Master Thesis

Validation of Interpersonal Stances Expressed by Virtual Suspect Characters in a Police Interview Setting

Qiuhua Fu

HUMAN MEDIA INTERACTION
ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

EXAMINATION COMMITTEE
Dr. Ir. Rieks op den Akker
Dr. Mariët Theune
Dr. Merijn Bruijnes

December 5, 2014

UNIVERSITY OF TWENTE.

Abstract

This thesis describes the research and development activities that were carried out during my final project in the course of the Human Media Interaction Master of Science study at the University of Twente. These activities contribute to a project collaborating with the police academy that aims at building a tutoring system in which police trainees can train interrogation strategies when interviewing crime suspects. The system consists of a virtual 3D interrogation room and the trainee interacts with a virtual character that plays the role of a suspect. It is important that the characters act in a believable and coherent way when interacting with the interviewer. In particular we want that these virtual characters are able to show behaviors and expressions that are recognized as expressing the interpersonal stances the suspect takes towards the police officer and in response to his way of approaching the suspect.

Research by many others showed that interpersonal stance can be modeled by two dimensions: controlling, ranging from dominant to submissive and affiliation, ranging from friendly to hostile. In this model we distinguish four stance categories corresponding with the four quadrants of the two-dimensional space: dominant, submissive, friendly and hostile.

The first question we try to answer is the following. Suppose we ask different people what the stance is that a human actor acts out in a video fragment. Do they agree? The second question is this: if various actors act out a stance that people perceive as dominant what are the typical behaviors, postures and facial expressions of these actors that possibly cause people to perceive this as dominant. The same questions are studied for the other three stances. We found a number of characteristics for each of the stances. Then we built three virtual characters that show the four stances. The final question is: do people recognize the stances acted out by the virtual characters? In order to answer this question we again did a perception study where we asked human judges to tell what stances the characters showed.

The results revealed that, except for the dominant stance, people recognized the stances expressed by the virtual humans when they show the behaviors and expressions that were found as typical for these stances. Audio improves the perception of stances but wearing an Oculus Rift (for a 3D effect) does not help.

Preface

First I wanted to find a topic in social signal processing to do my final project. That is how I met my first supervisor dr. ir. Rieks op den Akker. He is associated with a project in collaboration with the police academy, which aims at building an police interview game to help the police trainees to improve their interrogation skills. He offered me a very interesting topic that is to find typical behaviors that can express different interpersonal attitude during an interrogation. I always have a strong interest in psychology, especially in behavior science.

At the end of my final project, I managed to build a 3D interrogation environment in Unity 3D and generate real-time behaviors. This was a whole new field for me. I enjoyed the process of learning new 3D technology, such as 3D animation, programming in 3D game engine and virtual reality head-mounted display (Oculus).

Acknowledgement

I would like to thank my three supervisors for their kindly advice and guidance.

I didn't have any experience in Unity 3D which is a professinal game engine. Thanks to Rieks, who gave me a chance to work at a virtual reality company called CleVR. My supervisor Freek in CleVR offered me tremendous help technologically.

I am grateful for my supervisors' advice when I was doing an online survey for the research topics. They told me to be more precise and cautious when it comes to science. They also gave me many useful comments on my reports, and they were patient when I made mistakes. I learned a lot during the process. Hopefully, my writing skill has been improved.

Qiuhua Fu
Enschede, November 2014

Table of Content

Chapter 1 Introduction

1.1 Background

The ability to express and understand social signals (such as feelings, mental state, personality and other traits of people) is said to be perhaps the most important factor that leads to success in life [35]. Nonverbal behaviors are irrepressibly impactful and convey most of the social signals, since they can be difficult to suppress [19, 35, 8]. Nonverbal behaviors are more accessible to the people who observe them than to the people who produce them [11]. However non-verbal behavior alone doesn't determine our perception of another person. During an encounter, perceivers rapidly form impressions of other individuals through observed static and dynamic characteristics. Static characteristics refer to stereotypes formed by visual appearance such as height, clothing and body shape, while dynamic characteristic refer to verbal and non-verbal behavior [8].

The same rules apply for the interaction with embodied conversational agents [36]. Embodied conversational agents are embodied agents (usually with a graphical front-end as opposed to a robotic body) that are capable of engaging in conversation with one another and with humans employing the same verbal and nonverbal means that humans do (such as gesture, facial expression, and so forth). Psychology research has shown that humans are literally wired for extracting social information from nonverbal communication [36]. Any facial expression, vocal outburst, gesture or posture often triggers unconscious analysis of socially relevant information [11, 36]. In fact, this mechanism seems to be so deeply rooted in our brain, that we cannot escape it, even when we interact with embodied conversational agents [36].

1.2 Research Goals

This research is a part of the project that is in collaboration with the police academy which aims at building embodied conversational suspects in a game that helps to improve the interrogation skills of police trainees. In this research, we want to investigate the human judges' perception of virtual suspects. If people perceive virtual humans more or less in the same way as real human beings, then we can apply the typical behaviors of different stances found in a prior research [15] to build the virtual suspects in the project with the police academy.

In the prior research [15], behaviors that were acted out by human actors and can express different interpersonal stances have been identified. In this research, virtual

humans are employed to simulate those behaviors. The main research question is **how people perceive stances expressed by virtual humans in an interrogation setting**. Three related sub-questions based on the factors that may affect people's perception of virtual humans are proposed.

People tend to judge others by their physical appearance [27]. These judgments are always stereotypes, however are not necessarily true. During interrogations, the police are confronted with a variety of people with different social status and that have different physical appearance, and they are also not immune from those stereotypes. Through this experiment, we want to investigate people's perception of virtual suspects and the difference in the judgments resulting from various appearances. Specifically, we want to find out **a. if the stereotypes against different physical appearances are the same for virtual suspects**.

In the previous study, the fragments were all mute. However, other research has indicated that synchronized sound with the behavior can reinforce our perception of interpersonal attitude [29]. Therefore we added audio to this experiment to evaluate **b. how audio can affect users' perception of the virtual suspects**. For example, is it true that subjects perceive the virtual suspects as more realistic with the presence of audio, and that audio results in more agreement on subjects' judgments?

3D techniques are highly related to Virtual Reality. Virtual reality is often used to describe a wide variety of applications commonly associated with immersive, highly visual, 3D environments. The Oculus Rift is an upcoming virtual reality head-mounted display that offers an immersive virtual environment (IVE) and is capable of tracking the position and orientation of the head. Since the environment is completely built in a 3D environment, it is also interesting to test **c. the difference of playing with and without Oculus Rift.**

1.3 Research Approach

In this research, a quantitative method was applied. Hypotheses were proposed in order to answer the research questions. After building a 3D interrogation environment, participants were invited to an experiment. Questionnaires were used to investigate participants' perception of the stances of virtual suspects. The collected data was later analyzed using SPSS Statistics, which is a software package used for statistical analysis. The analysis result was used to verify if the hypotheses were correct. One should be able to reproduce the experiment and generalize and replicate the findings by following the steps described in this report.

During the development of the interrogation environment, the general rule is that the behaviors should be believable and generated in real-time. To build embodied conversational suspects in a game, it is essential for the system to be able to generate real-time behaviors, so that the virtual suspects can be more interactive. Believable behavior means that the behaviors should be realistic and well synchronized. Behaviors of the virtual agent such as facial expression, body posture and speech, should be synchronized and consistent. For example, the virtual agent should be talking with his mouth open. Since we don't have a behavior generation mechanism, the simple way to have believable behavior is to fully duplicate human behaviors. In this research, virtual suspects were made to simulate a sequence of behaviors acted out by human actors.

1.4 Structure of the Report

Chapter 2 is about related work. Chapter 2 firstly gives an overview of related researches by others, and then introduces two prior studies of this research.

Chapter 3 explains how the interrogation environment has been built. Firstly the avatars and animations have been built in blender. Then facial expressions and body postures have been controlled and generated in real-time in Unity 3D.

Chapter 4 elaborates the procedure of the experiment. Based on the result of data analysis, some conclusions were drawn. The shortcomings of this research are also discussed.

Chapter 5 gives some suggestions of future work.

Chapter 2 Related Work

Section 2.1 gives a summary of researches with similar topics done by others. The approach of our research is compared with other related researches. Section 2.2 gives references to theories that are relevant to the research questions. Later the hypotheses are also proposed based on these theories. Section 2.3 reviewed prior studies of this research.

2.1 Researches with Similar Topics

Embodied conversational agents are more and more widely employed to play different roles in different social interactions, such as museum guide [21], communication partner [23] and personal assistant and companion [33]. Virtual agents with social abilities are usually referred to as relational agents [4]. To make these roles acceptable for people who interact with them, it is essential for the virtual agents to be able to convey different interpersonal attitudes. This ability also increases the believability of the agent effectively [5].

Other researches [30, 10, 31] also focus on the implementation of building virtual agents that have the ability to convey interpersonal attitudes. Non-verbal behaviors that can express social attitudes with a particular communicative intention are also identified in [31]. Comparing to our prior researches [14, 15], the research approaches are different. In [31], users are asked directly to configure the non-verbal behavior of an agent with particular social attitudes (e.g. users are asked, how an agent should look like when he/she denies something with a submissive attitude). However in our prior researches [14, 15], users are required to judge the attitude of the suspect by viewing video fragments, and then behaviors are extracted from the suspect by annotating the fragments.

A computational model that enables an agent to convey social attitude through nonverbal behaviors has been proposed [30]. There is no social context or interactive situation specified in the model. However the perception of these behaviors is influenced by the context of the interaction [10]. This conclusion also implicates the necessity of conducting our research, which tries to find behaviors that can express different attitudes in a police interrogation setting. Comparing to our research, [10] is focused on finding non-verbal behavior sequences in a job interview setting using a multi-layer model.

2.2 Theories Related to Research Questions

The physical appearance includes natural characteristics such as height, body shape, physiognomy, skin and hair color, as well as artificial characteristics such as clothes, ornaments, make up, and other manifests used to modify/accentuate the facial/ body aspects. The main social signal associated to physical appearance is the attractiveness. Psychology studies have long shown that physical attractiveness can elicit positive personality attributions as well as positive behavior towards other people [22]. Attractive people are often judged as having high status and affiliation even if no objective basis for such judgments exists. However attractiveness is a subjective perception that can't be properly defined or compared. Tall individuals tend to be attributed higher social status and, in some cases, they actually hold a higher status [35]. For example, a survey has shown that the average height of the American CEOs of the Fortune 500 companies is around 7.5 cm higher than the average height of the American population. Moreover, 30% of the same CEOs are taller than 190 cm, while only 4% of the rest of the American population lies in the same range of height [35]. Attractive people are often judged as more friendly and dominant in social relations [35].

