TÉCNICO LISBOA

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Improving Believable Interactions in Real-time Multi-party Interactive Experiences

Exploration on a Multi-player Multi-agent Video-game

Ricardo Eugénio Proença Rodrigues

Supervisor : **Professor** Carlos António Roque Martinho

Thesis Proposal for the PhD Degree in
Computer Science and Engineering

April 2022

Abstract

Believable interactions between synthetic characters are an important factor defining the success of a virtual environment relying on human participants being able to create emotional bonds with such characters. Not only is it important for characters to be believable, but also interactions with and between them too. This work proposes a model for synthetic characters interactions based on anticipation and emotion and that deconstructs the traditional atomic action into three stages: anticipation, action, and follow-through. Anticipation is further subdivided into interruptible and uninterruptible stages, while follow-through is subdivided into uninterruptible and interruptible stages (order of execution matters). With these divisions, the model allows for precise non-verbal affective communication, which improves the believability of both synthetic characters and their actions.

We improve on the model by expanding on its emotion expression capabilities while adapting it to *Adfectus*, a 3D real-time video-game use case where two characters battle one another while being supported by two claques that enjoy the show. A plan is proposed to implement and evaluate the improvements.

Keywords

Virtual Agents, Synthetic Characters, Believable Interactions, Emotion, Anticipation

Resumo

Interacções credíveis entre personagens sintéticas são um factor importante no sucesso de um ambiente virtual que depende da capacidade de participantes humanos criarem laços emocionais com essas personagens. Não só é importante que as personagens sejam credíveis, como é importante que as interacções com elas e entre elas também o sejam. Este trabalho apresenta um modelo para interacções entre personagens sintéticas baseadas em antecipação e emoções que desconstrói a acção tradicionalmente atómica em três etapas: antecipação, acção e seguimento. A antecipação é ainda subdividida nas etapas interrompíveis e não interrompíveis, enquanto o seguimento é divido nas etapas não interrompíveis e interrompíveis (a ordem de execução importa). Com estas divisões, o modelo permite comunicação não-verbal precisa, o que melhora a credibilidade das personagens sintéticas e das suas acções.

Melhorámos o modelo expandido as suas capacidades de expressão emocional, enquanto o adaptamos para o *Adfectus*, um videojogo modelado em 3D que é usado como um caso de estudo onde duas personagens batalham entre si enquanto duas claques os apoiam e desfrutam o espectáculo. Um plano para o implementar é proposto, assim como um plano de avaliação.

Palavras Chave

Agentes Virtuais, Personagens Sintéticas, Interacções Credíveis, Emoção, Antecipação

Dedicated to my parents, Baltazar and Gena, for raising me as you did,
my best of friends, Inês and João, for clearing my head in tough times, and
my girlfriend, Raquel, for calming me even when I did not know I needed to.
Without you, none of this would have been possible.

***The oldest and strongest emotion of mankind is fear,
and the oldest and strongest kind of fear is fear of the unknown.***

H.P. Lovecraft

Acknowledgments

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with ref. UID/CEC/50021/2019, and FCT grant from project Tutoria Virtual with ref. TDC/IVCPEC/3963/2014.

A thank you to professor Carlos António Roque Martinho, for the long meetings that made this thesis such a good experience and for always being available, and for taking time from his family to help me.

Also, a thank you to the Masters' students I had the pleasure to work with, André Lima, Ricardo Silva, Ricardo Pereira, Taíssa Ribeiro, and João da Silva.

Contents

List of Figures

List of Tables

Acronyms

AAMAS International Conference on Autonomous Agents and Multi-Agent Systems. 27, 70

ACII Conference on Affective Computing and Intelligent Interaction. 27, 70

AI Artificial Intelligence. 13, 27, 51, 72

CoG IEEE Conference on Games. 27, 70

HAI International Conference on Human-Agent Interaction. 27

HUD Heads-up Display. 65, 66

IVA International Conference on Intelligent Virtual Agents. 27, 70

NPC Non-Player Character. 2, 4, 15, 16, 28, 55, 58

SRC Social Regulatory Cycle. 33

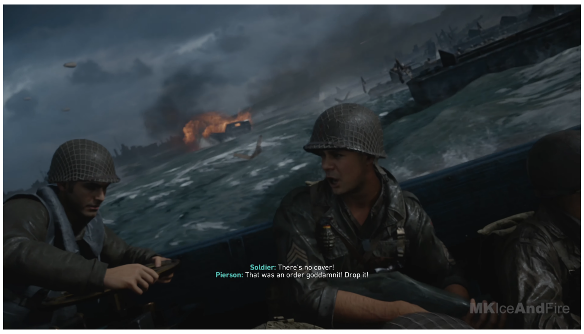
SUS System Usability Scale. 40

1 Introduction

Contents

1.1 Motivation

In today's cinematographic world, many movies conquer the audiences and create an imaginary world where the audience loses themselves. This is called immersion, a state of being deeply engaged or involved[1]. But how can movies create such immersion? Let's take for example movies like "The Lord of the Rings"[161] and their fervorous battles, when a character moves or attacks mid-combat it is always clear to the viewer their intention, either by the movement of their eyes or their body: when Legolas shoots his arrows to a distant target, the target is usually shown first, then the camera passes to Legolas preparing and shooting the arrow, and then the target again, being shot[1]. This flow creates anticipation in the audience and a sense of presence, where the audience can feel as being there, which refers to what is called the suspension of disbelief, the notion that the implausibility of something can be suspended for the sake of enjoyment[2].



(a) In-game cinematic showcasing two characters interacting in the game *Call of Duty: WWII*[163][2].

(b) In-game interaction with a plot-relevant character in the game *The Witcher 3: Wild Hunt*[156][3].

Figure 1.1: Examples of pre-scripted interactions in video games that showcase believable interaction between characters.

Many video games try and succeed in creating the suspension of disbelief by introducing pre-scripted scenes and narrowing the player's playable area and actions, giving it a more cinematic feel (see Figure 1.1). Take for example the first-person shooter game's genre (e.g. *Call of Duty: WWII*[163], *Battlefield V*[158], and *Halo Infinite*[153]), with a campaign where one is put in a soldier's perspective in the midst of a modern/futuristic war, with frenetic and over-the-top scenes on par with many of today's action movies. In this setting, the player can easily feel immersed and feel that they are cooperating with their synthetic companions, mainly because of the constraints of the game. Yet, in the Open-World Role-playing Games genre (e.g. *The Witcher 3: Wild Hunt*[156], *Horizon Zero Dawn*[159], and *Fable II*[160]) this kind of immersion is hard to achieve, mainly because of the interactions between synthetic characters or Non-Player Characters (NPCs) and the player. Often the player can only interact with a few plot-relevant characters and the way one can interact with them is by a set of dialogue options or predefined actions (see Figures 1.1b and 1.2), making such interactions appear unnatural and breaking the suspension of disbelief.

[1] Multiple examples of this character's behavior can be viewed in the video with the link `https://www.youtube.com/watch?v=Ousk2CSMvWI`.

[2] Image collected from the video of *Call of Duty: WWII* gameplay with the link `https://www.youtube.com/watch?v=F1Y2h2evV1Y`.

[3] Image collected from the video of *The Witcher 3: Wild Hunt* gameplay with the link `https://www.youtube.com/watch?v=F1Y2h2evV1Y`.

Figure 1.2: In Fable II[160], it is possible to interact with most characters using a set of predefined actions that will increase or decrease a value on an "attractiveness" scale.

The same issues apply when a more action-oriented interaction is underway (e.g. combat, racing, fleeing, parkour). Many times there is no verbal or non-verbal communication between characters, creating a mechanical and oversimplified battle sequence with no sense of immersion. It is good to point out that much of human communication is made in a non-verbal way, by the movement of the hands or eyes, indicating the other person's intent[3–5]. Emotion expression is also essential to correctly perceiving and delivering an intention [4], and, in many video games and interactive experiences, the expression of emotions is limited to pre-scripted scenes. In *The Lord of the Rings* movies, during a battle, characters frequently show their intentions through gestures, guiding their colleagues to safety and building on the audience's expectations. These interactions lead to the topic of Believability, for now, we will use its more obvious linguistic denotation that something can be believed by someone [6] and that when an interaction is believable, it is maintaining the suspension of disbelief.

Meanwhile, the study of Believability in the Academia is often catered to the study and replication of human behavior [6–8], also regularly connected with emotions [9–11]. In this area the exploration of multi-party interactions is common [10, 12] and focus is given to the simulation of behavior [13, 14]. While the role of real-time interactions is present in many works [7, 12, 15], the focus is not often on the enjoyment of the experience (as in the Game Industry), but rather on determining how believable the characters or the experience are.

Believable interactions are explored differently in Academia and the Game Industry and to different ends, we hope to join the real-time interactions more often seen in the Game Industry with the multi-party interactions more often explored in Academia to create Believable Interactions in an interactive environment.

1.2 Problem

Video games today strive to lead their player audience into a sense of presence by creating more believable characters and interactions, promoting the suspension of disbelief. Yet we are still far from achieving real-time interactions that the audience would classify as believable without any doubt. Elsewhere, the Academia is striving to explore multi-party interactions and is creating more believable characters interacting with each other, yet they are still far from achieving a level of believability seen in pre-scripted scenes. One problem stems from the lack of verbal and non-verbal communication between characters (being those controlled by the player or NPCs), especially on the duration of a single important action (sometimes emphasized by the use of slow-motion).

Faced with this problem, **how can we enable believable multi-party real-time interactions between human and synthetic characters?** Let us break the question down into its different components. *Believability* refers to interactions that give an illusion of life (it will be discussed in detail in Section 2.1). *Multi-party* refers to the presence of multiple parties in a scene (two or more) that interact with each other, for example having two NPCs interacting with each other and with the player character (note the importance of interactions between synthetic characters and not only with the player and their character). *Real-time* in this context refers to the dynamic development of interactions, where players experience the scene while actively interacting with it (as opposed to the use of pre-scripted scenes that remove control from the player while displaying an almost cinematic video). With these concepts in place, we can now tackle this problem in its proper context.

1.3 Hypothesis

Our hypothesis is that by *subdividing actions according to traditional principles of animation*[9] and allowing for *emotional reactions and behavior interruption* during their expression, the overall *real-time interaction will be perceived as more believable*. The division of an action results in three stages: anticipation, action, and follow-through. Each stage takes some time to play out, allowing for emotion expression and interpretation to be done under different contexts of the same action, either by the performer or anyone (human or synthetic) that spectates the action unfold (often also performing an action themselves). By explicitly modeling a split of an action into the three stages, we are able to communicate the intentions of a character in a clearer way and give a richer emotional context for all the characters involved in the scene, consequently improving believability.

Let's demonstrate a scenario showcasing the use of action subdivision. In this hypothetical fantasy-inspired scenario we will see two characters, Rua and Gorm, during a fight in an arena until one falls.

> Gorm is holding a two-handed axe and Rua a long spear. The battle starts and both fighters run toward each other. As they get closer Gorm decides to perform a very slow but highly destructive overhead attack by pulling his axe back, revealing his intention in the process. Rua perceives this and shows confidence because she believes the attack

will miss. She then decides to thrust her spear towards Gorm. As a result, and still, during the anticipation stage, Gorm becomes scared and decides to cancel his attack to defend himself. Rua's attack then advances to the action stage and is blocked by Gorm, leading Gorm to feel relieved and Rua irritated. Both fighters move away from each other, during the follow-through, to prepare for the next attack, and the battle proceeds.

As shown by this scenario, our model ensures that the characters express emotions at the correct timing, which we argue will make interactions more believable.

In a previous work [16], we proposed *3motion* as a model that implements action subdivision, we will expand on this model and use it to support the evaluation of our hypothesis. Furthermore, we plan on also implementing an approach with an atomic action (i.e. without the subdivision into stages), thus considering only the beginning of the action and its end. This approach would simulate a classic view on actions, seen in many interactive experiences nowadays. With these two approaches, we can compare how each performs in terms of believability in a future evaluation.

1.4 Contribution

The main contributions of this work are:

- Review of the state of the art on the concepts of believability and emotions and their expression, specifically in synthetic characters;
- Definition of a computational model for synthetic characters communicating through an action split between anticipation, action, and follow-through;
- Implementation of *Adfectus*, an arena game where two characters battle one another, as a use case to test our hypothesis;
- Evaluation with users to assert the viability of the proposed approach when compared with a classic approach.

1.5 Outline

In the next chapter, we start by defining believability and what are believable characters. We then move to emotions, their expression, and their connection with synthetic characters. Connecting both believability and emotions, we discuss character animation through the principles of traditional animation. We close the chapter by discussing recent works that, like ours, tackle emotional multi-party interactions among synthetic characters. We then move to a new chapter, discussing the work we did on the expression of emotions in virtual coaches and how it impacts the way perception of emotions should be interpreted, also discussing the implementation of the model used and its emotional system. A chapter detailing the model implementation is then presented, contemplating the action subdivision, an illustrative scenario, and preliminary evaluations on a text-based application use case. This is followed by a chapter presenting a discussion on the future work and the implementation of *Adfectus*. Finally, a conclusion chapter is presented.

2

Related Work

Contents

In this chapter we present different but intertwined subjects that are relevant to our implementation and the context it is in. We start by discussing believability, its definition, and what are believable characters. We then move to emotions, their expression, and their connection with synthetic characters. Connecting both believability and emotions, we discuss character animation through the principles of traditional animation. We then discuss other recent works that discuss emotional multi-party interactions among synthetic characters. All these topics help create the context for the definition of 3motion *and the creation of* Adfectus*.*

2.1 Believability

In its more common linguistic denotation, believability means that something can be believed by someone [6], yet there is still no generally agreed or precise definition of believability, instead there is a "family of related meanings denoted by the same word"[6]. The entertainment industry gradually linked believability with the audience's engagement during a performance, while in the context of Artificial Intelligence and video games, we can add that something about a character or even the character itself is believed to be living by someone[7, 8]. Some authors take a *realistic approach*, stating that believability is defined as "someone believes that some character or bot is real" [6, 17] and many recent works give attention to a realistic portrayal of virtual humans [18–21] aligned with this view on believability. Others take an *artistic approach*, focusing on the ability of a character to provide an *illusion of life* [8, 9, 11, 22], meaning "something [...] that appears to be living, that has an inner strength, a vitality, a separate identity – something that speaks out with authority – a creation that gives the illusion of life"[9]. Furthermore, this definition is regularly associated with the audience's *suspension of disbelief*[11, 13, 22] and it refers to the audience's ability to accept "limitations in the presented story, sacrificing realism, and occasionally logic and believability, as well as the media content's aesthetic quality for the sake of enjoyment"[2].

Togelius *et al.*[6] states that behind the definition of believability there are two broad classes of examples: *Player Believability*, "someone believes that the *player* controlling the character is real, i.e. that a human is playing as that character instead of the character being computer-controlled"[6]; *Character Believability*, someone believes that the character itself is real in a certain context.

Player Believability assumes the observer knows the character isn't real and that he "believes that a human has an ongoing input to and control over these processes, and that the human's control is interactive in the sense that the human is aware of what the character is doing in the game"[6].

Character Believability has its research mainly inspired by the roots of two other fields: *drama* and *animation*. When referring to *drama* we can go back to the ancient Greeks, where, according to Aristotle, a believable character "should be able to (1) capture, (2) represent, and (3) project believable states. Whether a believable character possesses any of the properties is irrelevant: it only needs to appear to have them"[7]. Additionally, Prendinger and Ishizuka[23] claim that realistic-looking characters' performance has high expectations from the audience, meaning that little deficiencies lead to user irritation and dissatisfaction, as opposed to an obvious synthetic embodiment. This is also discussed in the problem of the *uncanny valley* where almost, but not completely, real characters tend to be "creepy"[24] and elicit negative emotions in humans.

In *animation*, let us take as an example Chuck Jones' Road Runner[1], a cartoon where a coyote (Wile E. Coyote) repeatedly attempts to catch and subsequently eat the Road Runner without success. During a chase when the Road Runner runs off a cliff and Wile E. Coyote follows, they are both able to stop in mid-air, but it is only when the coyote looks down, becoming aware of his situation, that he falls, while the Road Runner continues running and escapes. "This is the audience's believable world, and the audience will never question the fact that it is not realistic. It is just the way how the world works"[7]. If the rules were broken, believability would be lost and the audience's suspension of disbelief as well: it wouldn't be realistic in the world they created.

Therefore a believable character has to be consistent with its world and also give the illusion of life, "the change of shape shows that a character is thinking, but it is the thinking that gives the illusion of life, and it is life that gives meaning to the expression"[9, 25].

2.1.1 Definition of Believability and its Components

Having explored some different views on Believability, we can now define what our working definition will be. Giving focus on the more artistic approach, we define ***believability*** **as "the belief that an environment, its characters, and their interactions give the illusion of life"**.

While Character Believability has been discussed, the *environment* and the *interactions* need to be detailed. An environment gives the audience a context and helps to create expectations of what is or isn't possible. A *believable environment* must create a consistent context where the characters will interact in and interact with. This context encompasses visual style – where matching it with the characters' design is important to place the characters in the world –but also encompasses the affordances it provides to the characters, even if they don't employ them all (e.g. if the setting is realistic, having a character do super-human actions should be justified, otherwise it will appear uncanny). Therefore, a *Believable Environment* is defined as "a place that establishes a context that allows for consistent believable interactions by characters and with characters that give the illusion of life."

This leads us to the question of what is the definition of Believable Interactions? Similar to the previous definitions, believable interactions should provide consistency, in this case of behavior. The interactions performed by a character should be consistent with the character and their intentions, if the interaction is with the environment, then it should respect the context provided, if more characters are involved, the interaction (or multiple concurrent interactions performed by all the characters) should be coherent with the characters, the context, and the relationships between them. Therefore, a *Believable Interaction* is defined as "an interaction that allows for the expression of an action while respecting the restrictions of the environment and being aligned with the characters' intentions and their relationships." A diagram depicting the connection between the three presented topics on believability can be seen in Figure 2.1.

[1] Wile E. Coyote and The Road Runner. Chuck Jones.

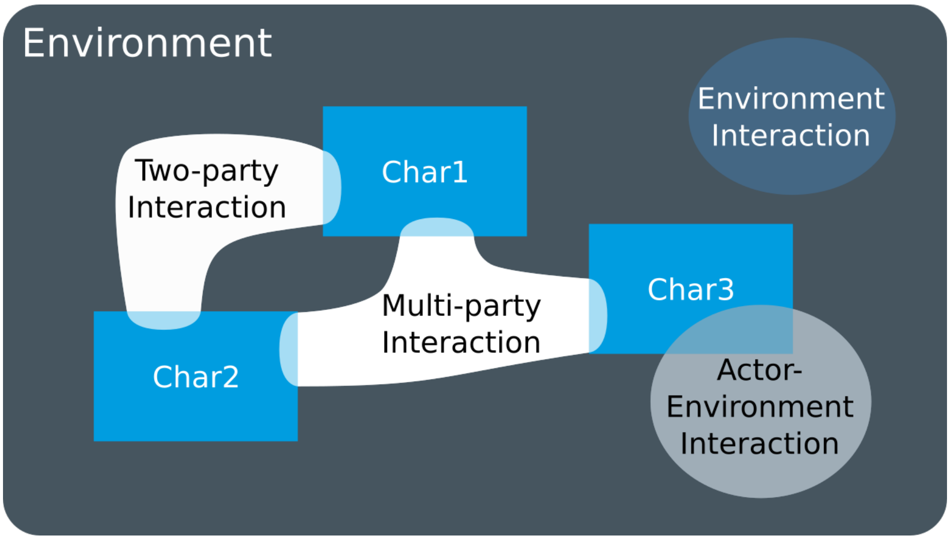


Figure 2.1: Diagram depicting the connection between believable environment (surrounding rectangle), characters (blue rectangles), and interactions (rounded shapes). A believable character exists within an environment, as without it there is no grounding to create the suspension of disbelief (the environment can be simple, but must ground the characters in it). Interactions exist between characters, not only between a couple (two-party interaction between Char1 and Char2) but also multi-party (between Char1, Char2, and Char3), and between them and the environment (actor-environment interaction with Char3). The model also considers interactions without characters, as if the environment interacts with itself (e.g. a tree falling in the forest without anyone being there).

In this work, we focus on interactions, specifically those between Believable Characters (independently of who controls the characters, human or machine) with special attention given to emotion expressions and non-verbal communication. The environment will be given less attention but should be able to provide enough context so that the audience can focus on the characters and their interactions (e.g. provide a floor and walls the characters can navigate and be situated in). Essentially, we focus on *emotional non-verbal multi-party believable interactions*.

2.1.2 Believable Agents

When one combines the artistic aspects of believability and artificial intelligence, the result is what is known as believable agents. Their goal is to simulate believable characters in a virtual environment. As in the definition of believability, several authors, from different fields, gave their definition of a believable agent [8, 26, 27]. Bates [26] states that believable agents require "only that they not be clearly stupid or unreal". Such broad, shallow agents must "exhibit some signs of internal goals, reactivity, emotion, natural language ability, and knowledge of agents [...] as well as of the [...] micro-world"[26]. Ortony states "Believability entails not only that emotions, motivations, and actions fit together in a meaningful and intelligible way at the local (moment-to-moment) level, but also that they cohere at a more global level – across different kinds of situations, and over quite long time periods"[27]. Nayak *et al.*[8] give us a definition more closely linked to the illusion of life, defining believability as "the ability of the agent to be viewed as a living, although fictional, character", furthermore, the authors link the illusion of life and the suspension of disbelief to the creation of believability, quote "only if the user views the agent as a living character will the user seriously mentally engage him/herself with the agent."[8].

These definitions point to an agent that in order to be believable must have its own goals, be reactive and emotional, and be aware of himself and its world while remaining consistent at a local and global level.

2.1.3 Awareness and Situatedness

Along with Believability, concepts emerge that bind the characters into their environment, two relevant concepts are those of Awareness [28, 29] and Situatedness [30–33]. Awareness can be defined as one's ability to "show they perceive the world around them"[28] and "is an essential aspect of believability for virtual characters capable of interacting with humans users"[29]. Situatedness has been defined by multiple authors [30–32]. We will be making use of the following definition as "a theoretical position that posits that the mind is ontologically and functionally intertwined within environmental, social, and cultural factors"[33]. For an agent to be believable it is required that it is perceived as having both awareness and situatedness. Independently of the agent's ability to perceive its environment "correctly", it should react according to the context. If an agent hears a door close-by opening unexpectedly it is expected that the agent will react by looking at it.

2.1.4 Theory of Mind and Empathy

Some different concepts in psychology, such as *Emotion Contagion*[34, 35], *Interpersonal Affect Regulation*[36], and *Imitation*[37], share some similarities that are also present in two larger concepts, that of *Theory of Mind* and *Empathy*. Theory of Mind can be defined as "the cognitive capacity to attribute mental states to self and others"[38, 39] and is traditionally associated with logical thinking and critical for simulating social emotions[40]. Empathy can be defined as an "affective response more appropriate to someone else's situation than to one's own" [41] and is more often associated with emotional interpretation. Both are critical for creating believable agents[10, 14, 42]. These concepts are important to the context of behaviors we will implement. Empathy is given a special focus when discussing Virtual Tutoring in Chapter 3.

2.2 Assessment of Believability

Different approaches arise when discussing the measurement of Believability. Hamdy and King [43] defend that "[w]hat can be evaluated [...] are players' opinions regarding the character's behavior, interaction, the relationship it managed to forge with them, and the whole experience." This "[p]layer experience can be measured through subjective, objective, or gameplay-based approaches" [6, 43]. Subjective approaches refer to the measurement through user reports [16, 22, 28, 44], objective approaches refer to the measurement of participants' physiological responses[45, 46], and gameplay-based approaches refer to measurements based on logging of gameplay statistical data[47–49]. In this work, we will focus on assessment through user-report and through gameplay.

2.2.1 Metrics for Character Believability

We will make use of a user-report assessment based on the work of Gomes *et al.*[28] that has participants play or watch gameplay of a game and then have them express their agreement with a set of metrics about their perception of the agents and how the agents perceived other agents. The metrics, or believable dimensions, are the following:

- **behavior coherence:** The audience will evaluate the coherence of a character's behavior, which is one key aspect of believability[50].
- **change with experience:** In the context of interactive narrative, it represents how an agent changed because of a story event, a significant change in a life value of a character.
- **awareness:** The audience should perceive the agent as being aware of his surroundings.
- **behavior understandability:** The audience must understand the agent's behavior, therefore the agent must express itself in a way that its thoughts and motivations are clearly understood.
- **personality:** The agent should be perceived as a unique individual. Its behavior should suggest unique personality traits.
- **emotional expressiveness:** The agent should be able to express its emotions so that the audience can perceive them correctly.
- **social:** The audience must be able to acknowledge a social relationship between the agents.
- **visual impact:** "The agent should draw the attention of the participant."[51]
- **predictability:** An agent's behavior must be moderately predictable to the audience, meaning that a very predictable agent will harm believability as much as an unpredictable one, affecting behavior coherence[50]. The extremes should be avoided.

The audience's perception is asserted using Likert scales, one scale per dimension (see the templates for the scales in Appendix A.1). As for emotional expressiveness, participants are asked what emotions are displayed by the agent in specific moments, such as joyfulness or sadness. If the participant's answer matches the emotion the system was trying to reproduce, this dimension would score higher, lowering if the answers did not match. More details on emotion expression are presented in Section 2.4.3.

2.3 Emotions

When talking about believable characters it is impossible not to talk about emotions, being one of the major factors that make a character believable[9] (with a work going as far as defining believability as "to what extent do you think the character is feeling the emotion"[52]). Unfortunately, there isn't an exact definition for emotion. Kleinginna *et al.*[53] compiled ninety-two disparate definitions for emotions into distinct categories pertaining to the more basic psychological theory they supported (e.g. affective, cognitive, physiological, adaptive) and although there is no concrete conclusion to the definition of emotion, there is a consensus of the view that emotion is considered by most theorists, "as a bounded episode in the life of an organism, characterized as an emergent pattern of component synchronization preparing adaptive action tendencies to relevant events as defined by their behavioral meaning and seeking control precedence over behavior"[54]. We use this definition as it bounds an emotion to anticipatory behavior (shown through action tendencies) and to previously experienced events, allowing for the creation of emotion through information given by the context one is in and their personality, aligning with definitions of believability.

