Examensarbete

HÖGSKOLAN I SKÖVDE

DIGITAL RÖSTMANIPULATION

En undersökning om digital pitchförändring och hur det påverkar upplevelsen av en röst.

DIGITAL VOICE MANIPULATION

A study on digital pitch change and how it affects the experience of a voice.

Examensarbete inom huvudområdet Medier, estetik och berättande
Grundnivå 30 högskolepoäng
Vårtermin 2020

Karl Boustedt
Fredrik Jansson

Handledare: Jamie Fawcus
Examinator: Anders Sjölin

Sammanfattning

Denna studie undersöker hur digital pitchförändring påverkar upplevelsen av en röst. En monoton röst har spelats in och olika pitch-versioner av rösten har skapats. Bakgrundskapitlet tar upp områden som psykologi, kontext, röst och demografi m.fl. För att få svar på frågeställningen har en undersökning genomförts. Ett digitalt spel skapades med syfte att undersöka deltagarnas upplevelser. I spelet stöter deltagaren på fem olika karaktärer vars tal endast består av vokaler. Varje karaktärs röst har pitchförändrats på olika sätt och deltagarna ombads anteckna sina tankar för vardera karaktär. Undersökningen hade 8 deltagare som efter spelsessionen deltog i en intervju. Svaren från deltagarna visar på att den pitchförändring som gjort påverkat deltagarnas upplevelser. Det finns vissa likheter mellan svaren men även en del skillnader. Resultatet visar på att pitchhöjning verkar vara ett effektivare sätt att förmedla en känsla än både vid oförändrad pitch samt vid pitchsänkning.

Nyckelord: röst, pitchförändring, upplevelse, uppfattning, digital.

Innehållsförteckning

Innehåll

1 Introduktion

Röster har en vital roll i hur kommunikation sker mellan människor. Det kan därför vara intressant att få en fördjupad kunskap om hur röster uppfattas av olika människor. Den forskning som finns idag visar på att många faktorer har en påverkan på hur en röst uppfattas. Flertalet tillvägagångssätt kan användas, exempelvis kan röster analyseras via teknisk data, spelats upp i olika kvalité, olika typer av inspelningar/inspelningstekniker etcetera. Ett tillvägagångssätt som flertalet tidigare studier använt är att spela in röstskådespelare som försöker gestalta olika sinnesstämningar, för att därefter se hur dessa inspelningar uppfattas. Det tillvägagångssätt som denna studie använder är att istället nyttja monotont inspelade vokaler som grund. Dessa inspelningar används sedan för att skapa flera versioner som pitchförändras på olika sätt för att därefter undersöka om uppfattningen förändras mellan de olika versionerna.

Det område som denna studie ämnar att undersöka är hur digital pitchförändring påverkar upplevelsen av en röst. Denna studie ämnar alltså inte att undersöka hur pitch i tal påverkar upplevelsen utan siktar helt och hållet in sig på hur en digital pitchförändring som gjorts i efterhand, efter inspelningstillfället, påverkar upplevelsen av en röst. Den frågeställning som denna studie försöker besvara är “Hur påverkar digital manipulation i form av pitchförändring upplevelsen av en röst i en tredimensionell spelmiljö från ett förstapersonsperspektiv?”.

För att få svar på frågeställningen genomfördes en kvalitativ undersökning i form av ett speltest samt en efterföljande intervju. Eftersom undersökningsområdet behandlar känslor och upplevelser ansågs det vara viktigt att låta deltagarna förklara och utveckla sina svar för att få en tydligare bild av deras upplevelser. Därefter jämförs deltagarnas svar för att hitta likheter och skillnader mellan svaren.

För att undersöka frågeställningen har en artefakt skapats, i form av ett digitalt spel, där deltagarna stöter på fem olika karaktärer som spelar upp stavelser från olika vokaler i en slumpmässig följd. Dessa vokaler har pitchförändrats på olika sätt för vardera karaktär och deltagaren uppmanas att skriva ner sina tankar om vardera karaktär vid testtillfället.

2 Bakgrund

2.1 Psykologi

För att få en grundläggande förståelse för människans förmåga att anpassa sig i olika situationer krävs en del bakgrundsforskning inom psykologi. Det kan vara viktigt att vara medveten om hur människors upplevelser förändras genom exempelvis, exponering av manipulerade ljud, hur upplevelsen av ett ljud förändras över tid, hur situationen som ljudet spelas upp i spelar roll etcetera.

Alexi Grinbaum (2015) beskriver hur människan har en förmåga att anpassa sig i olika sociala sammanhang. Grinbaum tar upp exempel på detta där interaktion med barn samt personer med funktionsnedsättning är ett av de tillfällen där social anpassning ofta sker för att göra det sociala mötet mer trivsamt.

Ett till exempel på hur anpassning kan ske kommer från en undersökning gjord av Patricia E.G. Bestelmeyer, Julien Rouger, Lisa M. DeBruine och Pascal Belin (2010). Resultatet av undersökningen visade på att deltagarnas uppfattning förändrats under testets gång vilket visar på att människan anpassar sitt lyssnande. Deltagarnas uppfattning av original-ljudet förändrades från första uppspelningen till andra uppspelningen under testet.

Vad som är viktigt att förstå är att människan har en förmåga att anpassa sitt lyssnande och att en testdeltagares uppfattning av ett ljud kan förändras över tid. Det är därför viktigt att utforma en undersökning med detta i åtanke vid bedrift av studier rörande människors uppfattning.

2.2 Kontext

Hur ett ljud uppfattas påverkas till stor del av vilken kontext det presenteras i. Exempelvis kan ett ljud som spelas upp ihop med presentation av en bild (Cox 2008b) eller ett ord i slutet av en mening (Roring, Hines & Charness 2007) uppfattas på ett annat sätt än om de skulle presenteras separat utan kontext. I en studie undersöker Trevor J. Cox (2008b) hur visuellt stimuli påverkar “hemska” ljud. En undersökning gjordes där deltagarna lyssnade på ett antal hemska/obehagliga ljud som antingen presenterades med ett associerat visuellt stimuli, ett icke-associerat visuellt stimuli eller en blank grön ruta. Associerat visuellt stimuli kunde exempelvis vara en bild på en person som biter i ett äpple tillsammans med ett ljud av ett äpple som äts, medan icke-associerat visuellt stimuli för samma ljud var en bild på pärlor i olika färger.

Slutsatsen av Cox (2008b) studie är att visuellt stimuli påverkar ljudupplevelsen, men hur det påverkar beror på vilket ljud och vilken bild som används. Exempelvis menar Cox (2008b) att ljudet

av en tandläkarborr förstärks av en bild på en tandläkare medan ljudet av en fiol som spelas dåligt förmildras av en bild på fiolen. Detta beror på relationen man har till den associerade bilden. De flesta har inte någon bra relation till tandläkare vilket förstärker den negativa upplevelsen av ljudet. Däremot i fiolens fall kan den låta hemsk utan någon bild när det är osäkert vad den kan relateras till men när bilden på en fiol finns där kan det konstateras att ljudet kommer från fiolen, vilket mildrar upplevelsen av ljudet.

Roring et al. (2007) undersökning visar att kontexten av att placera ett ord i slutet av en mening gör det enklare att identifiera ordet än om det presenteras separat. Detta gäller framför allt för äldre personer och Roring et al. (2007) resonerar att detta troligtvis beror på deras nedsatta hörsel. Dock skulle denna information kunna vara användbar även för personer med fullgod hörsel. Exempelvis vid tillfällen där bakgrundsbrus eller andra dåliga uppspelningsförhållanden skapar svårigheter att höra det uppspelade materialet.

Uppspelningen av ett ljud kan påverka hur det upplevs och tolkas. Även kvalité på ett ljud kan påverka upplevelsen. Mark Grimshaw (2009) beskriver hur ett ljud som har lägre kvalité än det visuella material som hör ihop med ljudet kan skapa obehagskänslor hos lyssnaren. Han nämner även att överdriven artikulation av en röst kan upplevas som obehaglig. Grimshaw (2009) tar upp ett exempel där munnen på en människolik karaktär rör sig ovanligt mycket men rösten är oförändrad.

Ytterligare en slutsats Grimshaw (2009) kom fram till är att kontexten även är mycket viktig för att en känsla av obehag ska uppstå hos lyssnaren. I vilket sammanhang uppspelningen av ett ljud sker har en stor betydelse för hur detta ljud kommer att uppfattas av lyssnaren. Han påstår att detta är sant både när det kommer till trovärdighet samt vilken sinnesstämning ljudet förmedlar till lyssnaren. Exempelvis kan kontexten spela stor roll huruvida ett ljud upplevs som komiskt eller skrämmande beroende på den värld som ljudet spelas upp i. Grimshaw (2009) tar upp ett exempel på detta och nämner dubbning av filmer som någonting som kan uppfattas som löjligt, fånigt eller till och med roligt, men vid andra tillfällen istället upplevas som väldigt obehagligt för lyssnaren.

Slutligen nämner Grimshaw (2009) att ett ljud som är familjärt för lyssnaren kan väcka obehagskänslor om detta ljud manipuleras. Däremot beror det på hur ljudet manipulerats samt var lyssnaren stöter på ljudet. Han tar upp ett exempel på detta i den japanska filmen *Ringu* (1998) där en telefonsignal introduceras tidigt i filmen med ett vanligt ringande. Senare i filmen manipuleras signalen vilket skapar obehagskänslor hos åskådarna. Här spelar kontexten stor roll igen för att en obehagskänsla ska uppstå.

Kontext verkar ha en stor påverkan på hur ett ljud uppfattas och kontext kan användas för att förmedla sinnesstämningar till åskådaren. Det kan vara viktigt att ha en förståelse för hur kontexten påverkar det vi hör, speciellt vid skapandet av en komisk eller skräckprodukt. Det kan vara viktigt vid

skapandet av en prototyp att ha kontext i åtanke och aktivt välja om någon kontext ska användas eller inte.

2.3 Röst och Demografi

Flertalet studier har undersökt hur förändring av röster, läten, vokaliseringar eller meningar påverkar det upplevda budskapet. I en undersökning gjord av César F. Lima, São L. Castro och Sophie K. Scott (2013) har en slutsats nåtts som visar att det är möjligt för människor att tolka olika sinnesstämningar och känslor genom enklare vokaliseringar. Deltagarna fick lyssna på 121 olika vokaliseringar och därefter kategorisera vilken sinnesstämning dessa vokaliseringar tillhörde. Ett intressant resultat som Lima et al (2013) fann var att kvinnor hade lättare för att urskilja sinnesstämningar från röster än män.

