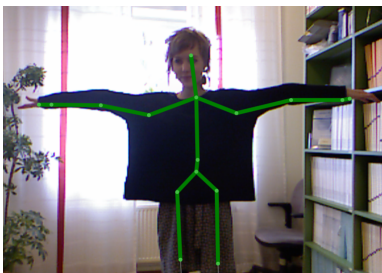
Med reliabilitet menas huruvida en metod visar samma resultat vid upprepade undersökningar. Reliabiliteten är med andra ord ett mått på tillförlitlighet i en mätning (Bell 2000). Vi gör bedömningen att reliabiliteten i fokusgruppen är relativt låg. Således bör resultaten från detta moment tolkas med viss försiktighet. Vi genomförde enbart en enda fokusgrupp, och det är mycket möjligt att resultaten från denna brainstorming berodde till stor del på rådande gruppdynamik. För att åstadkomma högre reliabilitet hade det kanske varit önskvärt att genomföra flera fokusgrupper där samma frågeställningar presenterades.

Vidare drar vi slutsatsen att reliabiliteten för prototyptesterna är relativt hög. Det är föga troligt att vi skulle få helt andra svar om vi genomförde dessa test flera gånger. Det är emellertid värt att notera att det i gruppen av testanvändare fanns en överrepresentation av personer med stort musikintresse och att detta kan ha påverkat de erhållna resultaten i positiv riktning. Vidare gäller att testen genomfördes i små grupper där användarna inte borde ha kunnat påverka varandra, och därmed också resultaten, i någon vidare utsträckning.

4 Resultat

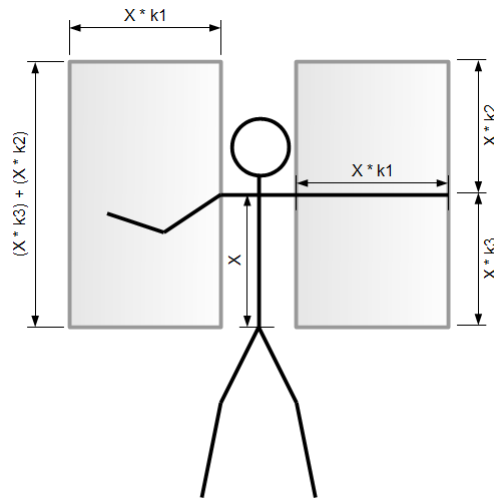
4.1 Prototypen

Prototypen utformades som ett loop-verktyg. Ett antal korta *samples* på 5-10 sekunder och 100 bpm spelas upp samtidigt i programmet. Dessa samples tilldelades volymen noll vid programmets start. Användaren kunde sedan sätta igång olika trumrytmer och samples genom att göra olika gester framför Kinect-kameran. Prototypen utvecklades i programmeringsspråket Pure Data. Positioneringsdata från Kinect-kameran erhöles med hjälp av ett modifierat exempelprogram från Microsoft vid namn Skeleton Tracker. Programmet medföljer i utvecklingspaketet för Kinect (Kinect SDK) och med detta kan man registrera alla större leder och kroppsdelar hos en person som står placerad framför Kinect-kameran. Skeleton Tracker ger även visuell återkoppling genom att rita upp ett skelett i realtid. Detta kan ses i figur 11. Koordinater för positionen på några utvalda leder kan med hjälp av Skeleton Tracker skickas över datorns nätverk till en given port, via TCP/IP. Genom att lyssna på denna port i Pure Data kan positionsdata effektivt överföras från Kinect. Med hjälp av videoinspelning kunde prototypens fördröjning mätas till att vara mellan 50-100 ms.

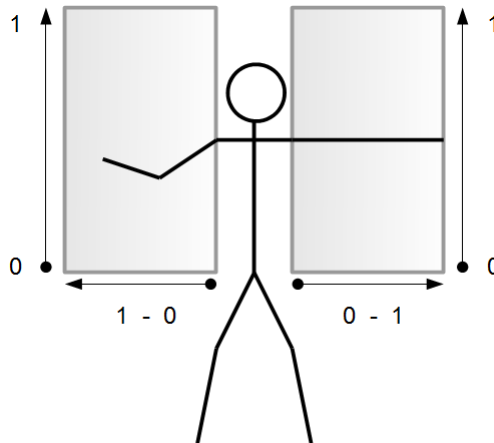


Figur 11: Visuell återkoppling från programmet Skeleton Tracker.

Positionen för höger och vänster handled, axel, midja samt nacke valdes som parametrar för att modifiera ljud i prototypen. Idén är att avståndet från handled till axel i horisontell ledd skall utgöra en skala där handen närmast axeln motsvarar minimum och en utsträckt arm motsvarar maximum. Samtidigt erhålls en skala i vertikalt led där en fullt utsträckt arm ovanför axeln är maximum och en utsträckt arm längs kroppen är minimum. En illustration av dessa skalor kan ses i figur 12.

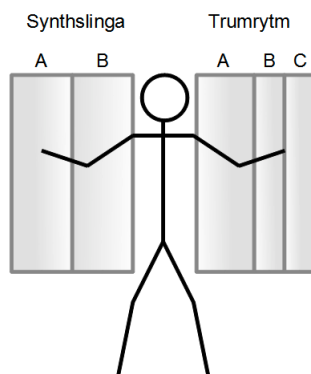


Figur 12: Dimensionering av prototypens koordinataxlar. Figuren visar ett koordinatsystem, ett för höger och ett för vänster hand. Koordinataxlarnas längd beräknas utifrån avståndet mellan midja och nacke. Konstanterna k_1 , k_2 och k_3 justeras så att en utsträckt arm hamnar i ytterkant av skalan.