Note that in this definition of emotion, there is no reference to the compartmentalization of emotions, yet in society and in many works (most notably Ekman *et al.*[55, 56]) we find multiple words that represent different categories of emotions, for example Ekman *et al.*[55] defined the six basic families of emotion: *Fear*, *Anger*, *Disgust*, *Surprise*, *Happiness*, and *Sadness* (see Figure 2.2). This act of categorization allows for a better understanding of emotions and their expression in different contexts. In her work, Barret[57] shows that these categorizations are not universal but bound to culture and people, and she hypotheses that "emotion words hold the key to understanding how children learn emotion concepts in the absence of biological fingerprints and in the presence of tremendous variation. Not the words in isolation, mind you, but words spoken by other humans in the child's affective niche who use emotion concepts. These words invite a child to form [...] concepts for 'Happiness,' 'Fear,' and every other emotion concept in the child's culture" [57]. From this hypothesis, we gather that understanding any emotion (defined by a word) is bounded by culture and language. Therefore, we know that the feeling of emotions is varied and takes many shapes depending on the episodes in our lives, and only when we name feelings into words do we compartmentalize and fit them into known "appropriate" words in our culture.

The use of the previously mentioned six basic families of emotion [55, 56] is common in Artificial Intelligence (AI) works[59–61] as they offer a good compromise between variability and expressivity. Many emotional expressions have been identified for each emotional family[58], and all contain variations in the various emotional modalities associated with them. And although the expression of emotion may not reflect the inner feeling of said emotion [57, 62], in western culture these basic emotions are well known (e.g. in movies such as *Inside Out*[165]) and accompanied by a prototype of the emotion expression, or a mental representation of the best example of emotion expression (e.g. the prototype of sadness expression may be of one crying and sobbing, yet not all instances of sadness are expressed that way)[57], allowing for an easier expression of emotion that will be familiar to those within the culture.

2.3.1 Emotion and Anticipation

Anticipation and emotions are closely related. One of the emotions' principal functions is precisely that of anticipating events, especially when those events involve the well-being of the organism. *"If I am walking in the woods and, suddenly, 'something' ahead on the path lets out a loud roar, my heart races, my muscles tense, I 'feel' afraid and ready to run away"*[25], in this example the emotions

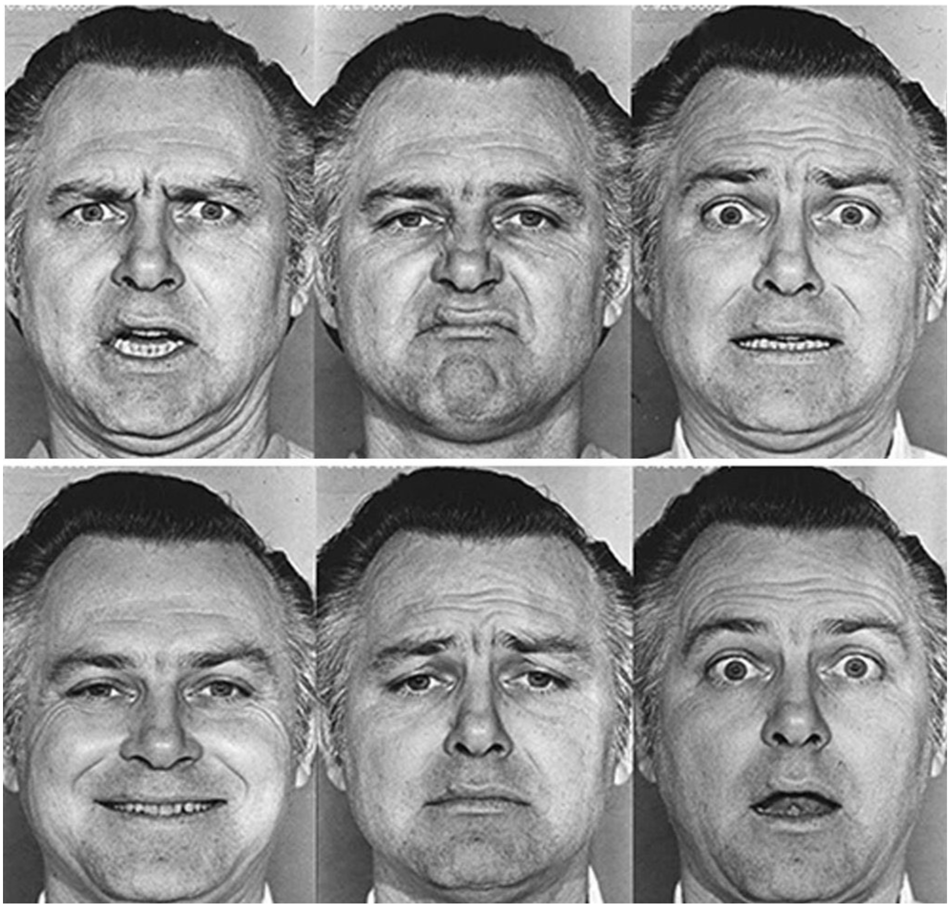


Figure 2.2: The expression of the six basic families of emotion [56, 58], from top left to bottom right: Anger, Disgust, Fear, Happiness, Sadness, and Surprise.

helped reduce the number of possible actions, by eliminating most of the consequences of each action from consideration a-priori. Therefore creating an **action tendency** or in other words a desire to behave in select communicative or important actions that are connected to a particular emotion.

Yet the anticipation of an event may also elicit an emotion. Let's rephrase the previous example, *If I am **hunting** in the woods and, suddenly, 'something' ahead on the path lets out a loud roar, my heart races, I found my prey, today I'll feed my family, I am happy*. In this case, the emotion was elicited by anticipation. These emotions are often related to expectations, commitment towards important goals, and the validation or invalidation of both expectation and goals. Therefore, the same outcome can lead to a wide range of emotional experiences, based on different types of expectations. "To be prepared for what is to come is a crucial factor in survival"[25].

2.4 Emotion in Synthetic Characters

When discussing emotion in synthetic characters, among others, there are two perspectives to consider, the *internal representation of emotions*, or how emotions are created, used, and/or processed internally without explicit visual cues, and the *expression of emotions*, or physical behavior that is associated with emotions. Both are important to consider when portraying emotions in a synthetic character.

2.4.1 Internal Representation of Emotions

To answer the question of how one can compute emotions, multiple authors proposed different works[50, 63, 64]. In an effort to incorporate these different affective models, Scherer *et al.*[54] created five categories: *Appraisal Theory Approaches*; *Anatomical Approaches*; *Rational Approaches*; *Communicative Approaches*; and, *Dimensional Theory Approaches*. Each category differentiates itself in what particularity it wishes to convey special relevance or the psychological theory they are backed by. A summary of each category is described in Appendix A.2. We give special attention to the Appraisal Theory Approaches and one in specific, the Emotivector.

Appraisal Theory Approaches postulates that "all emotions come mostly from our own interpretations of events"[63], where our appraisal of the situation is the emotional response. The theory is best used in connecting awareness with emotion, focusing on the individual and their psychological response, where his own judgment of a situation is to blame as the source of his emotional response.

Ortony *et al.*[50] describe an emotional classification that states that emotion is structured into the categories of *Fortune-of-others*, *Prospect-based*, *Well-Being*, *Attribution* and *Attraction*, or more largely grouped into *consequences of events*, *actions of agents*, or *aspects of objects*. This is known as the **OCC Model** and many virtual emotional models that opt for the appraisal theory often base themselves on this model. The OCC Model attempts to incorporate all emotions, but with no relationship between them other than categorical, but not all models based on the OCC model incorporate all emotions. One good example is the model proposed by Ochs *et al.*[13] which focuses on the believability of the NPCs. They attempt to improve the experience by focusing on the personality, social relations, and roles of the NPCs inside the game. The emotions modeled, using the OCC, were joy/distress, hope/fear, and relief/disappointment. There is also an emotional decay component implemented to revert the emotional state to a neutral state after some time period. This model focuses on the NPCs and tries to increase believability through simulation of social relations.

Another variation on the OCC model comes from He *et al.*[65], it is a fuzzy emotional model for virtual agents. The emotions of events modeled were hope/fear, satisfaction/fear-confirmation, and relief/disappointment, and were based on three variables, desirability, "if an event is beneficial to an agent, it is desirable otherwise it is undesirable", importance, "we equate goal's importance with motivation intensity", and likelihood, "we equate likelihood with the possibility of what other virtual agents will do". An important feature is "how the model accounts for relations between agents as a predictor of behavior for them"[63].

Using a partial OCC model implementation, the work of Jacobs *et al.*[66] focuses on the importance of the link between reward/punishment and emotional response. They propose the use of an emotion label system that converts each agent's state transition, an initial joy/distress mapping that converts into a hope/fear mapping over time, related to the agent's previous knowledge, hopefully allowing the gathering of useful information for planning and decision making of the agent. This model's objective is focused on the modeling of the agent behavior but is important to note how the modeling anticipatory behavior improved the tested agent's performance.

FatiMA[64] offers an interesting appraisal theory application, where the OCC model is implemented by storing appraisals (valence based) in a numeral intensity value (-10;10). This model acting along with goal mechanisms and perceived events, models the complete range of emotions inside the OCC, being able to give individual personalities and coping mechanisms to deal with specific goals. This model however not take into account anticipation of behavior or emotions.

The appraisal theory approaches show potential when used in a more static NPC or Environment emotional association, giving a simpler and robust emotional model to the in-game interactions. We now focus on Emotivector, a model used in our implementation.

2.4.2 Emotivector

Emotivector[7] is a system that focuses on the modeling of emotions by incorporating both anticipation and expectation. By splitting a sensorial input into several categories, according to sensations, a pair between expectation and valence, the model creates a range of emotions that an agent can express. "An increase in a positive sensation or a better reward than expected leads towards excitement, a decrease of a positive sensation or a worse reward than expected leads towards discontentment; a stronger punishment than expected leads towards depression, and a lower punishment leads to pleasure"[7]. Other categories can be made, such as expecting a reward and receiving a punishment can lead to sadness and frustration. This approach offers an appraisal model for virtual agents based on anticipation and is the model used for our implementation of an emotional model. Due to its importance to our work, we will describe it in more detail based on what is described in [7]. It is also important to note that in [67] we developed an emotional model that made use of Emotivectors, the implementation will be discussed in Chapter 3.

2.4.2.A Behavior of an Emotivector

An emotivector is an "anticipatory mechanism attached to a sensor" and associated with a "one-dimensional aspect of perception". When an event is perceived by the agent (anything that happens can be considered here, even the behavior of the agent itself), the associated emotivector of that event performs the following steps: (1) using the history of the event, the emotivector computes the next expected value for that event; (2) by confronting the expectation with the actual sensed value, the emotivector computes a preliminary salience for that instance of the event, and; (3) "a sensation is generated according to a model inspired in the psychology of emotion." This process allows for when a sensation of an event is processed, it carries recommendations such as "you should seriously take this signal into consideration, as it is much better than we had expected" or "just forget about this one, it is as bad as we predicted."

A – Salience or Attention To determine how relevant a particular emotivector is (how much attention should be given to it), a salience value is calculated. Two components, exogenous and endogenous, are computed to determine the final value. The exogenous component is based on the estimation error and reflects the principle that the least expected (or more surprising) is more likely to attract the attention. If no further information is given, the exogenous value is the only factor contributing to salience, however, if a "searched" value exists, the endogenous component is computed. It reflects whether the change in the distance to the "searched" value is better or worse than the expectation and the level of mismatch. The combination of both exogenous and endogenous components defines the salience. However, "an emotivector with a search value can also provide a qualitative interpretation of the percept", as explained in the next section.

B – Sensation or Emotion In [7], emotions are assumed to be "conditioned responses of primary sensations" and the model is focused on the generation of said sensations. These are defined across two dimensions: valence and change. "The emotivector estimation is used to anticipate a reward or punishment which, when confronted with the actual reward or punishment, triggers one of the [...] sensations. The intensity of each emotion is given by the endogenous component of the percept."

The four most simple sensations considered in the model were the following:

S+ (positive increase) If a reward is anticipated and the effective reward is stronger than the expected, an S+ sensation is thrown.

$+ (positive reduction) If a reward is anticipated but the effective reward is weaker than the expected, a $+ sensation is thrown.

S- (negative increase) If punishment is anticipated and the effective punishment is stronger than expected, an S- sensation is thrown.

$- (negative reduction) If punishment is anticipated but the effective punishment is weaker than expected, a $- sensation is thrown.

The original work (through the use of meta-predictor discussed below) and others that made use of Emotivectors [67–69] would go on to use more complex versions of the system with more sensations ([67] will be further detailed in Chapter 3).

With these sensations, we can generate emotions based on anticipation (e.g. feeling sad when punishment is greater than expected, S-), but it is also possible to compute emotions based on a combination of multiple sensations to create more believable characters (e.g. perceiving $+ and S- and feeling fear, because not only were you not rewarded as expected, but also you were punished harder than expected).

C – Prediction or Anticipation "The computation of the emotivector salience relies on the capacity of the emotivector to predict its next state." As there is no a-priori knowledge of the signal, [7] it is assumed that the intensity of a signal will change by a random small amount at each discrete time step, before suddenly changing to a random new value. "In other words, the sensed value will tend to remain constant except for certain points in time."

This assumption lead to the implementation of a "Simple Predictor" based on the Kalman filter, "a set of mathematical equations that provides efficient computational recursive means to estimate the state of a process in a way that minimizes the mean of the squared error." The authors defend that while "this predictor is not as optimal as the ones it is inspired by, it provides a good efficiency/adaptation relation that performs well in real-time over unpredictable signal dimensions, and does not require any previous fine-tuning to work."

Note that this predictor is applicable in scenarios where data is gathered in high frequency and has low variance (it is also useful when spacial prediction is involved, which was the case in [7]). Other works using Emotivector [67, 69] assume the signal changes with high variance, thus the previous predictor is not fit in those use-cases. In its place, these works use varied algorithms based on moving averages to deal with the difference in frequency.

2.4.2.B Emotivector Management

"[A]ll emotivectors are kept together in a salience module, responsible for their management as a whole". This manager also needs to determine which percepts are "*a-priori* more relevant to the [...] the agent." The following strategies are presented in the work: **(1) Winner-Takes-All** that selects the percept with the highest salience, while convenient it "hides" other percepts even if they are just "a little less salient"; **(2) Salience Ordering** creates a salience-descending ordered list that allows for more than one percept to be analyzed while prioritizing the most salient ones; and, **(3) Meta Anticipation** was the approach used in their use case and it makes use of a meta-predictor "whose role is to assess the salience of its associated predictor." Instead of receiving a percept to predict, it receives the prediction errors of another predictor. "If the prediction error is higher than the meta-predictor estimated error, the percept sensed by the emotivector is marked as relevant."

The meta-predictor also created the concept of uncertainty on each percept, or "how trustworthy the prediction is". This allowed the system to take into account other sensations that take into account if the prediction is as expected (e.g. "this percept is as bad as we expected it to be").

As a closing remark, this model is designed to support the creation of believable behavior and is not intended to "substitute any current approach but rather complement the agent architecture for life-like characters."

2.4.3 Expression of Emotions

An emotional expression is a behavior that communicates an emotional state or attitude (see example in Figure 2.3). Emotions can be conveyed through multiple modalities, such as facial expressions [9, 56, 67, 70], back-channeling [4, 5, 71], and body language [3, 8, 52, 72]. Synthetic characters can make use of most of these modalities and even exaggerate them for the benefit of believability. Furthermore, "it was found that humans tend to interact with computers as they do with real people"[52].

Facial expressions are one of the briefest signals associated with emotion expression, usually lasting only mere seconds[55]. Those that last for longer periods tend to be associated with more intense feelings. Although they do not necessarily correlate with actual emotional experience[57, 62], they are very useful at portraying each of the basic families of emotion [56, 58] (Fear, Anger, Disgust,

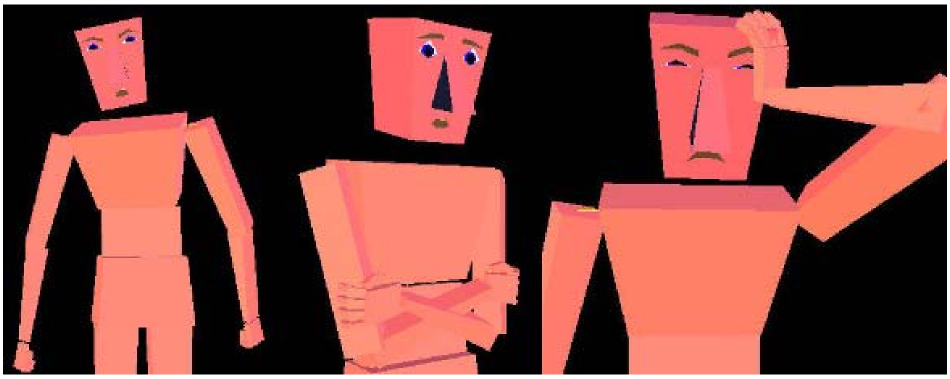


Figure 2.3: Example of expressions of the character used in Nayak's system [8] expressing (left to right) anger, defensiveness, and headache.

Surprise, Happiness, and Sadness; see Figure 2.2) and, in addition to enhancing the conversation, can also be used to convey dominance and affiliation. One of the more prominent techniques for coding facial expressions is the Facial Action Coding System (FACS) [73]. Developed for measuring facial movement, FACS can objectively score each one of the six families of basic emotions, effectively coding them as a series of Action Units (AUs). These have been built over the years. Notably, both Arya *et al.*[59] and Makarainen *et al.*[60] suggested models that combined the AUs from the six families of basic emotions in a way that would allow for the creation of perceptually valid facial expressions.

Table 2.1: The facial expressions prototypical characteristics of the six basic emotions.[56]

Emotion	Associated facial expression
Fear	The eyebrows appear raised and straightened. The eyes are opened and tense during fear, the upper eyelid raised, and the lower eyelid tense. The mouth opens in fear, but the lips are tense and may be drawn back tightly.
Anger	The eyebrows are drawn down and together. The eyelids are tensed, and the eye appears to stare out in a penetrating or hard fashion. The lips are in either of two basic positions: pressed firmly together, with the corners straight or down; or open, tensed in a squarish shape as if shouting. There is ambiguity unless anger is registered in all three facial areas.
Sadness	The inner corners of the eyebrows are raised and may be drawn together. Raising the lower lid increases the sadness conveyed. The corners of the lips are down or the lip is trembling.
Surprise	The eyebrows appear curved and high. The eyes are opened wide during surprise, with the lower eyelids relaxed and the upper eyelids raised. The jaw drops during surprise, causing the lips and teeth to part.
Happiness	Corners of lips are drawn back and up. The cheeks are raised. The lower eyelid shows wrinkles below it and may be raised but not tense.
Disgust	The brow is lowered, lowering the upper lid. The nose is wrinkled. The upper lip is raised. The lower lip is also raised and pushed up to the upper lip or is lowered and slightly protruding.

Another important non-verbal component of emotion expression, during social interactions, is *back-channeling* signals. These not only help us deliver our communicative intents [71], thus promoting effective communication [4], but also help us establish our personal tendencies, which in turn help define us as truly unique individuals [5]. Non-verbal behavior gives the listener a chance to provide basic communicative functions toward what the speaker is saying [5], and its absence could lead to negative experiences and incorrect interpretations of the situation [4]. The two most common back-channeling signals are gaze and head motion. *Gaze* establishes the rhythm of conversations, indicates interest in objects/people, improves listeners' comprehension, expresses complex emotions, and facilitates interpersonal processes [74]. It is worth noting that gaze is not simply relegated to eye movement. Head and body movement/position can be an active component in gazing. These are not independent of one another and, if treated as such, will appear random and disjointed[75]. *Head motion* is an integral part of communication. Nodding can be used to show our agreement/distaste of a subject matter, as a reinforcement when making a point, or even as a way to avoid unwanted gaze. It also plays an important role with regard to expression. An angry or distressed person will make erratic movements, while a calm, collected person will show very little head motion. This means it can be used to assist in discriminating between emotions[76].

"At the present, most researchers agree that words serve primarily to convey information and the body movements to form relationships"[3]. *Body Language* is often used to express emotion, specifically named *Emotional Body Language*, and is one of the most important forms of nonverbal communication [3]. Furthermore, it is interpreted similarly when portrayed by agents or humans in terms of recognition [3]. They include movements of hands, head, and other parts of the body that allow individuals to communicate a variety of feelings, thoughts, and emotions. Even "[b]reathing is known to participate to the expression of specific emotions"[52]. Do note that "to correctly interpret body language as indicators of emotional state, various parts of body must be considered at the same time"[3]. One can also assume that the correct expression of emotional body language is more important than that of facial expression, as when both "convey conflicting emotional information, judgment of facial expression is hampered and becomes biased toward the emotion expressed by the body"[77]. Noroozi *et al.*[3] provides an example of general movements protocol for six basic emotions presented in Table 2.2.

Another example of a generalization of emotional body language is provided by Nayak *et al.*[8] and chooses seven main characteristics to portray a person's mental state (see Figure 2.3): confidence, anxiety, interest, thought, anger, defensiveness, and pain. Each characteristic allows the programmer to set its value on a continuous scale and that affects the movement of the character. When multiple characteristics affect the same body movements, some of them can be averaged out to create a mixed expression. A small description of the general movements protocol for the seven characteristics is presented in Table 2.3.

2.4.3.A Secondary Channels

Aside from the aforementioned modalities, the dialogue itself can also be used to convey information and emotion. When considering synthetic characters, verbal dialogue can be communicated not only through audio (e.g. prerecorded dialogues or dictation technology), but also through text which

Table 2.2: The General Movement Protocols for the Six Basic Emotions.[3]

Emotion	Associated body language
Fear	Noticeably high heartbeat rate (visible on the neck). Legs and arms crossing and moving. Muscle tension: Hands or arms clenched, elbows dragged inward, bouncy movements, legs wrapped around objects. Breath held. Conservative body posture. Hyper-arousal body language.
Anger	Body spread. Hands on hips or waist. Closed hands or clenched fists. Palm-down posture. Lift the right or left hand up. Finger point with right or left hand. Finger or hand shaky. Arms crossing.
Sadness	Body dropped. Shrunk body. Bowed shoulders. Body shifted. Trunk leaning forward. The face is covered with two hands. Self-touch (disbelief), body parts covered or arms around the body or shoulders. Body extended and hands over the head. Hands kept lower than their normal positions, hands closed or moving slowly. Two hands touching the head and moving slowly. One hand touching the neck. Hands closed together. Head bent.
Surprise	Abrupt backward movement. One hand or both of them moving toward the head. Moving one hand up. Both of the hands touching the head. One of the hands or both touching the face or mouth. Both of the hands are over the head. One hand touching the face. Self-touch or both of the hands covering the cheeks or mouth. Head shaking. Body shift or backing.
Happiness	Arms open. Arms move. Legs open. Legs parallel. Legs may be stretched apart. Feet pointing to something or someone of interest. Looking around. Eye contact relaxed and lengthened.
Disgust	Backing. Hands covering the neck. One hand on the mouth. One hand up. Hands close to the body. Body shifted. Orientation changed or moved to a side. Hands covering the head.

can be used to reinforce the emotional state of a synthetic character. More than its intrinsic meaning, dialogue in such applications can be enriched by other elements, such as color, kinetic typography, and speech bubbles, which can be used in tandem with characters, as a secondary communication channel. Starting with *color*, Kaya et al. studied its properties in association with emotion [78, 79]. The main findings relate to how symbolism is dependent on each individual, and which aspects (objects or physical space) they associate each color with. Unfortunately, these associations are dependent on personal experience, as well as a social and cultural background such as age, gender, and nationality. Despite being a complex topic and difficult to properly utilize, the richness of their symbolism allows for the conveying of certain emotions [80]. The second property, *kinetic typography*, is a communicative medium that adds some of the expressive properties of film to static text [81]. Studies on such animations show that one of the best ways to convey emotion is by imitating body and sound movements. The results presented by Malik et al. showed that specific motions, such as shaking, fading, and bouncing are, depending on the intensity, able to convey emotion, regardless of context [82]. It is important to notice that kinetic typography cannot normally replace or nullify the intrinsic emotive content of a sentence. Instead, it should be used to reinforce it. Finally, *balloons* have been used in comics for a long time to specify which character is speaking and how they are talking. However, artists create their conventions when conveying the physical or emotional state of a character [83]. Despite the diversity, there are a few bubble shapes that are commonly used in the art form - speech,

Table 2.3: The General Movement Protocols for the Seven Characteristics of Nayak's work.[8]

Characteristic	Associated body language
Confidence	A complete lack of confidence will lead to a self-doubting hunched-over character. As confidence increases, the character's posture straightens with his chest out and chin up.
Anxiety	As anxiety increases it creates the effect of the character becoming increasingly nervous and fidgety. As anxiety increases, the character's shoulders gradually become increasingly tensed up or shrugged.
Interest	If the character has no feelings towards any object for an extended period of time, he gets bored and may yawn. However, if the character is interested in or is averse to an object, then it holds the character's attention and therefore his gaze, and rotates towards and away from the object, respectively. As the feelings become even stronger, the character begins to walk towards or away from the object.
Thought	The scale goes from lack of thought to deep thought. As the character becomes more deeply lost in thought, his natural fidgeting decreases. When maximally deep in thought, he is almost completely frozen.
Anger	The natural fidgeting of the character is directly proportional to the level of anger. As the anger level rises, the character's body starts trembling, gradually begins to gaze toward the entity at which he is directing his anger. If the anger level of the character rises above a certain point, the character performs the hostile gesture of clenching his fists tightly. If he is walking, he will swing his clenched fists as he walks.
Defensiveness	When defensive, the character will cross their arms and periodically gaze away from the entity of which he is defensive, preventing confrontation and avoiding the impression of aggressiveness. However, on high levels of defensiveness, the character is facing an imminent threat and will not gaze away, instead placing both hands in fists in the air in front of him to physically protect himself in self-defense.
Pain	Pain in the head, stomach, or feet is debilitating. Hence, the character's posture and, if he is walking, the walk speed are both inversely proportional to the level of pain. The character hunches over and slows down as pain increases, also holding the body part in pain.

whisper, thought, and scream. If properly used, these types of balloons can help to convey specific emotions to the user. The previous statements also apply to the use of *backgrounds*. In comics, the background of a scene is also constantly used to specify the state of mind of a character [83], through the use of color, icons, or a combination of both (see Figure 2.4). However, the more ambiguous and lesser-known a shape is, the less meaning users can derive from it.

All discussed elements have unique properties, which allow them to express or reinforce the emotion being expressed, creating different emotional communication channels. The challenge arises from attempting to utilize them in unison to effectively express emotional states. This topic is further discussed in Chapter 3.

(a) Expression of terror when a character cannot overcome its despair.

(b) Expression of confidence when a character overcomes a challenging situation.