En liknande studie har genomförts där det undersöks om icke-verbal kommunikation uppfattas på samma känslomässiga plan över olika kulturer (Gendron, Roberson, Marieta van der Vyver, Feldman Barrett 2014). Undersökningen visade på att personer från USA hade lättare att urskilja känslomässiga sinnesstämningar än testpersoner från Namibia.

Ytterligare en artikel som tar upp skillnader på hur personer från olika kulturer uppfattar betoningar och röster är skriven av Yanhong Zhang och Alexander Francis (2010). Zhang och Francis undersöker hur personer med engelska som modersmål, samt personer med kinesisk mandarin som modersmål, uppfattar vikten av betoning på vokaler. Studien visar att de deltagare med mandarin som modersmål har en förmåga att tyda betoning på vokaler, dock inte lika bra som deltagare med engelska som modersmål. Detta kan vara en brist i praktiskt användande och inte i uppfattning, men Zhang och Francis (2010) anser att det behövs mer forskning inom området för att tydligt kunna säkerställa detta.

Liknande skillnader mellan kulturer går att hitta i Mark Grimshaws artikel *The audio Uncanny Valley: Sound, fear and the horror game* (2009) där han nämner att frekvens kan ha en påverkan på hur obehagligt ett ljud upplevs. Han skriver även att ljud och melodiska stycken som inte följer en struktur eller ett mönster, upplevs som obehagligare än melodiska ljud och stycken. Detta gäller inom västerländsk kultur, däremot är det oklart om detsamma gäller för kulturer som inte är västerländska.

Det finns även andra faktorer som påverkar lyssnandet hos människor, bland annat ålder. I en undersökning om uppfattning av syntetisk röst visar Roy W. Roring, Franklin G. Hines och Neil Charness (2007) att äldre har svårare än yngre personer att uppfatta ord som läses upp av en syntetiskt skapad röst. Undersökningen visade även att vid användandet av syntetisk röst var det svårare för deltagarna att identifiera orden vid långsammare uppläsningshastighet. Detta gällde vid alla åldersgrupper. Roring et al. (2007) resonerar att detta kan bero på att den syntetiska rösten inte

innehåller lika tydlig prosodi (ljudläran som beskriver accentuering och längdförhållanden, (Nationalencyklopedin 2020)), vilket blir extra tydligt vid långsam uppläsning och skapar otydligheter och svårigheter för identifikation.

2.4 Röst, Ljud och Uppfattning

I både Lima et al. (2013) samt Gendron et al. (2014) undersökningar visar det att deltagarna har en förmåga att läsa av vilken sinnesstämning den uppspelade rösten försökt förmedla. Däremot visar dock Gendron et al. (2014) undersökning att deltagarna från Namibia haft större svårigheter än deltagarna från USA. Dock hade deltagarna från Namibia förmågan att tolka hur stark en viss känsla var (exempelvis om det var en mycket stark känsla eller en mer diskret känsla).

En mer teknisk undersökning har gjorts av Rainer Banse och Klaus R. Scherer (1996) där olika meningar spelats in och därefter har deltagarna fått svara på frågor angående deras uppfattning av meningarna. Inspelningarna genomgick även en teknisk analys där amplitud, frekvensinnehåll, formantfrekvenser samt uppläsningshastighet analyserats och dokumenterats. Därefter jämfördes den tekniska analysen och de mänskliga bedömningarna för att hitta likheter och skillnader. Resultatet visar på en likhet mellan de mänskliga bedömningarna och de statistiska analyserna. Detta visar att det kan vara möjligt att på förhand lista ut vilken typ av känsla ett ljud kommer ge genom att enbart tekniskt analysera ljudet.

I samma studie av Banse och Scherer (1996) visar resultatet att deltagarna haft en hög förmåga att kategorisera in de olika inspelningarna i olika sinnesstämningar. Studien visar även att de felsvar som angivits har legat nära det korrekta svaret vilket styrker antagandet att deltagarna lyckats identifiera sinnesstämningarna. Exempelvis gjordes en del förväxlingar mellan förtryckt ilska och okontrollerbar ilska. Intresse var även oftare förväxlat med stolthet och glädje än med de resterande känslorna. Detta visar på att deltagarna tolkat rätt grundkänsla, dock har inte alla deltagare kategoriserat känslorna på samma sätt.

I en studie gjord av Valentina V. Lublinskaja, Jaan Ross och Elena V. Ogorodnikova (2006) undersöks hur personer reagerar på digitalt manipulerade ljud. I studien ombads deltagarna att identifiera olika typer av syntetiserade, tal-liknande ljud, dock inte verkligt tal, för att sedan kategorisera vad de ansåg att de hört. Deltagarna skulle kategorisera de olika ljuden som vokaler, konsonanter, konsonant följt av en vokal eller en stavelse. Olika typer av modulation gjordes på ljuden och använde sig bland annat av AM (amplitude modulation) med hjälp av ett amplitude envelope för att förändra ljudet. Resultatet av studien visar att förändringar på ett ljuds amplitud förändrar uppfattningen av ljudet, speciellt vid icke-vokaliserade ljud och läten.

Detta visar att även ljud som är syntetiskt processade kan förmedla information till lyssnaren och människor har en förmåga att kunna tolka digitala ljud, speciellt vid förändring av ljudets amplitud. Detta kan även kopplas ihop med de artiklar under denna underrubrik och visar på att människan frekvent tolkar ljud, oavsett om det är röster, läten eller annat som skulle kunna vara mänskligt.

2.5 Metod och Tillvägagångssätt

Det finns olika tillvägagångssätt att bedriva en studie. Ett tillvägagångssätt är att genomföra en kvantitativ studie som använder en stor deltagarmängd. Detta för att dra slutsatser genom införskaffande av många svar på frågeställningen. En kvantitativ studie använder numerisk data. Ett annat tillvägagångssätt är att bedriva en kvalitativ studie som istället insamlar mer utvecklade svar från en mindre deltagarmängd. Denna metod använder data som är mer åsikts- och upplevelsebaserad.

Cox (2008a) genomför en kvantitativ internetbaserad studie där han jämför olika obehagliga ljud med varandra. Undersökningen hade ett högt deltagarantal och detta är en av fördelarna med en internetbaserad undersökning. Dock hävdar Cox (2008a) att det finns nackdelar med undersökningar som utförs över internet. Resultatet blir inte lika pålitligt som det resultat man får av ett experiment som utförs i en kontrollerad lab-miljö. Detta på grund av att det inte går att säkerställa hur lyssningen genomförs, om tillräcklig ljudnivå används eller om testpersonen förstått instruktionerna för testet.

César F. Lima, São L. Castro och Sophie K. Scott (2013) och Gendron, Roberson, Marieta van der Vyver, Feldman Barrett (2014) använder sig av en kvalitativ metod vilket resulterar i mer data att analysera med mer utvecklade svar från testdeltagarna.

Genom att analysera de olika metoder som tidigare undersökningar använt skapas en tydligare bild för vilka fördelar och nackdelar som var metod har. För denna undersökning har en kvalitativ metod vid datainsamlingen valts. Detta för att låta deltagarna ge mer utvecklade svar för att tydligare beskriva sin upplevelse.

3 Problemformulering

Då dagens röstmanipultionsteknik konstant utvecklas kan det vara intressant att undersöka hur denna teknik påverkar upplevelsen av röstinspelningar. Detta kan vara värt att undersöka för att få en fördjupad bild på vilket sätt denna teknik påverkar det lyssnaren hör.

För att ta reda på detta genomfördes en undersökning för att fördjupa kunskapen inom just digital manipulation av röster. Frågeställningen som denna studie ämnar att undersöka är "Hur påverkar digital manipulation i form av pitchförändring upplevelsen av en röst i en tredimensionell spelmiljö från ett förstapersonsperspektiv?".

3.1 Metodbeskrivning

För att få svar på frågeställningen skapades en artefakt. Artefakten är ett enkelt spel byggt i Unity Engine (2019). Spelet är ett 3D-spel i förstapersonsperspektiv. Spelet innehåller fem karaktärer som pratar ett påhittat låtsasspråk. Karaktärernas röster har samma ord-innehåll men är bearbetade på olika sätt i form av pitch-förändringar på vokalerna. Miljön och karaktärerna i spelvärlden är neutrala, enkelfärgade former och ytor. Detta för att minimera den påverkan som det visuella har på deltagarnas uppfattning av rösterna. Även ambient ljud och musik har valts att inte användas av samma anledning.

Testdeltagare genomförde en kort spelsession och lyssnade på de olika karaktärerna. Banan i spelet är utformad som en lång korridor där deltagaren stötte på vardera karaktär i en förutbestämd ordning. Detta gjordes för att det skulle vara enklare att sammanställa resultatet efter den genomförda studien samt hjälpa deltagaren att hålla reda på karaktärerna. Deltagaren ombads att skriva ner sina tankar på valfritt sätt under testets gång. Testet var ej tidsbaserat. Efteråt genomfördes en kort intervju med varje deltagare för att få en bättre uppfattning av deras upplevelse. Intervjun innehöll både utvecklande längre frågor men även kortsvarsfrågor för att få en snabb överblick över varje deltagares upplevelse. Därefter sammanställdes svaren för att få ett svar på frågeställningen. Eftersom det var uppfattning och upplevelse som undersöktes valdes främst insamling av kvalitativ data i form av intervjuer.

3.2 Datainsamling

Den datainsamlingsteknik som användes vid undersökningstillfället var en spelsession med efterföljande intervju. Valet att använda intervjuer som insamlingsmetod gjordes baserat på den frågeställning som valts att undersökas. Då undersökningen ämnat att ta reda på hur deltagarna

upplevt spelsessionen är det därför viktigt att låta deltagarna lämna mer utvecklade och beskrivande svar av sin upplevelse. Detta valdes för att kunna skapa en så tydlig uppfattning som möjligt.

Undersökningen var först planerad att utföras på Högskolan i Skövde i en av de studior som finns tillgängliga genom utbildningen. Detta valdes för att dessa lokaler ger både möjlighet till bokning samt avskildhet. Lokalerna är belägna på skolan vilket skulle göra det lättare att få tillgång till testpersoner. På grund av utomstående förhinder tvingades studien istället att genomföras på distans.