Gender stereotype is also common phenomenon we can't ignore. In fact, gender stereotype is deeply ingrained in human psychology, extending even to inanimate machines [26]. Regarding social attitude, women are perceived as more friendly and submissive than men [16].

In this research, we focused on three stereotyped somatotypes [13]. Endomorphs refer to individuals that are round, fat and soft, mesomorphic individuals are bony, muscular and athletic, and ectomorphs are characterized as tall, thin and fragile. From the perspective of interpersonal attitude, endomorphs tend to be perceived as friendly and submissive, mesomorphs tend to be perceived as dominant, and ectomorphs are perceived as submissive and inclined to be difficult [37].

It is often claimed that multimodal presentation of information, i.e., information presented to more than one sensory modality, provides more redundancy and is therefore easier to interpret correctly than unimodal presentation [9,39,6]. In prior research [15], only non-verbal behaviors have been considered, because they are most relevant to social attitude. However, it is suggested that multimodal integration of vocal verbal, prosodic and gestural means are important for the study of stance [9]. Several experimental studies found that integrating the information of audio and video leads to an improved performance of affective behavior recognition, and integrating audio and video can make use of the complementary information from the two channels [39, 6]. In addition, psychological studies have shown that the integration of information from multiple modalities (vocal and non-verbal behaviors

in this report) yield a coherent representation and inference of stances [2, 29].
Immersive virtual environments (IVEs) allow the simulation of real-world experience and scenarios that would otherwise be impractical to realize for reasons of cost or feasibility. Their primary advantages are affording more ecological validity without compromising experimental control and allowing the decoupling of variables that naturally covary [24, 38] (e.g. in the physical world, egocentric distance of an object and its angular size are highly correlated, and with IVE, it is assured that researchers can change one variable without changing the other). IVE can be used in research of areas of perception, spatial cognition, and social interaction, as well as many other areas of psychology [24]. IVE offers opportunities for understanding how users perceive and act in space. This advantage is beneficial for the project with the police academy. The limit of this research is within an interrogation room setting. However in real life, police are also required to question people on the street. IVE technology can stimulate this kind of situation. With the ability of position detection, the interviewer can walk around the street and talk to people.

2.3 Review of Stance Behavior

This part gives a summary of the literature study [14] and the Research Topics [15], which are prior studies of this research.

2.3.1Interpersonal Attitude

This research focused on interpersonal attitude expressed by verbal and nonverbal behaviors in order to build 3D virtual suspects. Argyle [1] identifies two fundamental dimensions of interpersonal attitudes that can account for a great variety of non-verbal behavior, which are affiliation and controlling. The scale of affiliation varies from hostile to friendly, while the scale of status varies from dominant to submissive. Affiliation (also called liking or appreciation) can be broadly characterized as liking or wanting a close relationship [16]. Status (also called controlling or power) shows the degree of control one has on another, and can be expressed in two main ways, space and relaxation. Posture is particularly associated with expressive relationships between people or their attitude towards each other, for example a close posture displays a liking while drawing up to full height shows dominant attitude [18].

Both in the previous research [34, 14, 15] and this research, the two-dimensional Leary's model, also known as the interpersonal circumplex (IPC), was used as a framework for classifying the interpersonal attitude. Consistent with the two

dimensions identified by Argyle, the horizontal dimension of Leary's model is affiliation, while the vertical dimension is status (Figure 1). Four interpersonal attitudes can be derived from this model, which are dominant-hostile (DH), dominant-friendly (DF), submissive-hostile (SH) and submissive-friendly (SF).

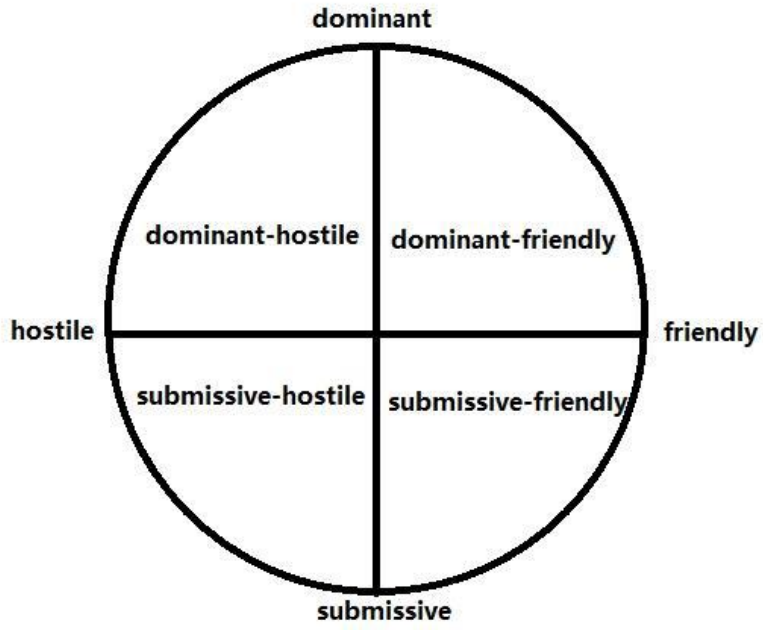

Figure 1. Leary's model of interpersonal attitude

2.3.2 Typical Non-verbal Behavior of Different Interpersonal Stances during Police Interrogations

Through a comprehensive literature research [14], we were able to find some typical nonverbal behaviors of different interpersonal attitudes (Table 1). Apart from behavioral cues, physical appearance such as height, body shape and physiognomy, as well as artificial characteristics such as clothes, ornaments and make-up also have an influence on the perception of others' interpersonal attitudes (Table 2).

The next step was to identify behaviors that express different attitudes in a police interrogation setting [15]. We were focused on the non-verbal behaviors, since they convey most social signals [35]. 16 interrogation video fragments were chosen from popular criminal TV series. They were selected and edited in the following manner: 1. the suspect was the only person in the fragments; 2. all the fragments were mute; 3. the suspects in the fragments were being interrogated in a room rather than police asking random witnesses on the street for information. Through an online survey, 63 subjects rated all 16 fragments on the attitude (affiliation and controlling) of the

suspects and the spontaneity of the actors with a 5-point Likert scale (Figure 2).

Table 1. Typical gesture and posture w.r.t. different attitude (literature review)

	dominant	submissive	hostile	friendly
Head movement	Tilt head up, Orient head toward other, Shake head[7]			Tilt head up, Orient head toward other[7]
Hand gesture	movements directed away [28], high gesture rate while talking [7],initiate hand shaking [7]	movements directed inward [28], object-adaptor, self-touch [30]	self-protection gestures [35], folding arms [16,30]	Touch other[30], object-adaptors [30], initiate hand shaking[7]
posture	Space filling &asymmetric postures[16,28], erected posture [7]	Shrinking postures[16]	distant postures, including postures that present some sort of barrier to interaction[16]	physically close postures, other close interaction or direct orientation[16,30]
Leg movement	wide stance of the legs[28]		rhythmically moving legs [35]	
Facial expression	Facial anger [18], self-assured expression [30], expressive face [30]	Facial sadness [30]	Facial disgust, facial anger [30]	Smile[16,8 ,18]
Gaze behavior	More gaze[8,30], gaze for a long time [30]	Avert gaze[30]	gaze for a long time [30]	Mutual gaze[8,30]
Focus of attention		Pay attention to other[30]		Pay attention to other [30]
Turn taking	Overlapping speech [35]	Pause often[28]	Overlapping speech [35]	
vocalization	Loud voice[28], high pitch [35], high rhythm [35]	Low voice[28]		

Table 2. Typical physical appearance w.r.t. different attitude

	dominant	submissive	hostile	friendly
appearance	Tall[35], neat[26]	Sickly[26], unstylish[26], Messy[26]		Healthy[21], neat[26]
attractiveness	Attractive/beautiful[35]			Attractive/beautiful[35]
Somatotypes	Mesomorph[35, 37]	Ectomorph, endomorph [35, 37]		

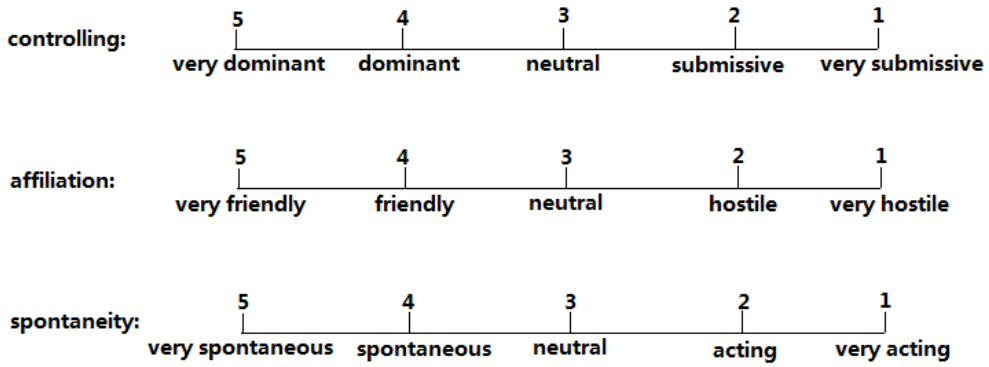

Figure 2. illustration of 5-point Likert scale in the previous online survey [15]

To find the typical behaviors of different stances in a police interrogation, we first need to identify the fragments, which people think are the most dominant, submissive, hostile or friendly, as the best fragments, and then annotate the best fragments to extract both facial expression and body posture. A linear regression model on the spontaneity revealed that actors with higher rating on spontaneity tend to result in less deviation participants' judgments (deviation on controlling = 1.906-0.279 * spontaneity, p=0.117, and deviation on affiliation = 1.876 - 0.319*spontaneity, p=0.010). Thus, the best fragments were selected according to the following criteria: 1. the average spontaneity rating must be above 3; 2. the average rating on the attitude is the highest or lowest of a certain category (highest when referring to dominant and friendly, lowest when referring to submissive and hostile).

By annotating the best identified fragments, we were able to find some nonverbal behaviors (as shown in Table 3) that match the findings of Table 1. In the interrogation fragments, the lower part of the body can't be seen so we can only derive leg movement from literature study. Since the fragments are mute, the vocal non-verbal behavior can also only be derived from literature review. Figure 3 gives some screenshots of typical posture and facial expression of different interpersonal attitude from the best fragments.