Figure 2.4: Two expressions of emotions in the game Darkest Dungeon[162]. This game uses a mechanic of accumulation of stress that can make the characters' despair (image on the left) or, in rare cases, overcome the stress and become confident (image on the right). To express either situation, the developers made use of a physical expression through the posture of the body and facial expression (when visible), but also through the use of the background, mixing color (red for negative motifs and white for good motifs) and symbols (circle with spikes expressing stress, circle modeled like clouds to express confidence).

2.5 Recognition and Assessment of Emotion Expression

When discussing emotion recognition it is important to discuss the concept of Emotional Intelligence, meaning the ability to know one's own and other individuals' emotions and intentions, and the capacity to use such information to adapt to certain environments[84, 85]. Many models for Emotional Intelligence have been proposed with different views on the base definition. A particular model was presented in the book "Executive EQ" and defined 4 cornerstones: "*emotional literacy*, which includes knowledge of one's own emotions and how they function; *emotional fitness*, which includes emotional hardiness and flexibility; *emotional depth*, which involves emotional intensity and potential for growth and *'emotional alchemy'*, which includes the ability to use emotion to spark creativity"[84]. This concept and model are important to take into account when discussing the evaluations with participants, as different people have different degrees of Emotional Intelligence and that can impact their perception of the expression of emotions.

Studies found that neither younger nor older adults are able to correctly identify most emotions with 100% certainty. Additionally, Mill *et al.*found that the correct recognition of negative valences decreases with age, starting at the age of 30 years old [86]. This suggests that recognition varies not only from emotion to emotion but also from individual to individual. Regarding probable causes, Bassili[87] presented a discussion in which he points to similarities/ambiguities between facial expressions, alongside poor actor performance, as the likely contributors to erroneous recognition. Additionally, Barrett defends that without context it is almost impossible to recognize emotion[57]. Another possible issue is the use of multiple emotional channels. Mower et al.[88] presented evidence that adding a second channel can create channel bias, which manifests itself when one channel is more predominant and therefore captures the receiver's attention, or when two channels transmit contradictory cues. This is an important concern when combining multiple distinct channels, like the ones discussed in Section 2.4.3.

Having discussed emotion recognition, we can now discuss the assessment of Emotion Expression. Lottridge *et al.*[89] reviewed relevant research on emotion and its measurement, discussing how one can measure a participant's affective state through physiological measurements [90–92] and self-report [52, 93–95]. Focusing on the participants' assessment of emotions of others, a short review of works revealed that the participants' input is often extracted through reports, either about other people [52, 96–98], virtual agents [52, 99], or anything in general (usually for labeling) [100, 101]. Other works extract the participants' assessment through their physiological responses to interactions with others or, more specifically, with virtual agents [102, 103]. In this work, we will focus on the assessment of emotions of others through self-report to evaluate our agents.

2.6 Principles of Traditional Animation

People are "remarkably adept at recognizing the actions performed by others" [104] and some works "point to the existence of neural mechanisms specialized for registration of biological motion" [104]. With no more than 12 point-lights portraying biological motion, "people can reliably discriminate male from female actors, friends from strangers" [104].

When we apply biological motion to animation, it requires work to appear believable. The animated character, while not needing to be realistic, should move and behave consistently with the way it looks[52, 105] and discrepancies in behavior were found to be disturbing[105]. The principles of traditional animation, first introduced in the book *The Illusion of Life: Disney Animation*[9], mitigate this problem by defining standardized practices followed by Disney's animators to create believable animations (see Figure 2.5), both traditionally drawn and digital[106]. Below follows a description of some of the principles.

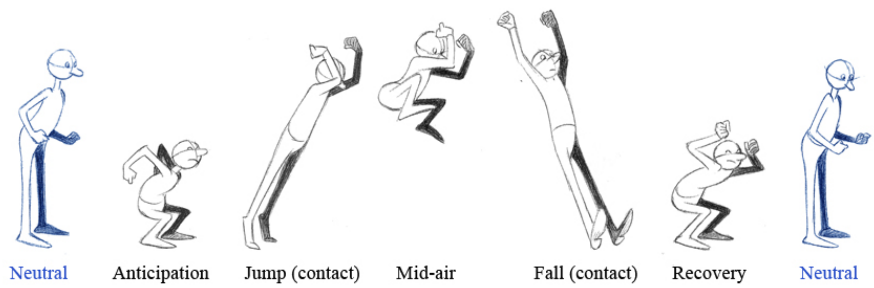


Figure 2.5: Stages of a jump animation. In this figure, we can see some of the principles of traditional animation, such as *Anticipation*, notice the stretching of the arms and legs in the anticipation, preparing the viewer for the jump, which also showcases the *Staging*. When landing, the *Follow-through* principle is shown in the exaggerated motions of the fall and its recovery back to the neutral position.

- ***Timing***, or speed of an action, defines how well the idea behind an action will be read by an audience. More importantly, timing defines the weight of an object, as in the example “a giant has much weight, more mass, more inertia than a normal man; therefore he moves more slowly. (...) he takes more time to get started and, once moving, takes more time to stop.”[106].

One can also define the emotional state of a character by its movement, where the varying speed of an action indicates whether the character is lethargic, excited, nervous, or relaxed.

- ***Anticipation*** is the preparation for the action, for example, if a character wishes to grab a cup of coffee they first raise their arm and stare at the cup, broadcasting their intention, which leads those watching to expect the character to pick up the cup before the action is done. Without anticipation, many actions are abrupt, stiff, and unnatural.

  An exaggerated anticipation can also emphasize the weight of an object when a person has to bend down to be able to pick up a heavy crate, or used to show a character’s emotional state when one is scared or anxious about doing something he must do.

- ***Staging*** “is the presentation of an idea so it’s completely and unmistakably clear”[106]. This principle declares that to clearly stage an idea the audience must be led to be paying attention exactly to what the creator wants them to, otherwise the idea will be missed.

  When staging an action, it is important that only one action be passed to those watching, to do that there should be a contrast between the object to focus on and the rest of the scene, for example, in a big crowd walking in the side-walk, a person standing still will attract the viewer’s attention.

- ***Follow-through and Overlapping Action*** – Most times an action does not come to a sudden stop after it is complete, in many movements like a jump there is the termination of the action or Follow-through. You can see an example in the recovery in Figure 2.5, where the action is carried past their termination point.

  An Overlapping Action can be variations added to the timing and speed of the loose parts of objects or an action that overlaps the previous one, which makes the objects seem more natural and maintains a continual flow between the phases of an action.

- ***Exaggeration*** is self-explanatory, but it has to be done with care. It can work with every component, but not in isolation. The exaggeration of various components must be balanced, where some elements are exaggerated and the others are used as natural elements for the viewer to use as a comparison so that the scene remains believable.

  When animating characters, exaggeration is very important to transmit their emotional state. If a character is sad, make him sadder; if he is wild make him frantic. The exaggeration allows for a better interpretation of the emotion being displayed[52].

- ***Secondary Action*** “is an action that results directly from another action”[106]. It is important since it adds realistic complexity to the scene, but must always be kept subordinate to the primary action. “Secondary action is something that is under the character’s control that embellishes or enhances the primary action in a way that adds character or personality.”[2] While the Overlapping Action is an involuntary response to the motions of the primary action, the secondary action is a voluntary response by the character to express personality.

Although secondary, this type of actions will be very important to this work, since we will consider the reply to the primary action, of the character and of those who watched, to be secondary actions.

There are other principles that were not described above as they are not a focus of this work: **Squash and Stretch** principle defines the rigidity and mass of an object by distorting its shape during an action; **Straight Ahead Action and Pose-To-Pose Action** are two approaches to the creation of movement; **Slow In and Out** principle that specifies the spacing between frames to achieve subtle movements and timing; **Arc** is a visual path of action for a believable movement; and, **Appeal** principle describes how to create an action that the audience enjoys watching.

All of these principles help create believability in the characters, their interactions, and the environment around them. To close this discussion, here is a quote from the authors of the book,

> *“It is more than a drawing and more than an idea, possibly it is the love we feel for characters so heroic, so tender and funny and exciting – all of them entertaining, yet each different, each thinking his own thoughts, and experiencing his own emotions. That is what makes them so real, and that is what makes them so memorable. It is also what gives them the astounding illusion of life.”*[9]

2.6.1 Anticipation and Execution

We will make use of the principles to define our action subdivision and emotion expression model. Our work will follow a similar approach to the work of Costa [51]. In his work, he proposes a new approach to agent communication and cooperation by dividing an action into two stages, *anticipation* and *execution*, following part of the principles of the traditional animation.

The *anticipation stage* serves the purpose of broadcasting the intent of an agent so that every other “is expecting it and can prepare accordingly”. After the broadcast, the agent may choose to either execute or cancel their action, based on the other agents’ responses. If the agent chooses to execute the action it enters the *execution stage*, where processes like path-finding and animation playing will allow the agent to execute the action. The agents’ stages may overlap at any time, thus the agents must be aware of each other’s current state of action, so they can effectively cooperate.

To determine if an agent continues from anticipation to execution, Costa proposed the use of a *Confidence value*. For each action, each agent has a confidence value that anticipates if the agent’s action will be successful. When the agent broadcasts his intention to perform an action, his confidence will increase or decrease depending on the other agents’ reactions. If the confidence value is below

[2] `https://medium.com/frame-by-frame/anticipation-follow-through-overlapping-action-and-secondary-action-f80069aba725`

a certain threshold, then the agent cancels the action, feels frustrated, and the process starts all over again.

The confidence threshold is dependent on the agent's personality. "Just as feedback impacts different people differently in the real world, it should impact different agents differently"[51]. Additionally, the outcome of the action can also influence the confidence associated with future intents, creating an adaptive agent.

This approach gives a good foundation to build upon, incorporating the notion of a multi-staged action and the broadcast of intentions.

2.7 Emotional Multi-party Interactions among Synthetic Characters

Until now, through the exploration of related work, we created a context that allows for the discussion of a set of topics, i.e. multi-party interactions, emotion expression, real-time interactions, and believability. These topics are present in both the problem we want to solve as well as in our proposed hypothesis (recall Section 1.2). Considering their relevance to our work, are there other works that tackle all or most of these topics?

We collected relevant works from AI-related conferences[3] held in 2021. From the collected works, those tackling multi-party interactions tend to focus on decision-making, but lack emotion expression [107–112], while those showing emotion expression tend to focus on human-agent interactions (with a single agent), but often lack in multi-party interactions [113–122]. Real-time interactions are regularly present in human-agent interaction works, but often it is not the object of study (e.g. in [115] a study portraying real-time interactions is conducted to examine "whether social agents can elicit the social emotion shame as humans do"). Believability is not commonly present in recent works, yet when present it appears as an objective of the work[108, 112, 123–125].

Works tackling most of the relevant topics are not common, but a few exist[107, 108, 126, 127]. In [107], communication between reinforced learning agents is explored and a new communication mechanism is presented to complement other existing ones. The authors ran simulations (in the game StarCraft II[154]) to measure the success of the application of the communication mechanism. This mechanism focuses on the communication of information (so that other agents learn from that) based on novelty, in other words "Communicate what surprises you"[107] (the common approach is to "Communicate what rewards you"). Results show a greater degree of success when agents use both approaches simultaneously, that is they share what rewards and what surprises them. This work showcases what information is relevant to share and its approach on surprise is similar to what is seen in the Sensation salience in Emotivector (see Section 2.4.2).

In [108], while focusing on decision-making, the authors take special attention to the believability of behavior and real-time interactions in multi-agent environments. In this work "a method that allows

[3] International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), IEEE Conference on Games (CoG), International Conference on Intelligent Virtual Agents (IVA), Conference on Affective Computing and Intelligent Interaction (ACII), International Conference on Human-Agent Interaction (HAI).

[for NPCs acting as] guards to exhibit a real-time search behavior for an opponent that better exploits awareness of the level geometry" was introduced. They explore a novel approach to path-finding and search behavior based on a skeletal graph representation to better propagate the probability of potential opponent locations. Their approach is shown to display interesting behavior for multi-guards searching for an opponent on several maps. While no believability evaluation was conducted, the concern for believable behavior (in favor of realistic or optimal behavior) is relevant to our work and their approach can later inform how we want to enable decision-making in our agents and portray believable behavior.

In [127], the authors offer a "platform that researchers and developers can use as a starting point to setup their own multi-IVA applications." In it, they implement multiple integrated components, such as topic section engines, dialogue engines, and behavior realizers, so that the agents can seamlessly interact with each other. Furthermore, "the viability of the platform has already been shown by uptake not only in student teaching, but also in at least 4 funded research projects that employ the full setup". This work helps us understand how a multi-agent platform is created and how every component is connected.

In [126], an exploration of how virtual agents react to non-verbal communication is performed. Three agents are depicted interacting with each other, one performs a wave (pre-programmed action of an agent waving his arm), while the others decide to wave back (or not) if they perceive that the wave is directed at them (through the analysis of body direction and other motion features). The work is still early in its development, but the authors say that in future works "[i]n a multi-agents scenario, [this approach] would allow for the design of not fully predefined non-player character reactive behaviors." The work offers an interesting view on non-verbal communication between synthetic characters, showcasing the potential of using pre-scripted animations in less scripted (or more spontaneous) interactions. These types of interactions appear to be closely linked to the expression of back-channeling signals discussed in Section 2.4.3, further motivating these almost instinctive (or less conscious) behaviors.

These works help us understand the current state of the art surrounding the previously mentioned topics and solidify the validity of previously mentioned subjects (such as back-channeling and salience). These works also showcase the existing gap we are filling with our work, further motivating investigation on the subject.

2.8 Summary

This chapter focused on the exposition and discussion of works that are relevant to our problem, hypothesis, and implementation. As a review, our problem is focused on how can one enable believable multi-party real-time interactions between synthetic characters. To address this problem in the correct context we presented related work on believability and its different definitions by different authors. We proposed our own definition for believability and also believable environments, characters, and interactions. Surrounding this topic, we delve into what believable agents are, how they need awareness and situatedness to be grounded in the world, and how theory of mind and empathy allows characters to bond and be more believable. We then discussed works that tackle the assessment of believability.

One of the crucial elements that make a character believable is emotion, thus we then explored how emotions are defined and how they are expressed. We discuss their link with anticipation and how one influences the other, through action tendencies, and vice-versa, through context. We then move to how emotions are used in synthetic characters, looking into their internal representation, how emotions are created, and their external representation, how emotions are expressed. Special attention was given to Emotivector, an emotion model that makes use of anticipation to generate emotions, and emotional body language, which describes how we view emotion expression in body movements. To close the study on emotions, a review on recognition and assessment of emotion expression is shown.

Connecting believability, emotion, and expression we discuss the principles of traditional animation. These principles are a set of standardized practices followed by Disney's animators to create believable animations and can be applied to digital environments. Afterward, we discuss the subdivision of an action applied to an interactive experience to help create more believable interactions and characters.

Finally, we review a selection of recent works that explore multi-party interactions through communication and emotion expression between synthetic characters.

3

A Case Study on Expression of Emotions

Contents

"Every word has consequences. Every silence, too."

Jean-Paul Sartre

This chapter discusses our work done to create a virtual coach used in Virtual Tutoring[1], a coaching mobile application to help university students in their studies. The virtual coach is relevant in the context of this work in its emotional model, which uses Emotivectors to help compute the emotions expressed, in the expression of said emotions, that make use of multiple channels to communicate a single emotion (e.g. facial expressions, backgrounds), and in the evaluations that point out common issues to be aware of when developing emotional synthetic characters. We will give an overview of the architecture of the virtual coach and discuss in detail any area that is relevant to the present work. More details on this topic were published in [67]. Other works have been published describing part of the implementation of or making use of these coaches [128, 129].

3.1 An Introduction to Virtual Coaches in an Online Learning Environment

In recent years, there has been a widespread availability of online learning environments, offering more flexibility in a student's learning process, by allowing non-regular study schedules and access to readily available resources. This led to the development of new teaching and learning methods [130–134]. Although new methods arise, some problems remain. One such problem is the implicit existence of external factors influencing the students (e.g. responsibilities, lack of time), which lead to disinterest in the learning environment. To help mitigate these problems, we created a virtual coach that accompanies the students, adapts to their needs, helps them in their tasks, and understands (as well as responds to) their intentions, motivations, and feelings [135]. This virtual coach employs emotions along with empathy and interpersonal affect regulation strategies[36] to better motivate students, using both verbal and non-verbal communication to better express emotions.

3.2 Empathy and Interpersonal Affect Regulation in Social Interactions

An important aspect to consider when creating social virtual agents is *empathy*, an "affective response more appropriate to someone else's situation than to one's own" [41] (also discussed in Section 2.1.4). There are two main categories of arousal mechanisms known to be responsible for empathic affective responses. One category is the preverbal, automatic, and involuntary arousal mechanisms, including motor mimicry. The other is the higher-order cognitive modes, associated with conscious mental states, which include *perspective-taking* [136] defined as "putting oneself in another's shoes and imagining how he or she feels" [41], this being considered a very important component of empathy [137].

This leads to one other important aspect of social interactions: *interpersonal affect regulation*, defined by Niven et al. [36] as "deliberate attempts to regulate the feelings of others". Based on a corpus

[1] https://sites.uab.pt/tutoria-virtual/

of strategies for interpersonal affect regulation, Niven et al. [36] reached a classification that "primarily distinguished between strategies used to improve versus those used to worsen others' affect, and between strategies that engaged the target in a situation or affective state versus relationship-oriented strategies" (see Figure 3.1). This research was further developed, and studies were conducted to see the impact of using these strategies, with the authors having confirmed [138] that “attempts to improve others' affect will be associated positively, with the target reporting friendship and trust with the regulatory agent.”

	Strategies to improve affect	Strategies to worsen affect
Engagement strategies	Positive engagement	Negative engagement
	Affective engagement: Directly trying to improve the way the target feels about a situation, e.g., allowing the target to vent	Affective engagement: Directly trying to worsen the way the target feels about a situation, e.g., explaining how the target has hurt someone
	Problem-focused strategies, e.g., listening to the target's problems	
	Target-focused strategies, e.g., pointing out the target's positive characteristics	
	Cognitive engagement: Trying to change the way the target thinks about a situation in order to improve the target's feelings, e.g., giving the target advice	Behavioral engagement: Trying to change the way the target behaves in relation to a situation in order to worsen the target's feelings, e.g., complaining about the target's behavior
Relationship-oriented strategies	Acceptance	Rejection
	Attention: Giving the target attention to communicate validation, e.g., making it clear that you care about the target	Rejecting the target's feelings: Rejecting the target's feelings to communicate snubbing, e.g., making it clear that you do not care how the target feels
	Valuing, e.g., making the target feel special	Confrontational strategies, e.g., being rude to the target
	Distraction, e.g., arranging an activity for the target	Nonconfrontational strategies, e.g., ignoring the target
	Humor: Being humorous towards the target to communicate validation, e.g., joking with the target	Putting one's own feelings first: Putting one's own feelings first to communicate snubbing, e.g., sulking around the target

Figure 3.1: Interpersonal Affect Regulation Strategies Classification from Niven *et al.*[36]. The strategies are categorized by either improving or worsening affect, and by being engagement or relationship-oriented strategies. For example, Distraction is categorized as a relationship-oriented strategy that improves affect.

Following these works, Reeck et al. [139] introduced the Social Regulatory Cycle (SRC), a psychological model created to better understand the processes for which one Regulator influences a Target, and the latter's processes for dealing with said influence. In the work, four stages are defined for the Regulator: identify the target's emotions; evaluate the need for regulation; select a strategy, and; implement the strategy. These stages are used in the application to guide the agent's design and implementation.

3.3 Virtual Coach

In [67], we created a virtual coach capable of taking into consideration the coachee's goals and emotions, by both placing itself into the learner's shoes, a key competence that all great coaches share [135], and motivating/engaging the user through the use of interpersonal affect regulation strategies (see Figure 3.2).

The behavior of a virtual coach is divided into four main steps (see Figure 3.3): *affective appraisal* where, by gathering user data and modeling expectations, the coach determines the affective state of the student; *strategy selection* where, using the computed affective state of all involved in the interaction, an intervention strategy is selected to improve the user's affective state; *dialogue selection* where, having the strategy determined, a dialogue branch that best suits the strategy context is se-

Figure 3.2: Prototype of the "Virtual Tutoring" coaching application. On the left we see Maria, talking to the user under the attentive eye of her fellow coach, John.

lected; and finally, *dialogue management*, where the dialogue is presented through the use of two and three-dimensional elements to create a more engaging experience. The steps are detailed in the following subsections.

3.3.1 Affective Appraisal

In this step, the virtual coach calculates an expectation for what is likely to happen within the context of the ongoing tasks. This generates expectations, to which the coach will later affectively react to, based on what is actually reported during the interactions with the coachee. Here, the coach infers what the coachee is feeling and, based on its individual personality, its own affective state as a response to that. This personality maps a mismatch between expectation and outcome, in the context of the current mood of the virtual coach, to a new emotion. As such, different coaches with different personalities will have different affective reactions to the same event.

Gathering Data

We gather both qualitative and quantitative information over time. This is done to get a better understanding of what the coachee is feeling in the current session/task context. *Qualitative* information refers to a user's personal opinion on a particular task. To get a better understanding of how we can adjust our training approach, three distinct qualitative metrics are collected by querying the user about how challenging, enjoyable (intrinsic motivation), and important (extrinsic motivation) a task is, at different points in time. *Quantitative* information is also collected and gives the coaches more objective data to understand the results of the approach over time. Again, three different quantitative metrics are recorded over time: *task performance*; *effort*, the amount of time spent on a task; and *engagement*, how frequently the coachee engages with the task. These generic task-independent metrics provide the context for understanding how the coachees are progressing over time on the

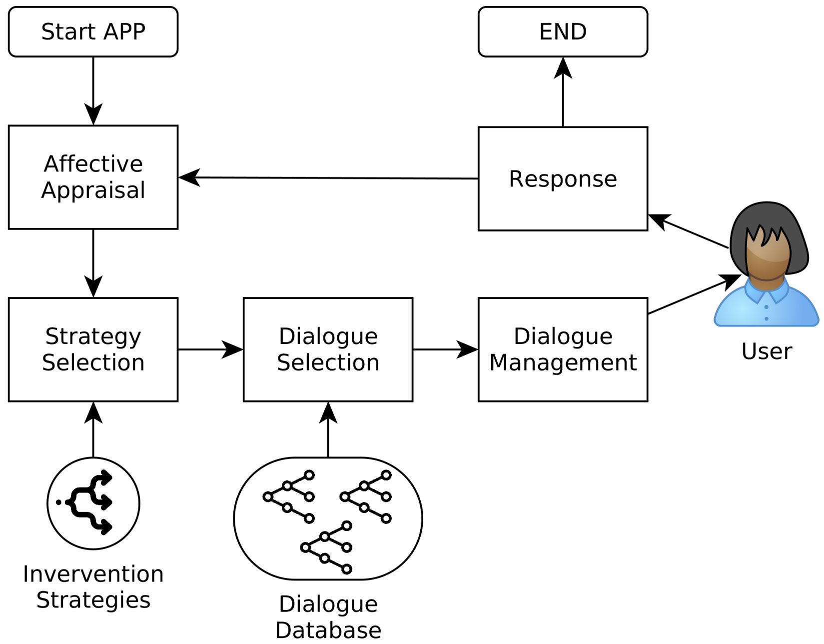


Figure 3.3: Overview of the Virtual Coach interaction loop: information is first collected and an empathic affective appraisal is made (*Affective Appraisal*), followed by the selection of an intervention strategy (*Strategy Selection*) that will dictate the dialogue that will be selected (*Dialogue Selection*). This dialogue is then interactively presented to the (*Dialogue Management*).

multiple tasks they are involved with.

Computing Expectations

The collected data provides us with past information, such as knowing the coachee's initial performance or attitude towards a certain task; more recent information, relevant for short-term actuation; and estimates for the collected metrics based on all the collected data. Using an approach based on the *Emotivector* model [7] (see Section 2.4.2), an affective response that takes into account the previous expectations is generated. In other words, the coach calculates an estimate for all metrics of the tasks at hand and reacts to an event by comparing the mismatch between the expected metric and the actual measured metric.

In [7], they made use of metrics that were collected with high frequency and low variance, yet our application domain collects data more scarcely and with high variance. As such, we approached the analysis of the time series differently, relying on a weighted moving average of the value and its first derivative. Consider that, at time t, a metric value is $x_t \in [0, 1]$, we predict $\hat{x}_{t+1}$, its value at time $t + 1$, using the Equation 3.1 below[2].

$$\begin{gathered} \hat{x}_{t+1} = \sum_{i=0}^{n-1} (x_{t-i} + \dot{x}_{t-i}) w_i \\ \text{where } \sum_{i=0}^{n-1} w_i = 1, \text{ and } \dot{x}_t = (x_t - x_{t-1}) \end{gathered} \tag{3.1}$$

[2]In our implementation, we used $n = 5$, and the following weights $w_0 = 0.5$, $w_1 = 0.2$, and $w_{(2,4)} = 0.1$, based on data collected from an empirical exploratory study.

Handling Multiple Metrics

When several metrics are used simultaneously, they all compete for attention in terms of their relevance to the interaction. We implemented the concept of exogenous salience presented in [7] (ignoring endogenous salience) and used estimation error to reflect the principle that "the least expected is more likely to attract attention" [7]. In essence, we select the metric with the greatest salience computed as: $(x_t - \hat{x}_t)^2$.

From Sensation to Emotion

Directly mapping a sensation (metric mismatch) to an affective expression based on the personality of the virtual agent (e.g. mapping a greater reward than anticipated to the expression of happiness) did not produce good enough results in our exploratory study. As such, we extended the approach to take into account the current mood of the virtual coach, itself based on the previous emotion expressed by the virtual agent, to enrich the expressive palette.

We began by defining a set of possible emotions to express based on Ekman's research [55] (Happiness, Surprise, Anger, Disgust, Fear, and Sadness, with the addition of a Neutral expression), and developed an animation system supporting the expression of each emotion on a continuous scale of intensity (i.e. $0;1$). Then we defined the *mood* as three possible states (positive, neutral, or negative) based on the valence of the previous emotion expressed by the virtual agent: Happiness and Surprise were mapped to positive mood; Anger, Disgust, Fear, and Sadness were mapped to negative mood; and Neutral was mapped to neutral mood. Based on the intended personality of the virtual agent, we then paired the mood with a sensation and connect them to the expression of an emotion, e.g. being in a negative mood and receiving a punishment better than expected is appraised as surprise (see Table 3.1). This approach allowed the creation of distinct and consistent synthetic personalities (defined as the matching of mood and sensation to an emotion) while being flexible enough to create distinct behavior, such as optimistic or pessimistic perspectives.