Figur 13: Prototypens koordinataxlar. Figuren illustrerar de resulterande koordinatsystemen efter dimensionering av konstanterna k_1 , k_2 och k_3 i figur 12. Vi ser att resultatet är skalor från noll till ett i vertikal respektive horisontell ledd.

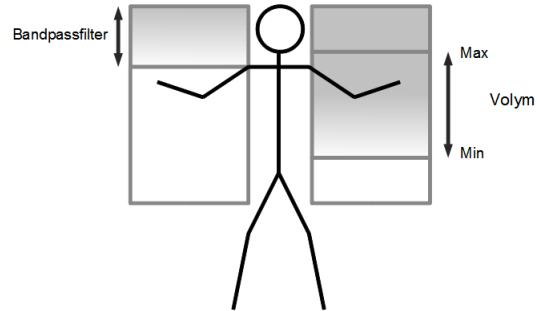
Positioneringsdatan behandlas vidare i ett steg där alla värden normaliseras och fyra skalor från noll till ett skapas; det vill säga en skala för varje hand i horisontell respektive vertikal ledd. Detta kan ses i figur 13. Genom att använda avståndet mellan midja och nacke som avståndsreferens kan skalan räknas ut korrekt oavsett rådande avstånd mellan användare och kamera. En arms längd räknas alltså ut som avståndet mellan midja och nacke multiplicerat med en bestämd multiplikator.



Figur 14: Prototypen: synth och trumrytmer. Här illustreras vilka områden som mappar till olika synthslingor respektive trumrytmer.

I figur 14 visas hur varje spalt har tilldelats samples. Det ljud som spelas upp styrs av vilken spalt som användarens hand befinner sig i. Vänster hand spelar synthslingor och höger hand spelar olika trumrytmer. Man kan på detta sätt kombinera en melodi med en trumrytm. Med höger hands position i vertikal ledd kan också volymen kontrolleras för trummorna.

I figur 15 illustreras hur man kan reglera volym och styra ett bandpassfilter med händerna. Av praktiska skäl sattes minsta volym till 0.3 och maximal volym till 0.6 på den vertikala skalan mellan 0 och 1. Man behövde således inte flytta armen varken längst ned eller högst upp på skalan för att kunna använda hela volymkontrollen. Vänster hands position i vertikal ledd styr ett bandpassfilter vilket aktiveras i höjd med axeln. Intervallet där filtret var aktivt placerades på ett sådant sätt att man kunde lägga på effekten medan man spelade upp en av melodislingorna.



Figur 15: Prototypen: volym och bandpassfilter. Figuren visar de områden som kan aktivera olika funktioner i prototypen. Här aktiveras ett bandpassfilter med en hand och den andra handen aktiverar en volymkontroll.

Bandpassfiltret konfigurerades så att det skulle ge en så hörbar effekt som möjligt när bandet panorerades. Detta testades fram under utvecklingen och resulterade i att bandpassfiltret bara är aktivt då vänster hand höjs över vertikalkvärde 0.7, ungefär i höjd med axeln. Då handen höjs över axeln tonas bandpassfiltret in från det ofiltrerade ljudet och när handen sänks igen så tonas ljudet över till det ofiltrerade. Bandpassfiltrets egenskaper ändras successivt utifrån hur högt handen hålls. Vid övergången från ofiltrerat ljud till filtrerat används 830 Hz som mittfrekvens och 5 som Q-värde, för att sedan nå sitt maximum med 15 kHz mittfrekvens och 15 som Q-värde då handen är i maxläge, strax ovanför huvudet.

Efter att prototyptesterna hade genomförts så modifierades prototypen utefter den feedback som fåtts. Dessa ändringar diskuteras vidare i kapitel 4.4.

4.2 Fokusgrupp och brainstorming

Fokusgruppen genomfördes i ett tidigt skede av undersökningen. Tanken med denna initiala fokusgrupp var främst att samla idéer, inspiration och tankar som främjade utvecklingen av den prototyp som beskrivs i tidigare kapitel. En Kaossilator (se kapitel 2.6 samt figur 7) samt ett antal videos från tidigare relevanta projekt (se kapitel 2.8) användes som inspiration för att förbättra deltagarnas förståelse för ämnet, innan en diskussion tog vid. De frågor som ställdes till deltagarna presenteras längre ner på denna sida.

Deltagarna i fokusgruppen utgjordes av sex civilingenjörstudenter i åldersgruppen 23-25 år. En översikt över deltagarna (här betitlade S1-S6) kan ses i tabell 1. För att kartlägga användarnas musikaliska bakgrund tillfrågades deltagarna om hur många instrument de spelade. Användarna bads också gradera sitt musikintresse på en skala från ett till fem, där fem definierades som det värde som motsvarade störst musikintresse. Det konstaterades att merparten av deltagarna spelade instrument. Medelvärdet för musikintresset i gruppen låg på 4.2. De flesta deltagare hade med andra ord ett relativt stort intresse för musik.