Comparing our findings (Table 3) with the literature review (Table 1), Table 3 only includes the nonverbal behaviors we can find in an interrogation setting, so behavior such as initiating hand shaking which is identified as typical dominant behavior through the literature study is not included in this table. Besides, Table 1 is more like a general outline, while Table 3 is more about specific behaviors. The behaviors in Table 3 give an instance to the general outline in Table 1. For example, the literature says space filling and asymmetric postures are dominant behavior, and it is observed in a dominant fragment that the suspect put one arm on the chair while the other arm is stretching out (Figure 3b); the literature says shrinking posture is submissive, and we noticed a submissive suspect bending his spine. Figure 3d, 3f, and 3g give examples of expressions of contempt, sadness, and happiness as mentioned in Table 1. Figure 3e explains what type of expression was considered as expressive face which is a dominant facial expression according to [30]. In this case, the suspect raised his eyebrows to an extreme extent with his eyes wide open.

Figure 3. a. dominant-hostile expression: tilt head up with a gaze down toward the interrogators; b. dominant-hostile posture: asymmetrical, space-filling and distant posture; c. submissive posture: shrinking posture with bended spine; d. hostile expression: expression of contempt; e. dominant expression: expressive face with extreme brow raising; f. submissive expression and posture: expression of sadness and self-touch; g. friendly expression: smile; h. hostile expression: cross-eye gaze with eyelid raising

One inconsistency is found regarding “tilt head up”. It is mentioned in [30] that this movement tends to show friendliness, while all the three fragments that are classified as hostile (two most hostile fragments and one dominant fragment which is also hostile) have been annotated with “tilt head up”. When interpreting social behaviors, the social context should be taken into account [10]. Non-verbal signals

can be interpreted with different perspectives. For instance, a smile is a sign of friendliness [16, 8, 18], however, a smile followed by a gaze and head aversion conveys embarrassment [20]. In the interrogation fragments, the suspects tilt heads up with gazes looking down toward the interrogators (as shown in Figure 3a). This combination determines that people perceive "tilt head up" as a hostile behavior.

Table 3. Typical nonverbal behavior observed from the best recognized fragments

	dominant	**submissive**	**hostile**	**friendly**
Head movements	Shake head, Tilt head up, Head oriented facing the other	Head tilts downward	Tilt head up with a gaze look down toward the interrogators	
Hand gesture	Arm movement away from body ; more movements while talking; asymmetrical arms action	Folding arms; Self-touch; Hold a coffee cup	Folding arms;	
posture	head and trunk oriented facing the other;; straight & erected trunk ; one arm on the chair while the other stretch out	Symmetrical arms posture; bend the spine	trunk lean backward	trunk lean forward
Facial expression	Extreme brow raiser; eyes wide open	Expression of sadness	Expression of contempt	Expression of happiness (smile)
Gaze behavior	More gaze	Averted gaze	Gaze for a long time with upper lid raiser; cross-eye	

These are the typical behaviors that can express different attitudes in a police interrogation when people judge human suspects. However, it is not certain how people perceive virtual suspects. For example, if people will still categorize the virtual suspect into the same attitude, when the virtual suspect is showing the same facial expression and body posture as the human suspect. With the question in mind, three

3D virtual suspects were made to simulate the suspects in the best fragments. Then participants were invited to an experiment. They watched the stances performed by the virtual suspects and judged their attitudes.

Chapter 3 Interrogation Environment

This chapter introduces the process of building the integration environment. We chose Blender[1] and Unity 3D[2] to be our developing tools. Both tools are professional and free. Unity accepts Blender files. Blender is capable of creating avatars with realistic facial expressions and body postures, and Unity can control real-time generation of facial expressions and body postures. The collaboration of the two software meets the rule of generating believable behaviors in real-time (mentioned in research approach in chapter 1).

The interrogation environment is used later in an experiment to investigate people's perception of virtual suspects. In the interrogation environment, there is no scenario or interaction between the participants and the virtual suspects. During the experiment, participants watched the scripted behaviors performed by the virtual suspects and judged their stances.

The technical terms used in this chapter are explained in Appendix A.

3.1 Modeling and Animation

This part explains the work that has been done in Blender, which is open-source 3D computer graphics software. The original models were downloaded online with the authorization of the authors to re-use them. They were further adapted to fit in the interrogation setting. All the 3D models used in this project were built and modified in Blender, later imported into Unity 3D.

Animations contain two parts, one is facial animation, and the other is body posture and action. Animations were made to simulate the suspects in the best fragments. They were built using different techniques. Blender shape keys were employed to make facial animation and blender rigify was used to build biped rig for virtual suspects. Then the avatars were made to imitate the best fragments from the previous study.

[1] Website of Blender: http://www.blender.org/
[2] Website of Unity 3D: http://unity3d.com/

3.1.1 Three Virtual Suspects

According to [37], there are three stereotyped somatotypes, which are endomorph, mesomorph and ectomorph. It is interesting to investigate if people have the same stereotypes against the virtual humans.

Three virtual suspects that stand for three stereotyped somatotypes are employed for the experiment. The man in the suit[3] (Figure 4) stands for mesomorphic body type. The lady[4] (Figure 5) represents ectomorphic body type, and the man in a blue shirt[5] (Figure 6) represents endomorphic body type.

Figure 4. suspect 1: mesomorph

[3] 3D model available at: http://tf3dm.com/3d-model/bruce-wayne-95619.html, 11 November, 2014
[4] 3D model available at: http://www.blendswap.com/blends/view/62580, 11 November, 2014
[5] 3D model available at: http://tf3dm.com/3d-model/obese-male-71456.html, 11 November, 2014

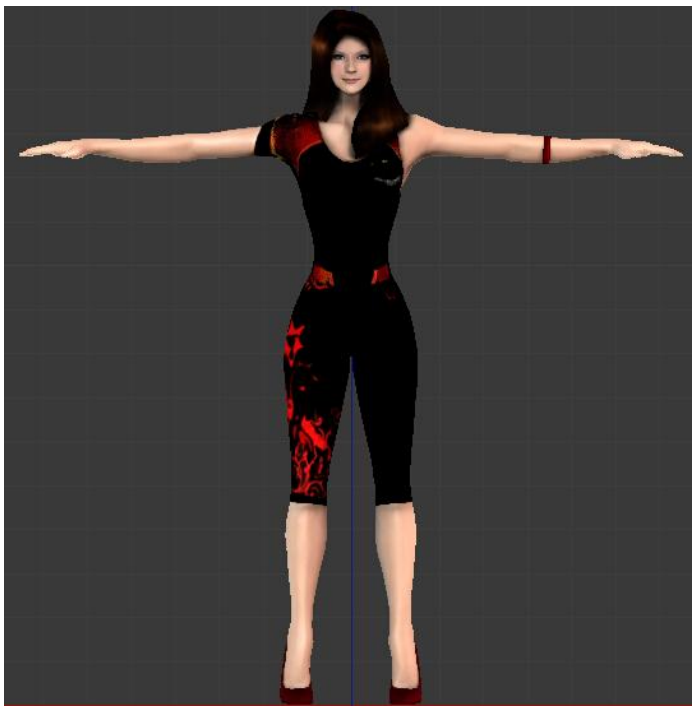

Figure 5. suspect 2: ectomorph

Figure 6. suspect 3: endomorph

3.1.2 Facial Animation

Blender shape keys store different shapes of a same object. In other 3D applications

they are called “morph targets”, “blend shapes”, or even “vertex keys” in older versions of Blender. They are the only way to directly animate the shape of objects. They can be used to directly animate a face, the folding of a sheet of paper, etc. But they can also be used as indirect animation tool – think for example in animating the shape of a curve used as bevel or taper for another one: this would allow developer to animate a pumping artery, or a pulsing heart, or a growing worm, or balloon being inflated, or a tree growing.

For each avatar, 38 facial shape keys were created so that the virtual suspects can imitate almost any facial expression. The layout of shape keys was the same for each avatar as listed in Table 4. A variety of expressions can be simulated by the combination of different facial shape keys. Examples are given in Figure 7. A smile is a combination of shape key 0, 1, 30 and 31 (smileLeft, smileRight, cheekLeftRaise and cheekRightRaise), a sad expression is a combination of shape key 11, 13, 14, 32 and 33 (browLeftLower, browRightLower, browMidRaiser, lipcornerLeftDown and lipcornerRightDown), and a facial anger is a combination of shape key 2, 4, 6, 8, 11, 13, 15, 28, 29, 36 and 37 (lidTopLeftUp, lidTopRightUp, lidBottomLeftUp, lidBottomRightUp, browLeftLower, browRightLower, browMidLower, lipcornerInLeft, lipcornerInRight, lipsTopIn, lipsBottomIn) [13]. The intensity of an expression can be adjusted by setting different weight of the shape key.

Table 4. layout of blend shapes

Sequence	Facial Shape Key	Sequence	Facial Shape Key
0	smileLeft	19	jawOut
1	smileRight	20	lipsTopOpen
2	lidTopLeftUp	21	lipsBottomOpen
3	lidTopLeftDown	22	lipsTopOut
4	lidTopRightUp	23	lipsBottomOut
5	lidTopRightDown	24	sneerLeft
6	lidBottomLeftUp	25	sneerRight
7	lidBottomLeftDown	26	lipcornerOutLeft
8	lidBottomRightUp	27	lipcornerOutRight
9	lidBottomRightDown	28	lipcornerInLeft
10	browLeftRaiser	29	lipcornerInRight
11	browLeftLower	30	cheekLeftRaise
12	browRightRaiser	31	cheekRightRaise
13	browRightLower	32	lipcornerLeftDown
14	browMidRaiser	33	lipcornerRightDown
15	browMidLower	34	lipsPuckerLeft
16	jawDown	35	lipsPuckerRight
17	jawLeft	36	lipsTopIn
18	jawRight	37	lipsBottomIn

Figure 7. left: expression of happiness, middle: expression of sadness, right: expression of anger

3.1.3 Body Posture and Action

The Blender add-on called Rigify provides models with a biped rig. Rigify generates a full rig for the model and by using the armature as a template you will be able to forge it into a nice simple rig for humanoid characters. The armatures need to be adjusted to fit the body of the avatars. A script (Appendix B) that makes the bone structure of Blender Rigify suitable for Unity Mecanim was run on each of the armatures.