Table 3.1: Example of a coach's personality that not only maps the sensation to an affective expression but also takes into account the current mood based on the valence of the previous emotion expressed. Emotions have an intensity ranging from 0 (zero) to 1 (one).

	Punishment		
Mood	Worse than Exp.	As Expected	Better than Exp.
Negative	Anger(0.5)	Sad(0.2)	Surprise(0.2)
Neutral	Sad(0.5)	Neutral	Happy(0.4)
Positive	Surprise(0.5)	Happy(0.2)	Happy(0.2)
	Reward		
Mood	Worse than Exp.	As Expected	Better than Exp.
Negative	Anger(0.2)	Neutral	Surprise(0.7)
Neutral	Sad(0.2)	Neutral	Happy(0.4)
Positive	Surprise(0.3)	Neutral	Happy(0.9)

3.3.2 Strategy Selection

Having both the user's perceived emotions as well as the coach's affective response, the virtual agent decides how to act by selecting the appropriate intervention strategy from a library that contains both task-oriented strategies and affect regulation strategies. *Task-oriented* strategies focus on the completion of a specific task, e.g. preparing for an important exam. Given their nature, these strategies are highly dependent on context, where some may depend on a particular event while others need to be presented in sequence. *Affect regulation* strategies, based on the work of Niven *et al.*[36], focus on improving/worsening affect while also differentiating between engagement and relationship-oriented strategies (see Figure 3.1). With these strategies, we encourage the creation of an emotional bond with the user through interaction.

The selection of a specific strategy is made through a map between a measured metric (see section 3.3.1) and a sensation (e.g. reward worse than expected) to generate a strategy. This mapping was informed by a pedagogical team from the institution where our experiments took place. Further details are presented in [67].

3.3.3 Dialogue Selection and Management

After a strategy is selected, the dialogue best suited for tackling said strategy is chosen from a database of pre-built dialogue trees[3]. Each tree is tagged with information regarding its assigned strategy (e.g. engagement strategy), as well as the emotional states both the coaches and the users should be experiencing in the given context.

The coaches then interact with the user by traversing the dialogue tree, while expressing the emotions previously selected, which are manipulated by both dialogue and user interactions. Note that coaches are able to react to user responses, but also to a lack of response (e.g. if the user takes too long to respond), giving attention to both active and passive user behavior. This dynamic gives the coach the ability to express multiple emotions during a single action, as well as being able to react to certain situations that break the regular flow of dialogue, such as interruptions or lack of engagement from the coachee. We created two subsystems to handle the expression of emotions in this critical stage of the interaction with the user: the *Bubble System*, which controls the graphical dialogue elements; and the *Avatar System*, which controls the synthetic characters and their animations. Both systems focus on the transmission of the six primary families of emotions [55], with the varying degrees of intensity elicited during the affective appraisal (see Section 3.3.1).

3.3.4 Bubble System

The Bubble System controls all 2D elements used to communicate emotion, where the main component is the balloons. The relevant aspects of the balloons are their shape, text, colors, and animations, which change accordingly to the emotion felt by the character. More details of each component are presented in Appendix B.1, and a full description of the Bubble System can be found in [128].

[3]In our implementation, we created the dialogues for a small set of exercises with the students in our use case and reviewed by a pedagogical team. This led to the removal of certain words considered negative from the first drafts.

Balloon Shape – We created one distinct speech bubble for each emotion, plus a neutral state, based on the ones traditionally used in Comics (e.g. spiked balloons for anger) as well as using design guidelines to convey the appropriate valence for the associated emotion. To that end, we followed a user-centric iterative process until reaching a desired response to a speech balloon (e.g. curves for positive valences, and lines/obtuse angles for negative valences). For example, the balloon for anger is defined in the standard comic way as a spiked shape[4], while disgust ended up having a wave format in order to convey the feeling of nausea.

Balloon Color – We decided to follow the same approach as Pires et al. [140], and draw inspiration from the Disney™ movie "Inside Out"[165]. The colors chosen follow the same representations as those present in the movie, and for those emotions that were absent, a mix of other emotions/colors was used, to help communicate the appropriate valence (e.g. surprise was defined as orange by mixing the aggressiveness of red and the positive valence of yellow).

Balloon Animation – We decided to create simple balloon animations as scaling/motion over time, following the pose to pose principle (see Section 2.6). In total, we ended up with 14 distinct animations, a pair (one to show the balloon, and another one to hide it) for each of the six emotions, plus another pair for the neutral state. To distinguish between different intensities, the speed of the animation was changed accordingly (slower speeds for lower intensities, faster speeds for higher ones).

Text Animation – Regarding the text component of the speech balloons, we went with a parametric approach using animation curves (curves that define the text animation functions), since this allows us to have generalized effects and apply different curves to them, thus producing different animations (e.g. appear over time, jump, fade in/out, and shake).

3.3.5 Avatar System

The Avatar System manipulates the animated responses from the synthetic characters. The system is implemented as an animation controller with parallel layers of state machines. When presented with an emotion to express, the system updates the significant animation parameters and transitions to the required animations states. This approach of having each layer as a separate machine allows stacking animations together (i.e. blending), with the benefit of being able to use simple animations to achieve greater ranges of expressiveness. Our controller focuses on three main aspects: Expressions; Moods; and, Events. These can be blended together when relevant (e.g. to mix motion and expression). A full description of the Avatar System can be found in [129].

Expressions – The Expression Module handles brief expression requests, to be interwoven to create varied behavior. The main design goal of this subsystem is to support a variable range of emotional behavior, only having access to a limited number of assets. To achieve this, we implemented a

[4] Representation of scream (anger) as seen in Spider-Man Annual #5 - A Day at the Daily Bugle: `https://imgur.com/a/lJOl2tw`

system of blending trees for each emotion. This allowed us to tune the resulting intensity by adjusting the weight of the clips assigned as the blend tree's nodes.

Moods – The Mood Module coordinates the idle animations that communicate a character's emotional state. These dominate most of the facial regions and had to be attenuated when composed, otherwise, the output of the expression module would go unnoticed. Additionally, we also needed to animate the full range of back-channeling modalities, such as the various eye and head movements. This involved adding another set of overlapping layers with these animations, to play alongside the already existing idle animations.

Events – The Event Module comprises all animations not directly linked to the expression of emotion (i.e. attention-driven gaze, and speech-driven mouth movements). For these event animations, we were once again required to add a damping mechanism, so that they could play smoothly with those of the previous modules[5]. Additionally, to ensure that these motions adapted to the emotional state of the character, we added a slight variation to the speed and frequency those animations play at[6].

3.3.6 System Summary

The behavior model of the virtual coach focuses on four main steps: *affective appraisal*, *strategy selection*, *dialogue selection*, and *dialogue management* (see Figure 3.3). In these steps, we analyze and gather both qualitative and quantitative information, so that the coach can deduce the emotional state of the user, as well as its own synthetic affective response, which may differ based on the implemented personality. This is achieved by contrasting each tracked metric with its predicted value. Once the strategy (be it task-oriented or affect regulating) is decided, the appropriate dialogue tree is selected based not only on said strategy but also on the emotions and expectations of the coach. Lastly, the appropriate emotional responses are expressed by the coach through the use of 2D and 3D visual feedback while, at the same time, gathering information based on the responses given by the user during the interactive dialogue.

3.4 Virtual Tutoring Use Case and Evaluation

Returning to the previously mentioned problem in the rise of disinterest of students in online learning environments, a preliminary evaluation was performed to better understand and appraise the approach described, where the virtual coach tutors in a real-world online distance learning environment scenario to better understand their impact.

A prototype was developed with two 3D animated virtual coaches (see Figure 3.2), that interacted with and assisted students, during a 10-day period, while preparing for the final exam of an online course. During these 10 days, a sample of 13 students (6 female) with ages ranging from 29 to 66 years old interacted with the virtual coaches on their mobile devices, directly providing them with both

[5]Difference in talking animations with and without damping: `https://goo.gl/a8wb5F`
[6]Speed/frequency difference between the moods of happiness and sadness: `https://goo.gl/Yx9dHo`

quantitative and qualitative data on how their study was progressing, and receiving comments and suggestions from them. Quantitative data was based on the amount of daily study time reported by the students, as well as the results from three formative assessments, while subjective data included perception of challenge, as well as endogenous and exogenous motivation. At the beginning and end of the evaluation, participants were asked to fill out a questionnaire on their beliefs regarding virtual agents. Additionally, a System Usability Scale (SUS)[141] questionnaire was given at the end of the evaluation to assess the general usability of the system and inform of eventual interaction problems that may have occurred.

Answers to the questionnaires show no changes were detected concerning the beliefs of the participants, in regards to the expression of emotion in virtual agents, during the conducted pre- and post-tests. An interesting fact, however, is that some students continued to interact with the application even after the trial period was over, possibly showing interest in continuing to use the features provided by the application. The median SUS score is 85 ($min = 57.5$, $Q_1 = 75$, $Q_3 = 95$, $max = 100.0$), which indicates an above-average score and near excellent rating [142]. These early results indicate a potential in this application, where users want to use it and seem to improve their studies, but improvements are needed in order to resolve inconsistencies and better the emotional and empathic impact of the tutors.

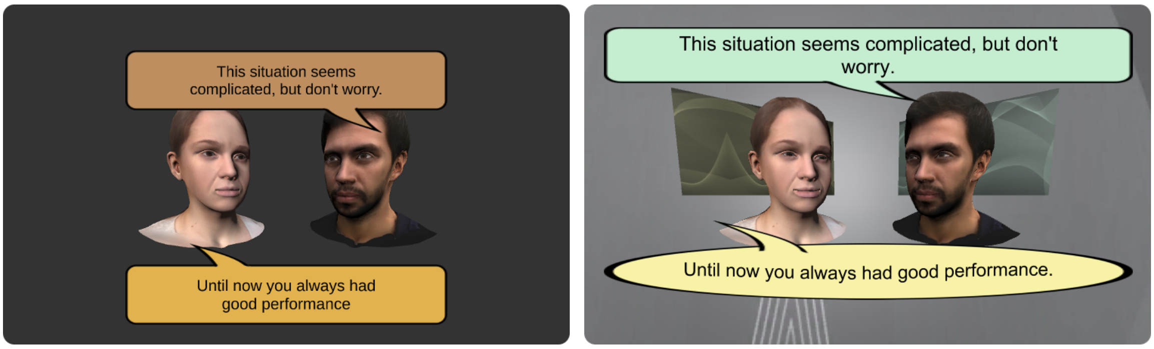
3.5 Evaluating Emotional Expression and Recognition

During development, concerns over the need for the clear expression of affective states led us to carefully consider the evaluation methods supporting our decisions. To that end, two versions of the application were built to perform comparative studies:

First version – This version features a visual expression mechanism using the characters and base animations provided by Didimo[7]. Dialogue-wise, text is presented via speech balloons featuring a brown hue, a rectangular shape, and two short fade animations (one to show the balloon, and another to hide it). The environment around the characters was comprised solely of a gray static background (see Figure 3.4a).

Second version – We created this version with the intent of creating more natural expressions and, at the same time, improve emotion expressions through 2D elements. It is capable of smooth transitions and varying intensities, allowing the characters to express emotions while talking and reacting to events. The speech balloons feature distinct colors, shapes, and animations for different emotions/intensities. The background environments also assist in conveying the mood of the characters, and in the portrayal of the conversation topic (see Figure 3.4b). Essentially, this version encapsulates the previously described components that interface with the user, the *Dialogue Management* and its subsystems (see Section 3.3.3).

[7] Didimo, Inc. `https://www.didimo.co/`



(a) *First* version **(b)** *Second* version

Figure 3.4: The two versions of the coaching application depicting Maria (left), talking to the user under the attentive eye of her fellow coach, John. The rendition of text shown here was replaced in the experiments.

Having the two versions, we conducted two separate experiments. The *first experiment* was aimed at understanding whether the changes added by our modules had improved the expression of emotion in our system. We enlisted the help of two separate population samples by having them compare, side-by-side, the first version of the application with the second version of the application. One sample was comprised of *experts*, namely, Ph.D. students and professors from the fields of intelligent virtual and robotic agents, while the other contained only non-experts. The intent was to analyze whether the data collected from the two samples would present any significant disparities.

The *second experiment* focused on emotion recognition, given the importance of understanding whether the participants were capable of recognizing the emotions being communicated by the application. We gathered a population sample mixing both experts and non-experts, with similar demographics to the first experiment, and uniformly divided it into two distinct groups. One group would experience only the first version of the application, while the other would experience only the second version. This would allow us to investigate the positive/negative impact our changes had on the users' ability to recognize the conveyed emotions.

During these experiments, all the dialogue text was replaced with sentences composed of a single letter (e.g. *"OOO O OOOOO"*) to avoid biased information elicited from interpretations of the dialogues.

3.5.1 First Experiment: Emotion Expression Comparison

Thirteen videos, featuring distinct expressions of emotion from each of the two systems, were recorded and randomly presented, side by side, in an online questionnaire[8]. Participants were asked to evaluate the success of each animation in communicating the specified emotions, for both systems, on a scale from 1 to 7, with 1 being the lowest rank. A total of 10 responses for the non-expert sample were collected, with 5 belonging to female participants. The average age was 27 years old, ranging from 23 to 40 years of age. The expert sample had the same number of participants and a similar distribution of sex and age, averaging 29 years old, on a range spanning from 22 to 36 years of age. Altogether, we collected a total of 260 comparisons between the two versions of the application.

[8]A version of the questionnaire presented to the participants of the first experiment can be found at `https://goo.gl/PmUztE`

Table 3.2: Scores given by participants (experts, non-experts, total), for both intensities and their union, divided into quartiles for the First and Second versions.

Scores given by participants		**Low Intensity**			**High Intensity**			**Total**		
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
First	Non-Experts	2	3	4	2	2.5	4	2	3	4
	Experts	1	2	3	1	2	3	1	2	3
	Total	2	2	3.25	2	2	4	2	2	4
Second	Non-Experts	3	5	6	4	6	7	3.75	5	6
	Experts	2	4	5	4	5	6	2.75	5	6
	Total	2	4.5	6	4	5	6	3	5	6

Looking at the median scores (Q2) from the non-expert sample, depicted in Table 3.2, we can observe that participants found the second version to be better at conveying the specified emotions. A Wilcoxon signed-rank test shows that this difference had statistical significance $(Z = -6.771, p < 0.0005)$. On a 1 to 7 scale, we have a median value of 3 for the first version (regardless of emotion intensity), while the second version has a median value of 5. If we separate between low and high intensities, the disparity in median values remains, with the gap between high intensities increasing by a small margin. These median scores were found to be of statistical relevance, for both low $(Z = -4.484, p < 0.0005)$ and high $(Z = -5.109, p < 0.0005)$ intensities.

For the expert sample, we could observe a wider gap, with a median value of 2 for the first version (when ignoring emotion intensity), while the second version features an identical median value of 5. A Wilcoxon signed-rank test shows that this difference also had statistical significance $(Z = -6.779, p < 0.0005)$. Taking into account emotional intensity, the tendency observed in the first condition remains, with the gap between the medians of high-intensity emotions showing a slight accentuation. These scores were also found to be statistically relevant, for both low $(Z = -4.032, p < 0.0005)$ and high $(Z = -5.276, p < 0.0005)$ intensities.

In summary, the second version was seen by experts and non-experts to be better at conveying emotions than the first version, independent of the intensity.

3.5.2 Second Experiment: Emotion Recognition

Twenty-six videos (13 for each version) featuring different emotional expressions (6 primary emotions x 2 intensities + neutral), were recorded and randomly presented in an online questionnaire[9]. For each video, participants were asked to identify which emotion they believed was being expressed, its intensity, and what aspects (characters/balloons/environment) influenced their decision. The questionnaire was based on the Geneva Emotion Wheel[143], with the participants selecting first the recognized valence, followed by the respective emotion (from a set of 20 possible). A total of 38 respondents participated in our survey, with a total of 170 videos watched for the second version and 156 for the first version.

Comparing the new data with the results from the previous experiment, it was possible to identify emotions that were perceived as better animated during the first experiment but suffered from poor

[9] A version of the questionnaire presented to the participants of the second experiment can be found at `https://goo.gl/8ZaZT6`

emotional recognition in the second experiment. An example is the emotion of sadness, specifically, its low-intensity expression (that we will refer to as sadness-low) had poor accuracy in both versions, with the second version having 0% of correct answers. Surprise-low was also more accurately perceived in the first version, despite having previously obtained favorable results in the comparative experiment. Nevertheless, we also have emotions that coincide with the favorable results obtained in the comparative experiment, that point to them as being better animated, namely anger-low (see Figure 3.5), fear-high, and surprise-high. The remaining emotions did not show significant signs of being better or worse in terms of emotional recognition, even though they were pointed as being better expressed in the second version during the first, side-by-side, experiment.

If we crosscheck our results with the findings of Bassili [87] on emotional confusion (Table 3.3), we can observe that most recognition mistakes falling outside said findings were corrected in the second version, and the remaining cases exhibit only singular occurrences. The exception here is fear-low, which began being mistakenly recognized as anger.

Table 3.3: Overlap (blue background) of Bassili's findings and the emotional recognition mistakes obtained during testing. Each number represents the difference in mistakes between the First and Second versions, with positive values representing more mistakes in the Second version (e.g. high-intensity disgust is confused with anger in the first version, but is confused with fear in the second version of the application).

. . . could have been confused with . . .

	happiness	surprise	sadness	fear	disgust	anger
happiness		1		1		
surprise				1		1
sadness					-1	1
Fear	-2	3			-2	3
disgust		1		1		-3
anger						

Low Intensity

	happiness	surprise	sadness	fear	disgust	anger
happiness						
surprise						1
sadness	1			-1		
Fear		-2	2		-1	-1
disgust				2		-4
anger			1	2	-2	

High Intensity

Regarding the dimensions of expression reported by the participants as influencing recognition, the animated characters were reported as being 1.45 times more influential than the balloons, and the balloons 3.55 times more influential than the environment. The more prominent reported features were facial expressions (29.02%), followed by the shape of the speech balloons (18.13%). It is important to note that the relative importance of each feature changes from one emotion to the next. For example, fear-high and anger-low were better recognized in the second version, but the character features were reported as more relevant for fear while the balloon features were pointed as more relevant for anger. This information is crucial to focus development on the features that are the most relevant for effective emotion expression.

3.5.3 Discussion

The two experiments offer conflicting results regarding emotion recognition. The first suggests that the second version of the prototype is better than the first at expressing any emotion. While the second study suggests that the second version isn't as good as it might have appeared.

A possible factor contributing to these differences is that side-by-side comparisons may lead to feature dominance, meaning that users can report preferring a specific version (e.g. for being more 'polished' than the other), despite being worse at communicating the specific emotion (e.g. for having contradictory cues, confusing users, etc.). In our prototypes, although both used the same models and animations, the second version had blending animations, animated backgrounds, richer text, and bubble animations, which may lead to feature dominance.

Furthermore, in the first experiment users were asked to rate the expressions for a given emotion (e.g. "How would you rate A_i and A_j, when comparing the two in the context of expressing a specific emotion E?") and in the second experiment users were asked to select which emotion best suits the animation (e.g. "What emotion did A_k expressed? E_1; E_2; E_3; ..."). Another possible factor may arise in which the first version leads to *forced-choice* – "an unintentional cheat sheet for the test subjects"[57] – which guides participants to an answer. In the first experiment, if the question mentioned *sadness*, the participant would only have to compare the animations to their concepts of sadness. In the second experiment, the selection pool was wider, thus the participant was not guided in their answer and had to try and recognize the expressed emotion, which resulted in less accurate emotion recognition. Although the experiments measured different metrics, we must be aware of forced-choice and its impact on the participants.

Because the experiments had a limited number of participants, and because several factors were changed from the first to the second version of the system supporting the expression of emotion, it is impossible to pinpoint a specific change that could explain the obtained results. A more interesting question this experiment raises is *how much information can incremental evaluation provide during the development of virtual agents able to express emotions?*

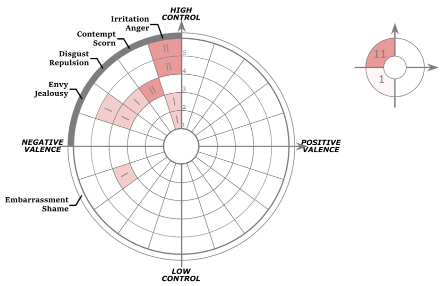
3.6 Summary

In this chapter, we presented an architecture that allows for the creation of intelligent virtual agents acting as coaches in distance learning environments. Based on (1) the objective and subjective metrics provided by the users, as well as the learning environment, (2) the affective state of the users, inferred from the evolution of these metrics over time, how they create expectations and how these expectations are actually met, and, (3) the synthetic affective response of the virtual coach based on its synthetic personality, the system is able to select from a repertoire of dialogue trees gathered from a human pedagogical team, the one that better conveys the interpersonal regulation strategy appropriate for that situation. We then provided an overview of all components built in a first implementation of the approach and described a real-world case study featuring two virtual coaches, running on a mobile device during a 10-day evaluation period, whose results support the interest of students in the continued interaction with the empathic virtual coaches, in the context of online distance learning education.

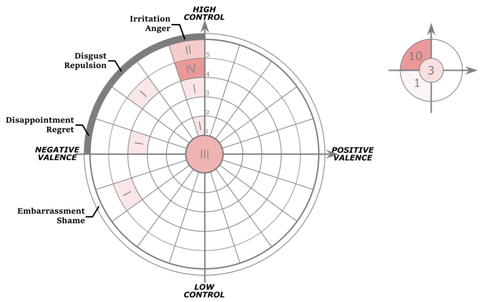
Finally, to better understand the impact on emotion expression/recognition of our prototype and the changes made to make said expression more natural, we conducted two sets of experiments. The first one directly compared the first version to the second version (side-by-side) while expressing specific emotions, to both expert and non-expert participants. Results clearly supported that the second version was better rated at communicating each individual emotion. The follow-up experiment asked participants to recognize individual emotions expressed by each version. The results showed that emotion recognition of the second version had not improved in all cases, as was expected based on the first experiment. We noticed that, although the expression for some emotions was rated as being better in the first experiment, the second version had a lower recognition rate than the first version, as we could confirm in the second experiment. We also noticed that particular elements were reported as influencing emotion recognition in the different emotion expressions, pointing to separate needs for each family of emotions.

This discussion exemplifies why, when choosing the system that will better support the expression of emotions for intelligent virtual agents, in the sense that the portrayed emotions will be easier to recognize, asking the users to compare two different versions of the system, side by side, may not be the most adequate way to proceed. This study, therefore, mirrors the incremental formative evaluations that often underlie virtual agent development, and offers a cautionary tale in evaluating aspects of a system during its development and its evolving context of use.

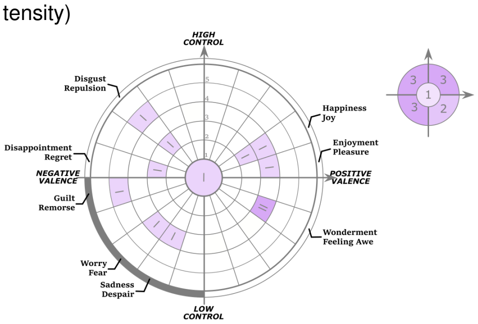
This chapter helps inform the development of specific empathic synthetic characters, we can extrapolate this knowledge to other contexts as it shares many common features of synthetic characters development. The emotion generation algorithm presented here will be used as the basis of the emotional model in the present work, as it provides anticipatory capabilities to the agents using Emotivectors. Furthermore, the lessons learned from the comparative studies will help guide future evaluations of the main work, not only as cautionary tales but also as comparative data when considering emotion recognition tasks.



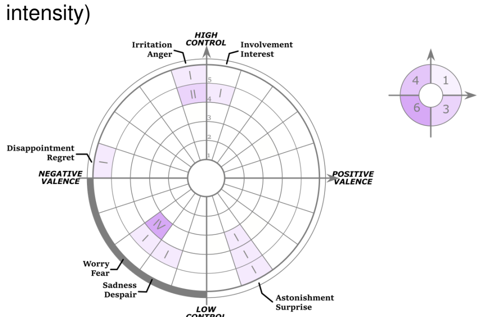
(a) Heatmap for the *First* version of Anger (low intensity)



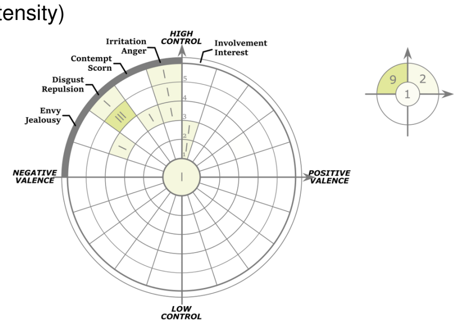
(b) Heatmap for the *Second* version of Anger (low intensity)



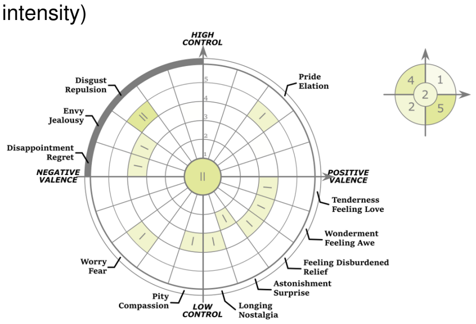
(c) Heatmap for the *First* version of Fear (low intensity)



(d) Heatmap for the *Second* version of Fear (low intensity)



(e) Heatmap for the *First* version of Disgust (low intensity)



(f) Heatmap for the *Second* version of Disgust (low intensity)

Figure 3.5: Heatmap for the emotions of anger, fear, and disgust (low intensity), for each application version. In the bigger wheel, the roman numerals indicate the number of times an emotion was recognized, and the Arabic numerals (one for each ring) indicate the intensity said emotion was recognized at. The middle circle indicates how many times users did not recognize any emotion. The smaller wheel indicates the total number of emotions recognized in each quadrant. The heatmaps for anger and fear exemplify emotions that were better recognized in the second version of the application (although fear is more often confused with other emotions). The heatmaps for disgust exemplify an emotion that was better recognized in the first version.

4

Multi-party interactions with *3motion*

Contents

In this chapter we discuss 3motion*, first developed in [144] and expanded here, is a model that splits actions into separate stages based on the principles of traditional animation[106] (see Section 2.6), aligned with an emotion expression model based on anticipation, it allows the creation of synthetic characters that can more easily react to others and their actions, making use of non-verbal communication. What follows is a discussion of its inner workings, as well as a summary of the first evaluation, performed using a text-based application use case.*

4.1 Action Subdivision

To enable believable multi-party real-time interactions between human and synthetic characters, we created *3motion*[1][16, 144], which approach deconstructs the traditional atomic action, generally used when implementing synthetic character behavior in virtual environments, into three distinct stages[2]: *anticipation*, *action*, and *follow-through* (see Figure 4.1).