Tabell 1: Deltagare i fokusgruppen

Deltagare	Ålder	Spelar instrument	Musikintresse
S1	24	ett	3
S2	23	ett	5
S3	25	inget	3
S4	25	flera	5
S5	23	flera	4
S6	23	flera	5

4.2.1 Frågor som ställdes vid fokusgruppen

1. Vilka funktioner och effekter skulle ni vilja se hos den här typen av produkt? Hur skulle ni vilja modifiera musik med kroppsrörelser?
2. Skulle det tillföra något att kunna ändra ljudet med händerna och kroppen? Eller är det likvärdigt att göra det på datorn?
3. Exempel på sådana funktioner som skulle kunna finnas är: att byta låt, ändra volym, tonhöjd, bpm, EQ-kurva, göra loopar samt att använda instrument likt i en MIDI-kontroller. Hur skulle man visa sådana effekter med händerna? Hur skiljer sig rörelserna på Kaossilatorn (två dimensioner) mot rörelserna i en tredimensionell rymd?
4. Tilltalas ni av idén om en kroppsrörelsebaserad musikprodukt?
5. I vilka sammanhang skulle ni vilja använda en produkt av denna typ? På fest, som instrument, som discjockeys eller för några andra syften?

Det första ämnet som diskuterades i fokusgruppen var vilka funktioner som är önskvärda hos en musikapplikation som styrs med hjälp av en Kinect-kamera, samt vad man skulle vilja använda motion capture-teknik till i ljudsammanhang. En mängd olika förslag diskuterades, exempelvis möjligheten att mixa musik med hjälp av kroppsrörelser. Flera av deltagarna ansåg att det vore roligt om Kinect kunde användas för att utveckla ett musikrelaterat tv-spel. Ett annat förslag var att använda tekniken för att utveckla ett Kinect-baserat loop-verktyg. Vissa av användarna ansåg att det skulle vara önskvärt att använda positioneringsdata från Kinect-kameran som input till andra musikprogram så som till exempel Cubase. Det framfördes även önskemål om att kunna använda Kinect för att påverka ljud på samma sätt som med en MIDI-kontroller.

En viktig poäng som diskuterades under brainstormingen var att man vid utvecklingen av den här typen av produkt bör reflektera över huruvida det ska gå att spela "dåligt", "falskt", eller inte. En av deltagarna poängterade att det kanske vore bäst om musikapplikationen, likt en MIDI-kontroller, skulle vara konfigurerad för att alltid låta någorlunda melodiskt och därmed aldrig skulle kunna låta dåligt. Flera deltagare ansåg att det vore önskvärt att den musik som genererades skulle låta "perfekt", trots att man inte rörde sig precis eller perfekt. Det framfördes flera önskemål om att systemet skulle vara enkelt att använda men ändå låta bra och avancerat.

Merparten av deltagarna ansåg att det var mer intressant att styra musik med kroppen som kontroll, istället för att styra musik med hjälp av ett

gränssnitt på en datorskärm. Ett flertal poängterade att det vore roligare att modifiera ljud i 3D, samt att man kan få betydligt fler parametrar att jobba med i en 3D-rymd jämfört med på tvådimensionell skärm. Idén att det är intuitivt att modifiera musik med gester framfördes av flera deltagare. Deltagarna trodde även att det skulle vara svårare att styra musik på ett precist sätt i en luftrymd, jämfört med att göra det på en tvådimensionell yta. Ett förslag på hur detta precisionsproblem skulle lösas var att erbjuda någon form av visuell återkoppling på en skärm.

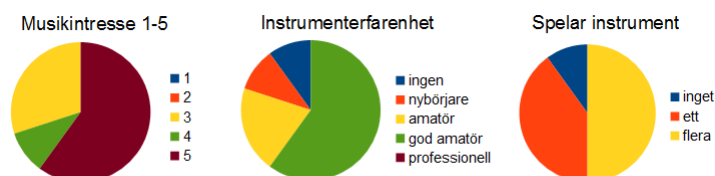
I fokusgruppen diskuterades även huruvida det fanns några rörelser som kändes naturligt förknippade med specifika ljudeffekter. Exempel på sådana rörelser som deltagarna ansåg naturliga var att röra händerna framåt och på så sätt aktivera systemet, att byta låt eller ljud med hjälp av en svepande handrörelse, samt att kunna generera snabb musik med hjälp av snabba handrörelser och vice versa.

Den allmänna inställningen till idén om rörelsebaserad musikproduktion bland deltagarna var positiv och flera deltagare påpekade att det känns intuitivt att modifiera ljud med rörelser. Det framfördes även ett flertal synpunkter angående att rörelser i luften är intressanta för att de introducerar nya metoder för att skapa musik. Paralleller drogs även till andra nytänkande gränssnitt för musikproduktion.

Det framkom att merparten av deltagarna i fokusgruppen skulle kunna tänka sig att använda en musikapplikation av denna typ vid sociala aktiviteter i underhållningssyfte. Många deltagare ansåg att tekniken skulle vara rolig att använda vid en fest, helst i formen av ett spel. Korrelationen mellan musik och dans diskuterades också, och flera av deltagarna ansåg att det vore roligt att kunna manipulera musik utifrån dansrörelser.

4.3 Prototyptester

Ett antal användartester av den prototyp som beskrivs i kapitel 4.1 genomfördes relativt tidigt i utvecklingsprocessen⁴. Testanvändarna utgjordes av 10 civilingenjörstudenter. Åldersfördelningen hos dessa användare kan ses i tabell 2. Deltagarnas musikaliska bakgrund illustreras i figur 16. Av de tio testanvändarna spelade nio ett eller flera instrument. Sex av dessa deltagare graderade sig själva som goda amatörer på sina förstainstrument. När testpersonerna bads gradera sitt musikintresse på en skala från ett till fem svarade alla tre och uppåt. Prototyptesterna var av typen observationsintervju. Användarna fick först en kort genomgång av hur prototypen fungerade, varefter de fick testa systemet. Efter testet fick användarna svara på ett antal frågor, vilka presenteras på nästa sida.