Even though we are required to adjust armature repeatedly for each avatar, since the body types are different, it is not necessary to build body posture and action for each character separately. We only made body animation for the lady, and retargeted the animation to the endomorphic and mesomorphic avatar. Pose libraries were created in order to better reuse the postures in real-time animation. In each pose library, there are behaviors of one attitude. These behaviors are the typical behaviors that express different attitudes based on previous findings (Table 3). There are some overlaps in the pose library. For example, folding arms is both a submissive and a hostile behavior, because it is a self-touch (submissive) and self-protection (hostile) gesture (see Table 1 in Section 2.3.2).

With the facial shape keys and Rigify done in Blender, now the avatars are ready to be imported into Unity to create real-time animation.

Table 5. Pose Library

Dominant	Asymmetric arms, asymmetric & spacing-filling legs, crossed legs, erected trunk, head tilt up, nod, open arms, point at, shake head
Submissive	Averted gaze, bended trunk, closed arms, folding arms, head and gaze down, symmetric & closed legs
Friendly	Arms reaching out, head oriented toward the other, nod, trunk lean forward
Hostile	Head tilt up, head turn sideways, point at, trunk lean backward, trunk turn sideways, folding arms
Basics	Head tilt up, head tilt down, head tilt left, head tilt right, head turn left, head turn right, idle, nod, point at, shake head

3.2 Controlling and Generating Real-time Animation

After importing the avatar (model and rig) and the animations separately into Unity, controlling and generating real-time animation have been done in Unity 3D. Unity 3D is a cross-platform game engine. Instead of showing the animations of the virtual suspects in video clips, the virtual suspects were imported into Unity 3D in order to generate real-time animation. An interrogation environment which is capable of generating real-time facial expressions and body postures was built in Unity 3D.

3.2.1 Animator Controller

The Animator Controller is the main component by which animation behavior is added to an object. The avatar defines an object's skeletal structure but an Animator Controller applies animations to the skeleton. An Animator Controller asset is created within Unity and allows you to maintain a set of animations for a character and switch between them when certain conditions occur. For example, you could switch from a walk animation to a jump whenever the spacebar is pressed. The controller manages the transitions between animations using a so-called State Machine. The main state machine in our interrogation environment is shown in Figure 8. The suspect is initially in an idle state, and can switch into a dominant, submissive, hostile or friendly state. The transitions between idle state to other states can be triggered by pressing keyboard 1, 2, 3 or 4. The virtual suspect will automatically switch back to idle state when the animations in other states finish playing. The same animator controller is applied to all three avatars.

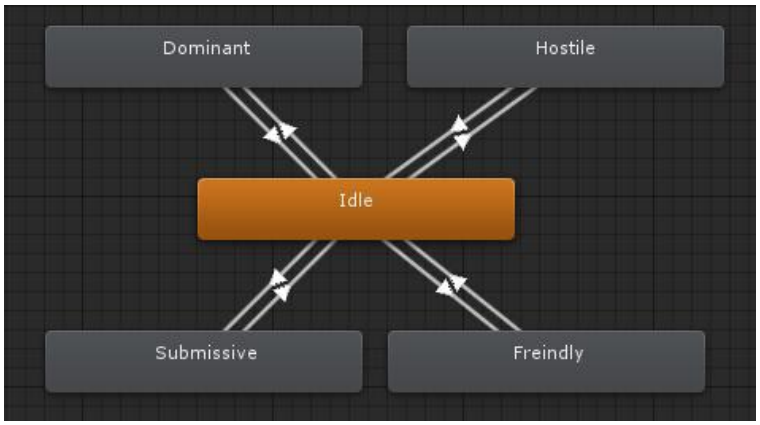

Figure 8. transitions in the state machine

3.2.2 Avatar Body Mask and Animation Layers

Unity uses Animation Layers for managing complex state machines for different body parts. Specific body parts can be selectively enabled or disabled in an animation using a so-called Body Mask. The Body Mask assets can be reused in Animator Controllers, when specifying animation layers. The body parts include: Head, Left Arm, Right Arm, Left Hand, Right Hand, Left Leg, Right Leg and Root.

In this system, there are five layers, which are base layer, head, arms, trunk and legs layer. Each layer has a corresponding body mask, except for base layer. All body parts can be animated in base layer. For example, a mask (Figure 9) that enables only the animation of the head was applied to head layer. The weight parameter acts as a switch of a layer. The layer is switched off, if the weight equals 0, and it is switched on, if the weight is above 0. The weight also controls the percentage one layer overrides the base layer. If the weight of head layer is 0.8 and the weight of other layers is 0, the Unity will blend 80% of the animation in head layer with 20% of the animation in base layer. The base layer is always 100% on, and the main state machine was located in the base layer. The postures in Table 5 (page 17) were re-assigned to the corresponding layer, according to the body part that needs to be moved. An example is given in Figure 10, showing all the states in the head layer. Each state is assigned with one animation. We can make the avatar change a posture or perform an action in real-time by turning on the intended layer and specify the state name.

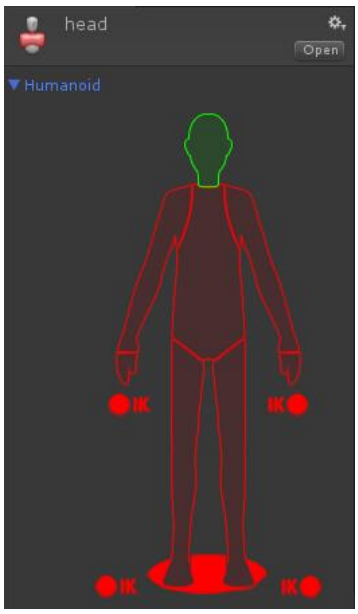

Figure 9. a body mask for head layer (green means enabled, red means disabled)

Figure 10. behaviors in head layer

3.2.3 Blend Shapes

Facial shape keys were imported into Unity as blend shapes. The facial animation is completely generated in real-time by setting the weight of blend shapes. The weight of blend shapes controls the intensity of an expression. One can adjust the influence of the blend shape to the default shape, 0 means the blend shape has no influence and 100 means the blend shape has full influence. It's also possible to set the blend weights through code using functions like GetBlendShapeWeight and SetBlendShapeWeight. The layout of the blend shapes for each avatar must be the same so that the code which controls facial animation can be re-targeted to different avatars.

3.2.4 Re-usable Source

Animation

Because of the similarity in bone structure, it is possible to map animations from one humanoid skeleton to another, allowing retargeting and inverse kinematics. With rare exceptions, humanoid models can be expected to have the same basic structure, representing the major articulate parts of the body, head and limbs. The Mecanim system makes good use of this idea to simplify the rigging and control of animations. A fundamental step in creating an animation is to set up a mapping between the simplified humanoid bone structure understood by Mecanim and the actual bones present in the skeleton; in Mecanim terminology, this mapping is called an Avatar. The animation and pose library can be retargeted to humanoid model with the same basic bone structure by applying the same animator controller.

Real-time Body Action and Facial Animation

The code to control the body posture and facial animation can be both reused. Appendix C is a function that can change the body posture. Developer must specify target layer and state, the start time, endurance, weight of the layer, speed of blend-in and blend-out in this function. It can be applied to the avatars with the same animator controller. The function in Appendix D is for real-time facial expression generation, which allows us to control when to show what facial animation for how much time and the intensity of the expression. This can be reused by avatars with the same layout of blend shapes as in Table 4 (Section 3.1.2).

Chapter 4 User Test

This chapter explains the purpose and detailed procedure of the user tests that investigates people's perception of virtual suspects. We managed to invite 30 participants to the first user test. 27 of them are university students. The data of the experiment were recorded carefully, and later processed to answer the research questions.

In the first user test, there were objects in the interrogation environment that might have an influence on participants' perception of the virtual suspects' attitudes. In order to evaluate the influence, a second user test was carried out.

4.1 Hypothesis

To answer the research questions mentioned in chapter 1.2, the following hypotheses are proposed based on the theories in chapter 2.2:
-H1: compared to the result of online survey in the previous study, the average perceived attitude of the virtual suspects is the same as that of the human actors.
-H2: The perceived controlling and affiliation of each suspect differ from that of another. While showing the same facial expression and body posture, people perceive (a) the mesomorphic suspect as more dominant, and (b) the endomorphic suspect as more friendly and (c) submissive. The stereotype against the ectomorphic suspect is hard to be determined in this case, because the suspect is ectomorphic, attractive and a female, multiple stereotypes must be taken into account.
-H3: The presence of audio results in (a) more agreement on subjects' judgments and (b) makes the virtual agents more believable, compared to mute fragments.
-H4: Comparing with and without Oculus Rift, (a) there is no significant difference in the judgments. (b) People think the suspect is more realistic when playing with Oculus.

4.2 Experimental Design

The scene started with a suspect sitting in front of the subjects across the table as in Figure 11. In the interrogation room, there is a table, a lamp on the table and a chair for the suspect to sit on. Four objects were chosen as a medium for the interaction between the subjects and virtual suspects. The four objects are knife, coffee cup, handcuffs and cupcake (as in Figure 11). They were created with the intention of offering a more interactive environment. However, some participants reported that

this might affect their perception of the suspects' attitudes. The influence introduced by the objects on the subjects' perception was evaluated in Section 5.4.

Figure 11a. Suspect 1 in idle state

Figure 11b. Suspect 2 in idle state

Figure 11c. Suspect 3 in idle state

Five sessions were set up (Table 6). All participants went through all five sessions. Suspect 1 who is the mesomorphic male was selected to be compared with and without audio. The sounds were collected from the original fragments from previous

study (in the previous study, all fragments were edited to be mute). They were all male voice that sounds more like mesomorphs than endomorphs. In order to avoid the learning effect, the suspects were designed to respond differently to each object. The animations (both facial expression and body posture) were exactly the same for each suspect, but they were displayed in different order with respect to different suspect. The response pattern is shown in Table 7. For suspect 1, the order of playing the attitudes is submissive, hostile, friendly and dominant. For suspect 2, the playing order of the animations is dominant attitude, submissive attitude, hostile attitude and friendly attitude. For suspect 3, the playing order is hostile attitude, friendly attitude, dominant attitude and submissive attitude.