Efter samtal under handledning, samt tidigare kurser, gjordes valet att undersökningen behövde innehålla minst sex testdeltagare för att kunna dra en slutsats av undersökningens resultat. Då tanken var att sökandet efter testdeltagare främst skulle ske på högskolan skulle demografin troligtvis bestå främst av högskoleelever. Eftersom undersökningen istället genomfördes på distans över internet såg demografin aningen annorlunda ut. Ingen preferens på kön togs, men undersökningen strävade efter att ha lika många kvinnliga som manliga deltagare. Undersökningen (spelsession samt intervju) tog ca 30 minuter per person att genomföra.

Till den efterkommande intervjun ställdes frågor för att få svar på den frågeställning som denna undersökning försöker besvara. Frågorna var inte ja/nej-frågor utan var utformade på ett sätt som får deltagaren att svara mer utvecklat. Frågor som ställdes efter testsessionen såg ut på följande sätt *"Hur upplevde du den första karaktären?"*, *"Uppfattade du någon av rösterna som vänligare? Om ja, vilken?"*. För att se alla frågor, se Appendix A.

Efter undersökningen och sammanställningen av deltagarnas svar på de intervjufrågor som ställdes jämfördes svaren med varandra för att se om det fanns tydliga likheter eller skillnader mellan de angivna svaren. Denna data lade sedan grunden för den slutsats som denna studie drar i samspel med tidigare forskningsresultat.

3.3 Etik- och forskningsaspekter

Denna studie följer vetenskapsrådets forskningsetiska principer (Vetenskapsrådet 2002) och använder de fyra grundkraven som bör tas hänsyn till vid bedrivande av all typ av forskning. De fyra grundkraven är:

Informationskravet - *Forskaren skall informera de av forskningen berörda om den aktuella forskningsuppgiftens syfte.*

Samtyckeskravet - *Deltagare i en undersökning har rätt att själva bestämma över sin medverkan.*

Konfidentialitetskravet - *Uppgifter om alla i en undersökning ingående personer skall ges största möjliga konfidentialitet och personuppgifterna skall förvaras på ett sådant sätt att obehöriga inte kan ta del av dem.*

Nyttjandekravet - *Uppgifter insamlade om enskilda personer får endast användas för forskningsändamål.*

4 Genomförande

För att få ett svar på frågeställningen har en artefakt skapats. Utgångspunkten, samt artefaktens funktion, har förändrats flertalet gånger under utvecklingen av denna studie. Flertalet versioner av artefakten har skapats och när studiens frågeställning ändrats eller tagit en ny riktning, har även artefakten behövts förändras därefter.

Vid starten av denna studie var grundtanken att undersöka ett helt annat område än det område som slutligen valts. Den första frågeställningen som denna studie ämnade att undersöka var huruvida efterarbete av inspelade röster kunde påverka hur obehagligt dessa röster uppfattades.

För att undersöka detta skapades tio olika versioner av en röstinspelning. De olika versionerna hade olika mängd efterarbete, från en effekt upp till nio effekter, för att kunna mäta var någonstans rösten upplevdes som mest obehaglig. Tillvägagångssättet för att skapa denna artefakt var att varje ny version skulle använda exakt samma effektkedja som den tidigare versionen. Första versionen var originalinspelningen, andra versionen var en pitchsänkt version av originalinspelningen, tredje versionen hade exakt lika mycket pitchsänkning samt saturation, fjärde versionen hade exakt samma effektkedja som version tre fast ytterligare ett reverb. Detta fortsatte upp till version 10. Syftet var att kunna mäta hur mycket efterarbete/effekter som krävdes för att nå ett så obehagligt resultat som möjligt. Utöver detta var syftet att se om det vid något tillfälle blev för mycket efterarbete och om upplevelsen ändrades på grund av detta.

Efter både handledning och diskussion inom gruppen togs valet att inte använda den frågeställning som var aktuell i detta läge. Beslutet togs dels på grund av att området kändes för smalt att undersöka, samt att det var väldigt svårt att hitta tidigare forskning inom området. Ytterligare ett problem var hur datainsamlingen från artefakten skulle sammanställas. Artefakten hade väldigt många olika variabler som kunnat påverka spelarens uppfattning och därför skulle det vara svårt att veta exakt vad det var som påverkade deltagarna.

Efter detta tog studien en ny inriktning och rörde sig iväg från obehag och valde istället att undersöka hur efterarbete av en röstinspelning kunde påverka deltagarnas uppfattning av inspelningen. Det nya målet med artefakten var att, genom efterarbete, få en röstinspelning att låta som motsatt kön från den originalinspelning som manipulerats (manlig röst till kvinnlig och kvinnlig röst till manlig). Syftet var att utforska om det var möjligt att förändra rösterna med hjälp av efterproduktion och göra det svårt för lyssnaren att veta om det var en naturlig kvinnlig/manlig röst eller inte.

Inspelningarna var från en radioteater där replikerna var inspelade av personer utan tidigare röstskådespelarerfarenhet. Därefter förändrades dessa röster med hjälp av efterarbete, både på pitch

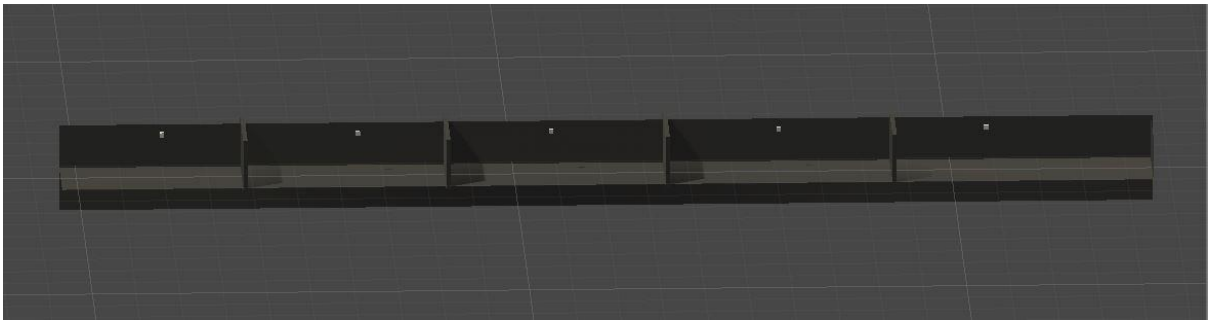
och formanter. Resultatet redovisades därefter vid en presentation under ett kurstillfälle där en del kritik och åsikter kom fram angående rösternas kvalité.

Åter igen efter både handledningstillfällen samt kritik och återkoppling från klassen togs beslutet att detta var för brett och för många variabler påverkade resultatet. För många variabler användes och detta gjorde det svårt att dra slutsatser. En åsikt inom gruppen var även att de redigerade inspelningarna inte riktigt höll måttet och hade brister i kvalité, vilket testdeltagarna troligtvis skulle påpeka efter testtillfället.

Själva grunden i frågeställningen förändrades inte drastiskt denna gång utan fokuset låg fortfarande på hur deltagarna uppfattar en mening. Däremot gjordes artefakten om ytterligare en gång. Denna gång använde sig artefakten av tre inspelningar där meningen "Jag tänkte att vi skulle åka på semester" lästes upp på olika sätt. Meningen spelades in av en person utan röstskådespelarerfarenhet som försökte gestalta en monoton, en glad och en arg sinnesstämning. Därefter användes en effektkedja likt den första versionen, fast denna gång med olika grundinspelningar. Effektkedjan steg gradvis och syftet var att se huruvida olika mängd effekter påverkade sinnesstämningarna på olika sätt.

Denna versionen led åter igen av samma problem som tidigare versioner där för många variabler påverkade resultatet. Den återkoppling denna version av artefakten fick visade på att hela idén var bristfällig och därför valdes en drastisk förändring att göras. Nu låg fokus på att göra en enklare och mer genomförbar artefakt.

Denna gång gjordes valet att inte använda en fullständig mening utan att istället endast applicera pitchförändringar på monotont inspelade uppläsningar bokstäver och siffror. Frågeställningen vid detta tillfälle är den nuvarande frågeställningen. Nu började artefakten närma sig den version som denna studie i dagsläget använder. Dock använde denna version mycket tydligare stavelser och hela ord (siffror) än den slutgiltiga versionen. Utöver detta togs även valet att skapa ett litet spel där deltagaren får stöta på olika karaktärer i en korridor (se figur 1) och föra anteckningar över sina tankar under spelsessionen.

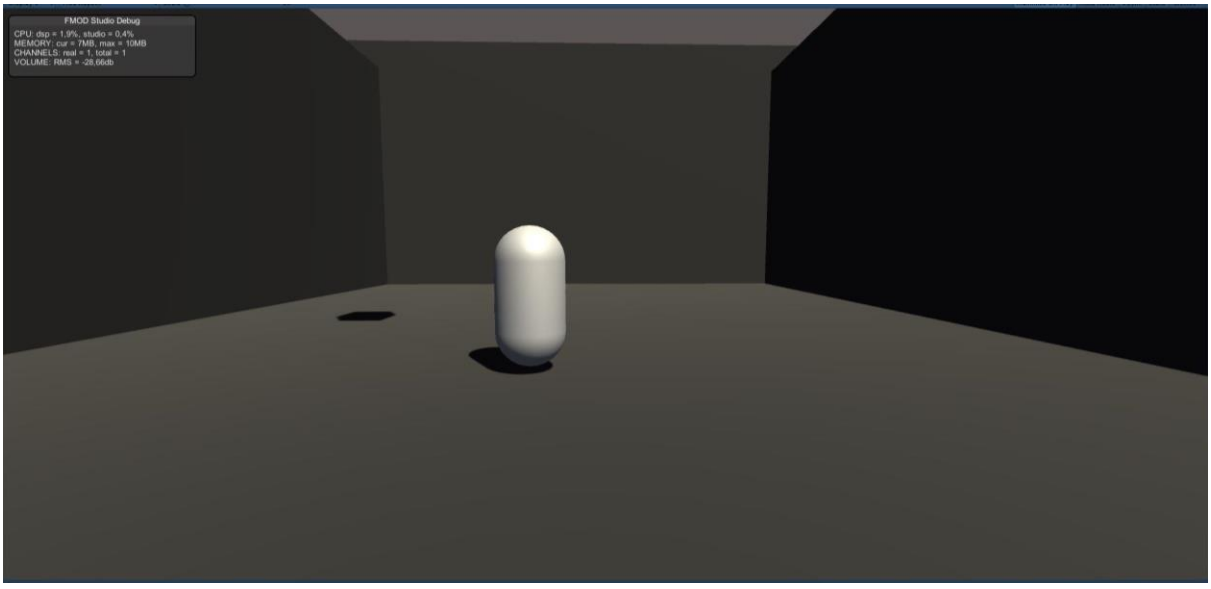


Figur 1. Artefaktens bandesign sett från sidan.