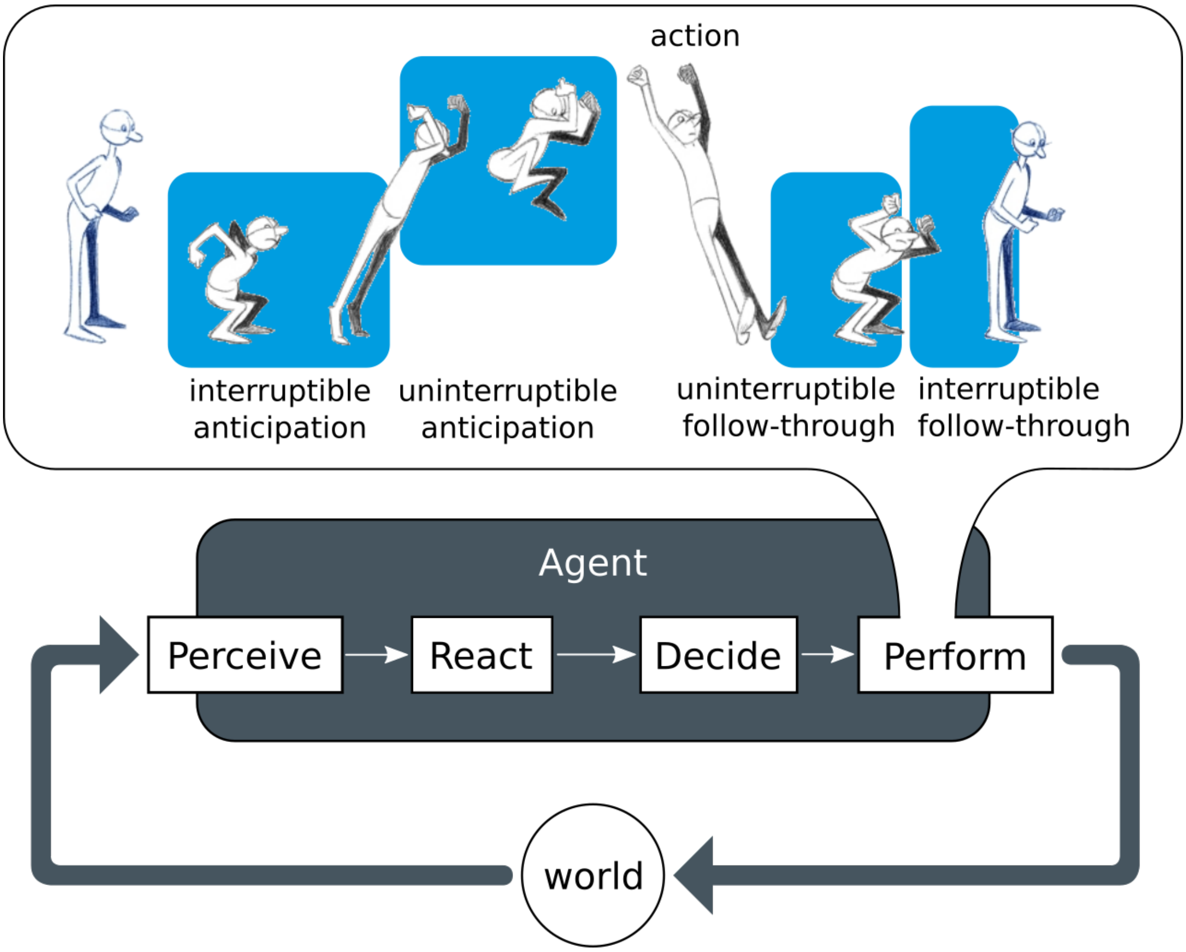


Figure 4.1: *3motion* conceptual approach depicting the four stages of an action and its connection to the agent's behavior. Note that there can be multiple iterations of the agent's behavior cycle happening while a single action is evolving. The 4 behavioral stages of an agent are also presented at the bottom of the figure. The agent perceives events from the world, that are then interpreted to create an emotional reaction and a decision on what action to perform, then the action starts or continues executing which impacts the world.

[1] The name *3motion* originates from the combination of Emotion and the three (3) stages of a subdivided action: *anticipation*, *action*, and *follow-through*.

[2] This approach was inspired by the work of Costa[51] where they subdivided an action into two stages: **anticipation** and **action**.

Each stage may take a certain time to play out, and interactions at different points in the sequence of events will have a different meaning for the other agents participating in the scene, as well as for the viewer passively watching or actively interacting with the scene, allowing for the creation of a rich interaction with multiple agents and/or users (certain stages can be interrupted, allowing for more unique interactions to be created).

The *anticipation stage* serves the purpose of communicating the intent so every other agent understands it and can expressively prepare for it[51]. In this stage, an agent broadcasts its intent and associated emotion and receives feedback in the form of emotions from the other agents that are aware of the expression. This allows an agent to interpret the emotional reactions in the context of his intentions and modulate his behavior accordingly. This stage is further subdivided into two sub-stages: an *interruptible stage* in which the agent is still able to cancel the initiated action, and an *uninterruptible stage* in which the action reached a point where it cannot be stopped. The precise timing of the expression and affective responses is crucial in these stages, otherwise, the audience may not be able to understand how decision-making took place in this shared affective context.

The *action stage* is instantaneous and only exists conceptually. This stage represents the moment the action is resolved and the state of the virtual world changes.

The *follow-through stage* is entered after the action is resolved and its result is broadcast, which will meet or challenge the expectations of the agents. As in the anticipation stage, the agent sends its affective appraisal to others, allowing them to feel happy, sorry, etc., and receives their affective feedback to perform a final appraisal of the action. This stage is also subdivided into two sub-stages: an *uninterruptible stage* in which the action is unable to be stopped, and an *interruptible stage* in which the agent regains the ability to terminate the action and start a new one. While in the interruptible stage, if the character does not start a new action (e.g. swing with a weapon) the action follows its natural follow-through, yet if a new action begins, this action is completed (its animations are interrupted, but its purpose has been completed). The follow-through stage marks the end of the action, asserting whether the acting and observing characters' expectations are fulfilled. As such, characters express their emotions according to how their expectations were or were not fulfilled regarding the expected result of the action.

Each of these action stages is associated with a different timing and duration, thus creating different moments for possible emotional responses. As a result, characters may have multiple reactions to a single action instead of reacting to an action as a whole.

4.2 The Endless Loop

To allow for the expression of an action and its stages, we modeled an agent's behavior in a 4-step *anticipation-based* cycle (see the bottom of Figure 4.1): (1) the agent *perceives* changes in the world based on its *expectations*, (2) *reacts* to them emotionally based on what was *anticipated*, (3) *decides* what to do next and *anticipates* what will happen, (4) and *performs part* of an on-going action (it is here that we apply the action subdivision), then repeats. This cycle occurs multiple times in the course of an action allowing the virtual agents to perceive events and express different emotions, among other expressions, while performing the same action.

4.2.1 Perceive

In the *perceive* step, the agent perceives events around them in the world. Special attention is given to two events, those from actions and emotions, that will help inform the next steps and give context to what is happening. Since characters will take some time to understand what others are doing and feeling, we explicitly model a delay in the perception of new events that represents the time required for the event to be understood by the virtual agent. The delay is used to emphasize the non-immediate response of agents, improving Believability, and its values are informed by research on vision, more specifically the Sustained System [145], and start at 200ms (humans take around 200 to 350ms to perceive new information, we can extend these values to improve readability and believability, see Appendix A.3). The delay, which varies from one event to the next and may be modulated by the perceiving agent, allows for the expression of interesting behavior such as the expression of surprise or being actively waiting for something. This also translates well to the viewers' understanding of what is happening in the scene, since the viewers themselves will generally take some time to understand what is going on [145].

Note that the agents in our model can always perceive the correct emotion and action performed by another, if they do not perceive it correctly it is a deliberate design choice. This approach is akin to theater, where all cast members know what the others will perform and react accordingly. This is done to favor the believability of behavior and reinforce others' expressions.

4.2.2 React

The *react* step selects the emotion the agent is feeling as a result of the perceived events as well as its expectations, in the context of the actions being performed at the time by each one of the virtual agents in the scene. This step is called throughout the execution of an action, allowing an agent to feel different emotions in the course of a single action, which allows the creation of a more believable behavior. The implementation of the appraisal of events and the generation of an emotion is based on Emotivector, the affective model by Martinho and Paiva [7] (see Section 2.4.2), where for each action they will expect a reward or a punishment, with this approach the agent can create expectations regarding the action's success and react upon them.

Emotivector allows the computation of Sensations (the mismatch of perceived and predicted events), which can then be interpreted and allow the creation of emotions. Our implementation can make use of several Emotivectors calculating predictions for continuous measures (e.g. perceived walking speed of opponent) and more discrete measures (e.g. a probability of defeat based on health and attacks received).

For continuous measures we assume that the "sensed value will tend to remain constant except for certain points in time"[7], thus we make use of the simple predictor proposed by Martinho and Paiva. Regarding discrete measures we assume the data collection is more scarce and with high variance (similar to what is seen on the Virtual Coach see Section 3.3.1), thus we make use of weighted moving averages and their first derivative. For both cases, a "searched" value might exist, depending on the context, in which case it will help generate more relevant sensations (e.g. expecting the opponent to be in a certain location to trigger a trap).

Having the emotivectors, the management of multiple emotivectors is highly dependent on the use case. It is feasible to assume one can use the solutions proposed by Martinho and Paiva[7], Winner-Takes-All (also used in the Virtual Coaches), Salience Ordering, and Meta Anticipation. We plan to make use of Meta Anticipation with additional rules that override the selection and allow the management to be more specific for the context (e.g. if the character is defensive, it would give more importance to emotivectors that refer to preserving its health and less to those referring to damaging the opponent).

Regarding the generation of emotions, based on the implementation of the Virtual Coach (see Section 3.3.1), we have an initial approach to the generation of emotions through a mismatch between the most salient Sensation and a previous mood. It is a simple solution that showed good results and is a good first implementation for this model.

4.2.3 Decide

It is in this step that an action is chosen to be executed. If the character is controlled by a human player, this step simply transforms the player's input into an action that can be later manipulated (if need be) to better express an emotion (e.g. having a character walk slowly if they are scared). If otherwise the character is being controlled by an AI, then a decision-making algorithm decides what action it wants to perform (in our first use case we implemented a simple rule-based algorithm to select actions, see Section 4.5).

4.2.4 Perform

The *perform* step allows the execution of an action and the expression of an emotion. It is at this step that the action subdivision is implemented and where emotion expression will be applied. This step will manage the action being portrayed and keep track of what stage it is in, meanwhile it will blend the action with the emotion expression. Given the high dependence on the setting of a use case, performing an action needs to be reviewed case by case and its implementation needs to take into account the perception of the user (similar to what is discussed in Section 4.2.1, but where the focus is on the expression itself).

4.3 Behavior Illustration

Going back to the illustrative scenario first introduced in the Hypothesis (see Section 1.3), here we will expand it and use it to showcase the action subdivision. The scenario, now depicted in Figure 4.2, is set in a fantasy-inspired setting where two characters, Rua and Gorm, fight in an arena until one falls. Gorm is holding a two-handed axe and Rua a long spear. The figure is followed by a text description of the same scenario with annotations of numbers that link back to specific moments in the figure.

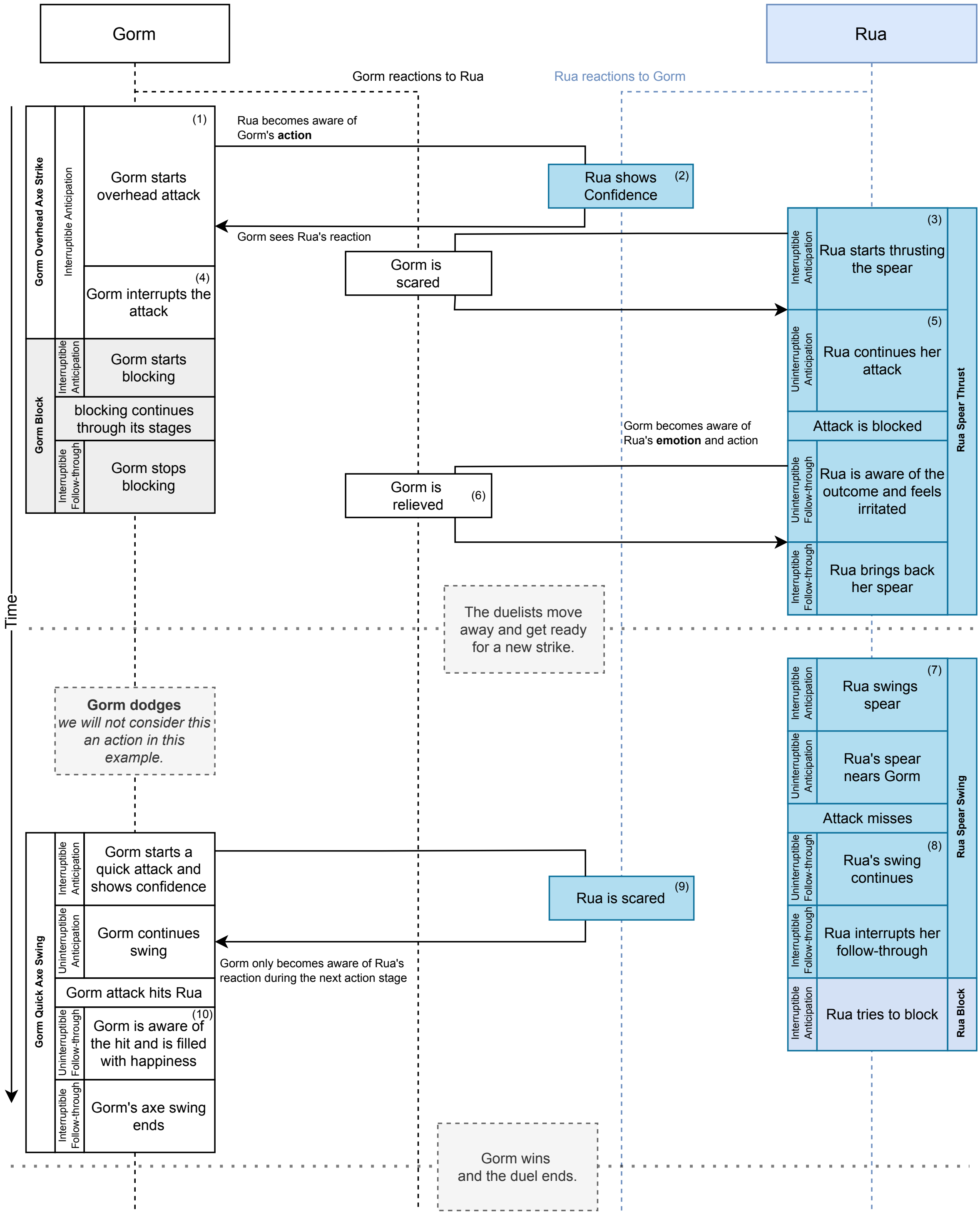


Figure 4.2: Action Flow of the Gorm dueling with Rua and their reactions. The figure depicts the subdivision of the action into several stages that allow for emotional non-verbal communication. Furthermore, it showcases the emotional response of the opponent during one's actions. Due to limited space and its similarity to the block action, the dodge action is not expanded in this example. Some blocks are annotated with numbers (e.g. "Rua shows Confidence" is annotated with the number *2*) to link them with the text description of the scenario.

The battle starts and both fighters run toward each other. As they get closer Gorm decides to perform a very slow but highly destructive overhead attack by pulling his axe back, revealing his intention in the process **(1)**. Rua becomes aware of this and shows confidence because she believes the attack will miss **(2)**. She then decides to thrust her spear towards Gorm **(3)**. As a result, Gorm becomes scared and decides to cancel his attack to defend himself **(4)**. Rua's attack continues **(5)** and is blocked by Gorm, leading Gorm to feel relieved and Rua irritated **(6)**. Both fighters move away from each other to prepare for the next attack.

The first one to attack again is Rua, now with a slower spear swing **(7)**. Gorm decides not to attack and focuses on dodging the attack. Unfortunately for Rua, she miscalculated the distance and the attack misses completely **(8)**. Gorm takes advantage of this opportunity and attacks with a much quicker axe swing leading Rua to feel scared for her life as she cannot recover quickly enough to defend himself from the attack **(9)**. Gorm's axe swing hits Rua defeating her in the process. Now Gorm feels happy since he has become the winner of this battle **(10)**.

In this scenario not only do two characters fight (common in video games) but also any wrong move can end the fight tragically (more common in movies), this creates tension and intense emotions during the combat, a good example where the use of *3motion* helps improve believability. The use of *3motion* ensures that the characters express emotions at the correct timing, which helps create more believable interactions. Another scenario depicting the use of *3motion* is shown in Appendix C.1.

4.4 Mental State

For the agent to be able to react to others' actions more effectively, each agent must have *theory of mind* and/or *empathy* (see Section 2.1.4). In our approach, each agent has some theory of mind and keeps mental states associated with itself and with other agents: each agent stores what actions are being performed by which agent at a certain time and what stage they currently are in, as well as the emotions felt and perceived at each stage of the action. This not only allows the agent to be able to predict certain behavior and express emotions accordingly but also allows the creation of rich and coherent behavior based on the dynamics of what was perceived by each agent.

4.5 Use Case: Text-based Application

As an early attempt of understanding the expressive power of our approach[16], a real-time text-based representation of a scene was developed with interacting agents and performed a user study with different variations of our model. Progress bars were used to represent playing an animation of an action and/or an emotion (see Figure 4.3). A small text next to each progress bar would describe the associated animation. Several progress bars would be displayed at the same time to mimic the concurrent execution of several actions in the scene.

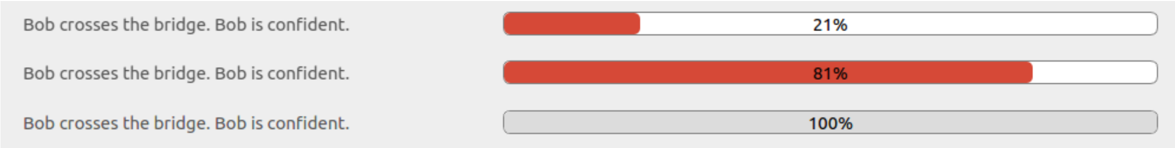


Figure 4.3: Example progression of an animation. Three moments of the execution are presented: just starting (top), almost finished (middle), and finished (bottom). The progress bar represents the time required to fully play the animation. The text depicts the action and associated emotion the agent is performing in the animation.

4.5.1 Interface Design Exploration

It was important to let the users understand what was happening. When the action started to unfold a message would appear informing the user about what was happening and, beside it, would appear a progress bar, informing the user of how long a stage had progressed, in a video game progression could be an animation being played (see Figure 4.4). When a stage was complete or interrupted the progress bar changed its color from red to gray, symbolizing that stage's end (see Figure 4.5a).

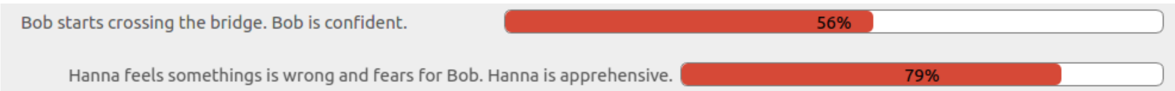
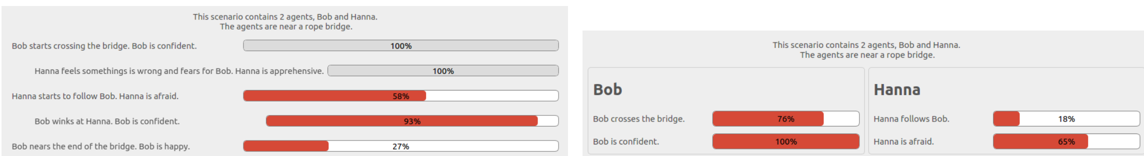


Figure 4.4: Action and response shown through text and progress bars. The indentation to the text of the response creates a visual bond to the previous message.

After having a way to express the progress of a stage of an action, we needed a way to express the progress of a scene. We created two interfaces to solve that issue, the *Log* view and the *Detail* view, in both an introduction is shown at the top, informing them of the scene and its agents.

In the *Log* view (Figure 4.5a), all new stages are placed below the previous stage, creating a log of what has happened and allowing the user to view the current state of the actions, where only the ones in red are active. In this view, responses to an action stage are placed below it with an indentation, creating a visual bond, without explicitly naming it (see Figure 4.4).

The *Detail* view (Figure 4.5b) shows all the agents in the scene and shows what they are doing and how they are feeling, this allows the user more easily see the state of every agent. In this view, the new action or emotion stages replace the previous one.



(a) *Log View:* New stages are shown below the previous.

(b) *Detail View:* New stages replace the agent's previous stage.

Figure 4.5: Ways of expressing a scene with progress bars.

After some informal evaluation, we decided that the *Log* view was best suited for expressing a scene, mainly because this view allows users to see what happened earlier and more easily recall previous events.

4.5.2 User Study

To measure the impact of the model regarding believability when compared to a more classic approach to behavior modeling, a user study was conducted in which participants were asked to fill out a questionnaire in which three videos were presented in a semi-random order[3], each containing the same scene of two agents about to cross a rope bridge using our real-time text-based representation (an illustration of the execution of this scenario can be seen in Appendix C.1). Each video shows a different algorithm for controlling the agents' behavior: *(a) Model*, the agents use the *3motion* model with action subdivision and correct timing of events, expressing coherent emotions; *(b) Classic*, the agents use the model without action subdivision and only signal the beginning and end of an action (emulating a virtual agent classic approach to behavior implementation); *(c) Misguide*, the agents use the model with action subdivision, but with incorrect timing (this is to determine if having more information shown, even if poorly timed, bias the perception of the scene as being more believable). After each video, the participants expressed their agreement with a set of statements about their perception of the agents and how the agents perceived other agents. Note that in this evaluation, the implementation did not take into account the subdivision of the follow-through, instead considering it an atomic state.

4.5.2.A Sample

Testing involved 52 participants with ages ranging from 18 to 55 ($M = 25.10$, $SD = 6.06$, 14 female). Each session had a duration of 15 to 20 minutes.

4.5.2.B Resources

A – Demographic Questionnaire: This questionnaire gathered generic demographic data, binary gender (Male and Female) and age, and specific data regarding interactions with Non-Player Characters (NPCs) (any agent or character in a virtual interactive experience) and animated environments (such as video games and animated movies). Regarding animated environments, information was gathered on how frequently one played video games and watched animated movies. Regarding NPCs, information was gathered on how preferred and important a NPC is in a video game.

B – Believability Questionnaire: First introduced by us in [146], the questionnaire we used to assess believability is structured in two sets of statements: the *participants perception of the agents* and the *participants perception of how the agents perceive other agents*. It is important to note that in the questionnaire agents are referred to as characters, as such, we can consider them synonyms in this context. The participants were asked to rate their agreement to the following statements using a 5-point Likert scale from *1 - Strongly Disagree* to *5 - Strongly Agree*.

- From your (the participant) point of view:

[3] The ordering of the videos is discussed below and was designed to avoid bias between videos while limiting the number of questionnaires needed.

– Q1 - I understood what the characters were doing.
– Q2 - I could predict the characters' actions.
– Q3 - I understood what the characters were feeling.
– Q4 - I could predict the characters' feelings.
– Q5 - I understood the characters' intentions.

• From the virtual agents' point of view:

  – Q6 - The characters were aware of each other.
  – Q7 - The characters were aware of each other's actions.
  – Q8 - The characters could predict each others' actions.
  – Q9 - The characters were aware of each other's feelings.
  – Q10 - The characters could predict each others' feelings.
  – Q11 - The characters were aware of each other's intentions.

• Q12 - The interaction between characters in this scene was believable.

The dimensions in both sections are based on the works from [28] and [51]: awareness, behavior understandability, predictability, behavior coherency, change with experience, and social metrics. A dimension measuring the ability of an agent to perceive and interact with other agents was added. While the previous works focused on the participant's point of view, our approach also explores the participant's understanding of the virtual agents' points of view, which is one contributing factor to the creation of believable interactions (linking back to topics such as Theory of Mind and Empathy, shown in Section 2.1.4). The final question (Q12) helps us collect how the participant consciously evaluates the scene's believability, and what scene is considered the most believable.

C – Videos: We developed three videos to represent the same scene depicting two agents, Bob and Hanna, crossing a bridge (see Figure 4.6). As previously stated, what differs in each video is how the actions are subdivided and the timing used.

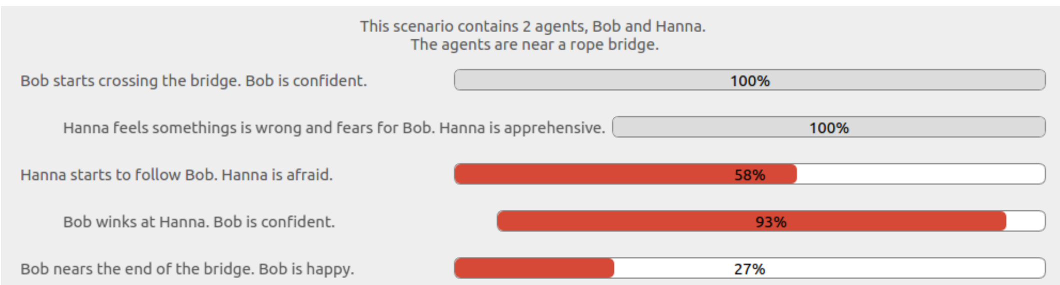


Figure 4.6: Screen capture of the *(a) Model* video depicting a moment where Hanna started following Bob while being afraid, and Bob is nearing the edge of the bridge being is happy and confident.

(a) Model video[4]: The agents use the *3motion* model with action subdivision and correct timing of events, expressing coherent emotions. This video represented the correct usage of *3motion*, thus we expected it to be the one containing more believable interactions.

(b) Classic video[5]: The agents use the model without action subdivision and only signal the beginning and end of an action, emulating a virtual agent classic approach to behavior implementation that neglects anticipation. We expected this video to represent a believable scene, but to be hindered by the lack of emotion expression present.

(c) Misguide video[6]: The agents use the model with action subdivision, but with incorrect timing. This video's purpose is to determine if having more information shown (the *(b) Classic* video contains less information than the *(a) Model* video), even if poorly timed, creates more believable environments or not. We hypothesized that poor timing breaks the suspension of disbelief and thus we expected this video to be the least believable (see Section 2.1 on Believability).