Table 6. five sessions in the experiment

	Suspect	Oculus	Audio
Session 1	Suspect 1	Yes	No
Session 2	Suspect 2	Yes	No
Session 3	Suspect 3	Yes	No
Session 4	Suspect 1	Yes	Yes
Session 5	Suspect 1	No	Yes

Table 7. intended stance of each suspect w.r.t different object

	Knife	Coffee Cup	Handcuffs	Cupcake
Suspect 1	Submissive	Hostile	Friendly	Dominant
Suspect 2	Dominant	Submissive	Hostile	Friendly
Suspect 3	Hostile	Friendly	Dominant	Submissive

Participants were first briefly introduced about the procedure of the experiment. With Oculus covering the subjects' eyes, they were not able to press the keyboard and select the objects themselves. The researcher chose the objects one by one from left to right (Figure 11). Each time one object was selected, one animation was played. After watching one animation, participants were asked to rate the suspect's attitude from very dominant to very submissive, and then to rate from very friendly to very hostile. This process was repeated for every animation. Participants judgments were measured and collected with a 5-point Likert scale (as shown in Figure 12). The responses were recorded by the researcher. By the end of session 4, they were asked about their subjective feelings about the presence of sound (Appendix E). In session 5, they were asked to compare playing with and without Oculus (Appendix E). After playing the animations in session 4, the Oculus was removed from participants' head. The questions of participants' feelings about the sound and the questions in session 5 were filled in by participants themselves.

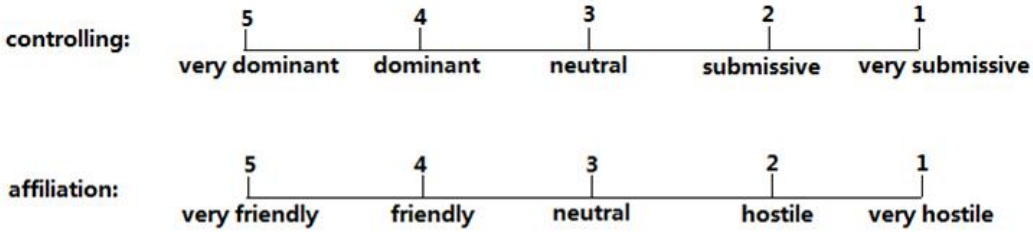

Figure 12. 5-point Likert scale of attitudes

During the experiment, the participants were allowed to change the orientation of the view by rotating their heads (when playing with Oculus Rift) or move the mouse (when playing without the Oculus).

4.3 Result

This Study vs. Previous Study [15]

H1: comparing with the result of online survey in the previous study, the average perceived attitude of the virtual suspects is the same as that of the human actors.

Participants judge the attitudes of the suspect with a 5-point Likert scale (Figure 12). The attitude of the suspect in a fragment is determined by the average rating of all judgments on this fragment. If the average rating on affiliation is bigger than 3 (neutral), then the attitude is classified as friendly, otherwise as hostile. If the average rating on controlling is bigger than 3 (neutral), then the attitude is classified as dominant, otherwise as submissive. The higher the rating is, the more dominant or friendly the suspect is, and the lower the rating is, the more submissive the suspect is. Table 8, 9 and 10 show the average rating and the attitudes that subjects classify each animation into. Table 11 shows the result of previous study. Since the fragments in the previous study are all mute, only the sessions that subjects play without audio are compared to the previous study. The ratings on the attitudes of each suspect and the result of the previous study are visualized in Figure 13, which includes the results in Table 8, 9, 10 and 11. The difference between the ratings of each suspect and the previous study is quite obvious.

As mentioned before, the animations were made to imitate the best identified fragment from the previous study. It is assumed that the identified category of interpersonal attitude is the same as in the previous study. Stances are combinations of attitudes on two dimensions. There are four categories of stances that the fragments can be categorized into. They are dominant-hostile (DH), dominant-friendly (DF), submissive-hostile (SH) and submissive-friendly (SF). However, when we tried to identify the best recognized fragment in the previous study, we only marked the fragments as most significant in one dimension. Four fragments were identified as most dominant, submissive, friendly and hostile

respectively. This is because one fragment might be significant in one dimension, while not significant in the other dimension. Take the most submissive fragment in Table 11 for an example. The rating on controlling is the lowest as 2.17, which indicates people perceive this fragment as quite submissive, while the rating on affiliation is close to neutral (3.06). This also makes it easier to categorize the typical behaviors into each attitude (e.g. it is easier to identify a dominant behavior than a dominant-hostile behavior). The intended attitude in this research is also measured in one dimension, since the intended attitude is the attitude that the fragments are found to be most significant in. There are also consequences of identifying the best recognized fragments in one dimension. As can be seen from Table 11, the best fragments include two DF stances, however no SH stance is included. The outcome is the categorized stance according to the participants' ratings on two dimensions.

Table 8. classification result of suspect 1

Intended Attitude	Dominant	Submissive	Friendly	Hostile
Controlling	2.83	2.60	3.17	3.47
Affiliation	3.60	3.43	3.63	2.43
Outcome	SF	SF	DF	DH

Table 9. classification result of suspect 2

Intended Attitude	Dominant	Submissive	Friendly	Hostile
Controlling	2.93	2.37	3.33	3.23
Affiliation	3.60	3.73	3.93	2.60
Outcome	SF	SF	DF	DH

Table 10. classification result of suspect 3

Intended Attitude	Dominant	Submissive	Friendly	Hostile
Controlling	2.37	2.17	2.90	3.07
Affiliation	3.10	3.47	3.33	2.17
Outcome	SF	SF	SF	DH

Table 11. best fragments w.r.t different attitude (result from previous study)

Best in	Dominant	Submissive	Friendly	Hostile
Controlling	4.02	2.17	3.21	3.73
Affiliation	3.16	3.06	3.41	2.17
Outcome	DF	SF	DF	DH

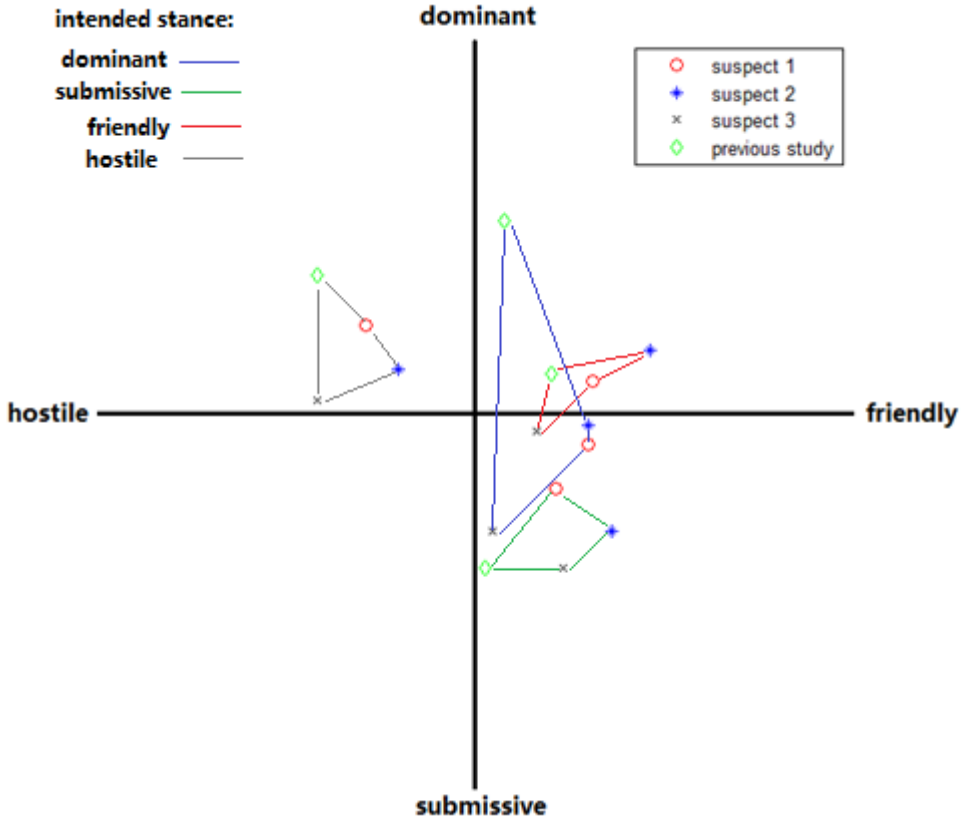

Figure 13. ratings of each suspect's attitude

The most significant difference was found in the intended dominant fragment (marked with blue line in Figure 13), where all three suspects were categorized as submissive. While the most dominant fragment (with highest average rating on controlling, 4.02) in the previous study, the ratings in this study for three suspects are all lower than 3, with the lowest rating for suspect 3 which is only 2.37. Suspect 3 is the endomorphic male which has a stereotype to be perceived as submissive [37]. The other three fragments were categorized into the same stance as in the previous study. Another observation is that the rating gap on controlling between the most dominant and most submissive fragment is as large as 1.85 in the previous study. However in this study, the average rating gap on controlling between the dominant and submissive fragment for suspect 1 and 3 is about 0.2, and for suspect 2 is 0.6, which are all smaller than the rating gap in the previous study. The difference can be a stance of regression towards the mean. In statistics, regression toward the mean is the phenomenon that if a variable is extreme on its first measurement, it will tend to be closer to the average on its second measurement—and, paradoxically, if it is extreme on its second measurement, it will tend to have been closer to the average on its first. However, this may also tell us the perceived difference in controlling is smaller, comparing virtual suspect to real human actors. Research has shown that when employing virtual humans, one should exaggerate their expressions [17].

Comparing Between Suspects

H2: *People perceive (a) the mesomorphic suspect as more dominant, and (b) the endomorphic suspect as more friendly and (c) submissive.*

As can be seen from Figure 13, participants have different perception of different suspects' attitudes regarding the same intended stance. To verify this difference, statistical analysis is applied. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test that is suitable for comparing two related Likert scale data samples. It is used to compare two sets of scores that come from the same participants. In our case, a Wilcoxon signed-rank test was applied to verify whether there is a difference in participants' perception of different suspects. With 30 participants and 4 animations per virtual suspect, there are 120 judgments per suspect on one dimension.

Table 12. Mean of all the judgments on each suspect

	Suspect 1	Suspect 2	Suspect 3
Mean on controlling	3.02	2.97	2.63
Mean on affiliation	3.27	3.55	2.54

suspect 1 (mesomorph) vs. suspect 2 (ectomorph)

The Wilcoxon signed-rank test showed that there is no significant difference (Z=-0.552, p=0.581, two-tailed) in the judgments on controlling comparing suspect 1 and 2. The difference is bigger in the judgments on affiliation, however, still not statistically significant (Z=-1.56, p=0.119, two-tailed).

suspect 2 (ectomorph) vs. suspect 3 (endomorph)

Suspect 2 is significantly more dominant (Z=-2.796, p=0.005, two-tailed) and friendly (Z=-6.572, p=0.000, two-tailed) than suspect 3.

suspect 1 (mesomorph) vs. suspect 3 (endomorph)

The result is the same as before. Suspect 1 is significantly more dominant (Z=-2.914, p=0.004, two-tailed) and friendly (Z=-4.980, p=0.000, two-tailed) than suspect 3.