Ett problem som uppstod denna gång var att inspelningarna förmedlade en känsla då de var inspelade av en människa och inte en robot. Det skulle därför vara svårt att klassa dessa inspelningar som neutrala. Ett alternativ som diskuterades var huruvida användandet av en “text-to-speech”-funktion skulle användas eller inte. Ett av de problem som diskuterades internt inom gruppen var att “text-to-speech”-tillvägagångssättet skulle betyda att undersökningen inte skulle kunna appliceras på den frågeställning som valts, då denna frågeställning specifikt ville undersöka hur mänskliga röster uppfattades. Det diskuterades även om det var viktigt för studien att inspelningen var neutral eller inte. Det gruppen kom fram till i slutändan var att studien skulle gynnas av en monoton grundinspelning då detta skulle göra analysarbetet mer tydligt.

Gruppen ansåg även att det inte var bra att använda riktiga ord då detta kunna påverka hur testdeltagarna uppfattar ordet. Valet att använda siffror togs därför bort och därefter blev gruppen inspirerad av en annan student som enbart använde sig av vokaler i sin artefakt, vilket är ett bra tillvägagångssätt för att undvika att lyssnaren kopplar bokstäverna till olika ord eller meningar.

Den aktuella versionen använder nu sig endast av vokaler vars pitch antingen höjs eller sänks i början eller slutet av vokalen. Först användes vokaler som betonades på det sätt vokalerna betonas vid uppläsning av alfabetet, det vill säga inte vokalernas stavelser. Detta gjorde att det fanns en pitch i vokalerna redan vilket ledde till att den efterarbetade pitchförändringen inte var det enda som påverkade lyssnaren. För att lösa detta problem spelades långa vokaler in, utdragna a, e, i, o och u:n och sedan klipptes den delen där pitchen var stabil ut och användes som grund. På detta tillvägagångssätt har originalinspelningarna en monoton pitch och det enda som påverkar pitchen är det manuella efterarbetet som gjorts på inspelningarna.



Figur 2. Karaktärerna som deltagaren stöter på i artefakten.

Detta är den slutgiltiga versionen som användes i denna studie. Deltagaren fick gå igenom en kort bana där denne stötte på karaktärer (se figur 2) med olika pitch på de uppspelade vokalerna. Spelaren kunde endast stöta på karaktärerna en gång då det inte var möjligt för deltagaren att gå tillbaka till tidigare karaktärer. Testet genomfördes över internet.

4.1 Artefaktskapande

När tillvägagångssätt satts påbörjades utformningen av en artefakt. Vokalerna a, e, i, o och u spelades in i samma längd och ton. Partier av inspelningarna där ton och intensitet var stabil klipptes sedan ut. Därefter redigerades vokalerna med hjälp av programvaran FL Studio (2020). Det bestämdes att endast pitch skulle ändras i rösterna. Detta gjordes med ett pitchförändringsverktyg, Little Alterboy (2019), och rösterna redigerades till fem olika variationer:

Neutral (oredigerad)

Pitch-sänkning på slutet

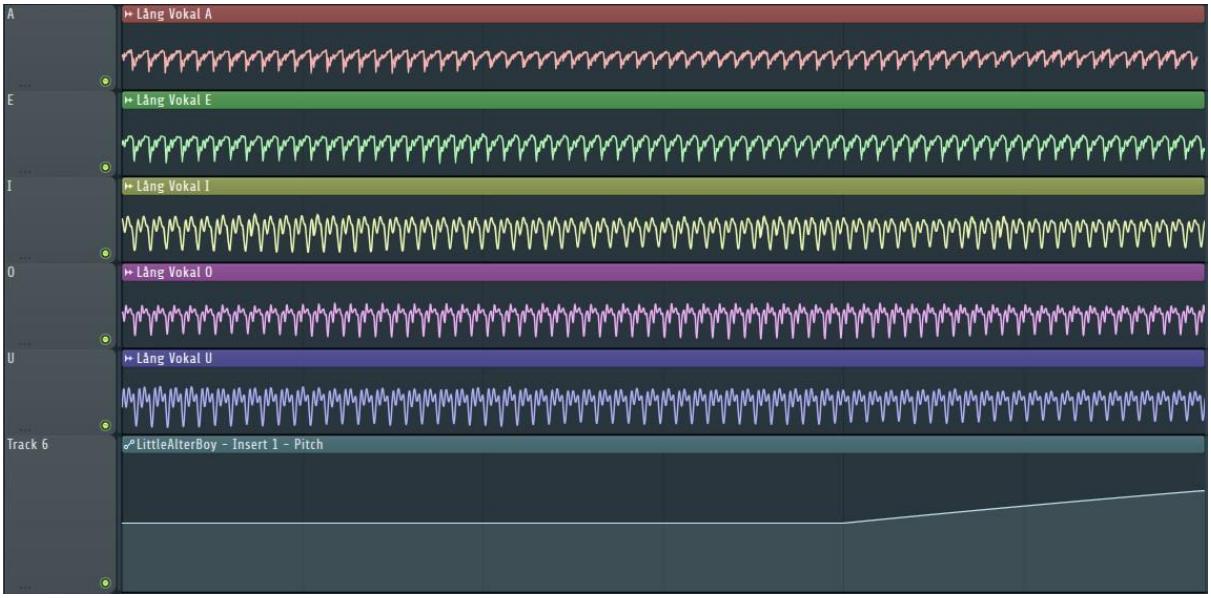
Pitch-höjning på slutet

Pitch-sänkning i början

Pitch-höjning i början

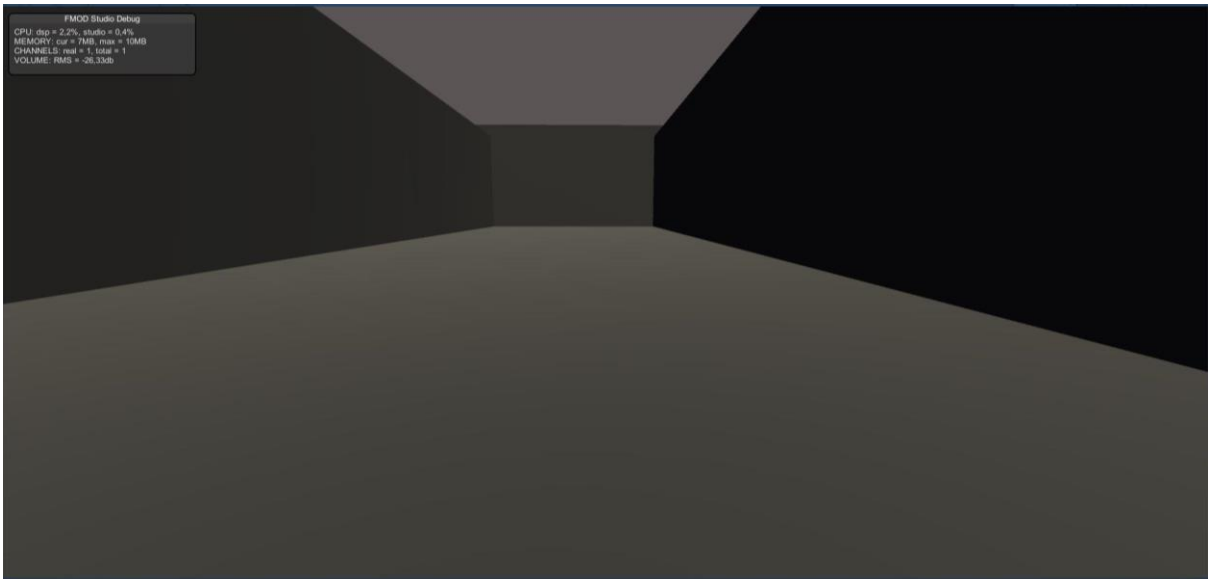
Figur 3. Little Alterboy (2019) med en pitchförändring på 5.6 semitoner.

De pitchförändringar som gjordes var en förändring på 5.6 semitoner (se figur 3) över en 20 millisekundersperiod och gjordes med en automationskurva (se figur 4). Dessa inställningar samt samma automationskurva användes för alla variationer. Både för de variationer med förändringar upp respektive ner i pitch samt för de variationer med förändringar i början respektive slutet av ljuden.



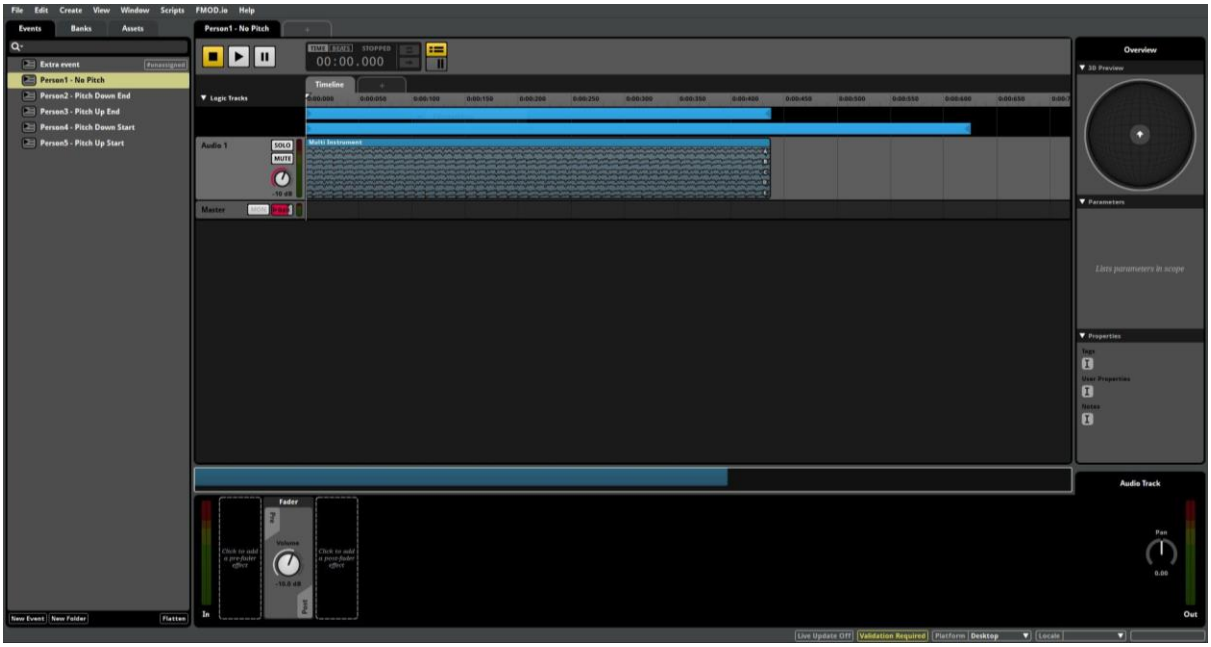
Figur 4. Vokalerna A, E, I, O och U med en automationskurva för pitchhöjning.