Figur 16: Deltagarnas musikaliska bakgrund. En illustration av den musikaliska bakgrunden hos de olika testpersonerna. Testpersonerna bads gradera sitt allmänna musikintresse från 1 (litet) till 5 (stort). De användare som spelade ett eller flera instrument bads även gradera sin färdighetsnivå på sina förstainstrument.

⁴Film från prototyptesterna: <http://www.kth.se/blogs/axelha/2012/03/ljud-kakor-och-en-funkande-andreas/>

4.3.1 Frågor som ställdes vid prototyptesterna

1. Har du TV-spel hemma? (Xbox, Kinect?)
2. Tyckte du att det var roligt att använda prototypen?
3. Tyckte du att du fick bra respons från systemet när du rörde dig?
4. Upplevde du att systemet var svårt eller rörigt?
5. Finns det någon/några funktioner som du saknade hos prototypen?
6. Skulle du vilja ha den här typen av system hemma? Skulle du kunna köpa det?
7. Vad skulle du vara beredd att betala isf?
8. Känns det naturligt/relevant att modifiera musik med handrörelser och gester? Är det bättre än att reglera parametrar genom att trycka på en skärm?
9. Pondera att man skulle kunna koppla systemet till andra program, till exempel Spotify. I vilka sammanhang skulle du i så fall vilja använda det här systemet?

Testanvändarna och dessas svarsfördelning presenteras i tabell 2. Av de tio personer som deltog i testet var det sju som ägde någon form av TV-spel. Endast tre stycken av personerna hade använt en Kinect-kamera förut, och samma antal hade en egen Xbox. Nio personer tyckte att det var roligt att använda prototypen. Nio av personerna svarade också "Ja" på frågan om huruvida de tyckte att det var underhållande och roligt att påverka musik med händerna.

När testanvändarna tillfrågades om de tyckte att prototypen gav bra respons svarade fem av tio personer att systemet hade för mycket fördröjning och att den fördröjning som existerade mellan en handrörelse och ljudeffekt var störande. Ytterligare fyra sa att systemet gav bra respons men var lite långsamt. Två av dessa ansåg att fördröjningen inte var något större problem. Fem personer påpekade att det är viktigt att programmets fördröjning är minimal för att det ska vara kul att använda systemet. En person svarade att systemet gav bra respons utan att säga något om fördröjningen.

Tabell 2: Deltagare vid prototyptesterna

Deltagare	Ålder	Spelar instrument	Nivå	Musikintresse	TV-spel
S1	24	Ett	3	Nybörjare	Nej
S2	23	Ett	3	Amatör	Ja
S3	25	Inget	3		Ja
S4	25	Flera	3	God amatör	Ja
S5	23	Flera	3	Amatör	Ja
S6	28	Flera	3	God amatör	Nej
S7	21	Flera	3	God amatör	Ja
S8	23	Ett	3	God amatör	Nej
S9	21	Ett	3	God amatör	Ja
S10	24	Flera	3	God amatör	Ja

Fyra av tio personer ansåg att systemet var lätt att använda. Övriga deltagare var av åsikten att programmet var ganska svårt i början, men att det snabbt blev enkelt. Sex personer ansåg att det vore roligt om det fanns fler ljud och ljudeffekter i programmet. Två personer ville kunna kontrollera volymen för alla loopar samtidigt.

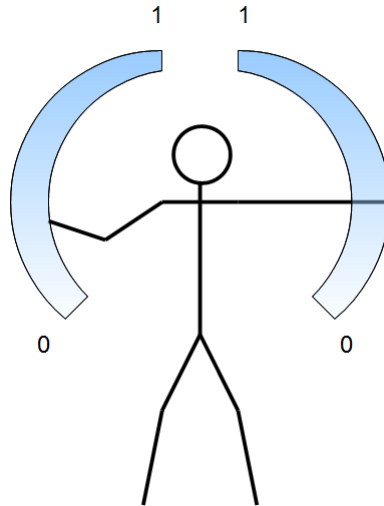
Fem av de tio testpersonerna svarade ja på frågan om huruvida de skulle vilja ha ett färdigtvcklat system av denna typ hemma. Fyra personer svarade att de kanske skulle vilja ha systemet hemma, och en person svarade "nej" på samma fråga. Omkring hälften av personerna uttalade att de gärna skulle köpa systemet om de redan hade en Kinect-kamera hemma. Av de som svarat ja på frågan så angav de flesta att de skulle vara beredda att betala ett antal hundralappar för produkten. Några användare ansåg att de skulle kunna betala ungefär tusen kronor för Kinect-kameran och programvaran i fråga.

Som nämnt ovan tyckte alla utom en av testpersonerna att det var underhållande och roligt att modifiera musik med händerna. De ansåg att det förhöjde upplevelsen, och att det var givande att röra sig till musiken på ett sätt som inte är givande när man sitter framför en datorskärm. Tre av tio användare föreslog också att systemet skulle kunna användas som lärandemedel för barn.