Stereotypes against Mesomorphic Suspect

Mesomorphic suspect is more dominant than endomorphic and ectomorphic suspects regarding the average rating on controlling (Table 12). The difference is significant when comparing to mesomorphic suspect, while not significant when comparing to ectomorphic suspect (**H2a is rejected**). The possible reason that endomorphic suspect is not significantly more dominant than the ectomorphic suspect is that the ectomorphic suspect in this experiment is attractive and attractiveness elicits dominance.

stereotypes against ectomorphic suspect

There is no difference in the perceived dominance of mesomorphic and ectomorphic suspects. The ectomorphic suspect is perceived as friendlier than the mesomorphic suspect and the endormorphic suspect, even though ectomorph is inclined to be hostile. This may be because of the other two appearance factors that the suspect is a female and good-looking. Both of the two factors tend to elicit high friendliness.

stereotypes against endomorphic suspect

Endomorphic suspect is significantly more submissive and hostile than the mesomorphic and ectomorphic (**H2b is rejected, H2c is supported**). The assumption that endomorphic individual is considered as friendly doesn't stand in this experiment. The cause might be the neutral face that has no added expression of the endomorphic suspect looks like a frown face, while the ectomorphic suspect's neutral face is with a slight smile (as shown in Figure 14). According to [8], smiling agents are perceived as remarkably more friendly. This also explains why the etcomorphic suspect has the highest rating on affiliation (Table 12).

Figure 14. neutral faces of ectomorphic and endomorphic virtual suspects

Mute vs. Sound

H3: *The presence of audio results in (a) more agreement on subjects' judgments and (b) makes the virtual agents more believable.*

As mentioned before suspect 1 who is the mesomorphic male was selected to be compared with and without audio, because the sounds collected from previous study were all male voice. Table 13 explains the difference in the judgments comparing mute fragments and fragment with sound. There is a significant difference between the judgments of playing with and without sound, except for fragment 4 (Table 13), which is the intended dominant stance. The average ratings of fragment 4 stay the same on controlling (Table 14) and have a 0.1-difference on affiliation (Table

15). The most significant change is found in the affiliation rating of the intended friendly fragment. This fragment is identified as friendly (score 3.63 on affiliation in Table 15) when it's mute, while as hostile (score 2.67 on affiliation in Table 15) when it's with sound. Apart from that, people categorized the fragments into the same stance comparing the two conditions. However, with audio, participants judged dominant fragments as more dominant, while submissive fragments were rated as more submissive, same rule applies for ratings on affiliation. Other research has also indicated that synchronized sound with the behavior can reinforce our perception of interpersonal attitude [28]. It can be seen from Table 16 and 17 that the standard deviations of judgments on mute fragments are all higher than fragments with sound. The presence of audio results in more agreement on subjects' judgments (H3a is supported).

27 out of 30 participants thought that it was easier to judge the interpersonal attitude when there was audio, and 29 participants thought the virtual suspect was more realistic (H3b is supported).

Table 13. Mute vs. Sound

	Fragment1	Fragment 2	Fragment 3	Fragment 4
Controlling	Z=-2.231, p=.026	Z=-2.627, p=.009	Z=2.570, p=.010	Z=-1.24, p=.901
Friendliness	Z=-2.839, p=.005	Z=-1.098, p=.272	Z=-3.168,p=.002	Z=-.465, p=,642

Table 14. the average rating on controlling

	Fragment1	Fragment 2	Fragment 3	Fragment 4
Intended Attitude	Submissive	Hostile	Friendly	Dominant
Mute	2.60	3.47	3.17	2.83
Sound	2.00	3.97	3.73	2.83

Table 15. the average rating on affiliation

	Fragment1	Fragment 2	Fragment 3	Fragment 4
Intended Attitude	Submissive	Hostile	Friendly	Dominant
Mute	3.43	2.43	3.63	3.60
Sound	3.97	2.27	2.67	3.50

Table 16. standard deviation of the rating on controlling

	Fragment1	Fragment 2	Fragment 3	Fragment 4
Intended Attitude	Submissive	Hostile	Friendly	Dominant
Mute	1.33	1.02	1.11	1.14
Sound	0.98	0.86	0.77	0.96

Table 17. standard deviation of the rating on affiliation

	Fragment1	Fragment 2	Fragment 3	Fragment 4
Intended Attitude	Submissive	Hostile	Friendly	Dominant
Mute	0.72	0.92	1.29	1.01
Sound	0.65	0.73	1.15	0.67

With Oculus vs. Without Oculus

H4: *Comparing with and without Oculus Rift, (a) there is no significant difference in the judgments. (b) People think the suspect is more realistic when playing with Oculus*.

The Wilcoxon signed-rank test showed that there is no significant difference (Z=-0.311, p=0.756, two-tailed) in the judgments comparing playing with and without Oculus (H4a is supported). Even though 23 participants reported they preferred to play with Oculus and 26 participants thought that the suspects were more vivid (H4b is supported), it is not certain that if Oculus makes it easier to observe the suspects' behaviors. Especially for people who wear glasses, it is difficult to wear glasses and Oculus at the same time.

4.4 Evaluation of the Influence Introduced by the Objects

Since there might be uncertain influence introduced by the objects in front of the suspects, a second experiment was carried out. Cupcake and coffee cup more or less represent a friendly attitude, while knife and handcuff represent a hostile attitude. When participants regard these symbols as the attitude of the police and think that the suspect is responding to their offers, these items may affect subjects' judgments. In this experiment, we want to find out if there is an influence on the judgments caused by the items. If there is, then what the influence is (e.g. do friendly items result in a higher rating on friendliness). This user test is done one month after the first user test, so the participants' judgments shouldn't be affected by the first user test.

There are two sessions in this user test. 10 people who also took part in the first user test joined this user test. All 10 participants went through two sessions. The ectomorphic suspect was selected for the use of this experiment. In session I, participants went through the same procedure as in session 2 of the first user test. In session II, the objects were all removed as shown in Figure 15. The animations presented to the participants were the same as in session I. Both sessions were done with the Oculus Rift. The subjects were asked to judge the interpersonal attitude after watching each animation, and the responses were recorded by the researcher using the same 5-point Likert scale as in Figure 12.

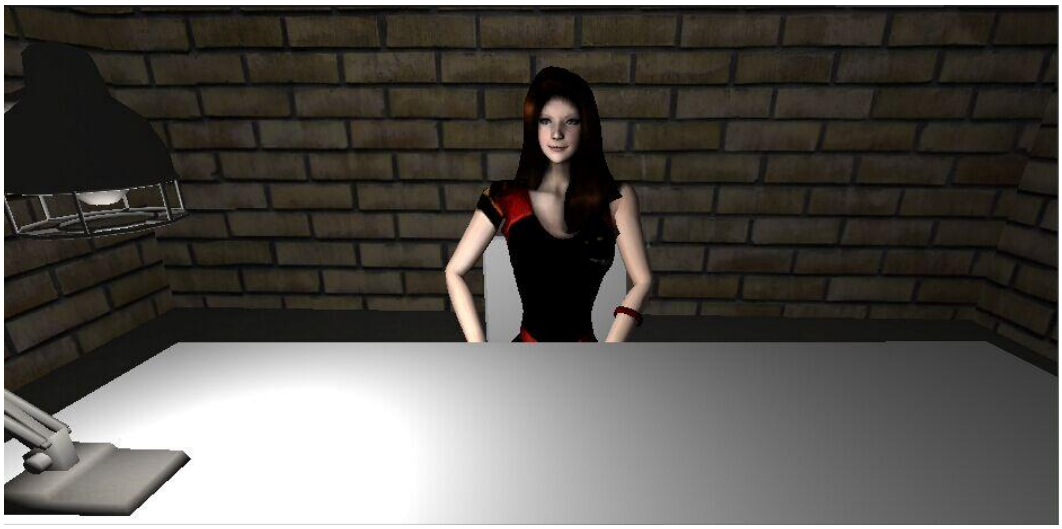

Figure 15 objects were removed in session II of user test II

The results of the two sessions in user test II were compared. For each session, there were 80 judgments (4 fragments, 10 participants and 2 questions per fragment). The Wilcoxon signed-rank test shows that the difference between judgments of the two sessions is not significant at all (p=.925, Z=-0.94). However this doesn't say much about if the objects actually have an influence on subjects' judgments, since the friendly items (coffee and cupcake) and the hostile items (knife and handcuffs) might cancel the influence on each other if we compare all the judgments together. The judgments with objects are compared separately to no objects in Table 18. Table 19 compares hostile and friendly objects to no object. Both Table 18 and 19 indicate no significant difference.

Table 18. compare each object to no object in user test II

	Controlling	Friendliness
Knife vs. no objects	Z=-1.414, p=.157	Z=-1.897, p=.058
Coffee cup vs. no objects	Z=-1.186, p=.236	Z=-1.604, p=.109
Handcuffs vs. no objects	Z=-9.66, p=.334	Z=-.108, p=.914
Cupcake vs. no objects	Z=-.073, p=.942	Z=-.264, p=.792

Table 19. compare friendly and hostile objects to no object in user test II

	Controlling	Friendliness
friendly items vs. no objects	Z=-.265, p=.791	Z=-1.290, p=.197
hostile items vs. no objects	Z=-.328, p=.743	Z=-.942, p=.346

The data collected in user test II is independent from user test I, because user test II took place one month after user test I. To compare the second user test to the first user test, a Mann-Whitney U test is applied. Mann-Whitney U test is suitable for non-normal distributed and independent data. The result (Table 20 and 21) showed that there is no statically significant difference between the judgments of two

experiments. We can conclude that the items don't have an obvious influence on subjects' perception of suspects' attitude.

Table 20. compare each object (user test I) to no object (user test II)

	Controlling	Friendliness
Knife vs. no objects	U=503.5, p=.262	U=475.5, p=.136
Coffee cup vs. no objects	U=452.5, p=.089	U=771.0, p=.648
Handcuffs vs. no objects	U=116.0, p=.241	U=116.5, p=.263
Cupcake vs. no objects	U=558.0, p=.077	U=608.0, p=.803

Table 21. compare friendly and hostile objects (user test I) to no object (user test II)

	Controlling	Friendliness
friendly objects (user test I) vs. no objects (user test II)	U=452.5, p=.089	U=561.0, p=.648
hostile objects (user test I) vs. no objects (user test II)	U=503.5, p=.262	U=475.5, p=.136

4.5 Discussion

Overall, people tend to categorize the virtual suspects into the same attitudes as human actors, regarding the ectomorphic and mesomorphic suspects. The fragments were categorized to the same stance comparing to the previous study. Except for the first fragment, which was identified as the most dominant fragment in the previous study while identified as submissive in this study. A possible cause is that the suspect is with a straight erected posture in idle state. When transiting from idle state to dominant state, the body seems to slightly bend over. However with real actors, people don't have this reference (i.e. an idle state) to compare to. Regarding further development, we should pay more attention to the movement from idle state to intended state when we make new animations.