För att sedan kunna genomföra en undersökning och testa ljuden skapades en spelprototyp i spelmotorn Unity (2019). Här skapades en tredimensionell miljö utformad som en lång korridor med fem karaktärer (se figur 2) utplacerade en efter en. Tanken här var att testdeltagaren skulle ta sig fram i korridoren och stöta på och lyssna på karaktärernas röst en i taget. Ordningen på karaktärerna var som listat ovan. Karaktärerna programmerades med ett triggersystem vilket aktiverade dem när spelaren närmade sig och avaktiverade dem när spelaren rörde sig ifrån dem. För att korridoren inte skulle kännas för lång och tom implementerades väggar mellan karaktärerna. Även dessa använde ett triggersystem för aktivering och avaktivering (se figur 5). Väggarna förhindrade också spelaren från att gå tillbaka till de tidigare karaktärerna i korridoren. Detta förhindrade även att deltagaren skulle bli förvirrad och gå till samma karaktär flera gånger, då det är brist på visuellt material att navigera efter.



Figur 5. Dörren och karaktären försvinner när deltagaren rör sig framåt.

För att ljuden skulle spelas upp i spelprototypen användes FMOD Studio (2020) för att skapa en icke-linjär uppspelning. I FMOD Studio skapades fem olika event, ett för varje karaktär och variation. För att göra rösterna mer intressanta att lyssna på och för att få dem att likna faktiskt tal användes ett slumpmässigt loop-system i FMOD-eventen (se figur 6). Två loop-regioner med olika stor uppspelningssannolikhet användes för att skapa en slumpmässighet i uppspelningen.



Figur 6. Logiken i FMOD (2020). De två looparna tillåter förändring och tillfälliga stopp i talet för att simulera mer naturligt tal.

Vid första iterationen användes de neutrala rösterna i alla event. Olika vokaler av den neutrala rösten spelades då först upp ett slumpmässigt antal gånger för att sedan spela upp respektive pitchförändrad vokal i slutet. Detta gjordes för att efterlikna en så naturlig mening som möjligt.

Efter en del lyssnande och rådfrågande ändrades dock denna iteration. I den nya iterationen spelades endast en variation av rösterna upp i varje event. Denna iteration var den som användes vid pilotstudien.

4.2 Pilotstudie

När frågeställningen och artefakten var färdigställda genomfördes en pilotundersökning för att se huruvida den artefakt som skapats gav svar som kunde besvara frågeställningen. Grundtanken var att testet skulle genomföras i en av skolans lokaler med en av gruppmedlemmarna närvarande. Detta för att vara tillgänglig för frågor samt för att säkerställa att testet genomfördes på rätt sätt. Genom att utföra testet i en lokal, och att varje deltagare genomförde testet på samma plats, hade flertalet variabler kunnat räknats bort från svarsresultaten. Alla testdeltagare hade då genomfört testet under samma förutsättningar och samma utrustning hade använts. Utöver detta hade även kroppsspråk och spontana reaktioner varit enklare att dokumentera. Därefter skulle en efterföljande intervju ske.

Detta blev dock inte fallet då det, på grund av utomstående faktorer, inte gick att genomföra testet på det sätt som från början var planerat. Beslutet togs då att genomföra testet online istället, och föra kommunikation till testdeltagaren via kommunikationsprogrammet Discord (2020). Detta ledde dock till att kontroll av både utrustning, volym, prestanda på hårdvara och andra faktorer nu blivit variabler i undersökningen. För att minska dessa variabler ombads deltagarna att genomföra testet med hörlurar och behaglig volym.

Testdeltagarna deltog frivilligt och fick hjälp att både ladda ner artefakten och installera den på den dator som testet skulle genomföras på. Därefter följde en kortare beskrivning av vad undersökningen skulle handla om, däremot nämndes inte exakt det som undersöktes. Testdeltagarna blev ombedda att skriva ner sina tankar och idéer under testets gång för att lättare komma ihåg detta efter testets slut. Testdeltagarna fick även information om att deltagandet var frivilligt, att de kunde avsluta testet när de ville samt att information som samlades in inte skulle användas till någonting annat än denna undersökning.

Testdeltagarna genomförde testet och hade möjlighet att ställa frågor under hela sessionens gång. Efter genomfört test gjordes en intervju, även denna via programmet Discord (2020). Testdeltagarna blev informerade om att intervjun spelades in och fick därefter svara på ett kort frågeformulär där ålder, kön och spelvana dokumenterades (se Appendix B). Detta gjordes för att få en större överblick av deltagarnas demografi.

För att ta reda på deltagarnas upplevelse ställdes ett antal intervjufrågor (se Appendix A). Frågorna var formulerade på ett öppet sätt för att få deltagarna att förklara hur de upplevt varje röst, samt för

att ge utrymme till mer utvecklade svar. Deltagarna blev heller inte tvingade att lämna ett svar om de ansåg att de inte upplevt någonting från någon av karaktärerna.

Efter att ha genomfört pilotstudien framkom det att studien går att genomföra på distans via internet. De förändringar som gjordes efter, samt under, genomförandet av pilotstudien rörde främst den efterföljande intervjun. Från början ställdes flertalet upprepade frågor med små förändringar, exempelvis “Vad kände du från karaktär A?”, “Vad kände karaktär A?”. Detta togs bort redan efter första deltagaren på grund av att det blev väldigt svårt för deltagaren att förklara sitt svar på frågorna. Deltagaren svarade naturligt på flera frågor på samma gång vilket lade grunden till beslutet att istället ställa en bredare fråga från början och låta deltagaren tala fritt och därefter smalna av frågorna för mer specifika svar.

En annan förändring som gjordes var hur testet presenterades. Denna förändring gjordes även den efter första testdeltagaren. Från början förklarades syftet med testet innan testet genomfördes vilket kan ha påverkat hur deltagaren uppfattade rösterna. En mindre detaljerad förklaring gavs därefter till testdeltagarna för att inte avslöja för mycket vad testet undersökte. Den information som deltagarna fick i den slutgiltiga undersökningen var att hen kommer stöta på flera karaktärer och därefter anteckna sin upplevelse av karaktärerna (se Appendix C).

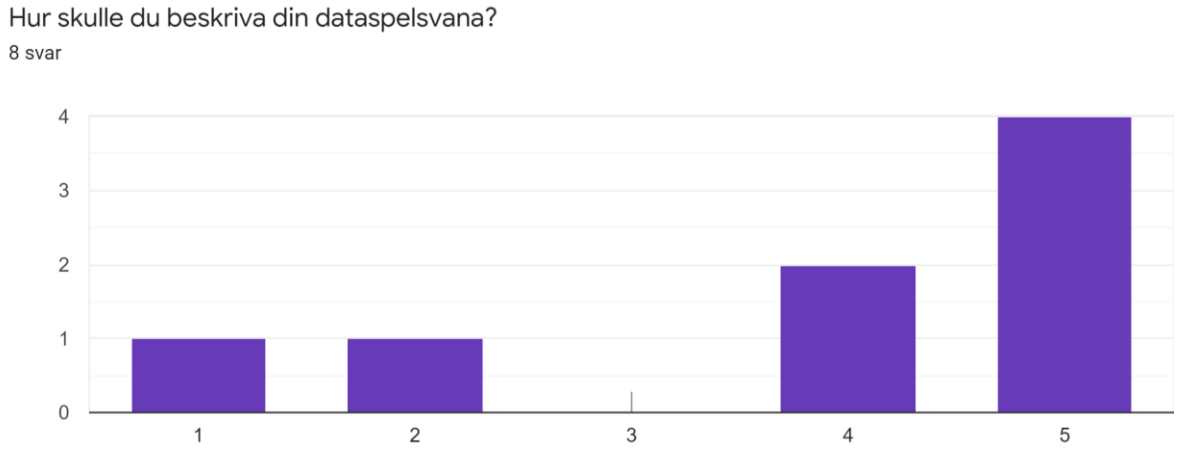
En del intervjuteknik anpassades då till en början mycket instämmande och jakande skedde från testledarnas sida vilket även det kan ha påverkat hur deltagaren angivit sina svar. Mer fokus lades istället på att inte färga deltagarens svar.

5 Utvärdering

5.1 Presentation av undersökning

Den undersökning denna studie analyserar genomfördes av 8 deltagare varav 7 var män och en var kvinna. Av dessa 8 deltagare genomförde 2 deltagare undersökningen under pilottest-stadiet. Ett val togs om dessa två deltagares svarsresultat skulle tillåtas att räknas med i sammanställningen eller inte, där det ansågs vara acceptabelt då endast mindre korrigeringar gjorts mellan pilotstudien och den slutgiltiga undersökningen. Dessa förändringar ansågs inte vara tillräckligt markanta för att påverka hur dessa två deltagare upplevt testet i jämförelse mot de deltagare som genomfört det slutgiltiga testet.

Fem av deltagarna befann sig i åldersgruppen 20-29 år och tre i åldersgruppen 30-39 år. Deltagarna svarade även på vilken dataspelsvana de hade. På en femgradig skala där 1 var “Spelar inte alls” och 5 var “Spelar mycket” svarade hälften (4st) av deltagarna att de “Spelar mycket” och resten av svaren var jämnt fördelade över skalan. En deltagare svarade “4”, en deltagare svarade “2”, en deltagare svarade “1” och ingen deltagare svarade “3”. (se figur 7)



Figur 7. Fördelning av deltagarnas dataspelvana. 1 = “Spelar inte alls” 5 = “Spelar mycket”

Undersökningen genomfördes på distans med hjälp av verktyget Discord (2020) där deltagarna kopplades upp i ett samtal tillsammans med testledarna. Deltagarna ombads på förhand att ha tillgång till en dator och ett par hörlurar.