To avoid bias regarding the order of the videos, three versions of the questionnaire were made, each presenting the videos in a different sequence, also assuring they do not repeat the same position in any questionnaire. The ordering follows a Latin-square distribution[7], where each row represents a version of the questionnaire and each column represents the position of a video in the order of each version. The videos were then distributed so that each video appears exactly once in each row and exactly once in each column. In table 4.1, we can see that *(a) Model*, for example, never repeats its position, favoring non-bias results.

Table 4.1: Questionnaire ordering. Rows represent a version of the questionnaire and the columns represent the order in which the videos are presented.

	1st	**2nd**	**3rd**
Version A	a	b	c
Version B	c	a	b
Version C	b	c	a

The distribution of the questionnaires followed a linear pattern (i.e. A, B, C, A, B . . .) to have all versions have the same number of participants. Unfortunately, our implementation could not account for the case where a questionnaire was opened, but not filled, allowing inconsistencies in the number of replies in each version. Fortunately, the number of replies was almost identical in any version of the questionnaire.

4.5.2.C Procedure

A participant would be provided a link to a web page that would redirect them to one of the versions of the questionnaire. They would fill out the *Demographic Questionnaire* first and then would be shown the three videos. After each video, they would fill out a *Believability Questionnaire* in the context of

[4]To view the Model video follow this link: `https://youtu.be/_ZLP-wv2yUo`
[5]To view the Classic video follow this link: `https://youtu.be/ONBGq8cpQR0`
[6]To view the Misguide video follow this link: `https://youtu.be/_rQ-gHsRIGY`
[7]Latin-square distribution: `https://en.wikipedia.org/wiki/Latin_square`

that video. When all videos were shown and respective questionnaires were filled out, the participant would submit the answers, closing their participation.

4.5.3 Results and Discussion

Testing involved a sample of a total of 52 participants. All the tests on the participants' data are presented in greater detail in [16]. Recall the three versions of the questionnaire, A, B, and C, with the videos in different orders (see Table 4.1), we will use these version naming during this section.

Of the 52 participants that filled the questionnaire, 18 participants replied to version A, 18 participants replied to version B, and 16 replied to version C. Regarding their gender, 14 participants identified as female (38 remaining participants identified as male). As for their age, 37 participants were aged between 18 and 25 years old, 14 were aged between 26 and 40 years old, and only 1 participant was aged between 41 and 55 years old. Regarding habits on interactions with NPCs and animated environments, a low percentage of participants (13.5%) do not usually play video games. Most participants (69.2%) have no preference for watching animated movies when compared with other movies. This lack of preference also applies to their choice (53.8%) of games that show interactions with NPCs compared with other games, although many (30.8%) do prefer games showcasing such interactions over other games. Lastly, interactions with NPCs were not considered more important than other features ($55,8\%$), for some participants (5.8%) this feature has no value.

For each video, we aggregated the responses given on the Believability questionnaire, which analysis we present henceforth. As previously mentioned, each statement was presented as a Likert scale, which values ranged from *1 - Strongly Disagree* to *5 - Strongly Agree*. Using the Shapiro-Wilk test we determined that none of the *data sets*, the answers given to the same statement in each video, were normally distributed and we are therefore only able to use non-parametric tests on our data. We then applied the Friedman test to each statement data set, determining if the answers for the three videos were or were not significantly different. For fine-grain, we proceeded to use the Wilcoxon signed-rank test, which allowed us to compare how the answer to a statement in two different videos differed. An analysis of each statement is detailed in [16].

The analysis of the collected data led us to conclude that the *(a) Model* video (the video that correctly uses the model, paying attention to timing) ranked higher in *Q12* ($\chi^2(2) = 21.798$, $p = 0.000$) and in almost every statement (all except *Q4*), meaning that participants perceived this video to contain the most believable scene. The *(b) Classic* and *(c) Misguide* were generally indistinguishable from each other, yet, for those statements where differences arose (*Q1*-*Q4*), the *(b) Classic* was ranked higher.

The statements regarding the participant's perception of the agents (*Q1*-*Q5*) were where we expected to see more similarities between the *(a) Model* and the *(b) Classic* videos. The expectations were confirmed and statements, such as "Q2 - I could predict the characters' actions", were similar in value between these two videos ($Z = -0.469$, $p = 0.639$). A broken expectation was that of the statement "Q4 - I could predict the characters' feelings", where we hoped the new information given by the action subdivision would allow participants to more easily predict the agent's emotions ($Z = -1.009$,

$p = 0.313$).

A weird phenomenon happen in the statement “Q5 - I understood the characters’ intentions”, where the three videos had no statically significant difference ($\chi^2(2) = 3.823$, $p = 0.148$). One can suppose that the intentions of the agents are easy to perceive in any of the videos or even that after watching the video the intentions were made clear. Another possible supposition is that the participants meant that they were capable of perceiving that the intentions had not changed.

The statements regarding the agent’s perception of other agents and their actions, feelings, and intentions gave results that always favored the *(a) Model* video, indicating that the correct usage of the model improves believability.

4.6 Summary

In this chapter, we presented our solution to enable believable multi-party real-time interactions through the sub-division of actions and usage of non-verbal emotional communication. We first introduced the concept of action subdivision into three stages: anticipation stage, action stage, and follow-through stage. The anticipation stage can be further subdivided into interruptible and uninterruptible sub-stages, and the follow-through stage can be further subdivided into uninterruptible and interruptible sub-stages. This solution allows for a range of emotions to be communicated in the duration of a single action, which deepens the possible interactions and allows for a bigger emotional involvement of the agents. After that, we discussed how the action subdivision integrates with the agents’ internal behavior loop, which is split into four steps: Perceive, React, Decide, and Perform. Delving into each step, we discussed the perception of the environment, the emotional responses, the decision-making processes, and finally the performance of actions and emotions using the action subdivision.

Following, an illustrative scenario is shown depicting how the model works and some of its novelties compared with models that do not consider action subdivision.

A text-based application was then detailed as one of the use cases of *3motion*. Text and progress bar elements were used to display unfolding scenes with agents interacting. A user study was performed where 52 participants were asked to evaluate the believability of three videos showcasing a scene of application, differentiated by their use of action subdivision and of different timings. Results suggest the video with action subdivision and correct timing, essentially depicting the correct usage of *3motion*, was rated as more believable in almost every statement compared to the other two videos.

Having discussed the model we now understand how it works and where we can apply it. The next chapter will delve into a use case of this model and the short-term future work planned for it.

5

Designing Believable Interactions in *Adfectus*

Contents

In this chapter, we will discuss how we designed a use case for 3motion *that mimics common game development settings and assets. The use case is a video game named* Adfectus *developed using modern technologies and set in a 3D environment the player can interact with. First, we detail the game and its components, focusing on the different characters' emotional expression capabilities, followed by a description of the planned evaluations to determine if more believable interactions and experiences were achieved, finishing with the presentation of a one-year plan on when to implement and evaluate the work along with possible publication avenues.*

5.1 Adfectus

Adfectus is an arena game where two characters battle one another (see Figure 5.1). Each candidate has a claque that supports them and helps them throughout the game. To win the game, a character has to better his opponents, but also please its crowd that will only be satisfied with a good show (e.g. if there is a big gap between player skills and the player with better skills is unforgiving and greedy, not allowing the other player to hit back or improve, it will create a bad show). This game is being developed to be a testbed for *3motion*, showcasing mechanics that are often used in action games in a 3D environment and, at the same time, offering new mechanics that can be improved further through the use of *3motion*, such as interruptible and uninterruptible stages of attacks, and interactions with the claques [1].

Figure 5.1: Showcase of Adfectus' characters and arena. Two playable characters can be seen in the foreground holding hammers and ready to face each other on the ground floor of the arena. On the balcony, two claques are seen supporting their preferred duelist.

Visually the game uses a low-poly aesthetic[2], this allows us to make use of exaggerated expres-

[1] A gameplay video of an early version of Adfectus can be seen at `https://youtu.be/sjsV4uG3Eag`.
[2] `https://en.wikipedia.org/wiki/Low_poly`

sions, such as scaling parts of the body, without compromising the believability of the characters. The setting is reminiscing of medieval fantasy genre[3], a genre of fiction set in a fictional universe that specifically pertains to or is influenced by the Medieval period of European history. In it, an arena will be the main place portrayed, it will be composed of a ground floor and an upper round balcony that surrounds the floor. The balcony will be populated by two *claques*, each will support one of the two characters on the floor below, these are the *duelists* that will face each other to win the battle and are the ones controlled by players.

5.1.1 The Duelist - Player Character

The duelist is the main character of Adfectus and the main focus of our work. It is the character that battles another duelist and determines the winner or loser of the battle. To battle, the duelist will make use of melee-type weapons, such as swords or hammers, along with armor and a shield (see Figure 5.2). During the battle, the duelist can perform attacks with their weapon to damage the opponent[4], block incoming attacks with its weapon and shield, and move or run around the floor. Additionally, the duelist can also interact with certain objects (e.g. picking up a new weapon from the ground) that might be scattered around the floor. To clearly distinguish the two duelists, different armor is worn by each one[5]. The set is chosen mainly due to a big color hue difference or a unique piece that differentiates itself (see Figure 5.1). The same can be applied to the weapons.

Figure 5.2: Showcase of a character of Adfectus holding a sword while wearing metallic armor. Diverse items can also be seen, such as a hammer, a dagger, and a shield. The setting in the background exemplifies the expressive potential of the environment.

Some actions, such as attacking or blocking, will be able to be interrupted. This mimics the anticipation stage of *3motion* where an action is interruptible until a certain threshold is reached[6] and afterward the action is followed until its end or an external factor interrupts it (e.g. a duelist is performing an attack but misses and they are struck by a counterattack). This introduces the ability to

[3] `https://aesthetics.fandom.com/wiki/Medieval_Fantasy`

[4] The current implementation provides a light attack, that is quicker but does less damage, and a heavy attack, that is slower but does more damage on hit.

[5] In the current implementation, the armor does not provide any mechanical feature (i.e. a set of armor does not provide any additional protection against attacks).

[6] In the current implementation, the interruptibility threshold is defined through animation events that determine which state an action is in.

perform feints, whereby canceling an attack (or block) before the anticipation becomes uninterruptible the duelist can start performing a new action sooner than expected.

The duelist's implementation will make use of *3motion*, thus enabling interactions with other characters through the use of emotional non-verbal communication. The emotional processing and expression will be handled through *3motion*, while the decision-making will be done by a player (see Chapter 4). The character is controlled by a player from a third-person perspective making use of a controller. It is important to note that the execution of an action (decided by the player) can be modified by *3motion* to reflect an emotional response (e.g. walking slowly because one is scared).

5.1.1.A Emotion Expression Capabilities

As discussed in Section 2.4.3, there are multiple channels to communicate emotions. We plan to explore the expression of emotion through the following channels: *Facial Expressions*, more specifically iconic representations (such as emojis, see Figures 5.3 and 2.3), which are simplified but recognizable faces, easing development and increasing recognizability; *Emotional Body Language*, through the use of prototypical posture to display the emotion felt (see Table 2.2 and Figure 5.4); and, *color and backgrounds*, a secondary channel reinforcing the emotions felt, such as making use of the halos and colors seen in the game Darkest Dungeon (see Figure 2.4).

Figure 5.3: Example of different expression of emotions through emojis[7], as a representation of iconic imagery.

Figure 5.4: Example of animation adapting to the state of the character in *Tomb Raider*[157]. It shows the character walking normally (left) and walking while it is hurt (right), note the arm holding the area where the character is hurt. Lighting is also used to highlight the gravity of the situation, using darker tones to portray a bad feeling.

Due to the third-person perspective used to control the duelists, the player's character's facial expressions may be obstructed, therefore an exploration on the use of secondary channels or of

[7] Faces by Devendra Karkar from NounProject.com.

Heads-up Display (HUD)[8] elements might be needed to understand the best way to showcase the character's expressions through different channels. Moreover, the nature of the game also has to be considered. Adfectus is a fighting game, hence players will very likely be immersed in combat, focused on defeating the opponent and so any subtle expression of emotion, will probably be missed. For this reason, we will focus more on body posture, backgrounds, and HUD elements (e.g. making use of emojis in such elements) which are effective ways of communicating emotion during a battle.

5.1.2 Claque

The claques are the entities that will determine who is the winner of a battle. Each claque will be represented by a crowd of characters with similar looks to that of the duelist they are supporting and will be present in the higher balcony of the arena. Claque's behavior is implemented using *3motion* and its addition to the work helps showcase the expressive potential of the model by presenting different opposing and collaborative entities (the duelist has an opponent in the other duelist and supporting characters in the claque). Its goal is to enjoy the battle. The enjoyment is calculated by measuring (a) Duelists' performance and (b) attendance to "personal" preference (e.g. they wish to see a specific weapon being used). Each member of the claque will have different preferences but will share some of them with the remaining claque. They will react emphatically to their duelist and antagonistically towards the opponent duelist.

The decision-making is different from the duelists (that are controlled by the players) and will have more intelligent behavior. The claques will analyze the player performance and thus favor a winner (information that will be made available to the players) and also will have personal preferences that they will provide to their favored duelist (e.g. use an axe; do a heavy attack; block an attack), if the player decides to please its crowd, he will be rewarded with a greater probability of winning (the probabilities algorithm is still to be defined).

Regarding player performance, one can make use of Emotivectors from the emotional model to provide expectations of the player, they can be interpreted in different ways. The claque may want to have a metric follow the expected values (e.g. they might want the player to avoid being hit by an enemy attack) or differ from the expected (e.g. always missing an attack will probably be seen as bad). Another source of information may be recorded through the implementation of N-Grams for Player Prediction[147], which "can predict the player's next move with surprising accuracy" thus offering the claques a chance to judge how constant the player's behavior is and act accordingly (e.g. getting bored if the player keeps repeating the same moves).

In the current stage of implementation, no decision has been made towards which features to implement (some, all, or another approach), but this discussion helps inform how a claque can be designed to support a player while having their own goals.

[8] In video-games, the heads-up display is the method by which information is visually relayed to the player as part of a game's user interface.

5.1.2.A Emotion Expression Capabilities

Similar to the duelist, the individual members of the crowd can make use of multiple channels of communication (moreover they will share the same features, such as model, skeleton, and animations), yet the distance to the camera and the focus being on the duelists the expression of emotions needs to be adjusted. The use of emotional body language needs to be exaggerated and almost iconographic (in the sense that the expression can be viewed easily and easily associated with the respective emotion). The same applies to the facial expressions used. Video games make use of Emotes that portray the characters in iconic and identifiable poses to express emotions or moods (see Figure 5.5). The selection of some emotes to reflect emotional states might improve the readability of the claque by the player.

Figure 5.5: Emote of a character in *Overwatch*[155] expressing a burst of happiness. The exaggerated display makes it recognizable even if far from the character.

Figure 5.6: Screenshot from *Assassin's Creed Odyssey*[164] highlighting a radial element of the HUD that indicates an incoming ranged attack. This element provides a visual cue to otherwise hidden interactions.

Furthermore, the use of secondary channels needs to be adapted to consider a crowd and not a single element. Background imaging is either removed, as they might not be seen, or adjusted to not overwhelm other expressions of emotions by the duelists. The use of HUD elements to facilitate the display of claque information provides a good alternative to communicate their emotions (see Figure 5.6).

5.1.3 Winning the Duel

To win the duel, duelists will need to please their claque. As previously described, the claque will provide active cues for the duelist to satisfy them, that is the preferences of the crowd (see Section 5.1.2), but the player will also need to be dynamic in that repetitive behavior will lead to boredom in the claque. We also plan to have the duelist be able to present emotes (e.g. raising his arms in confidence) to boost the claque's satisfaction, thus presenting another active behavior for the player to interact with the audience.

Mechanically, we will have a satisfaction meter, representing the number of pleased claque members, that when high enough (using a predefined threshold) will make the claque very excited during a predefined time. During this time, if the player successfully strikes their opponent they win the match. This mechanic offers the attacking player the opportunity to deliver the final blow in a cinematographic way while allowing the defending player the opportunity to surprise everyone with a well-timed defense and/or counter-attack.

Currently, no described behavior or winning mechanic is implemented.

5.1.4 Illustrative Scenario with Claques

We have already portrayed the use of *3motion* in a setting with two duelists, which helps showcase the interactions possible without the presence of a claque (see Section 4.3). Now, to better showcase a desired interaction between duelists and claques, we present an illustrative scenario set in the world of *Adfectus*. We start by presenting the setting and characters, followed immediately by a combat between the previously presented characters *Rua* and *Grom*.

> The summer air brings no wind, yet the sand and stones of the arena seem to bring dust to life. The two tribes, the Rubrims and the Livens, join to watch their heroes battle. They are strange folk for they need not fight. They live in peace, yet find this show of strength not only entertaining but also enriching to their social lives. The heroes will not battle to the death but will fight until one falls. They are all there for a good show.
>
> The two tribes split the arena in half and await the entrance of their heroes. Rubrims, always showing their crimson colors, offer Rua, a young and vigorous woman. She brings her pink hair in a ponytail and on her head rests a crown adorned with feathers. Her body is covered in a bronze armor with reddish tones. The armor is light and her reflexes are quick. She is ready! Livens, with their blue colors, bring Gorm, an older and experienced man. Although his head is covered with a silly blue and white top hat, his body is covered in beautiful silver-blue plate armor. He is strong and wise. He is ready!
>
> Silence falls in the arena as the heroes walk in and face each other. . . The bell rings! The fight is underway.
>
> Although the bell rang, the duelists hold their ground, studying each other and planning their moves. Rua decides to take advantage of her quick reflexes and rushed to pick up a spear from the walls of the arena. Gorm rushes for a big two-handed axe and grabs it just in time to see Rua's spear fall on him. The weapons clash and the crowd roars. The Rubrims applaud Rua for her strike, while the Livens, although surprised by the strike, applaud Gorm's block.
>
> The duelists take a step back staring at each other with confidence in their eyes. It is time for Gorm to attack. He takes a step forward and starts swinging his axe with force. Although heavy the axe swings fast, forcing the surprised Rua to dodge backward, barely

escaping the blade. Taking advantage of the dodge, Rua prepares to strike once again, bringing her arm back to swing at the opening provided by her opponent. She opens a smile that soon fades as she notices her opening closing and a downward strike falling on her.

The time seems to slow down... Gorm is confident the attack will finish the battle, Livens cheer the hero onward. Rubrims gaps in fear and disbelief, yet Rua's confidence seems unshaken. Gorm can no longer stop the fall of his axe, yet Rua's confidence is unnerving. He seems less sure of his attack, but can no longer stop it. In the blink of an eye, Rua cancels her attack and dodges the strike. The axe collides with the floor, giving her time to swing her spear. The crowd goes silent in surprise. Gorm is unable to parry or dodge her attack, and Rua knows it, happy and confident, striking him in the chest.

Gorm falls... Livens remain perplexed, their hero has fallen. The Rubrims cheer louder than ever! Rua has won!

This scenario justifies the combat in-world as a show between two neighboring tribes and a tense and emotional duel where each duelist is representing their tribe. While it does not showcase an active role of the claque, it does showcase the expressive potential of interactions with their presence.

5.2 Planed Evaluations

Multiple evaluations should be considered to correctly test this use case. The most important will be a study on the impact on believability by the application of the model, yet we must consider other aspects first. Two considerations to take into account are that of the *emotion expression capabilities* and how they are perceived, and that of the *influence of the crowd* or lack thereof.

Emotion Expression Validation – To validate our emotion expression modalities, we plan to show participants videos of characters expressing different emotions and then asked them to describe which emotion they think was expressed in the video. Then we check whether our characters can correctly convey the emotions we desire. This approach is similar to the one described in Section 3.5.2. Given the results of these tests, if need be, we adjust our animations to better express the desired emotion. Multiple iterations of this evaluation might be necessary so that the animations reach a satisfactory level of emotion recognition.

Speak of PANAS questionnaire (10 items).

https://psycnet.apa.org/fulltext/1988-31508-001.pdf?auth_token=01c87b20f6078c53956309a36a9ee8f7f07
returnUrl=https%3A%2F%2Fdoi.apa.org%2FdoiLanding%3Fdoi%3D10.1037%252F0022-3514.54.6.1063

[148]

Evaluation on Believability – Similar to the previous use case using a Text-based Application (see Section 4.5), we want to measure the impact on believability of introducing *3motion* in *Adfectus*, which is the focus of our work. To do so we will have two versions of the game, one with atomic actions (the *classic* version) and one with *3motion*'s action subdivision (the *model* version)[9]. The use of a similar evaluation approach seems feasible and allows for comparison of results, although we intend to have a more interactive approach by allowing the participants to play the game (instead of watching a video of gameplay). The evaluation will take into account **(a)** the perspective of two players playing against each other, and **(b)** the perspective of a bystander watching situation (a) unfold. These conditions will allow us to have a better understanding of how believability can differ from an active perspective (that of the player) and a passive perspective (that of the bystander). We will make use of the Believability Questionnaire used previously to help us measure believability (see Section 4.5.2.B).

To focus on the interactions between only the characters battling, we propose a scenario with the environment stripped of content and without the presence of the claque (see Figure 5.7). To allow for depth perception and clear positioning in the environment, the floors and walls are shown with a grid-like pattern and a light will cast shadows from the characters onto the environment. By placing the characters in such environments, we hope that players and spectators avoid connecting the believability of the interactions with the environment and focus on the interactions between characters.

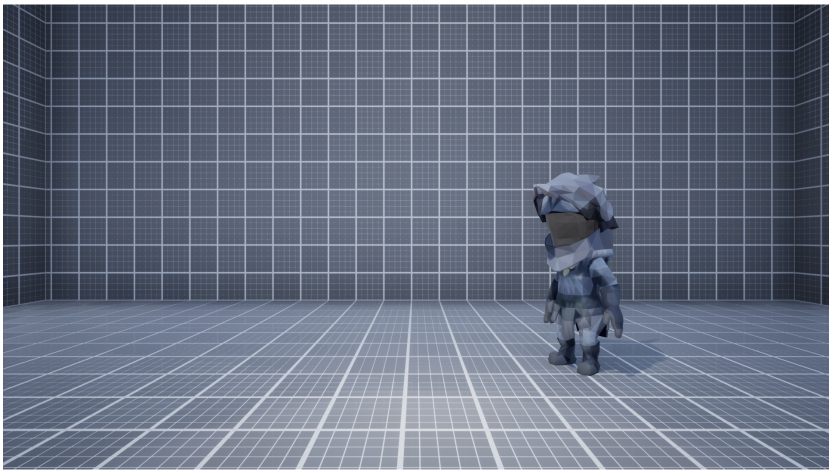


Figure 5.7: Presentation snapshot of Adfectus with the environment stripped of content, only presenting a grid-like pattern and a character.

Evaluation on Believability with Claques – Having evaluated the duelists and determining a baseline of believability, we can now add the environment and the claque, and perform a new evaluation. Using data collected from the first evaluation (e.g. videos of gameplay, and the values of the emotivectors during the interactions) and asking participants to determine how a character would feel (using for example the basic families of emotions), we can use this data-set to inform the emotion system on the best emotions to generate for a given context. We will use a revised emotion system on the duelists and the claque, improving believability. A new evaluation, similar to the previous one, will take into account **(c)** the perspective of two players playing against each other *supported by two claques*, and the previously described **(b)** but now watching situation (c) unfold.

[9]Unlike the previous evaluation, we will not include a *Misguide* version, as earlier results point to low believability in that version, showing that it is not better than the *classic* version, and another version of *Adfectus* would impact the already limited resources and time.

This setting showcases multi-party real-time interactions where different characters interact differently with each other (having opponents and allies), thus exploring more non-verbal communication opportunities offered by *3motion*. The evaluation, therefore, allows us to understand the impact of a complete system using *3motion* on the believability of *Adfectus*.

5.3 Implementation Planning

To achieve the completion of the thesis, we will focus on developing *Adfectus* (see Table 5.1). The current implementation offers a base game where to implement the model. We will start by focusing on the development of the Perform step and the subdivision of actions, giving special attention to their perception by the user. This requires a first implementation of the agent's behavior loop, Perceive, React, and Decide. The Decide step is especially important for the Claque behavior (the duelists are controlled by the player). This first implementation will result in a Workshop paper describing the architecture and early findings (Task 5, 3rd Month). It is fit for publication in game-related conferences with artificial intelligence tracks, such as, IEEE Conference on Games (CoG).

Table 5.1: Timeline of the planned activities for the closure of the thesis. Note that the publication row is detailed in the text.

	Months											
Task Description	1	2	3	4	5	6	7	8	9	10	11	12
1. First Agent's Implementation												
2. Preliminary Evaluation and Review												
3. Implementation' Iteration and Evaluation												
4. Dissertation Writing												
5. Publications												

At this point the Emotion Expression Validation will be performed, followed by the Believability Evaluation of the Duelists. After this point, the work developed will result in a full paper describing the architecture, the user studies, and its results (Task 5, 6th Month). It is fit for publication in artificial intelligence conferences, such as International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), Conference on Affective Computing and Intelligent Interaction (ACII), and International Conference on Intelligent Virtual Agents (IVA).

The evaluations and reviewers' comments will help inform improvements on the first implementation that will be further developed for a second iteration of the system which will include the claque. In this iteration, more focus will go into the emotion model and which values it should measure (react step), and the decision-making processes to allow for improved claque behavior (decide step). After this iteration, we will evaluate the Believability of the system including the claques, thus evaluating the complete system. A full paper will encompass this user study and its results (Task 5, 10th Month). It is for publication in both artificial intelligence and game-related conferences, such as CoG, AAMAS, ACII, and IVA.

Lastly, a paper encapsulating all of the work is fit for publication in artificial intelligence journals, such as Transaction of Affective Computing and Transaction on Games (Task 5, 12th Month).

6

Conclusion

In this document, we started by presenting the problem of how can one enable believable multi-party real-time interactions between synthetic characters. We hypothesize that by subdividing actions according to traditional principles of animation[9] and allowing for emotion reaction and behavior interruption during their expression, the overall real-time interaction will be perceived as more believable.

We then present some related work on Believability and its definition, Emotions and how it is used in synthetic characters, and the Principles of Traditional Animation, among other subjects. These topics help create the foundation for the work and are interlinked topics that give the context in which we can evaluate our work.

Before showcasing our proposed solution, we discuss our work on the Virtual Coach and its usage of emotional-filled real-time dialogue-based interactions. A particularly important detail is how the system makes use of Emotivector in the emotional system and how it explores the evaluations of emotion expression in synthetic characters and possible shortcomings of certain evaluations.