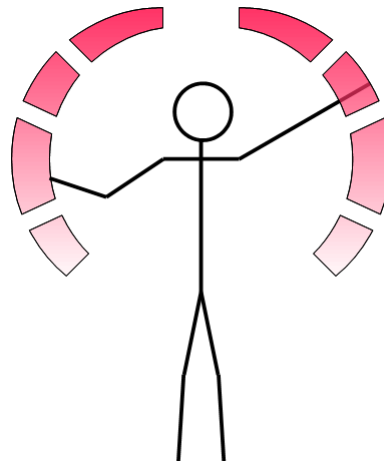
Det komplexa förhållandet mellan kroppsrörelser och dessas matchning till effekter blev uppenbart vid prototyptesterna. Ett flertal testanvändare utförde olika typer av gester och förväntade sig att dessa skulle ge ljudeffekter, trots att dessa inte hade implementerats i systemet. Många användare utförde cirkulära svepande rörelser och förväntade sig att dessa skulle ge hörbara utslag. Vi såg också att det verkade svårt för många att röra händerna utefter den mall vi skapat.

4.4 Slutgiltig Prototyp

Efter prototyptesterna så genomfördes ett antal modifikationer av prototypen. Den största förändringen var att en ny typ av mappning används i den modifierade prototypversionen. I den första prototypen kontrollerades valet av ljud genom handrörelser i rektangulära fält, vilket illustrerades i figur 13 och 14. Dessa rektangulära fält har i den modifierade versionen istället ersatts med cirkelsektorer. Vidare gäller att ljudslingorna triggas igång av rörelser som påminner mer om att sätta på och stänga av knappar än att röra sig i ett specifikt intervall. Detta medför att användaren kan låta ett ljud fortsätta spela, trots ändrad kroppslig position. Den modifierade prototypen introducerar också två olika lägen. I det ena läget håller man med raklång arm över en cirkelsektor och aktiverar på så sätt en ljudslinga. En eller flera slingor kan aktiveras samtidigt. I det andra läget kontrollerar man volymen för slingorna på respektive sida. Man byter mellan olika lägen genom att antingen stå med benen ihop eller med benen isär. Figur 17 och 18 visar prototypen med cirkelskala och cirkelsektorer som knappar.



Figur 17: Den vidareutvecklade prototypens volymkontroll. Den utvecklade prototypen innehåller cirkelsektorer istället för rektangulära områden. Då handen flyttas inom en cirkelsektor ändras volymen för ljudsingorna som valts på motsvarande sida. Detta läge är aktivt med benen brett isär.



Figur 18: Den vidareutvecklade prototypens olika ljudsingor. Här illustreras de cirkelsektorer som aktiverar olika ljudsingor. Detta läge är aktivt när användaren står med benen ihop.

5 Diskussion

Resultaten från fokusgruppen och prototyptesterna styrker uppfattningen om att en Kinect-kamera lämpar sig väl för musikprodukter som utvecklats i underhållningssyfte. Det framkom i den initiala fokusgruppen att merparten av deltagarna inte hade något större intresse av att utnyttja denna teknik för att spela instrument, utan att det främst var intressant att använda tekniken i en produkt som utformats för att användas som ett spel eller för sociala aktiviteter där flera personer kan samarbeta eller turas om att spela. Det föreslogs till exempel att man kunde låta människor "styra" musiken, ej vidare specificerat, på en fest, eller samarbeta i ett musikspel. Även vid prototyptesten uttalade merparten av testanvändarna att de hellre skulle använda sig av riktiga instrument eller musikprogram för att producera musik än att utnyttja ett instrument baserat på Kinect.

I vår prototyp behandlades enbart fallet rörelsebaserad musikproduktion i två dimensioner. En Kinect-kamera kan dock med fördel användas för att generera ljud i tre dimensioner. Genom att lägga till ännu en koordinataxel i parameterrymden möjliggör man att ännu fler effekter kan användas. Förutom den uppenbara fördelen att man kan placera ut fler parametrar om man introducerar en ny dimension så tycktes merparten av testanvändarna anse att det skulle vara roligt och intressant att kunna förändra musik genom att röra sig i tre dimensioner. När användarna tillfrågades om huruvida de fann det intressant att styra ljud i en rymd istället för på en tvådimensionell skärm så svarade merparten att detta var både önskvärt och roligt. Det tycktes finnas en utbredd uppfattning om att det är mer underhållande att modifiera ljud med hjälp av rörelser i en rymd, jämfört med att göra det samma på en skärm.

Utformningen av prototypen samt efterföljande användartester gav värdefulla insikter om hur en musikprodukt baserad på Kinect bör vara utformad för att tillfredställa en användares önskemål. Exempel på faktorer som tycktes viktiga för användarna var att systemet reagerade snabbt samt att den fördröjning som existerade var minimal. I kapitel 2.3 hänvisade vi till en studie om motion capture system (Mäki-Patola et al. 2004) där det kunde konstateras att så pass små fördröjningar som 30 ms kan detekteras av en lyssnare. Detta gränsvärde bör tas i beaktande när en produkt baserad på rörelseigenkänning utvecklas. Fördröjningen hos vår prototyp uppskattades till mellan 50-100 ms. Det vore önskvärt om denna tid istället låg på maximalt 30 ms, för att användaren inte skall kunna detektera eller störas av

systemets fördröjning.