Deltagarna fick läsa igenom ett samtyckesformulär (se appendix C) där information om hur den insamlade datan skulle behandlas samt kort information angående testet. Deltagarna ombads även att använda penna och papper för att föra anteckningar under spelsessionen. I formuläret listades även deltagarnas rättigheter.

Deltagarna fick därefter muntliga instruktioner från testledarna och deltagarna hade även möjlighet att ställa frågor om någonting var otydligt. De instruktioner som tilldelades behandlade endast instruktioner för spelets kontroller och vad deltagarens uppgift var. Däremot nämndes inget om undersökningens syfte.

Deltagarna genomförde därefter testet i en individuell kanal på Discord (2020) för att kunna fokusera på uppgiften utan att känna sig stressade av testledarnas närvaro. Deltagarna hade möjlighet att kontakta testledarna under hela testets gång, antingen via skrift eller genom tal, för att ställa frågor.

Efter genomfört test fick deltagaren svara på ett deltagarformulär och därefter genomfördes en intervju som spelades in. Deltagarna gav sitt medgivande om att bli inspelade muntligt vid intervjutillfället. Dessa intervjuer transkriberades därefter för att få en bättre överblick av deltagarnas upplevelse.

5.2 Analys

Syftet med studien var att undersöka hur olika pitchförändringar påverkar upplevelsen av monotont upplästa vokaler. Om resultatet skulle visa på ett tydligt samband mellan pitchförändringarna och den upplevda känslan, skulle denna teknik kunna appliceras på sammanhängande meningar och därmed hjälpa till med utvecklingen av framtida multimedia.

Efter genomförd undersökning har vissa likheter hittats mellan deltagarna svar. Men även en del skillnader. Resultatet av undersökningen visar på att pitchförändring förändrar upplevelsen av en inspelning. Däremot kan resultatet variera stort beroende på vilket sätt pitchen förändras.

För att sammanställa resultatet och göra det mer läsbart kommer karaktärerna att tilldelas förkortningar. Karaktärernas förkortningar är som följande:

Karaktär 1, ingen pitchförändring = K1

Karaktär 2, pitch-sänkning i slutet av vokalen = K2

Karaktär 3, pitch-höjning i slutet av vokalen = K3

Karaktär 4, pitch-sänkning i början av vokalen = K4

Karaktär 5, pitch-höjning i början av vokalen = K5

Resultatet visar att den inspelning (oförändrad pitch, K1) som legat grund för alla varianter var den version som testets deltagare ansåg förmedla minst känsla/ingen känsla. På frågan *"Var det någon karaktär som inte förmedlade någonting eller som du inte fick en känsla av?"* (se appendix A, fråga

6), svarade fyra av åtta deltagare denna karaktär (K1). Därefter var det tre deltagare som svarade att karaktär fyra (K4) inte förmedlade någon känsla.

En observation som gjorts var att ett antal deltagare beskrev vid intervjutillfället att det är möjligt att de inte uppfattat någonting från denna karaktär (K1) då det är den första karaktären som deltagaren stöter på. Eftersom begränsad information utdelats vid testtillfället nämner flera deltagare att de blivit ställda av hur karaktären låter och därför inte fått en känsla utan fokuserat på denna aspekt istället.

Deltagare 4 svar på frågan: Var det någon av karaktärerna som inte förmedlade någonting?

"Ja K1. Asså K1 var ju, eller ja iochförsig, nu ska jag inte vara för snabb här. K1 är ju den första som man får och då tänker man ju "Vad i helsike?". Och då försöker man ju inte, då kan man ju inte koppla någonting för då är det ju bara å-ö-i-o-å-ä-ö och det går väldigt fort. Nja, jo jag skulle nog säga K1."

Deltagare 5 svar på frågan: Vilken sinnesstämning fick du från den första personen?

"Det kändes väldigt onaturligt. det var lite robotliknande ljud. visste ju inte vad jag hade å förvänta mig. kändes konstigt å lyssna på"

Deltagare 6 svar på frågan: Var det någon av karaktärerna som inte förmedlade någonting?

"Det är väl den första då. Men det är väl lite för att jag inte visste vad testet gick ut på kan det ju ha med att göra. Men när jag kom till den andra så fattade man ju att det var ändringar å sådär."

Den karaktär som deltagarna ansåg förmedla mest/tydligast känsla var karaktär fem (K5) där fem deltagare svarade denna karaktär. Karaktär tre (K3) var det tre deltagare som svarade. En av deltagarna svarade både K3 och K5. Den sista deltagaren svarade K2. Många utav de svar som angavs visade att K5 stack ut från mängden och att denna karaktär var betydligt tydligare än de resterande fyra karaktärerna.

Deltagare 7 svar på frågan: Var det någon av karaktärerna som uttryckte en viss känsla väldigt tydligt eller gav dig en starkare känsla än de andra karaktärerna?

"Femman tyckte jag var tydligast. Den va lätt å separera från dom andra."

Deltagare 6 svar på frågan: Var det någon av karaktärerna som uttryckte en viss känsla väldigt tydligt eller gav dig en starkare känsla än de andra karaktärerna?

"Det var väl den sista då. den med bokstäverna där. den urskilde sig. Första och andra tyckte jag var ganska svåra att höra nån jättestor skillnad på men den sista stack ut tycker jag."

Däremot var även karaktär tre (K3) en karaktär som nämndes ofta i samma sammanhang. Denna karaktär var den som fick mest enhetliga svar när deltagarna bads beskriva karaktärerna med ett eller två ord. Tre deltagare beskrev karaktären som "frågandes" och deltagarna hade möjligheten att

beskriva karaktären hur de ville. Karaktär fem (K5) fick endast beskrivningar som var snarlika varandra på denna fråga men ingen av deltagarna beskrev K5 på samma sätt som en annan deltagare.

Den enhetliga beskrivningen av K3 reflekteras även i de svar där deltagarna lämnat på frågan "*Vilken sinnesstämning fick du av den tredje personen?*".

Deltagare 8:

"Var lite frågande, om man tänker på sättet som ljuden utforma sig så var det ju lite frågande ton på det. Så det blev för mig lite gladare."

Deltagare 2:

"Aa den var ju frågande, är det enda jag kan säga om den. Den var ju väldigt tydlig med att den var frågande."

Deltagare 3:

"Där, okej vad är frågeställningen? Ska jag bara berätta allmänt? Jag upplevde den som, jag petade ner "siren" och det tror jag var för att den var, att jag upplevde att det var en höjande tonart på varje, a, eller e, eller ö, eller vad det var. Eh, och i och med det så har jag skrivit ner att den var lite smått frågande. Likt tvåan (K2) som var lite gladare så är den här istället lite frågande liksom."

Även om deltagarna svarade att K1 var den karaktär som ansågs minst förmedla en känsla visar intervjusvaren att även K4 var väldigt otydlig. Vid sammanställning av deltagarnas svar hittas inga direkta likheter.

Deltagare 8 svar på frågan: Vilken sinnesstämning fick du från den fjärde personen?

"Påminde om tvåan. Oxå hackig. Lite samma känsla."

Deltagare 7 svar på frågan: Vilken sinnesstämning fick du från den fjärde personen?

"Här inget speciellt. Den kändes ganska neutral. "Finlandssvenska?" har jag skrivit. Hur vokalerna uttalades. Lät som en dialekt."

Deltagare 6 svar på frågan: Vilken sinnesstämning fick du från den fjärde personen?

"Det kändes nästan som ett beat ett tag. nästan som om att det skulle kunna bli en låt, av nån anledning. för den hade jag svårast att greppa..vad jag egentligen hörde men. å så kändes det som om det var lite frågande lite undrande."

Deltagare 2 svar på frågan: Vilken sinnesstämning fick du från den fjärde personen?

"Eh, den var lite som ettan att det var svårt att placera den i någon sinnesstämning så. Men jag tyckte att det ändå fanns någonting lite hotfullt, typ aggressivt, det kändes som att den typ utmanades eller något sånt."

På frågan *"Var det någon karaktär som inte förmedlade någonting eller som du inte fick en känsla av?"* var det tre deltagare som svarade karaktär fyra (K4), ett svar efter K1. Det finns en likhet mellan deltagarnas svar på denna fråga och andra frågor angående K4 från intervjun.

Dessa var de tydligaste resultaten av undersökningen däremot finns det även en ganska stor spridning i svaren som deltagarna angivit. På ingen utav frågorna gav alla deltagare samma svar och på de frågor där mönster hittats, fanns det deltagare vars svar stack ut från mängden.

5.3 Slutsats

Resultatet av undersökningen visar tydligt att deltagarnas upplevelse förändras mellan de olika karaktärerna. Detta visar på att pitchförändringar påverkar hur lyssnaren uppfattar en röst. Resultatet visar även på att vissa typer av pitchförändring skapar en tydligare uppfattning för lyssnaren än andra typer av pitchförändring.

Studiens resultat visar att pitchförändringar där pitchen höjs verkar förmedla tydligare känslor till deltagarna. K5 och K3 hade betydligt mer likartade svar från deltagarna vilket tyder på att deltagarna hade upplevt dessa karaktärer på liknande sätt. Karaktärer där pitch-sänkning sker verkar inte förmedla lika precisa känslor, eller så är fallet att det är svårare för deltagarna att sätta ord på sina upplevelser för att beskriva dessa karaktärer. Den karaktär som haft mest spridda svar var karaktär K4 där en pitch-sänkning sker i början av vokalen.

Resultatet visar även att deltagarna reagerat på K1 och dess avsaknad av pitchförändring då många av deltagarna beskrivit karaktären som artificiell eller robotisk. Detta stärker påståendet att de efterkommande karaktärerna med pitchförändring (K2-K5) påverkar hur deltagaren upplever dessa karaktärer.

En aspekt som kan ha påverkat testet mer än förväntat är den uppspelningsordning som valdes för karaktärerna. Exempelvis reagerade deltagarna starkt på K1 för att det var den första karaktären de stötte på och var osäkra på testets syfte. En teori är att deltagarna reagerat kraftigare på K1 med detta tillvägagångssätt än om karaktärsordningen vore slumpmässig. Hade inte K1 varit den första karaktären som deltagaren stött på varje gång, kan detta ha påverkat hur denna karaktär uppfattats. En spekulering är då att K4 hade varit den karaktär som upplevts som mest otydlig istället för K1. Detta baseras på att det finns en viss enhetlighet i deltagarnas beskrivningar av K1, där beskrivningarna av K4 är mer spridda.