Our proposed solution is then presented in *3motion*, a model for virtual agents that allows for the subdivision of actions and non-verbal communication in each stage. We first introduced the concept of action subdivision into three stages: anticipation stage (further subdivided into interruptible and uninterruptible sub-stages), action stage, and follow-through stage (further subdivided into uninterruptible and interruptible sub-stages). Then we described the internal behavior loop of the agents of *3motion*: Perceive, React, Decide, and Perform. The last step is where the action is performed and the subdivision occurs.

A first use case of the model is described then, showcasing a text-based application where text and progress bar elements were used to display unfolding scenes with agents interacting. A user study was performed where 52 participants were asked to evaluate the believability of three videos showcasing a scene of application, differentiated by their use of action subdivision and of different timings. Results suggest the video with action subdivision and correct timing, essentially depicting the correct usage of *3motion*, was rated as more believable in almost every statement compared to the other two videos.

Lastly, *Adfectus* was introduced as a use case for an interactive experience that introduces game mechanics and a 3D environment, in which *3motion* will be implemented. We described the duelists, characters controlled by the players, and the claques, controlled by AI, that compose the game. Special attention was given to these characters' emotion expression capabilities and it was also described how the gameplay flows, where the duelists need to satisfy their respective claque in order to win. Several evaluations are then proposed to determine the believability of *Adfectus* and of its characters that use *3motion*. Furthermore, a proposal for its development and execution of the evaluations was discussed, along with possible papers that can originate from these developments.

Bibliography

[1] E. Brown and P. Cairns, “A grounded investigation of game immersion,” in *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*. New York, New York, USA: ACM Press, 2004, p. 1297. [Online]. Available: http://portal.acm.org/citation.cfm?doid=985921.986048

[2] S. Böcking, “Suspension of Disbelief,” in *The International Encyclopedia of Communication*. Chichester, UK: John Wiley & Sons, Ltd, 6 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/9781405186407.wbiecs121https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405186407.wbiecs121https://onlinelibrary.wiley.com/doi/10.1002/9781405186407.wbiecs121

[3] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, “Survey on Emotional Body Gesture Recognition,” *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 505–523, 4 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8493586/

[4] R. M. Maatman, J. Gratch, and S. Marsella, “Natural behavior of a listening agent,” in *International Workshop on Intelligent Virtual Agents*, Springer. Kos Greece: Springer, 2005, pp. 25–36.

[5] E. Bevacqua, M. Mancini, and C. Pelachaud, “A listening agent exhibiting variable behaviour,” in *Intelligent Virtual Agents*, Springer. Tokyo Japan: Springer, 2008, pp. 262–269.

[6] J. Togelius, G. N. Yannakakis, S. Karakovskiy, and N. Shaker, “Assessing believability,” in *Believable Bots: Can Computers Play Like People?*, P. Hingston, Ed. Springer Berlin Heidelberg, 2012, vol. 9783642323, pp. 215–230.

[7] C. Martinho and A. Paiva, “Using Anticipation to Create Believable Behaviour,” *Proceedings of the national conference on Artificial Intelligence*, vol. 1, pp. 175–180, 2006. [Online]. Available: http://www.aaai.org/Papers/AAAI/2006/AAAI06-028.pdf

[8] V. Nayak and M. Turk, “Emotional Expression in Virtual Agents Through Body Language,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, 2005, vol. 3804 LNCS, pp. 313–320. [Online]. Available: https://link.springer.com/chapter/10.1007/11595755_38http://link.springer.com/10.1007/11595755_38

[9] F. Thomas, O. Johnston, and W. Rawls, *Disney animation: The illusion of life*. Burbank, California, USA: Disney Editions; Rev Sub edition, 1981, vol. 4.

[10] J. Mooney and J. M. Allbeck, “Rethinking NPC intelligence - A new reputation system,” *Proceedings - Motion in Games 2014, MIG 2014*, pp. 55–60, 2014.

[11] J. Bates, “The role of emotion in believable agents,” *Communications of the ACM*, vol. 37, no. 7, pp. 122–125, 1994.

[12] J. Orkin, “Symbolic representation of game world state: Toward real-time planning in games,” in *AAAI Workshop - Technical Report*, vol. WS-04-04, 2004.

[13] M. Ochs, N. Sabouret, and V. Corruble, “Simulation of the dynamics of nonplayer characters’ emotions and social relations in games,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 4, pp. 281–297, 12 2009.

[14] T. Bosse and F. P. de Lange, “On virtual agents that regulate each other’s emotions,” *Web Intelligence and Agent Systems: An International Journal*, vol. 9, no. 1, pp. 57–67, 2011. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/WIA-2011-0205

[15] K. Ijaz, A. Bogdanovych, and S. Simoff, “Enhancing the Believability of Embodied Conversational Agents through Environment-, Self- and Interaction-Awareness,” in *Proceedings of the Thirty-Fourth Australasian Computer Science Conference - Volume 113*, ser. ACSC ’11. AUS: Australian Computer Society, Inc., 2011, p. 107–116.

[16] R. Rodrigues and C. Martinho, “Towards Believable Interactions Between Synthetic Characters,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, J. Beskow, C. Peters, G. Castellano, C. O’Sullivan, I. Leite, and S. Kopp, Eds. Cham: Springer International Publishing, 2017, vol. 10498 LNAI, pp. 385–388. [Online]. Available: http://link.springer.com/10.1007/978-3-319-67401-8_48

[17] C. Pacheco, “PAGAN for Character Believability Assessment,” 2020. [Online]. Available: https://davidmelhart.com/projects/mazing/index.html

[18] M. Gawron and U. B. C. G. Innovations, “Implementation of the hair lighting model in Unreal Engine 4 based on the state-of-the-art,” *Computer Game Innovations*, p. 29, 2018. [Online]. Available: https://www.researchgate.net/profile/Adam-Wojciechowski-3/publication/335524620_Computer_Game_Innovations_2018/links/5d6a9bbb92851c85388334b8/Computer-Game-Innovations-2018.pdf#page=29

[19] T. Alldieck, G. Pons-Moll, C. Theobalt, and M. Magnor, “Tex2Shape: Detailed Full Human Body Geometry from a Single Image,” in *{IEEE} International Conference on Computer Vision ({ICCV})*. IEEE, 2019.

[20] K. Zhou, B. L. Bhatnagar, and G. Pons-Moll, “Unsupervised Shape and Pose Disentanglement for 3D Meshes,” in *European Conference on Computer Vision (ECCV)*, 2020.

[21] A. Ranjan, T. Bolkart, S. Sanyal, and M. J. Black, “Generating 3D Faces using Convolutional Mesh Autoencoders,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[22] Q. Reynaud, J.-Y. Donnart, and V. Corruble, “Evaluating the Impact of Anticipation on the Efficiency and Believability of Virtual Agents,” in *Intelligent Virtual Agents*, T. Bickmore, S. Marsella, and C. Sidner, Eds. Cham: Springer International Publishing, 2014, pp. 360–363. [Online]. Available: http://link.springer.com/10.1007/978-3-319-09767-1_47

[23] H. Prendinger and M. Ishizuka, *Life-Like Characters: Tools, Affective Functions, and Applications*, ser. Cognitive Technologies, and and H. P. Mitsuru Ishizuka, Eds. Springer, 2004.

[24] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley,” *IEEE Robotics and Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.

[25] C. Martinho, “Emotivector - Mecanismo Afectivo e Anticipatório para Personagens Sintéticas,” Ph.D. dissertation, Instituto Superior Técnico, 2007.

[26] J. Bates, “The nature of characters in interactive worlds and the Oz project,” Carnegie Mellon University, Tech. Rep. October 1992, 1993. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.6500

[27] A. Ortony, “On making believable emotional agents believable,” in *Emotions in Humans and Artifacts*. Cambridge, Massachusetts, USA: MIT, 2003, pp. 189–211. [Online]. Available: https://books.google.es/books?hl=es&lr=&id=jTgMIhy6YZMC&oi=fnd&pg=PA189&dq=On+making+believable+emotional+agents+believable&ots=l9fIhgy2AP&sig=Xd4a0oPqk679Kjk0s1gyhCeldfs

[28] P. Gomes, A. Paiva, C. Martinho, and A. Jhala, “Metrics for Character Believability in Interactive Narrative BT - Interactive Storytelling,” in *Interactive Storytelling*. Springer International Publishing, 2013, vol. 8230, pp. 223–228. [Online]. Available: http://link.springer.com/10.1007/978-3-319-02756-2_27%0Apapers3://publication/doi/10.1007/978-3-319-02756-2_27

[29] A. Bogdanovych, T. Trescak, and S. Simoff, “What makes virtual agents believable?” *Connection Science*, vol. 28, no. 1, pp. 83–108, 2016. [Online]. Available: https://doi.org/10.1080/09540091.2015.1130021

[30] S. Chandrasekharan and B. Esfandiari, “Software Agents and Situatedness: Being Where,” in *Proceedings of the Eleventh Mid-west Conference on Artificial Intelligence and Cognitive Science, 2000, AAAI Press, Menlo Park, CA*. Press, 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.6488

[31] J. Lindblom and T. Ziemke, “Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond,” *Adaptive Behavior*, vol. 11, no. 2, pp. 79–96, 2003. [Online]. Available: https://doi.org/10.1177/10597123030112002

[32] K. J. Rohlfing, “Situatedness: The interplay between context (s) and situation,” *Journal of Cognition and Culture*, vol. 2, no. 2, pp. 132–157, 2003. [Online]. Available: http://www.ingentaconnect.com/content/brill/jocc/2003/00000003/00000002/art00002

[33] M. Costello, “Situatedness,” in *Encyclopedia of Critical Psychology*, T. Teo, Ed. Springer New York, 2014, pp. 1757–1762.

[34] M. A. Hogg and R. S. Tindale, *Blackwell Handbook of Social Psychology: Group Processes*, M. A. Hogg and R. S. Tindale, Eds. Oxford, UK: Blackwell Publishers Ltd, 1 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470998458http://doi.wiley.com/10.1002/9780470998458

[35] G. Pereira, J. Dimas, R. Prada, P. A. Santos, and A. Paiva, “A Generic Emotional Contagion Computational Model,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6974 LNCS, no. PART 1, pp. 256–266. [Online]. Available: http://link.springer.com/10.1007/978-3-642-24600-5_29

[36] K. Niven, P. Totterdell, and D. Holman, “A Classification of Controlled Interpersonal Affect Regulation Strategies,” *Emotion*, vol. 9, no. 4, pp. 498–509, 2009.

[37] P. Gaussier, S. Moga, M. Quoy, and J. P. Banquet, “From perception-action loops to imitation processes: a bottom-up approach of learning by imitation,” *Applied Artificial Intelligence*, vol. 12, no. 7-8, pp. 701–727, 10 1998. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/088395198117596

[38] A. I. Goldman, *Theory of Mind*, E. Margolis, R. Samuels, and S. P. Stich, Eds. Oxford University Press, 5 2012, vol. 1. [Online]. Available: http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780195309799.001.0001/oxfordhb-9780195309799-e-17

[39] D. Premack and G. Woodruff, “Does the chimpanzee have a theory of mind?” *The Behabioral and Brain Sciences*, 1978.

[40] M. Si, S. C. Marsella, and D. V. Pynadath, “Modeling Appraisal in Theory of Mind Reasoning,” in *Intelligent Virtual Agents*, H. Prendinger, J. Lester, and M. Ishizuka, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 334–347.

[41] M. L. Hoffman and Z. Raboteg-Saric, “Empathy and Moral Development: Implications for Caring and Justice,” *Contemporary Sociology*, vol. 30, no. 5, p. 487, 2001. [Online]. Available: http://www.jstor.org/stable/3089337?origin=crossref

[42] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll, "Learning by feeling: Evoking empathy with synthetic characters," *Applied Artificial Intelligence*, vol. 19, no. 3-4, pp. 235–266, 2005.

[43] S. Hamdy and D. King, "Affect and believability in game characters - A review of the use of affective computing in games," in *18th International Conference on Intelligent Games and Simulation, GAME-ON 2017*. EUROSIS, 2017, pp. 90–97.

[44] H. Warpefelt, "Evaluating the believability of agents in virtual worlds," in *Poster Session of ACAI 2013 and EASSS 2013*. ACAI, 2013.

[45] G. N. Yannakakis, J. Hallam, and H. H. Lund, "Entertainment capture through heart rate activity in physical interactive playgrounds," *User Modeling and User-Adapted Interaction 2007 18:1*, vol. 18, no. 1, pp. 207–243, 9 2007. [Online]. Available: https://link.springer.com/article/10.1007/s11257-007-9036-7

[46] P. Rodrigues, A. Paiva, and J. Dias, "An Affect-Aware Intelligent Tutoring System for EmoRegulators A Restricted-Perception Wizard-of-Oz Approach," p. 102, 2018.

[47] C. Pedersen, J. Togelius, and G. N. Yannakakis, "Modeling player experience for content creation," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 2, no. 1, pp. 54–67, 3 2010.

[48] J. Pardal and C. Martinho, "Holiday Knight: a Videogame with Skill-based Challenge Generation," p. 92, 2019. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/283828618790385

[49] M. Marques and C. Martinho, "Profiling Players Through In-Game Animations," p. 80, 2019.

[50] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge, UK: Cambridge University Press, 1990.

[51] N. Costa, "Believable Interactions Between Synthetic Characters," 2015.

[52] A. Beck, B. Stevens, K. A. Bard, and L. Cañamero, "Emotional body language displayed by artificial agents," *ACM Transactions on Interactive Intelligent Systems*, vol. 2, no. 1, pp. 1–29, 3 2012. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2133366.2133368https://dl.acm.org/doi/10.1145/2133366.2133368

[53] P. R. Kleinginna and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and Emotion*, vol. 5, no. 4, pp. 345–379, 12 1981.

[54] K. R. Scherer, T. Banziger, and E. B. Roesch, *Blueprint for affective computing : a sourcebook*, 1st ed. New York, NY, USA: Oxford University Press, Inc., 2010.

[55] P. Ekman, *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, New York, USA: Macmillan, 2007.

[56] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. SAN JOSE CA USA: Ishk, 2003.

[57] L. F. Barrett, *How emotions are made: The secret life of the brain.* Boston, USA: Pan MacMillan, 2017.

[58] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[59] A. Arya, S. DiPaola, and A. Parush, "Perceptually Valid Facial Expressions for Character-Based Applications," *International Journal of Computer Games Technology*, vol. 2009, pp. 1–13, 2009. [Online]. Available: http://www.hindawi.com/journals/ijcgt/2009/462315/

[60] M. Mäkäräinen and T. Takala, "An Approach for Creating and Blending Synthetic Facial Expressions of Emotion." in *Intelligent Virtual Agents*, Springer. Netherlands: Springer, 2009, pp. 243–249.

[61] R. Rodrigues, P. Ferreira, R. Prada, P. Paulino, and A. M. V. Simao, "Festarola: A game for improving problem solving strategies," in *2019 11th International Conference on Virtual Worlds and Games for Serious Applications, VS-Games 2019 - Proceedings*, 2019, pp. 1–8.

[62] C. Crivelli and A. J. Fridlund, "Inside-Out: From Basic Emotions Theory to the Behavioral Ecology View," *Journal of Nonverbal Behavior*, vol. 43, no. 2, p. 161–194, 2019.

[63] P. Pimentel, "Gameplay Aware Expression for Virtual Characters in Video Games," 2015.

[64] J. Dias and A. Paiva, "Feeling and Reasoning:A Computational Model for Emotional Characters," in *Progress in Artificial Intelligence*, C. Bento, A. Cardoso, and G. Dias, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 127–140.

[65] S. He, Z. Liu, and W. Xiong, "An emotion model for virtual agent," in *Proceedings of the 2008 International Symposium on Computational Intelligence and Design, ISCID 2008*, vol. 1, 10 2008, pp. 511–514.

[66] E. Jacobs, J. Broekens, and C. Jonker, "Joy, distress, hope, and fear in reinforcement learning," in *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, ser. AAMAS '14, vol. 2. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1615–1616. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84911369211&partnerID=40&md5=cf583e64ab4a206b7f03d9ef9b80df11

[67] R. Rodrigues, R. Silva, R. Pereira, and C. Martinho, "Interactive Empathic Virtual Coaches Based on the Social Regulatory Cycle," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. Cambridge, United Kingdom: IEEE, 9 2019, pp. 69–75. [Online]. Available: https://ieeexplore.ieee.org/document/8925522/

[68] E. Doirado, “DogMate: Intent Recognition through Anticipation,” 2009.

[69] I. Leite, “Long-term Interactions with Empathic Social Robots,” Ph.D. dissertation, Instituto Superior Técnico, 2013.

[70] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, *Handbook of emotions*. New York, New York, USA: Guilford Press, 2010.

[71] J. Lee, Z. Wang, and S. Marsella, “Evaluating models of speaker head nods for virtual agents,” in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, International Foundation for Autonomous Agents and Multiagent Systems. Richland, SC: ACM, 2010, pp. 1257–1264.

[72] T. Dahle, “Culture and 3D animation : A study of how culture and body language affects the perception of animated 3D characters,” Ph.D. dissertation, University of Skövde, 2019. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-17242

[73] P. Ekman and W. Friesen, “Facial action coding system: A technique for the measurement of facial movement.” in *Differences among unpleasant feelings. Motivation and Emotion*. New York, New York, USA: Plenum Publishing Corp, 1978.

[74] S. Andrist, T. Pejsa, B. Mutlu, and M. Gleicher, “Designing effective gaze mechanisms for virtual agents,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM. Austin Texas USA: ACM, 2012, pp. 705–714.

[75] B. Lance and S. C. Marsella, “Emotionally expressive head and body movement during gaze shifts,” in *International Workshop on Intelligent Virtual Agents*, Springer. Paris, France: Springer, 2007, pp. 72–85.

[76] C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan, “Rigid head motion in expressive speech animation: Analysis and synthesis,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1075–1086, 2007.

[77] H. K. M. Meeren, C. C. R. J. van Heijnsbergen, and B. de Gelder, “Rapid perceptual integration of facial expression and emotional body language,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 45, pp. 16 518–16 523, 11 2005. [Online]. Available: https://www.pnas.org/content/102/45/16518https://www.pnas.org/content/102/45/16518.abstracthttp://www.pnas.org/cgi/doi/10.1073/pnas.0507650102

[78] N. Kaya and H. H. Epps, “Color-emotion associations: Past experience and personal preference,” in *AIC 2004 Color and Paints, Interim Meeting of the In ternational Color Association, Proceedings*. Porto Alegre, Brazil: Association Internationale de la Couleur, 2004, pp. 31–34.

[79] N. Kaya and H. H. Epps, “Relationship between Color and Emotion: A Study of College Students,” *College student journal*, vol. 38, p. 396, 2004.

[80] C. M. de Melo and J. Gratch, “The effect of color on expression of joy and sadness in virtual humans,” in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. Netherlands: IEEE, 2009, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/5349585/

[81] J. Forlizzi, J. Lee, and S. Hudson, “The kinedit system,” in *Proceedings of the conference on Human factors in computing systems - CHI ’03*. New York, New York, USA: ACM Press, 2003, p. 377. [Online]. Available: http://portal.acm.org/citation.cfm?doid=642611.642677

[82] S. Malik, J. Aitken, and J. K. Waalen, “Communicating emotion with animated text,” *Visual Communication*, vol. 8, no. 4, pp. 469–479, 2009. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1470357209343375

[83] S. McCloud, *Understanding Comics: The Invisible Art*. New York, New York, USA: Harper-Perennial, 1993.

[84] N. S. Schutte, J. M. Malouff, L. E. Hall, D. J. Haggerty, J. T. Cooper, C. J. Golden, and L. Dornheim, “Development and validation of a measure of emotional intelligence,” *Personality and Individual Differences*, vol. 25, no. 2, pp. 167–177, 8 1998. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0191886998000014

[85] P. Salovey and J. D. Mayer, “Emotional Intelligence,” *Imagination, Cognition and Personality*, vol. 9, no. 3, pp. 185–211, 1990. [Online]. Available: http://journals.sagepub.com/doi/10.2190/DUGG-P24E-52WK-6CDG

[86] A. Mill, J. Allik, A. Realo, and R. Valk, “Age-related differences in emotion recognition ability: A cross-sectional study.” *Emotion*, vol. 9, no. 5, pp. 619–630, 2009. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016562

[87] J. N. Bassili, “Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face.” *Journal of Personality and Social Psychology*, vol. 37, no. 11, pp. 2049–2058, 1979. [Online]. Available: http://content.apa.org/journals/psp/37/11/2049

[88] E. Mower, M. J. Mataric, and S. Narayanan, “Human Perception of Audio-Visual Synthetic Character Emotion Expression in the Presence of Ambiguous and Conflicting Information,” *IEEE Transactions on Multimedia*, vol. 11, no. 5, pp. 843–855, 2009. [Online]. Available: http://ieeexplore.ieee.org/document/4907039/

[89] D. Lottridge, M. Chignell, and A. Jovicic, “Affective Interaction: Understanding, Evaluating, and Designing for Human Emotion,” *Reviews of Human Factors and Ergonomics*, vol. 7, no. 1, pp. 197–217, 2011.

[90] I. Lefter, C. M. Jonker, S. K. Tuente, W. Veling, and S. Bogaerts, “NAA: A multimodal database of negative affect and aggression,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 21–27.

[91] B. Cheng, Z. Wang, Z. Zhang, Z. Li, D. Liu, J. Yang, S. Huang, and T. S. Huang, "Robust emotion recognition from low quality and low bit rate video: A deep learning approach," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 65–70.

[92] G. Chanel, S. Avry, G. Molinari, M. Bétrancourt, and T. Pun, "Multiple users' emotion recognition: Improving performance by joint modeling of affective reactions," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 92–97.

[93] L. Tian, M. Muszynski, C. Lai, J. D. Moore, T. Kostoulas, P. Lombardo, T. Pun, and G. Chanel, "Recognizing induced emotions of movie audiences: Are induced and perceived emotions the same?" in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 28–35.

[94] S. Ghosh, N. Ganguly, B. Mitra, and P. De, "Evaluating effectiveness of smartphone typing as an indicator of user emotion," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 146–151.

[95] S. Afzal, B. Sengupta, M. Syed, N. Chawla, G. A. Ambrose, and M. Chetlur, "The ABC of MOOCs: Affect and its inter-play with behavior and cognition," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 279–284.

[96] S. Kumano, R. Ishii, and K. Otsuka, "Comparing empathy perceived by interlocutors in multiparty conversation and external observers," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 50–57.

[97] ——, "Computational model of idiosyncratic perception of others' emotions," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 42–49.

[98] C. Sikora and W. Burleson, "The dance of emotion: Demonstrating ubiquitous understanding of human motion and emotion in support of human computer interaction," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 548–555.

[99] Y. Ding, L. Shi, and Z. Deng, "Perceptual enhancement of emotional mocap head motion: An experimental study," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 242–247.

[100] X. Lu, R. B. Adams, J. Li, M. G. Newman, and J. Z. Wang, "An investigation into three visual characteristics of complex scenes that evoke human emotion," in *2017 Seventh International*

Conference on Affective Computing and Intelligent Interaction (ACII). San Antonio, Texas, USA: IEEE, 2017, pp. 440–447.

[101] A. Hernández-García, “Perceived emotion from images through deep neural networks,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 566–570.

[102] N. V. Mudrick, M. Taub, R. Azevedo, J. Rowe, and J. Lester, “Toward affect-sensitive virtual human tutors: The influence of facial expressions on learning and emotion,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 184–189.

[103] J. Deanna Bailey, “Avatar and participant gender differences in the perception of uncanniness of virtual humans,” in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, Texas, USA: IEEE, 2017, pp. 571–575.

[104] E. Grossman, M. Donnelly, R. Price, D. Pickens, V. Morgan, G. Neighbor, and R. Blake, “Brain Areas Involved in Perception of Biological Motion,” *Journal of Cognitive Neuroscience*, vol. 12, no. 5, pp. 711–720, 9 2000. [Online]. Available: http://mitprc.silverchair.com/jocn/article-pdf/12/5/711/1758799/089892900562417.pdfhttps://direct.mit.edu/jocn/article/12/5/711/3471/Brain-Areas-Involved-in-Perception-of-Biological

[105] R. McDonnell, “Appealing Virtual Humans,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, 11 2012, vol. 7660 LNCS, pp. 102–111. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-34710-8_10http://link.springer.com/10.1007/978-3-642-34710-8_10

[106] J. Lasseter, “Principles of traditional animation applied to 3D computer animation,” *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 35–44, 1987.

[107] C. Sun, B. Wu, R. Wang, X. Hu, X. Yang, and C. Cong, “Intrinsic motivated multi-agent communication,” in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 3. Online, 2021, pp. 1656–1658. [Online]. Available: www.ifaamas.org

[108] W. Al Enezi and C. Verbrugge, “Skeleton-based multi-agent opponent search,” in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2021-Augus, 2021. [Online]. Available: https://streamable.

[109] A. Hosokawa and J. Morita, “Cognitive models leading to behavior-reading in a card game,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 337–341.

[110] A. R. Maina-Kilaas, G. D. Montanez, C. Hom, K. Ginta, and C. Lay, “The Hero’s Dilemma: Survival Advantages of Intention Perception in Virtual Agent Games,” in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2021-Augus, 2021.

[111] S. G. Ware and C. Siler, “The sabre narrative planner: Multi-agent coordination with intentions and beliefs,” in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 3, 2021, pp. 1686–1688. [Online]. Available: www.ifaamas.org

[112] S. Makri and P. Charalambous, “Towards a multi-agent non-player character road network: A Reinforcement Learning approach,” in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2021-Augus, 2021. [Online]. Available: https://youtu.be/-qwFddCkyDo.

[113] T. Shintani, C. T. Ishi, and H. Ishiguro, “Analysis of role-based gaze behaviors and gaze aversions, and implementation of robot’s gaze control for multi-party dialogue,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 332–336.

[114] D. M. Olson and Y. Xu, “Building Trust over Time in Human-Agent Relationships,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 193–201. [Online]. Available: https://doi.org/10.1145/3472307.3484178

[115] T. Schneeberger, M. Scholtes, B. Hilpert, M. Langer, and P. Gebhard, “Can Social Agents elicit Shame as Humans do?” *2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019*, 2019.

[116] L. A. Dennis and N. Oren, “Explaining BDI agent behaviour through dialogue,” in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 1, 2021, pp. 429–437. [Online]. Available: www.ifaamas.org

[117] I. Torre, S. Tuncer, D. McDuff, and M. Czerwinski, “Exploring the Effects of Virtual Agents’ Smiles on Human-Agent Interaction: A Mixed-Methods Study,” in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 9 2021, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/9597445/

[118] L. Yang, C. Achard, and C. Pelachaud, “Interruptions in Human-Agent Interaction,” in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA 2021*, vol. 1, no. 1. Association for Computing Machinery, 2021, pp. 206–208.