The stereotypes against different physical appearance weren't obvious for the ectomorphic suspect, because multiple stereotypes were added to the same suspect. Gender stereotype has already been proven to the same for virtual agents [26]. It has also been revealed that a strong effect of attractiveness stereotype stands still for virtual agents [22]. Users tend to form and maintain a better evaluation of attractive agents than of unattractive ones independent of actual interaction with the agent [22]. For further investigation about stereotyped somatotypes, other physical appearance factors that affect social perception should be excluded. It is recommended to keep the face and clothing of the suspects the same, and only change the body type.

Sound plays a very important role in people's perception of social signals. It is not the actual words that matters, but the vocal nonverbal behavior or prosody that surrounds the verbal message and can change the meaning conveyed by the words. The effect of vocal nonverbal behavior is particularly evident when the tone is ironic. In this case, the actual meaning of the words is changed into its opposite by just using the appropriate vocal intonation [35]. Sound doesn't just make the virtual suspect more believable, also makes it easier for subjects to judge the interpersonal attitude.

The Oculus makes people feel more present and makes the virtual human more vivid and convincing. However, one third of the participants agreed on "it is more difficult to judge the interpersonal attitude of the virtual suspects when playing with Oculus". It is uncertain if the Oculus helps the subjects better observing the virtual suspects especially for people who wear glasses. Oculus is not recommended if the system requires user input that needs user's visual ability. In this research, the Oculus is not recommended. Because it is required for the user to press the keyboard, but their eyes were covered with the Oculus. If the system supports more intuitive input such as speech input, then the Oculus is recommended because of the immersive virtual environment it brings.

Chapter 5 Future Work

In the future, it is interesting to see if presenting more behaviors of a certain attitude in one fragment would actually result in more agreement on the judgments of the attitudes. If it does, different levels can be set to the interrogation game by adding different postures. For instance, if the intended stance is dominant and more dominant postures or expressions can be added, then it is an easy level. By setting levels to the game, it will be possible to verify if the interrogation game actually helps police trainees with their ability to grasp the social signal conveyed by suspects' behaviors. However adding planned posture or expression in real-time introduces complexity in synchronization of the speech, posture and facial expression, which is essential for the genuine believability of virtual humans. A sophisticated computational model of behavior planner is required.

In order to build a computational model of behavior planner, the interaction pattern between the police and the virtual suspect should be further investigated. According to [16] and [8], the interpersonal attitude is a combination of the personality of the person and their relationship to the other person. The response of an interpersonal attitude can take two forms, compensation and reciprocation [16]. Compensation means that the other person responds in an opposite way or attitude, for example, if one person takes a close posture the other reacts with a distant posture. The same compensation also holds for dominant behavior, which may result in a submissive attitude of the other person. However, people may also respond in a reciprocal manner, for example, during conflicts, a dominant—hostile behavior may lead to dominant—hostile behavior of the spouse [16].

The behavior planner should present a combination outcome of personality and attitude [16, 3], since personality traits have a major influence on social attitude [8]. Based on the posture generation process in [16], the behavior generation model in Figure 16 is proposed.

In an interactive system, users can choose their response to the conversational agent. The combination of user input, agent's personality and current attitudes (affiliation and controlling) determines the new attitude. The scheme of interaction pattern between the police and the virtual suspects should be applied when generating new attitude. The new attitude immediately becomes the current attitude. Interpersonal attitude can be conveyed through different modalities. Depending on the agent's personality, agents have different behavior types to show their attitudes. There are six behavior types, which are close (high affiliation), distant (low affiliation), space filling (high controlling), shrinking (low controlling), relaxation (high controlling) and nervousness (low controlling) [16]. The behavior type gives feedback to user input, so

users have options to choose from. The concrete behavior is also determined by behavior types.

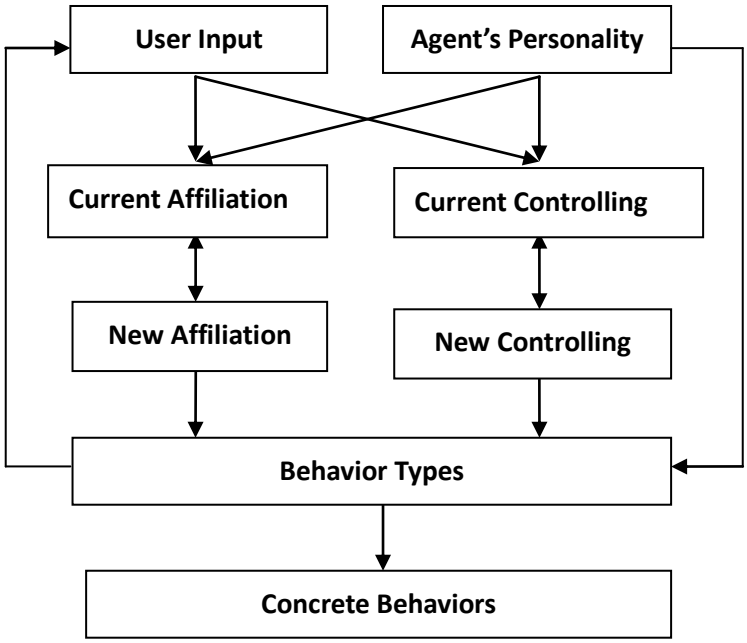

Figure 16. the behavior generation model based on [16]

In order to have a more systematic and comprehensive facial animation system, the system can be further adjusted according to Facial Action Coding System (FACS) [13] by matching the blending shapes to Action Units. There are 46 action units in FACS. Each action unit can be simulated using blend shapes, so all the facial expression in FACS can be simulated by different combination of blend shapes.

Summary

The purpose of this research is to investigate how people perceive the stances expressed by the virtual humans. In this report, we presented an interrogation environment that is capable of generating believable behaviors in real-time. Pose libraries were built for each attitude. One can import their own 3D avatars into the system, if the avatars have the same layout of the blend shapes. These avatars can be built in 3D modeling software such as Blender, Maya, and Cinema 4D.

The result of the user tests reveals that people tend to categorize the virtual suspects into the same attitudes as human actors and audio helps people judge suspects' attitudes. There are weaknesses in this study. The objects may have an influence on people's perception of suspects' attitudes. Even though the result of user test II indicates no significant difference in people's judgments after removing the objects, the number of participants is too small to ensure the conclusion. Another weakness is the lady suspect who introduces more than one stereotype. For parallel comparison, the lady suspect should be replaced at least with a male avatar.

There is still work to be done in order to fulfill the goal of the project that aims at building conversational agents for interrogation games in which police trainees can train interrogation strategies. In the future, a computational model that encompasses behavior generation scheme as in Figure 16 should be developed. The interaction pattern between the police and the suspects is still to be determined. Formulas which give weights to the user input, agent's personality and current attitude when calculating the new attitude should be introduced. Integrating the behavior generation model with the interrogation environment, we can have a conversational agent that function as a virtual suspect in a game.

Bibliography

[1].Argyle, Bodily communication, 2nd edition. Methuen, New York , 1988.
[2] N. Ambady and R. Rosenthal, Thin Slices of Expressive Behaviors Predictors of Interpersonal Consequences. A Meta-Analysis: Psychological Bull, vol. 111, no. 2, pp. 256-274, 1992.
[3] I. Ajzen, Attitudes, Personality and Behabior. McGraw-Hill International, pp 34, 2005
[4] T. W. Bickmore, Relational agents: Effecting change through human-computer relationships. Doctoral dissertation, Massachusetts Institute of Technology, 2003.
[5] T. W. Bickmore and R. W. Picard, Establishing and Maintaining Long-term Human Computer Relationships. ACM Trans. Comput.-Hum. Interact., 12(2): pp.293–327, 2005.
[6] G. Beattie and H. Shovelton, An Exploration of the Other Side of Semantic Communication: How the Spontaneous Movements of the Human Hand Add Crucial Meaning to Narrative. Semiotica., 184, pp.33-51, 2011.
[7] D. Carney, J. Hall, and L. LeBeau, Beliefs about the Nonverbal Expression of Social Power. Journal of Nonverbal Behavior, pp. 105–123, 2005.
[8] A. Cafaro, H. Vilhjálmsson, T. Bickmore, D. Heylen, K. R. Jóhannsdóttir, and G. Valgarðsson, First Impressions: Users' Judgments of Virtual Agents' Personality and Interpersonal Attitude in First Encounters. Intelligent Virtual Agents, pp. 67–80, Springer Berlin Heidelberg, 2012.
[9] M. Chindamo, J. Allwood, and E. Ahlsén, Some Suggestions for the Study of Stance in Communication. In Privacy, Security, Risk and Trust, 2012 International Conference on Social Computing, pp. 617-622. 2012.
[10] M. Chollet, M. Ochs, C. Pelachaud. Investigating Non-Verbal Behaviors Conveying Interpersonal Stances. First European Symposium on Multimodal Communication, 2013.
[11] B. M. DePaulo, Nonverbal behavior and self-presentation. Psychological Bulletin 111(2), pp 203–243 , 1992.
[12] A.H. Eagly and A. Mladinic, Gender Stereotypes and Attitudes Toward Women and Men. Personality and Social Psychology Bulletin, pp. 543-558, 1989.
[13] P. Ekman, W. V. Friesen, and Joseph C. Hager. Facial Action Coding System: The Manual on CD ROM. A Human Face, 2002.
[14] Q. Fu, R. op den Akker, M. Bruijnes, A Literature Review of Typical Behavior of Different Interpersonal Attitude, Capita Selecta HMI, University of Twente, 2014
[15] Q. Fu, R. op den Akker, M. Bruijnes, Typical Non-verbal Behavior of Different Interpersonal Stances during Police Interrogations, Research Topics report, HMI, University of Twente, 2014
[16] M. Gillies and D. Ballin, A Model of Interpersonal Attitude and Posture Generation. Intelligent Virtual Agents, pp. 88–92, Springer Berlin Heidelberg, 2003.