K1 upplevdes även som mest hotfull, där tre deltagare beskrev K1 som mest hotfull. Även detta tros bero på ordningen av karaktärerna. Detta är dock spekulering och det är svårt att säga vilken karaktär

som skulle ta dennes plats, då resterande svar kring hotfullhet är utspridda över de andra karaktärerna.

Ett annat område som verkar ha påverkat deltagarnas upplevelse är tidsintervallen mellan uppspelningen av vokalerna. Tanken med artefakten var att simulera tal genom att slumpmässigt ändra tidsintervallen mellan vokalerna. Detta verkar ha påverkat deltagarnas upplevelse mer än väntat då svar från deltagarna vid flera tillfällen nämnt denna tidsintervall. Följande är ett par exempel på dessa svar:

"Kändes hackigare. Om den första hade ett flöde så kändes den andra som lite hackigare, lite hårdare. Blev brott på ett annat sätt. Känns mer stressande."

"Det kändes nästan som ett beat ett tag. nästan som om att det skulle kunna bli en låt"

"Här har jag definitivt skrivit ner en lite mer stressande sinnesstämning blev jag i. Lite stressande. Och då menar jag mer stressande än K3 för i trean hade jag inte riktigt en sån direkt känsla utan det var mer… det som jag kom på efteråt.

Men där fick jag en direkt känsla utav stress och det tror jag beror på att direkt när jag kom in så var det väldigt snabba byten mellan dom här tonarterna. Det var inget mellanrum mellan a-na och e-na och ä-na och det var snabba tonövergångar och det upplevde jag som mer stressande. "

"Mindre väntetid mellan vokalerna. Det var ingen paus i princip av det jag kommer ihåg."

Studiens resultat visar att digital pitchförändring av en röst påverkar hur lyssnaren uppfattar rösten. Däremot är det svårt att förutsäga på vilket sätt pitchförändringen påverkar lyssnaren. Det verkar finnas ett tydligare mönster i hur en pitch-höjning påverkar lyssnaren men resultatet visar att detta mönster inte gäller för samtliga deltagare. En pitchsänkning verkar inte vara lika förutsägbar och detta tillvägagångssätt verkar vara bristfälligt.

Intervallen mellan vokalerna verkar även ha påverkat deltagarnas uppfattning i större grad än förväntat. Dock är det oklart om den pitchförändring som gjorts förstärkte upplevelsen av slumpmässigheten i uppspelningsintervallen eller inte.

6 Avslutande diskussion

6.1 Sammanfattning

Denna studie undersöker hur pitchförändringar påverkar upplevelsen av en röst. Arbetet ämnar undersöka huruvida en persons uppfattning av en förinspelad röst kan förändras genom att använda digital pitchförändring på rösten i efterhand. Den frågeställning som studien undersöker är "Hur påverkar digital manipulation i form av pitchförändring upplevelsen av en röst i en tredimensionell spelmiljö från ett förstapersonsperspektiv?".

För att få svar på detta har en artefakt skapats i form av ett kort digitalt spel som innehåller fem olika karaktärer vars röster har pitchförändrats på olika sätt. Rösterna består av korta inspelade vokaler. Pitchen på vokalerna höjs på två av karaktärerna, sänks på två karaktärer och är oförändrad på en karaktär. Med hjälp av denna artefakt genomfördes en undersökning där 8 personer deltog. Deltagarna bestod av 7 män och 1 kvinna i åldersgruppen 20-39 år med blandad dataspelvana. Deltagarna genomförde undersökningen via programvaran Discord (2020) där testdeltagare och testledare kunde kommunicera under testets gång. I spelet stötte deltagarna på karaktärerna i en förutbestämd ordning och ombads anteckna sina tankar om vardera karaktärer under speltillfället. Efter genomfört test intervjuades deltagarna och ombads förklara sina upplevelser till testledarna.

Resultatet av undersökningen visar att deltagarnas uppfattning förändrades mellan de olika karaktärerna, dock varierar resultatet beroende på hur rösten pitchförändrats. De röster med en artificiell pitch-höjning har gett ett mer samlat svarsmönster än både de röster med pitch-sänkning samt den röst med oförändrad pitch. Även områden som, tidsintervall (tid mellan vokalernas uppspelning) samt turordning (den ordning deltagarna stött på karaktärerna), har påverkat deltagarna upplevelser i större grad än väntat.

6.2 Diskussion

Efter genomförd studie kan en del slutsatser dras och dessa slutsatser kan jämföras mot andra studier i liknande forskningsområden. Studiens undersökning får fram ett resultat men validiteten av detta resultat kan diskuteras. Då undersökningen endast haft 8 deltagare resulterar detta i att varje deltagares svar påverkar utfallet av resultatet i stor grad. Ett större deltagarantal hade kunnat bidra till en tydligare översikt av deltagarnas upplevelse. Dock går det att se mönster i de angivna svaren och genom dessa mönster kan slutsatser dras från undersökningen. Även om ett större deltagarantal skulle ge en klarare bild av deltagarnas upplevelse är det oklart om denna data hade kunnat analyseras noggrant nog inom studiens tidsram. Eftersom en kvalitativ metod valts har därför mer

tid lagts på att analysera de svar som deltagarna angivit istället för att öka kvantiteten av testdeltagare.

En aspekt att ha i åtanke är att undersökningens demografi är aningen smal. Deltagarna var i majoriteten män och endast en kvinna deltog. Det hade varit önskvärt om lika många kvinnor som män genomförde undersökningen. Lima et al (2013) fann ett intressant resultat i sin undersökning där deltagarna ombads kategorisera olika vokaliseringar, där kvinnor hade lättare att urskilja sinnesstämningar från de röster som spelades upp än de män som deltog. Det hade varit intressant att kunna jämföra svaren från männen i vår undersökning med de hypotetiska kvinnliga svaren för att se likheter och skillnader. Tyvärr fanns inte möjligheten att hitta fler kvinnliga deltagare till denna undersökning.

Valet att använda en kvalitativ metod har gett svar på de frågor som ställdes under undersökningen. Det var användbart att under intervjutillfället ha möjligheten att be deltagarna utveckla och fördjupa sina svar vilket har lett till en tydligare bild av deras upplevelse. Däremot på grund av särskilda omständigheter tvingades undersökningen att genomföras på distans över internet. Detta, enligt Cox (2008a), medför nackdelar då kontrollen över testmiljön förloras och resultatet blir inte lika pålitligt. En skillnad mellan Cox (2008a) undersökning och denna undersökning var dock att Cox (2008a) samlade in kvantitativ analytisk data och var inte närvarande under deltagarnas testtillfälle medans testledare och testdeltagare hade kontakt under hela testtillfället under denna studie. Detta medförde att en större kontroll på testmiljön var möjlig för denna studie då tydligare information kunde ges. Däremot skulle studiens oönskade variabler kunna minskas om undersökningen genomfördes i en och samma lab-miljö och därmed ge större kontroll över undersökningen.

Bestelmeyer et al. (2010) och Grinbaum (2015) undersökningar påvisar att människor anpassar sitt lyssnade. Dessa studier visar att människan både anpassar sig efter sociala situationer samt att en persons uppfattning av ett ljud kan förändras vid högre exponering av ljudet. Då deltagarna i denna undersökning stöter på karaktärerna i en förutbestämd ordning kanske detta påverkar hur deltagarna upplever karaktärerna genom exempelvis anpassning eller familjaritet. Om ordningen, som deltagarna stöter på karaktärerna i genereras slumpmässigt skulle detta kunna säkerställa att ordningen inte påverkar deltagarnas upplevelse i lika hög grad och skapa ett klarare resultat. Då denna studie inte ämnat att undersöka anpassning hade det kunnat vara bra att använda slumpmässighet för karaktärernas ordning för att ta bort en variabel från undersökningen.

En annan aspekt i denna studie är användandet av slumpmässighet vid uppspelningsintervallen mellan karaktärernas vokaler. Detta hade en större påverkan på studien än vad som beräknats. Denna slumpmässighet implementerades för att simulera tal bättre än om vokalerna enbart spelades upp utan avbrott. Efter att studien nu genomförts anses att mer förarbete inom detta område varit

till stor hjälp för att redan i tidigt skede inse denna tidsintervalls påverkan på människors uppfattning. Mer eftertanke hade då lagts på denna variabel vid artefaktskapandet.

Då den artefakt som skapats för denna undersökning har lagt stort fokus på att minimera variabler, både auditivt samt visuellt, togs flertalet beslut för att förhindra oönskade variabler som skulle kunna påverka studien. Dessa beslut togs med artiklar som Cox (2008b), Grimshaw (2009) samt Roring et al. (2007), i åtanke som visar på att visuellt stimuli, kontext respektive var i en mening pitchförändring sker, påverkar lyssnarens uppfattning av ljuden. Den visuellt sterila miljön i artefakten verkar inte ha påverkat denna studies deltagare då ingen av de svar som samlats in under intervjuerna varken påpekat eller nämnt miljön som testet skett i. Deltagarna har dock reagerat på de olika pitchförändringar som sker.

Publikationer från både César et al. (2013) samt Banse och Scherer (1996) visar på att människor har en förmåga att kategorisera ljud i olika sinnesstämningar samt känslor vilket överensstämmer med resultatet från denna studie. Resultatet visar tydligt att deltagarnas upplevelser förändrats mellan karaktärerna och majoriteten av de svar som lämnats har varit relaterat till känslor. En intressant iakttagelse är dock att vissa av deltagarna lagt större fokus på det känslomässiga i vokalerna där andra deltagare har blandat mellan känslor och mer beskrivande förklaringar av ljudets karaktär. Detta skulle kunna förhindras genom att ge deltagarna en tydligare beskrivning av vad de skulle lyssna efter i testet. Dock valdes att ge så lite information som möjligt till deltagarna innan testet för att minimalt påverka deras upplevelse av undersökningen.

En intressant upptäckt i undersökningen som gjorts för denna studie är att den karaktär som upplevdes mest robotisk eller artificiell var den enda karaktären som inte använt någon form av digitalt efterarbete. Detta är intressant eftersom ingen annan av karaktärerna beskrevs som icke-mänsklig även fast dessa inspelningar är de som är minst mänskliga då digital manipulation använts. Detta visar att det är svårt för människor att avgöra vad som är mänskligt och inte, vilket kan kopplas till en studie genomförd av Lublinskaja et al. (2006) där det framgår att människor har en förmåga att kategorisera syntetiskt tal-liknande ljud. Detta visar att sättet som en röst eller läte spelas upp på är viktigare än det objekt som ligger till grund för ljudet. Detta kan vara ett intressant område för fortsatt forskning och bör genomföras som en egen undersökning för att få fram mer trovärdig data.