Det framgick att många av användarna ansåg att det skulle krävas att musikprodukten hade en stor mängd av olika effekter för att produkten skulle vara intressant. En annan synpunkt som presenterade av många användare var att det skulle vara bra med visuell återkoppling. Detta skulle kunna tolkas på flera olika sätt; antingen är det generellt sett bra att ge visuell återkoppling till ett sådant system som det vi testat, eller så efterfrågades visuell återkoppling eftersom vår prototyp brister i den intuitiva kopplingen mellan gest och ljud. Det är också möjligt att användare i allmänhet tycker att det är roligt att se sin kropp skissas på skärm, så som i fallet med Skeleton Tracker. Sammanfattningsvis konstaterar vi att en musikapplikation baserad på Kinect bör erbjuda en mängd olika ljudeffekter, snabb reaktion och detektion av användarens rörelser samt möjligtvis även ett grafiskt gränssnitt.

De slutsatser som kunde dras efter fokusgrupp och prototyptester var bland annat att det fanns ett intresse av att använda den här tekniken i ett spel som var utformat som en loopmaskin eller MIDI-kontroller. Det fanns inte något nämnvärt intresse för att använda Kinect-tekniken vid uppspelning av musik, om inte detta skulle vara i form av ett DJ-vertyg som fungerade som ett spel.

Ett mål med denna undersökning är att kartlägga vilken målgrupp som en musikprodukt baserad på Kinect borde rikta sig mot. De utvärderingsmetoder som vi använt i denna undersökning har uteslutande inriktat sig på testanvändare i åldersgruppen 20-28 år på Medieteknikprogrammet på KTH. Majoriteten av testanvändarna förhöll sig positivt till idén om en musikprodukt baserad på Kinects teknik, och flertalet var intresserade av att köpa en musikprodukt av detta slag om det skulle existera en sådan på marknaden. Man kan således dra slutsatsen att testanvändarna utgör en del av målgruppen, och att målgruppen till viss del består av unga människor med visst teknikintresse. Flera av testanvändarna var musikintresserade och spelade någon typ av instrument, vi anser därför att vår testgrupp som helhet var kvalificerad att bedöma huruvida vår prototyp, eller en produkt liknande den, fungerade tillfredställande som musikinstrument. Vi kunde inte se någon skillnad i hur väl prototypen mottogs av de med mer musikalisk erfarenhet kontra de med mindre. Slutsatsen man kan dra av det är möjligtvis att det ställs ungefär samma krav på en musikapplikation av detta slag oavsett musikalisk bakgrund. Detta kan kanske i sådant fall förklaras med att det som tilltalar i en sån här typ av applikation är enkelheten och möjligheten

att med enkla medel producera musik som låter bra. Med andra ord, vår prototyp ger inte en musikalisk utmaning, som ett instrument gör, utan snarare en interaktiv musikupplevelse.

Vid prototyptesterna framkom att de flesta av testanvändarna skulle vara mer benägna att investera i produkten om denna var kompatibel med redan existerande spelkonsoler. Med andra ord skulle merparten av testanvändarna vara intresserade av att köpa produkten om den fanns som ett "spel" till en spelkonsol. Som vi nämnt ovan verkar det ställas ungefär samma krav på produkten oavsett musikalisk erfarenhet och den gemensamma faktorn för de personer som tyckte att det var roligt att använda prototypen var ett generellt musikintresse. Vi anser det troligt att personer med ett medelstort till stort musikintresse som redan använder sig av spelkonsoler, så som exempelvis Xbox eller Wii, är en lämplig målgrupp för denna typ av produkt. Möjligtvis kan också tilläggas att målgruppen bör inbegripa personer som spelar tv-spel socialt.

5.1 Analys av mappning

Att besluta hur olika de effekterna skall kopplas ihop med kroppsliga rörelser är ett av de viktigaste momenten i utvecklingen av en produkt av detta slag. Problemet med mappning av denna typ är att det inte existerar någon självklar korrelation mellan den kroppsliga rörelsen och dess motsvarande ljudeffekt. Vi observerade under prototyptesterna att flera användare utförde cirkulära och svepande rörelser för att interagera med gränssnittet. Vi kan konstatera att det inte faller sig naturligt att röra kroppen efter rektangulära områden såsom vår prototyp fungerar. Alla kroppsdelar rör ju sig naturligt i cirkulära rörelser. Att förflytta handen från ett ställe till ett annat måste i vår prototyp göras strikt i vertikal eller horisontell ledd, annars ändras flera parametrar samtidigt. Resultatet är att prototypen mappar till flera olika funktioner för det som för kroppen i princip är samma gest. Det leder kanske till att det blir svårare att använda gränssnittet, eller att det inte känns intuitivt. Det finns inte heller någon naturlig position att vila armarna i. För att fortsätta spela musik med prototypen måste händerna hållas uppe i luften, oftast med något utsträckta armar. Efter ett tag blir man ganska trött i armarna, man kan ju inte stödja sig emot något i luften.

På grund av tidsbrist så har vi inte kunnat följa upp det första prototyptestet med ett nytt test av den förbättrade prototypen. Nedan följer därför en analys av vilka problem som påträffades hos den första prototypversionen

samt förslag på olika lösningar för att komma till rätta med dessa problem.