[17] A. Garc´ıa-Rojas, F. Vexo, D. Thalmann, A. Raouzaiou, K. Karpouzis, S. Kollias, L. Moccozet and N. Magnenat-Thalmann, Emotional Face Expression Profiles Supported by Virtual Human Ontology. Computer Animation and Virtual Worlds, pp. 259-269, 2006.
[18] U. Hess and P. Thibault., Why the same expression may not mean the same when shown on different faces or seen by different people. Affective Information Processing. pp. 145–158. Springer London, 2009.
[19] M. L. Knapp and J.A. Hall, Nonverbal Communication in Human Interaction.Harcourt Brace College Publishers, 1972.
[20] D. Keltner, Signs of Appeasement: Evidence for the Distinct Displays of Embarrassment, Amusement, and Shame. Journal of Personality and Social Psychology, 68:pp.441–454, 1995.
[21] S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth, A Conversational Agent as Museum Guide–Design and Evaluation of a Real-world Application. Intelligent Virtual Agents, pp. 329-343. Springer Berlin Heidelberg, 2005.
[22] R. Khan and A. De Angeli, The attractiveness stereotype in the evaluation of embodied conversational agents. In *Human-Computer Interaction–INTERACT 2009* (pp. 85-97). Springer Berlin Heidelberg, 2009.
[23] N. Leßmann, A. Kranstedt, and I. Wachsmuth, Towards a Cognitively Motivated Processing of Turn-taking Signals for the Embodied Conversational Agent Max. In Proceedings Workshop Embodied Conversational Agents: Balanced Perception and Action. IEEE Computer Society, pp. 57–64, 2004.
[24] J. M. Loomis, J. J. Blascovich, and A. C. Beall, Immersive Virtual Environment Technology as a Basic Research Tool in Psychology. Behavior Research Methods, Instruments, & Computers, *31*(4), pp.557-564, 1999
[25] M. McRorie, I. Sneddon, G. McKeown, E. Bevacqua, E. de Sevin, and Pelachaud, C, Evaluation of Four Designed Virtual Agent Personalities. *Affective Computing, IEEE Transactions on*, *3*(3), pp. 311-322, 2012
[26] C. Nass, Y. Moon and N. Green, Are Machines Gender Neutral? Gender-stereotypic Responses to Computers With Voices. Journal of Applied Social Psychology, volume 27, Issue 10, pp. 864-876, 1997.
[27] L.P. Naumann, S. Vazire, P.J.Rentfrow, S.D.Gosling, Personality Judgments Based on Physical Appearance. Personality and Social Psychology Bulletin 35(12), pp. 1661–1671 , 2009.
[28] M. Neff, Y. Wang, R. Abbott, and M.Walker, Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents, 10th International Conference, IVA 2010.
[29] J. A. Russell, J. Bachorowski, and J. Fernandez-Dols, Facial and Vocal Expressions of Emotion. Ann. Rev. Psychology, vol. 54, pp. 329-349, 2003.
[30] B. Ravenet, M. Ochs, and C. Pelachaud, A computational model of social attitude effects on the nonverbal behavior for a relational agent, Proc. of Workshop Affect Campaign on Artificial Interaction, 2012.

[31] B. Ravenet, M. Ochs, C. Pelachaud, From a User-created Corpus of Virtual Agent's Non-verbal Behavior to a Computational Model of Interpersonal Attitudes. Intelligent Virtual Agents, pp. 263-274, Springer Berlin Heidelberg, 2013
[32] J. Sanders, T. Smith, and J. Alexander, Type A Behavior and Marital Interaction: Hostile-Dominant Responses During Conflict. Journal of Behavioral Medicine, Volume 14, Issue 6, pp. 567-580, 1991.
[33] A. Schmeil and W. Broll, Mara: An Augmented Personal Assistant and Companion. In ACM SIGGRAPH 2006 Sketches, pp. 141, 2006.
[34] S. Spitters, M. Sanders, R. op den Akker, M. Bruijnes, The Recognition of Acted Interpersonal Stance in Police Interrogations, 4th IEEE International CogInfoCom 2013.
[35] A.Vinciarelli, M.Pantic and Herv´e Bourlard, Social Signal Processing: Survey of an Emerging Domain. Image and Vision Computing , *27*(12), pp. 1743-1759 , 2008.
[36] A. Vinciarelli, H. Salamin and M. Pantic, Social Signal Processing: Understanding social Interactions through Nonverbal Behavior Analysis. Computer vision and Pattern Recognition Workshops, pp 42-49, 2009.
[37] W. D. Wells and B. Siegel, Stereotyped Somatotypes, Psychological Reports, pp. 77-78, 1961.
[38] M. K. Young, G. B. Gaylor, S. M. Andrus, and B. Bodenheimer, A Comparison of Two Cost-differentiated Virtual Reality Systems for Perception and Action Tasks. In Proceedings of the ACM Symposium on Applied Perception , pp. 83-90, 2014 .
[39] Z. Zeng, Pantic, M., G. I. Roisman, and T. S. Huang, A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, *31*(1), pp. 39-58, 2009.

Appendices

Appendix A Glossary of Technical Terms

Armature
An "armature" is a type of object used for rigging. Armature object borrows many ideas from real life skeletons.

Rig
A rig is also a type of object used for rigging. It can be customized skeleton for any object. However armature only refers to human character rig.

Mesh
Mesh is also called polygon mesh. It is a is a collection of vertices, edges and faces that defines the shape of a polyhedral object in 3D computer graphics and solid modeling.

Rigging
After completing your character, you need to manipulate it for animation or just for posing. Rigging is the process of attaching a skeleton/armature to your character mesh object so you can deform and pose it in different ways.

Rigify
Rigify helps the process of rigging and automate the creation of character rigs. It is based around a building-blocks approach, where you build complete rigs out of smaller rig parts (e.g. arms, legs, spines, fingers). The rig parts are currently few in number, but as more rig parts are added to Rigify it should become more capable of rigging a large variety of characters and creatures.

Unity Mecanim
Unity has a rich and sophisticated animation system called Mecanim. Mecanim provides: easy workflow and setup of animations on humanoid characters, animation retargeting, Simplified workflow for aligning animation clips, convenient preview of animation clips, transitions and interactions between them, management of complex interactions between animations with a visual programming tool, and animating different body parts with different logic.

Blend Shapes
Blend Shapes, also called morph target animation, per-vertex animation, or shape interpolation, is a method of 3D computer animation. In a blend shape, a "deformed"

version of a mesh is stored as a series of vertex positions. In each key frame of an animation, the vertices are then interpolated between these stored positions.

Appendix B Blender Rigify to Unity Mecanim

```
import re
import bpy
porg = re.compile('ORG-*')
for object in bpy.context.object.data.bones:
    object.use_deform = False
for object in bpy.context.object.data.bones:
    if porg.match(object.name):
        object.use_deform = True
bpy.context.object.data.bones['ORG-heel.L'].use_deform = False
bpy.context.object.data.bones['ORG-heel.02.L'].use_deform = False
bpy.context.object.data.bones['ORG-heel.R'].use_deform = False
bpy.context.object.data.bones['ORG-heel.02.R'].use_deform = False
```

Appendix C Facial Animation in Real-time

```
IEnumerator ChangePosture(int stateNameHash, int layerIndex, float target,     float
blendIn, float blendOut, float duration, float delay=0f)
    {
        float startTime = Time.time + delay;
        float holdTime = startTime + blendIn;
        float outTime = holdTime + duration;
        float totalTime = outTime + blendOut;
        float weight=0;
         while (Time.time < startTime ) {
            yield return null;
        }
         anim.CrossFade(stateNameHash,0f,layerIndex,0.1f);
        while(Time.time < holdTime)
        {
            weight += target/blendIn * Time.deltaTime ;     //speed of blend in
            anim.SetLayerWeight (layerIndex,weight);
                yield return null; // wait for the next frame
        }
        weight=target;
        anim.SetLayerWeight (layerIndex,weight);
        while(Time.time < outTime){
                yield return null;
            }
        while(Time.time<totalTime){
            weight -= target/blendOut * Time.deltaTime ;//speed of blend out
            anim.SetLayerWeight (layerIndex,weight);
                yield return null; // wait for the next frame
        }

        anim.SetLayerWeight (layerIndex, 0     );

    }
```

Appendix D Body Action in Real-time

```
IEnumerator AnimateBlendShape(int blendShapeKey,float blendIn, float hold, float
blendOut, float target, float delay = 0f)
    {
        float startTime = Time.time + delay;
        float holdTime = startTime + blendIn;
        float outTime = holdTime + hold;
        float totalTime = outTime + blendOut;
        float value = 0;

            while (Time.time < startTime) { //delay
            yield return null;
        }
            while(Time.time < holdTime)
        {
            value += target/blendIn * Time.deltaTime ;    //speed of blend in
            sRenderer.SetBlendShapeWeight(blendShapeKey,value);
            yield return null; // wait for the next frame
        }
        value=target;
        sRenderer.SetBlendShapeWeight(blendShapeKey,value);
        while(Time.time < outTime){
            yield return null;
        }
        while(Time.time < totalTime){
            value -= target/blendOut * Time.deltaTime ;//speed of blend out
            sRenderer.SetBlendShapeWeight(blendShapeKey,value);
            yield return null; // wait for the next frame
        }
        sRenderer.SetBlendShapeWeight(blendShapeKey , 0);
    }
```

Appendix E Questionnaires in User Test I

Do you prefer to play the game with audio?

strongly prefer audio ○ prefer audio ○ neutral ○ prefer without audio ○ strongly prefer without audio ○

The virtual suspect in the game is more realistic with audio.

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

It is easier to judge the interpersonal attitude of the suspect with audio.

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

I feel more concentrated when there is audio.

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

It is easier to judge the interpersonal attitude in the game setting than the video acted out by real actors. (Only applicable if you have joined the previous survey. If not, please skip this question)

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

It is more difficult to judge the interpersonal attitude of the virtual suspects when playing with Oculus.

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

Do you prefer to play the game with Oculus?

strongly prefer Oculus ○ prefer Oculus ○ neutral ○ prefer without Oculus ○ strongly prefer without Oculus ○

The suspect in the game is more vivid when playing with Oculus.

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

The suspect in the game is easier to observe when playing with Oculus.

strongly agree ○ agree ○ neutral ○ disagree ○ very disagree ○

It is easier to judge the interpersonal attitude of the virtual suspects when playing with Oculus.

strongly agree	agree	neutral	disagree	very disagree
○	○	○	○	○