Den röst som använts för vokalerna är inspelad av en man men deltagarna blev aldrig specifikt informerade om att det var en mansröst de hörde. Dock kan det ha varit underförstått då rösten är relativt mörk och därför kopplats till en man. Det hade varit bra för studien om även en kvinnoröst spelats in för att undersöka skillnaden i upplevelsen mellan en kvinnoröst och en mansröst. Dessa röster hade sedan kunnat slumpmässigt spelats upp i undersökningen eller om två olika undersökningar gjorts med vardera röst separat för att se om upplevelsen förändras beroende på vilken röst deltagarna hört.

Ur en etisk synvinkel ansågs inte testet vara stötande på något sätt men för att informera deltagarna om deras rättigheter fick de läsa igenom och godkänna ett samtyckesformulär (se appendix C).

6.3 Framtida Arbete

Vid en potentiell fortsättning av arbetet hade en del förändringar behövt göras för att förbättra studien och eliminera potentiella oönskade variabler, i både artefakt som undersökningsmetod.

En potentiell förändring som hade varit intressant för studien hade varit att inkludera en kvinnlig röst, som även den spelat in vokaler som blivit efterarbetade på samma sätt som de vokaler som finns med i den slutgiltiga artefakten. Det hade då varit intressant att se huruvida kvinnorösten och mansrösten uppfattades annorlunda från varandra eller om likheter skulle finnas. Det hade även varit intressant att kunna undersöka pitchförändringar på hela ord eller fullständiga meningar som ett annat tillvägagångssätt att få svar på frågeställningen.

En annan förändring som skulle kunna göras vore att ändra ordningen på karaktärerna mellan deltagarna för att förhindra att följden blir densamma för alla deltagare. Detta för att undersöka om den ordning man stöter på karaktärerna har någon påverkan på hur dessa karaktärer uppfattas.

Ett område som påverkade deltagarna mer än väntat var de oregelbundna tidsintervallen mellan vokalerna. Flertalet deltagare nämnde rytm och tempo i karaktärernas tal, som styrdes av denna tidsintervall. Detta är troligtvis en variabel som skulle behöva tas bort vid fortsatta studier. Ett förslag kan vara att antingen inte ha någon tystnad alls, eller ha en bestämd tid mellan varje vokal som är lika lång oavsett vilken pitchförändring som sker.

Idealiskt hade undersökningen genomförts i en kontrollerad lab-miljö och inte över internet. Detta hade troligtvis förändrat undersökningen på flertalet plan och hade gett mer kontroll över testmiljön. Utöver detta hade mer information kunnat samlas in vid speltillfället, samt intervjutillfället, då kroppsspråk och andra diskreta signaler kunnat noteras av testledarna.

Det hade även varit i undersökningens intresse om mer fokus hade lagts på att få till en jämn fördelning mellan kvinnor och män på testets deltagare. Inom de tidsramar som studien tilldelats upptäcktes detta för sent och kan vara bra att ta med in i fortsatta studier. Vid ett större deltagarantal skulle deltagarna även kunna delas in i flera testgrupper. Detta för att kunna undersöka fler variabler, exempelvis karaktärsordning, uppspelningstempo osv.

Ett område som diskuterades under studien var hur mycket information som skulle lämnas till deltagarna. Valet att inte ge all information vid testets start skulle kunna ändras och istället skulle deltagarna bli informerade om testets syfte för att se om resultatet förändras drastiskt.

I ett större perspektiv kan kunskapen från denna studie appliceras på många användningsområden, specifikt inom spel- och annan media-utveckling. Spelindustrin växer och det är numera möjligt att på väldigt få personer skapa intressanta spel. Dock kan kostnaderna stiga snabbt i utvecklingsstadiet och framförallt mindre spelbolag kan vara i en situation där varken tid eller pengar finns för att göra om röstinspelningar som inte blivit bra vid första inspelningstillfället. Istället för att ytterligare ett inspelningstillfälle med röstskådespelaren skulle behövas kan möjligtvis ljuddesignern förbättra dem redan inspelade tagningarna med hjälp av den kunskap som införskaffas i denna studie, samt likande studier. Detta skulle kunna spara spelstudion både tid och pengar och därför kan detta vara användbar kunskap.

Ett annat område som denna kunskap kan appliceras på är inom ”text-to-speech” och förbättringen av detta. Det kan även hjälpa framtida forskning med en fördjupad förståelse av hur uppfattningen av röster påverkas av digital manipulation.

Referenser

Banse, R. & Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology*, 70(3), ss. 614–636.

Bestelmeyer, P. E. G., Rouger, J., DeBruine, L. M. & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, 117, ss. 217–223. doi:10.1016/j.cognition.2010.08.008

Chion, M. (1994). *Audiovision, Sound on Screen*. New York: Columbia University Press.

Cox, T. J. (2008a). Scraping Sounds and Disgusting Noises. *Applied Acoustics*, 69(12), ss. 1195–1204. doi: 10.1016/j.apacoust.2007.11.004.

Cox, T. J. (2008b). The Effect of Visual Stimuli on the Horribleness of Awful Sounds. *Applied Acoustics*, 69(8), ss. 691–703. doi: 10.1016/j.apacoust.2007.02.010.

Discord (2020). *Discord* (version 10.0.18362) [programvara]. Tillgänglig: https://discordapp.com/

FMOD (2020). *FMOD Studio* (Version 2.00.08) [programvara]. Tillgänglig: https://www.fmod.com/download

Gendron, M., Roberson, D., Marieta van der Vyver, J. & Feldman Barrett, L. (2014) Cultural Relativity in Perceiving Emotion From Vocalizations. *Psychological Science*, 25(4), ss. 911–920. doi: 10.1177/0956797613517239.

Grimshaw, M. (2009). The audio Uncanny Valley: Sound, fear and the horror game. *Games Computing and Creative Technologies*: Conference Papers.

Grinbaum A (2015). Uncanny Valley Explained by Girard's Theory. *IEEE Robotics and Automation Magazine*, 22(1), ss. 149–150. doi: 10.1109/MRA.2014.2385568.

Image Line Software (2020). *FL Studio 20* (Version 20.6.2) [programvara]. Tillgänglig: https://www.image-line.com/flstudio/

Lima, C. F., Castro, S. L. & Scott, S. K. (2013). When Voices Get Emotional: A Corpus of Nonverbal Vocalizations for Research on Emotion Processing. *Behavior Research Methods*, 45(4), ss. 1234–1245. doi: 10.3758/s13428-013-0324-3.

Lublinkskaja, V., Ross, J. & Ogorodnikova, E. (2006). Auditory Perception and Processing of Amplitude Modulation in Speech-like Signals: Legacy of The Chistovich–Kozhevnikov Group. I Divenyi, P., Greenberg, S. & Meyer, G. (red). *Dynamics of Speech Production and Perception*. Amsterdam: IOS Press (NATO Science Series Series I. Life and Behavioural Sciences). Tillgänglig: http://search-ebscohost-com.libraryproxy.his.se/login.aspx?direct=true&db=nlebk&AN=176064&site=ehost-live (Hämtad: 29 December 2019). ss 87-99.

Nationalencyklopedin (2020). Prosodi. Tillgänglig: http://www.ne.se/uppslagsverk/encyklopedi/lång/prosodi-(i-antik-ljudlära) (hämtad 2020-01-05)

Ringu (1998) [Film]. Regissör: Hideo Nakata. Japan: Ringu/Rasen Production Committee.

Roring, R. W., Hines, F. G. & Charness, N. (2007). Age Differences in Identifying Words in Synthetic Speech. *Human Factors*. 49(1), ss. 25–31. doi: 10.1518/001872007779598055.

Soundtoys (2019). *Little Alter Boy* (Version 5.3.1.15178) [programvara]. Tillgänglig: https://www.soundtoys.com/product/little-alterboy/

Unity Technologies (2019). *Unity* (version 2019.2.17f1) [programvara]. Tillgänglig: https://unity.com/

Vetenskapsrådet (2002). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning* Stockholm: Elanders Gotab. Tillgänglig: http://www.codex.vr.se/texts/HSFR.pdf

Zhang, Y. & Francis, A. (2010). The Weighting of Vowel Quality in Native and Non-Native Listeners' Perception of English Lexical Stress. *Journal of Phonetics*, 38(2), ss. 260–271.

Appendix A - Intervjufrågor

1. Vilken sinnesstämning fick du av den första personen? Andra, osv?

2. Uppfattade du någon av rösterna som vänligare? Om ja, vilken?

3. Uppfattade du någon av rösterna som hotfullare? Om ja, vilken?

4. Hur skulle du beskriva röst nummer 1? 2, osv? (Ok att säga “vet inte”.)

5. Var det någon av karaktärerna som uttryckte en viss känsla väldigt tydligt eller gav dig en starkare känsla än de andra karaktärerna?

6. Var det någon karaktär som inte förmedlade någonting eller som du inte fick en känsla av?

Appendix B - Deltagarenkät

En deltagarenkät skapades för att lättare få en överblick över studiens demografi. Länk: https://docs.google.com/forms/d/e/1FAIpQLSdgLzIYFIGKsZ42LLJmAqqAB5DmgiG8YuguOZiOgg67CkVbMw/viewform

Appendix C - Information till deltagaren

Samtycke för medverkan av deltagande

Denna text ämnar att ge dig den information du bör vara medveten om innan deltagande i denna studie. Det är viktigt att du läser igenom texten noggrant och ställer frågor till testledarna om någonting skulle vara otydligt.

I det test du snart kommer genomföra är det viktigt att du använder ett par hörlurar, gärna hörlurar med god kvalité, och kan sitta ostört i ca 20 minuter. Vi önskar även att du har tillgång till någonting att anteckna dina tankar och idéer på, antingen ett papper eller valfritt annat tillvägagångssätt.

I testet kommer du stöta på olika karaktärer och det som undersöks är hur du upplever dessa karaktärer.

Information om rättigheter

- Ditt deltagande är frivilligt och du kan avbryta testet när du vill utan att behöva förklara varför.

- Jag ger mitt medgivande om att den information som samlas in får användas i denna studie.

- Jag ger detta medgivande med förutsättningarna att den information som samlas in endast kommer användas och delas med den studentgrupp, samt handledare och examinator för studien.

- All information är anonym.

- Jag har fått möjligheten att ställa de frågor som jag har och har fått dessa frågor besvarade.

- Medgivande sker muntligt över internet då testet sker på distans.