Till att börja med så insågs det tidigt under prototyptesterna att de olika musikslingorna borde kunna aktiveras i ett område, för att sedan fortsätta spelas när gesten var utförd. Hos den första prototypen slutade ljudslingorna att spelas upp då händerna rört sig bort från de rektangulära områdena som definieras i figur 14 a) och b). Detta problem kan exempelvis lösas genom att implementera virtuella knappar i luften. Knapparna bör befinna sig på armlängds avstånd för att inte kunna komma åt av misstag. Att låta armarna hänga utmed kroppen i en viloposition, bör inte heller kunna aktivera någon funktion. Rörelsen att flytta armarna till denna viloposition bör helst inte heller aktivera någon funktion. Vidare så bör kontinuerliga funktioner så som volym och bandpassfilter inte ligga i en skala i vertikal eller horisontell ledd. Dessa bör istället utgå ifrån rörelser som är naturliga för människan, så som exempelvis svepande cirkelrörelser. En halvcirkelrörelse med utsträckt arm skulle kunna fungera bra i detta sammanhang. Med denna implementation skulle alltså en stor cirkelsektor tilldelas en skala.

För att undvika problemet med att en gest mappar till för många olika funktioner så kan det behövas flera olika lägen i systemet. Med flera lägen för manipulation så kan fler parametrar regleras utan oönskad dubbelmappning. Exempel på hur detta kan genomföras är att man byter mellan olika lägen beroende på var man står i rummet i förhållande till Kinectkameran, eller att ett läge aktiveras när man står med benen isär och ett annat med benen ihop. Ytterliggare en förbättring som kan implementeras i systemet är att låta olika kroppsdelar agera som knappar. Man skulle kunna låta huvud, mage eller kanske armbågar fungera som aktiva knappar på så sätt att en rörelse mot dessa kroppsdelar aktiverar en viss funktion. Alla dessa förslag är fullt möjliga att åstadkomma med hjälp av Skeleton Tracker. Något som skulle kunna göra prototypen ännu bättre vore att implementera igenkänning av kontinuerliga gester. Att man till exempel ritar en symbol i luften som aktiverar en funktion, skulle kunna göra prototypen betydligt mer intressant och dessutom ge utrymme för fler funktioner. Det skulle vara intressant att fortsätta utveckla prototypen i just detta avseende, men det ligger utanför tidsramen för detta arbete. Beskrivning av den lösning vi implementerat i denna studie finns beskrivet i det tidigare kapitlet 4.4.

5.2 Slutsats

Med hjälp av en Kinect-kamera kan man extrahera positioneringsdata från mänskliga rörelser med relativt god precision. Vissa begränsningar finns dock med denna typ av motion capture-system. Efter att ha testat Kinect-kameran tillsammans med den prototyp som beskrivs i kapitel 4.1 drog slutsatsen att Kinect-kameran i sig genererar relativt grov positioneringsdata. Kinect-kameran lämpade sig väl för att reglera funktioner som var utspridda över en relativt stor luftrymd i rummet. När kameran förväntades detektera små skillnader i rörelse i x eller y-led tenderade emellertid värdena ofta att hoppa inom olika spann. Det är möjligt att den mjukvara som vi använt vid undersökningen, Skeleton Tracker från Windows SDK, introducerar en mängd begränsningar beträffande precision. Ett val av annan mjukvara skulle kunna generera mer precisa och stabila positioneringsdata. Det finns andra typer av motion capture-system som kan fånga kontinuerliga kroppsrörelser mer exakt och precist än vad Kinect-kameran kan. Mer precisa motion capture-system lämpar sig förmodligen bättre vid utformningen av exakta instrument för musikproduktion, medan Kinect-kameran kan fungera bra för att trigga ljudeffekter som placerats ut i olika större zoner i en rymd. Vår slutsats är därför att Kinect-kameran fungerar bäst i produkter som utvecklats för underhållningssyfte, där man kan få god behållning av ett system med lägre precision.

Våra resultat pekar på att en musikleksak baserad på Kinect är en intressant produkt för målgruppen unga teknikintresserade med ett medelstort till stort musikintresse. Kraven som ställs på denna produkt skiljer sig inte mellan de användare som har liten och de som har stor erfarenhet av att spela musikinstrument. Vår samlade bedömning utifrån prototyptester och egna reflektioner är att denna musikleksak bör vara enkel att använda, alltid producera välljudande och för användaren tillfredställande musik, samt att interaktionen bör utgå ifrån naturliga kroppsrörelser. Man bör sträva efter att minimera fördröjningen mellan användarens gest och från leksaken utgående ljud. Det bör även tas i beaktning att precisionen för Kinect-kameran inte tillåter igenkänning av alltför små, finkänsliga, gester.

Vi drar också slutsatsen, utifrån resultaten erhållna från fokusgrupp och prototyptester, att en framtida produkt bör implementera igenkänning av kontinuerliga gester, som till exempel att rita symboler i luften. Musikleksaken får gärna efterlikna ett spel på så sätt att detta har inslag av tydliga mål och kanske även möjligheten att samarbeta mellan flera personer sam-

tidigt. Våra resultat bör, med viss försiktighet, kunna generaliseras till den undersökta gruppen, men inte nödvändigtvis till ungdomar och/eller teknikintresserade i allmänhet.

5.3 Framtida forskning och tillämpningsområden

Det finns många olika situationer och sammanhang där en Kinect-kamera kan användas för att reglera eller producera musik. Vår prototyp skulle kunna utvecklas vidare till att fungera som en avancerad musikapplikation för elektronisk musik. Om prototypen vidareutvecklas så att den introducerar mindre fördröjning skulle man kunna använda systemet för att spela trumljud eller melodier i realtid, mer likt ett riktigt instrument. Man kan tänka sig att artister och diskjockeys skulle kunna ersätta något eller några av de kontrollorgan som de använder vid liveuppträdanden med ett Kinect-gränssnitt för att tydligare visa för publiken hur de kontrollerar musiken. Tekniken skulle också kunna användas för utveckling av leksaker för barn eller i pedagogiska syften.