[119] T. Takahashi, K. Tanaka, K. Kobayashi, and N. Oka, “Melodic Emotional Expression Increases Ease of Talking to Spoken Dialog Agents,” in *Proceedings of the 9th International Conference on Human-Agent Interaction*. New York, NY, USA: ACM, 11 2021, pp. 84–92. [Online]. Available: https://doi.org/10.1145/3472307.3484180https://dl.acm.org/doi/10.1145/3472307.3484180

[120] E. B. Cloude, F. Wortha, D. A. Dever, and R. Azevedo, “Negative emotional dynamics shape cognition and performance with MetaTutor: Toward building affect-aware systems,” in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 9 2021, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/9597462/

[121] R. Asraf, C. Rozenshtein, and D. Sarne, “On the effect of user faults on her perception of agents faults in collaborative settings,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 372–376.

[122] Q. Roy, M. Ghafurian, W. Li, and J. Hoey, “Users, Tasks, and Conversational Agents: A Personality Study,” in *HAI 2021 - Proceedings of the 9th International User Modeling, Adaptation and Personalization Human-Agent Interaction*. Association for Computing Machinery, Inc, 11 2021, pp. 174–182.

[123] D. S. Carvalho, J. Campos, M. Guimarães, A. Antunes, J. Dias, and P. A. Santos, “CHARET: Character-centered approach to emotion tracking in stories,” in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 3, 2021, pp. 1457–1459. [Online]. Available: www.ifaamas.org

[124] A. Antunes, J. Campos, J. Dias, P. A. Santos, and R. Prada, “EEG Model: Emotional Episode Generation for Social Sharing of Emotions,” in *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*, vol. 1, no. 1. New York, NY, USA: ACM, 9 2021, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.1145/3472306.3478342

[125] D. Rato, M. Couto, and R. Prada, “Fitting the Room: Social Motivations for Context-Aware Agents,” in *Proceedings of the 9th International Conference on Human-Agent Interaction*. New York, NY, USA: ACM, 11 2021, pp. 39–46. [Online]. Available: https://doi.org/10.1145/3472307.3484185https://dl.acm.org/doi/10.1145/3472307.3484185

[126] P. Raimbaud, A. Jovane, K. Zibrek, C. Pacchierotti, M. Christie, L. Hoyet, J. Pettré, and A. H. Olivier, “Reactive Virtual Agents: A Viewpoint-Driven Approach for Bodily Nonverbal Communication,” in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA 2021*, 2021, pp. 164–166. [Online]. Available: https://doi.org/10.1145/3472306.3478351

[127] T. Beinema, D. Davison, D. Reidsma, O. Banos, M. Bruijnes, B. Donval, A. F. Valero, D. Heylen, D. Hofs, G. Huizing, R. B. Kantharaju, R. Klaassen, J. Kolkmeier, K. Konsolakis, A. Pease, C. Pelachaud, D. Simonetti, M. Snaith, V. Traver, J. Van Loon, J. Visser, M. Weusthof, F. Yunus, H. Hermens, and H. Op Den Akker, “Agents United: An Open Platform for Multi-Agent Conversational Systems,” in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA 2021*. New York, NY, USA: ACM, 9 2021, pp. 17–24. [Online]. Available: https://dl.acm.org/doi/10.1145/3472306.3478352

[128] R. Pereira, “Expressing Emotions through Animated Speech Balloons,” Master’s thesis, Instituto Superior Técnico, University of Lisbon, 2018. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/1972678479054259

[129] R. Silva, “Supporting Affective Expression in Multi-party Interactions,” Master’s thesis, Instituto Superior Técnico, University of Lisbon, 2018. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meic-t/dissertacao/1409728525632118

[130] R. Garrison, “Implications of Online and Blended Learning for the Conceptual Development and Practice of Distance Education,” 2009.

[131] T. Anderson and J. Dron, “Three generations of distance education pedagogy,” *International Review of Research in Open and Distance Learning*, vol. 12, pp. 80–97, 2011.

[132] A. Andrade, F. Nogueira, C. V. de Carvalho, P. Russo, and D. Gouveia, “Games to support problem-based learning,” in *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*. New Jersey, NJ, USA: IEEE, 2015, pp. 1–4.

[133] R. Leitão, J. M. F. Rodrigues, and A. F. Marcos, “Game-based learning: augmented reality in the teaching of geometric solids,” *International Journal of Art, Culture and Design Technologies (IJACDT)*, vol. 4, pp. 65–77, 2014.

[134] P. Rodrígues and J. Bidarra, “Transmedia storytelling and the creation of a converging space of educational practices,” *International Journal of Emerging Technologies in Learning*, vol. 9, p. 42, 2014.

[135] B. Cooper, P. Brna, and A. Martins, “Effective affective in intelligent systems–building on evidence of empathy in teaching and learning,” *Affective Interactions Towards a New Generation of Computer Interfaces*, vol. 1814/2000, p. 21–34, 2000. [Online]. Available: http://www.springerlink.com/index/j8v0l230t3503367.pdf

[136] F. Heider and M. Simmel, “An experimental study of apparent behaviour,” *The American Journal of Psychology*, vol. 57, no. 2, pp. 243–259, 1944. [Online]. Available: http://www.jstor.org/stable/1416950?origin=crossref%5Cnhttp://www.jstor.org/stable/10.2307/1416950

[137] N. D. Feshbach, “Studies of empathic behavior in children,” *Progress in experimental personality research*, vol. 8, p. 1–47, 1978.

[138] K. Niven, D. Holman, and P. Totterdell, “How to win friendship and trust by influencing people’s feelings: An investigation of interpersonal affect regulation and the quality of relationships,” *Human Relations*, vol. 65, no. 6, pp. 777–805, 2012.

[139] C. Reeck, D. R. Ames, and K. N. Ochsner, “The Social Regulation of Emotion: An Integrative, Cross-Disciplinary Model,” pp. 47–63, 2016.

[140] A. Pires, P. Alves-Oliveira, P. Arriaga, and C. Martinho, "Cubus: Autonomous Embodied Characters to Stimulate Creative Idea Generation in Groups of Children," in *Proceedings of the 17th International Conference on Intelligent Virtual Agents - IVA 2017*. Cham, Switzerland: Springer International Publishing, 2017, pp. 360–373. [Online]. Available: http://link.springer.com/10.1007/978-3-319-67401-8_46

[141] J. Brooke, "SUS - A quick and dirty usability scale," in *Usability Evaluation in Industry*. Taylor & Francis Group, 2&4 Park Square, Milton Park, Abingdon, OX14 4RN: London: Taylor and Francis, 1996.

[142] ——, "SUS: A Retrospective," *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, 2013.

[143] V. Sacharin, K. Schlegel, and K. R. Scherer, "Geneva Emotion Wheel Rating Study," 2012. [Online]. Available: https://archive-ouverte.unige.ch/unige:97849

[144] R. Rodrigues and C. Martinho, "Expectancy and Emotions in Synthetic Characters," 2016. [Online]. Available: https://sotis.tecnico.ulisboa.pt/record/1e9c2b70-07cd-4e68-93d8-885590d47b09

[145] B. G. Breitmeyer and L. Ganz, "Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing." *Psychological Review*, vol. 83, no. 1, pp. 1–36, 1 1976. [Online]. Available: /record/1979-20169-001http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.83.1.1

[146] R. Rodrigues and C. Martinho, "Measuring Believable Interactions, a Case Study: 3Motion," *Interaction with Agents and Robots: Different Embodiments, Common Challenges (Workshop at IVA 2017)*, 2017.

[147] J. Vasquez, "Implementing N-Grams for Player Prediction, Procedural Generation, and Stylized AI," in *Game AI Pro : Collected Wisdom of Game AI Professionals*, Taylor & Francis Inc, Ed. Natick, United States: Taylor & Francis Inc, 2013, pp. 567–580. [Online]. Available: http://www.gameaipro.com

[148] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales." *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 1063–1070, 1988. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3397865/http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.54.6.1063

[149] V. Nowlis and H. H. Nowlis, "THE DESCRIPTION AND ANALYSIS OF MOOD," *Annals of the New York Academy of Sciences*, vol. 65, no. 4, pp. 345–355, 1956.

[150] H. Schlosberg, "Three dimensions of emotion," *Psychological Review*, 1954.

[151] S. L. Franconeri, A. Hollingworth, and D. J. Simons, "Do New Objects Capture Attention?" *Psychological Science*, vol. 16, no. 4, pp. 275–281, 4 2005. [On-

line]. Available: https://journals.sagepub.com/doi/abs/10.1111/j.0956-7976.2005.01528.xhttp://journals.sagepub.com/doi/10.1111/j.0956-7976.2005.01528.x

[152] F. I. Parke, “Perception-based animation rendering,” *The Journal of Visualization and Computer Animation*, vol. 2, no. 2, pp. 44–51, 4 1991. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/vis.4340020204https://onlinelibrary.wiley.com/doi/abs/10.1002/vis.4340020204https://onlinelibrary.wiley.com/doi/10.1002/vis.4340020204

Video Games and Media References

[153] *Halo Infinite* [Video game]. (2021). 343 Industries, Xbox Game Studios.

[154] *StarCraft II* [Video game]. (2010). Blizzard Entertainment.

[155] *Overwatch* [Video game]. (2016). Blizzard Entertainment.

[156] *The Witcher 3: Wild Hunt* [Video game]. (2015). CD PROJEKT RED.

[157] *Tomb Raider* [Video game]. (2013). Crystal Dynamics, Square Enix.

[158] *Battlefield V* [Video game]. (2018). DICE, Electronic Arts.

[159] *Horizon Zero Dawn* [Video game]. (2020). Guerrilla, PlayStation PC LLC.

[160] *Fable II* [Video game]. (2008). Lionhead Studios, Microsoft Game Studios.

[161] *The Lord of the Rings*. (2003). New Line Cinema.

[162] *Darkest Dungeon* [Video game]. (2016). Red Hook Studios.

[163] *Call of Duty: WWII* [Video game]. (2017). Sledgehammer Games, Raven Software, Activision.

[164] *Assassin's Creed Odyssey* [Video game]. (2018). Ubisoft.

[165] *Inside Out*. (2015). Walt Disney Pictures, Pixar Animation Studios, Walt Disney Studios Motion Pictures.

A

Additional Related Work

Contents

A.1 Notes on Metrics for Character Believability

In Gomes *et al.*[28], the audience's perception is asserted using Likert scales, one scale per dimension. The templates for the phrases to be rated, except for emotional expressiveness, by the subjects are:

- awareness: <***X***> perceives the world around him/her.
- behavior understandability: It is easy to understand what <***X***> is thinking about.
- personality: <***X***> has a personality.
- visual impact: <***X***>'s behavior draws my attention.
- predictability: <***X***>'s behavior is predictable.
- behavior coherence: <***X***>'s behavior is coherent.
- change with experience: <***X***>'s behavior changes according to experience.
- social: <***X***> interacts socially with other characters.

A.2 Summary of Affective Models Categories

In [54], the author describes five categories that incorporate different affective models: *Appraisal Theory Approaches*; *Anatomical Approaches*; *Rational Approaches*; *Communicative Approaches*; and, *Dimensional Theory Approaches*. Here follows a small description of each approach.

Anatomical Approaches try to emulate the neural structure that is behind an emotional response. They tend to specialize in a single emotion, since they see emotions as separate entities with their own systems, giving great importance to the systems that create the emotion. Although a very detailed approach, they focus on a more raw and basic emotional response and tend to specialize in only one emotion.

Rational Approaches "ponder what adaptive function does emotion serve"[63], attempting to incorporate an abstract version of an emotion from its implementation in humans into a model of intelligence. This approach is typically associated with artificial intelligence research, where models using this approach are usually used to further develop machine intelligence theories.

Communicative Approaches focus on the social component of emotion, that "serves an empathic objective to aid in communication and to transmit non-verbal cues"[63]. This approach is more usually used in social studies, crowd dynamics, and multi-agent systems, focusing on the outward emotional display, often simplifying internal workings for creating an emotion.

Dimensional Theory Approaches focuses on the view that emotions shouldn't be viewed as discrete and unrelated – a concept introduced by Nowlis *et al.*[149] showing that there were between six and twelve independent affective states (ie. sadness, anger, anxiety, etc.) – but as an end product of a system of undisclosed variables[150], where affective states are connected and their origins is an n-dimensional vector. Furthermore, "a big advantage of the dimensional approach is one can attempt to code the seemingly complex nature of human emotions as a combination of simpler internal factors"[63].

Appraisal Theory Approaches postulates that "all emotions come mostly from our own interpretations of events"[63], where our appraisal of the situation is the emotional response. The theory is best used in connecting awareness with emotion, focusing on the individual and their psychological response, where their own judgment of a situation is to blame as the source of their emotional response.

Ortony *et al.*[50] describe an emotional classification that states that emotion is structured into the categories of *Fortune-of-others*, *Prospect-based*, *Well-Being*, *Attribution* and *Attraction*, or more largely grouped into *consequences of events*, *actions of agents*, or *aspects of objects*. This is known as the **OCC Model** and many virtual emotional models that opt for the appraisal theory often base themselves on this model. The OCC Model attempts to incorporate all emotions, but with no relationship between them other than categorical, but not all models based on the OCC model incorporate all emotions. One good example is the model proposed by Ochs *et al.*[13] which focuses on the believability of the NPCs. They attempt to improve the experience by focusing on the personality, social relations, and roles of the NPCs inside a game. The emotions modeled, using the OCC, were joy/distress, hope/fear, and relief/disappointment. There is also an emotional decay component implemented to revert the emotional state to a neutral state after some time period. This model focuses on the NPCs and tries to increase believability through simulation of social relations.

Another variation on the OCC model comes from He *et al.*[65], it is a fuzzy emotional model for virtual agents. The emotions of events modeled were hope/fear, satisfaction/fear-confirmation, and relief/disappointment, and were based on three variables, desirability, "if an event is beneficial to an agent, it is desirable otherwise it is undesirable", importance, "we equate goal's importance with motivation intensity", and likelihood, "we equate likelihood with the possibility of what other virtual agents will do". An important feature is "how the model accounts for relations between agents as predictor of behavior for them"[63].

Using a partial OCC model implementation, the work of Jacobs *et al.*[66] focuses on the importance of the link between reward/punishment and an emotional response. They propose the use of an emotion label system that converts each agent's state transition, an initial joy/distress mapping that converts into a hope/fear mapping over time, related to the agent's previous knowledge, hopefully allowing the gathering of useful information for planning and decision making of the agent. This model's objective is focused on the modeling of the agent behavior but is important to note how the modeling anticipatory behavior improved the tested agent's performance.

FatiMA[64] offers an interesting appraisal theory application, where the OCC model is implemented by storing appraisals (valence based) in a numeral intensity value (-10;10). This model acting along with goal mechanisms and perceived events, models the complete range of emotions inside the OCC, being able to give individual personalities and coping mechanisms to deal with specific goals. This model however not take into account anticipation of behavior or emotions.

The appraisal theory approaches show potential when used in a more static NPC or Environment emotional association, giving a simpler and robust emotional model to the in-game interactions.

A.3 Perception of Behavior

We live in a dynamic visual world: "Objects often move about in our environment and we, as observers, are often moving as we look at those events"[104], but the "[r]eflected light carries too much information for the human visual system to process at once"[151]. Instead, some locations or objects are selectively prioritized at the expense of others. Often, priority is task or goal dependent (e.g. drivers might preferentially attend to red objects because of the importance of brake lights, stop lights, and stop signs), but keep in mind that "the actual cues that capture visual attention are luminance-based transients [e.g. motion], not new objects"[151]. Furthermore, "[t]he visual system seems designed to give highest priority to moving objects, though at low resolution"[152].

At the turn of the century, vision research demonstrated that visual perception is divided into two major systems which operate in parallel and semi-independently. These two channels are referred to as the transient and the sustained neuron systems [145, 152]. The *sustained system* has poor temporal response but has a good spatial resolution. It is specialized in pattern and detail detection and processing of structural or figural information. The *transient system* has poor spatial resolution but has a good temporal response. It is specialized for motion detection, being part of an "early warning system"[145]. While the *sustained system* can help recognize shapes, it requires time to do so, this is counterposed with the transient system that only detects motion, but does so quickly. In other words, "[f]or any neuron, the better its capacity to resolve details in space, the poorer its capacity to resolve details in time"[145].

In practice, in humans "the critical duration for brightness discrimination [...] is on the order of 100ms, whereas for form identification or acuity the critical duration can range from 200-350ms"[145]. Parke also states that "[a]n abrupt change in the scene caused by motion will mask detail perception for about 250ms"[152]. These times give us a good understanding of how long it takes for someone to perceive the motion (100ms) and to actually understand what it is (200-350ms).

B

Appendix on the Expression of Emotions in Virtual Coaches

Contents

B.1 Bubble System in Detail

This section expands on the Bubble System presented previously. As previously stated, the Bubble System controls all 2D elements used to communicate emotion and is subdivided into two components: balloons and backgrounds. For the backgrounds, the relevant aspects are texture and color, while for the balloons, the sprite, text, effects, colors, and animations will change accordingly.

Balloon Shape – We created one distinct speech bubble for each emotion, plus a neutral state, based on the ones traditionally used in Comics (e.g. spiked balloons for anger) as well as using design guidelines to convey the appropriate valence for the associated emotion. To that end, we followed a user-centric iterative process until reaching a desired response to a speech balloon (e.g. curves for positive valences, and lines/obtuse angles for negative valences).

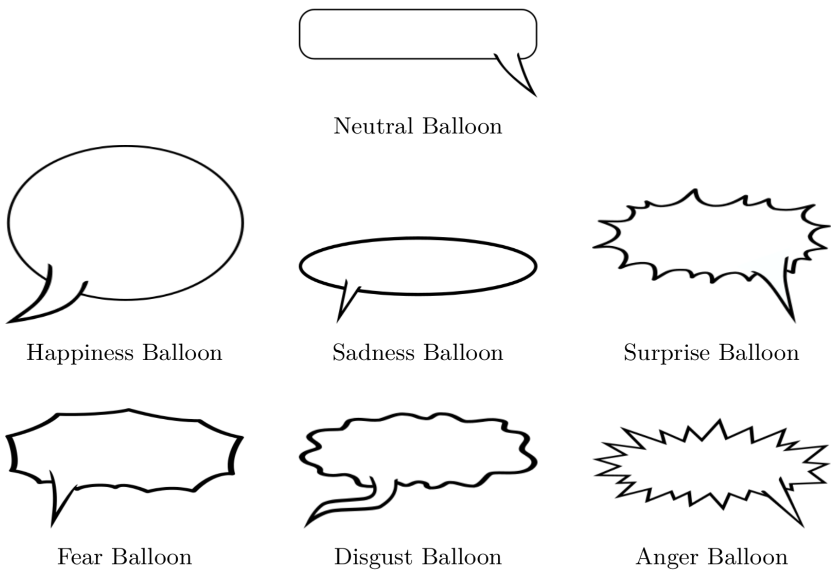


Figure B.1: Balloons shown with associated emotion. These balloons were colored and animated when appearing and disappearing.

We defined the neutral and anger balloon the standard comic way, as a rectangular and spiked shape respectively[1]. For happiness, we went with a more circular form, since curvy shapes tend to be perceived as positive, while sadness ended with an oval shape. Since surprise can either be positive or negative, we chose a mixture of balloons that had such valence. Therefore, it became similar to the anger balloon, but with fewer and more curved spikes. Disgust ended up having a wave format in order to convey the feeling of nausea. Finally fear, due to its negative valence, was created as multiple lines, creating obtuse angles.

Balloon Color – We decided to follow the same approach as Pires et al. [140], and draw inspiration from the Disney™ movie “Inside Out”[2]. The colors chosen follow the same representations as those

[1] Representation of scream (anger), narration (neutral), and dialog speech balloons as seen in Spider-Man Annual #5 - A Day at the Daily Bugle: `https://imgur.com/a/lJ0l2tw`

[2] Inside Out, Walt Disney Pictures, Pixar Animation Studios (2015)

present in the movie, and for those emotions that were absent, a mix of other emotions/colors was used, to help communicate the appropriate valence.

Table B.1: Matching between colors and corresponding emotion as used in the application. The hexadecimal code used for the color is also shown.

Emotion	Color	Hex
Neutral		#8CDBA1FF
Happiness		#F0E64DFF
Surprise		#FFC358FF
Sadness		#1D33CEFF
Fear		#AE52ECFF
Disgust		#C5D137FF
Anger		#FF0000FF

Anger was represented with red, happiness with yellow, disgust with yellow-green, sadness with blue, and fear with purple. Surprise was defined as orange (aggressiveness of red and the positive valence of yellow), and for the neutral state, we decided upon a mix between green and blue, since it is usually seen as relaxed. For readability, the text color changes between black and white according to the balloon color.

Balloon Animation – We decided to create simple balloon animations as scaling/motion over time, following the pose to pose principle (see Section 2.6). In total, we ended up with 14 distinct animations, a pair (one to show the balloon, and another one to hide it) for each of the six emotions, plus another pair for the neutral state. To distinguish between different intensities, the speed of the animation was changed accordingly (slower speeds for lower intensities, faster speeds for higher ones).

Happiness was achieved with bounces, and sadness with a lower speed, due to how it is used in traditional animation; for surprise, we added a pendulum motion; for fear, we created a palpitation over time, since fear is usually accompanied by a strong heartbeat; disgust was achieved with a wave motion, with the idea of seasickness; and, finally, anger was defined with a chaotic path.

Text Animation – Regarding the text component of the speech balloons, we went with a parametric approach using animation curves (curves that define the text animation functions), since this allows us to have generalized effects and apply different curves to them, thus producing different animations (e.g. appear over time, jump, fade in/out, and shake).

The neutral animation was made with characters appearing over time; happiness ended up defined as a wave with a bell curve, which creates a jumping motion; surprise was defined as a swing, just like the pendulum motion of its balloon; sadness was created using a fade effect, and fear as jitters; disgust was created as a warp over a curve, which is just a displacement; and finally anger was defined with a shake.

Background Icons – Finally, we created icons for the backgrounds, and colored and animated the corresponding planes. For this, we either chose generic patterns, or specific icons, depending on whether we wanted the plane to be abstract, or to convey one of the specific application metrics discussed previously.

For abstract backgrounds, we wanted an image with any unrecognizable pattern, as long as we could define its color. For icons, we based them on the application metrics. The backgrounds were also given animations, mostly to change between them, and their respective color. However, while the colors used reflected the emotion felt, the iconography was not pursued extensively.

C

Appendix on 3motion

Contents

C.1 3motion Illustrative Scenario

This scenario is depicted in Figure C.1, where Hanna and Bob are about to cross a bridge. The boxes in white represent the action stage and emotion an agent is feeling at a given time, the arrows represent events being passed to the other agent.

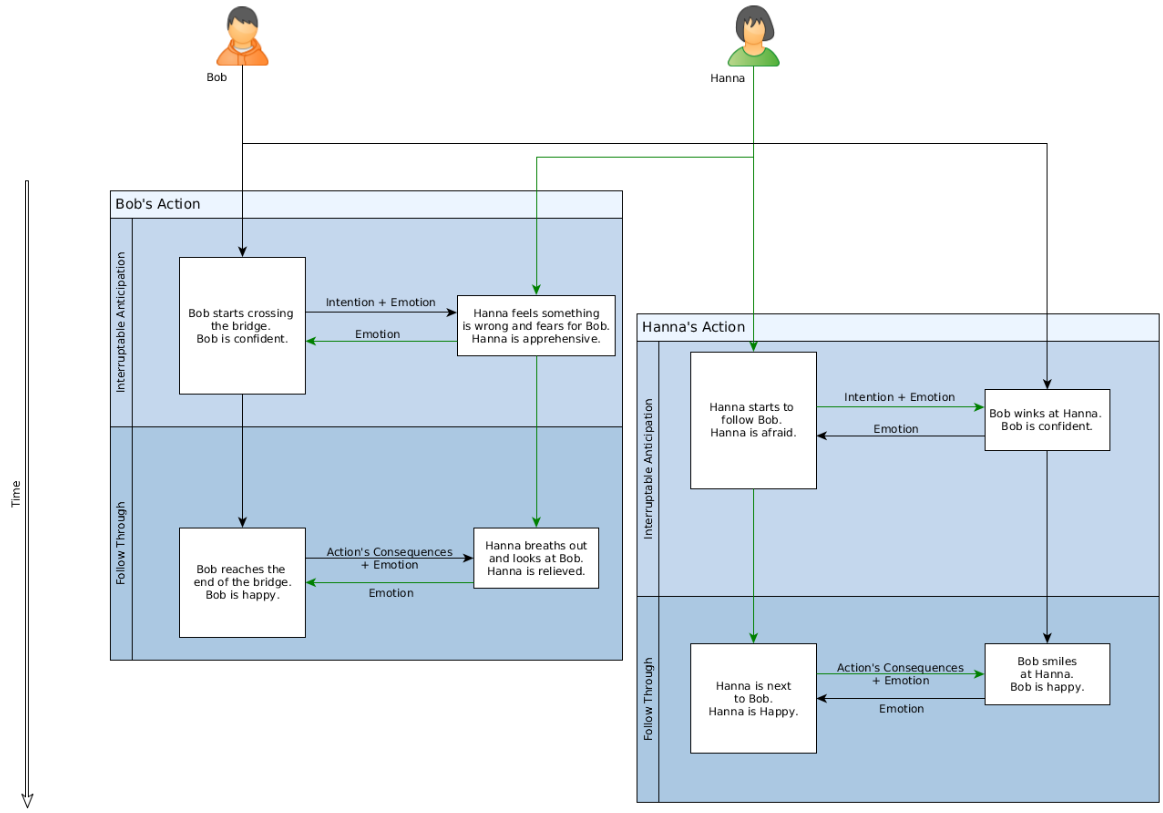


Figure C.1: Illustrative scenario where agents cross a bridge.

Bob confidently starts crossing the bridge, broadcasting his intention and emotion, some time later Hanna perceives the event and shows apprehension, but decides to follow Bob, showing fear in crossing the bridge. To this event Bob replies with confidence, maybe to try and calm Hanna down. Up to this point, the actions haven't left the interruptible anticipation stage and the agents could have canceled the action, but decided not to. It's important to note that in most systems, where the anticipation stage is not considered, these important non-verbal messages are completely ignored.

The first to reach the end of the bridge is Bob, showing happiness. Hanna is relieved that everything went fine and breaths out, and a little time later, she also reaches the end of the bridge next to Bob showing happiness, to which Bob replies with the same emotion.

Although with just two actions, this scene can be complex and help create more believable scenarios by using *3motion*.