Referenser

- Bell, J. (2000), *Introduktion till forskningsmetodik*, Studentlitteratur.
- Berg, T., Chattopadhyay, D., Schedel, M. & Vallier, T. (2012), 'Interactive music: Human motion initiated music generation using skeletal tracking by kinect', *SEAMUS*.
- Bevilacqua, F. (2002), 3d motion capture data: Motion analysis and mapping to music, *in* 'In Quentin Stout and Michael Wolfe, editors, The Sixth Distributed Memory Computing Conference', IEEE, IEEE Computer Society Press, pp. 562–569.
- Dyer, S., Martin, J. & Zulauf, J. (1995), 'Motion capture white paper'.
URL: ftp://ftp.sgi.com/sgi/A|W/jam/mocap/MoCapW_Pv2.0.html
- Embodied Sensor Technology (2011), 'Microsoft Kinect'. Bild tagen från:
URL: <http://embodied.waag.org/?p=38>
- Frisson, C., Dupont, S., Leroy, J., Moinet, A., Ravet, T. & Siebert, X. (2011), Loopjam: a collaborative musical map on the dance floor, *in* T. Dutoit, ed., 'QPSR of the numediart research program', Vol. 4, numediart Research Program on Digital Art Technologies, pp. 37–40.
URL: http://www.numediart.org/docs/numediart_2011_s14_p4_report.pdf
- Furniss, M. (1999), 'Motion capture', *MIT Communications Forum*.
URL: http://web.mit.edu/comm_forum/papers/furniss.html
- Godøy, R. I., Haga, E. & Jensenius, A. R. (2005), Playing air instruments: Mimicry of sound-producing gestures by novices and experts, *in* 'Gesture Workshop', pp. 256–267.
- Godøy, R. & Leman, M. (2009), *Musical Gestures: Sound, Movement and Meaning*, Taylor & Francis.
- Gulliksen, J. & Göransson, B. (2002), *Användarcentrerad systemdesign*, Studentlitteratur.
- Henze, N., Löcken, A., Boll, S., Hesselmann, T. & Pielot, M. (2010), Free-hand gestures for music playback: deriving gestures with a user-centred process, *in* 'Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia', MUM '10, ACM, pp. 16:1–16:10.
URL: <http://doi.acm.org/10.1145/1899475.1899491>

- Jensenius, A. R. (2007), ‘Action-sound: Developing methods and tools to study music-related body movement’.
URL: <http://folk.uio.no/alexanje/phd/jenseniusphd.pdf>
- Kaltenbrunner, M. (2010), ‘Kinect theremin’. Bild tagen från:
URL: <https://code.google.com/p/therenect/>
- Kinect Aku (2012a), ‘Kinect fun labs air band’.
URL: http://www.kinectaku.com/games/xbla/kinect_fun_labs_air_band
- Kinect Aku (2012b), ‘Kinect fun labs musical feet’.
URL: http://www.kinectaku.com/games/xbla/kinect_fun_labs_musical_feet
- KORG (2012a), ‘The Korg KP3 Effects Sampler’.
URL: <http://korg.com/kp3>
- KORG Electronics (2012b), ‘Kaossilator dynamic phrase synthesizer’.
URL: <http://korg.com/KAOSSILATOR>
- Kvifte, T. & Jensenius, A. (2006), ‘Towards a coherent terminology and model of instrument description and design’, *New Interfaces for Musical Expression Conference 2006* .
- Leonard, J. & Ng, K. (2011), ‘Music via motion: A distributed framework for interactive multimedia performance’.
URL: <http://www.ksi.edu/seke/dms11/DMS/51Ngmvm-dms2011-sub20110628.pdf>
- Microsoft (2012), ‘Kinect for windows’.
URL: <http://www.microsoft.com/en-us/kinectforwindows/>
- Mäki-Patola, T. & Hämäläinen, P. (2004), Latency tolerance for gesture controlled continuous sound instrument without tactile feedback, in ‘Proc. International Computer Music Conference (ICMC)’, pp. 1–5.
- Pakarinen, J., Puputti, T., & Valimäki, V. (2008), ‘Virtual slide guitar’, *Computer Music Journal* **32**(3), 42–54.
- Preece, J., Rogers, Y. & Sharp, H. (2002), *Interaction Design*, 1st edn, John Wiley & Sons, Inc., New York, NY, USA.
- Robson, C. (2002), *Real World Research*, Malden, MA: Blackwell Publishing.
- Vik, C. (2012), ‘Kinectar performance platform’.
URL: <http://kinectar.org/>

Wanderley, M. M. (2000), ‘Gestural control of music’.

Wikipedia (2012a), ‘Motion capture’.

URL: http://en.wikipedia.org/wiki/Motion_capture

Wikipedia (2012b), ‘Kinect’.

URL: <http://en.wikipedia.org/wiki/Kinect>

Yoo, M.-J., Beak, J.-W. & Lee, I.-K. (2011), Creating musical expression using kinect, *in* ‘Proceedings of the International Conference on New Interfaces for Musical Expression’, pp. 324–325.

URL: <http://www.nime2011.org/proceedings/papers/H03-Yoo.pdf>

Yosha, O. (2012), ‘Air piano’. Bild tagen från:

URL: <http://www.airpiano.de/>

