Supporting Affective Expression in Multi-party Interactions

Ricardo Filipe Fonseca Silva
Instituto Superior Técnico, Lisboa, Portugal

Abstract—**When designing synthetic characters, it is imperative to consider the expression of affective states, in order to achieve believable communications. This work implements an animation module, using Unity 3D's Mechanim system, capable of supporting interactions between two synthetic characters and a human. To accomplish this, we coordinated head motion and emotional expression, via the use of blending and layering of simple animations, to create more dynamic and believable dialogs. Two sets of experiments were conducted on the system, which revealed that while the animation quality was improved, not all occurrences of emotional ambiguity were able to be resolved, and further work needs to be done to balance emotional believability and recognition.**

Index Terms—**Believability, Expressiveness, Head Motion, Interaction, Virtual Tutoring**

1. Introduction

Synthetic characters are becoming more realistic with each passing day, with 3D artists being able to create characters whose expressions and emotions faithfully mimic those of an actual human.

In cinematography, one can find a suitable example in the 2009 movie "Avatar". Thanks to state of the art motion capture technology, the director James Cameron accomplished his goal of "*making the audience believe the Na'vi* (movie's extraterrestrial race) *were emotional creatures"* [1]. This concept of **Believability** is used to describe characters that appear to be alive, and are able to emotionally affect the audience [2].

Video games also frequently capitalize on believability. This is particularly noticeable in role-playing games, given that the player frequently engages in dialog with non-playable characters. An example is the video game "Star Wars: Knights of the old Republic" (LucasArts 2003)[1]. Despite being visually unimpressive, it manages to offer emotionally engaging 3D characters to its players, cementing the crucial role that properly conveying emotion plays in creating believability. Another example is the video game L.A. Noire (Rockstar Games 2011)[2], which focuses heavily on interactions between multiple characters, making it a necessity to have them be as believable as possible. In fact, characters who are emotionally disconnected may end up tarnishing the experience, as can be observed in the video game Mass Effect: Andromeda (Electronic Arts 2017)[3]. With confusing facial expressions, out-of-place back-channeling and emotionally ambiguous characters, it was met with unfavorable critics at launch. These examples support how important it is for characters to adequately display emotions and mannerisms that follow the active narrative.

While straight-forward in movies and video games, this becomes hard to accomplish when intelligent agents are involved, given their tendency to act in accordance to their models. Overcoming this hurdle is the key for creating synthetic characters capable of adequate affective expressiveness. In this work we bring forth an animation module, capable of supporting interactions between two characters and a human. It features an animation controller aimed at enhancing the believability of said characters. We hypothesized that using this module to synchronize the motions of the characters, with the remaining emotional modalities, would allow us to achieve believable dialogs, and promote more coherent and credible interactions. This module is also integrated into the domain of a larger application, so that it can be used for the purposes of remote tutoring for university students.

In the following section, we will discuss the concepts and findings that we consider important to our work, such as emotion, believability, animation, expressive modalities and virtual characters. Following, we will be presenting the architecture of our system, and present its most critical components. Lastly, we will analyze the data and observations of two experiments, focusing on the emotional quality and accuracy of our implementation, as well as discuss the benefits and limitations we uncovered as a result of said experiments.

2. Related Work

2.1. Emotion

Emotions are integral to creating alive characters, and are fundamental when trying to achieve believability [2] [3]. A character that cannot express believable emotions can never truly claim to have achieved true believability [4] [5]. Various aspects contribute to the way a character transmits emotion. From the way it moves and talks, to the gestures

[1] Bioware 2003, Star Wars: Knights of the Old Republic, computer program, LucasArts

[2] Team Bondi 2011, L.A. Noire, computer program, Rockstar Games

[3] Bioware Montreal 2017, Mass Effect: Andromeda, computer program, Electronic Arts

and facial expressions it conveys while engaged in conversation. These all help in conveying the correct emotion to the audience, which in turn makes the emotions more believable, and by proxy, makes the characters more believable. [3] [4] Some factors also influence what emotions a character should be feeling. The mood of a character is one of these factors. It both influences and is influenced by emotions, and the feelings involved tend to last for longer periods of time, when compared with emotions which can, at times, even be instantaneous. [6]

In our work, we will focus on what is knows as the six basic families of emotion; Fear, Anger, Disgust, Surprise, Happiness and Sadness [7] [6]. We say families, because more than one universal expression has been identified for each emotional family [8], and all contain variations in the facial expressions, movements, gazes and gestures associated with them.

Emotional Intelligence. Emotional Intelligence is the ability to recognize different emotions, and use that information to adapt to specific environments [9]. Since the main focus of our work is to correctly transmit emotions, it is important to understand how emotional responses differ between different people.

Neither younger nor older adults are able to correctly identify most emotions with 100% certainty. Accuracy varies not only from emotion to emotion, but also from individual to individual. Bassili [10] presented a discussion on which emotions could be mistakable with one another, and pointed to similarities/ambiguity between facial expressions, alongside poor actor performance, as the likely contributors to the mistakes. Another issue is the use of multiple emotional channels. Mower et al. [11] presented evidence that adding a second channel can impact emotion recognition, if the information provided is conflicting and/or ambiguous. This is an important aspect in the context of this work, given that the parent application will feature a combination of two distinct visual channels: a set of animated 3D characters, accompanied by (also animated) 2D elements.

2.2. Believability

The term *Believable* is used to describe characters that are able to elicit emotions from the viewer [2]. These do not need to be , and often simple characters turn out to be more believable than perfectly modeled ones [3], given that the latter may end up falling into the *uncanny valley* [12].

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. While doing so, they must take into account context of their emotions, given that believability is also a highly contextual notion [13]. **Believable Social Agents** is a subset of believable agents, used to reference agents that interact with each other, as well as with the user.

2.3. Fundamentals of Animation

Believability cannot be achieved without proper use of animation. Given how familiar and accustomed we are to seeing other human faces, incorrectly animating them is a quick way to shatter the audience's *suspension of disbelief.*

To understand what is involved in correctly animating characters, we must look at traditional animation for help. In their book *The Illusion of Life: Disney Animation* [4], O. Johnston and F. Thomas discuss how the *Fundamental Principles of Animation* came to exist. These would later be adapted to 3D animation by J. Lasseter [14].

All principles of animation are important when it comes to creating believable characters, but we will only be presenting the ones that are of significant relevance to our work.

1) **Squash and Stretch** - Objects exhibit considerable distortion in its shape during action. This allows us to show the relations between emotion and various sections of the face [14]. During anger, the eyebrows lower and become stiffer. During sadness, the corners of the mouth and eyes lower.
2) **Slow in and slow out** - Few actions happen over a constant frame. The bulk of the action usually happens at the start, with a bigger spacing between frames towards the end. This plays a role in achieving natural transitions between expressions.
3) **Arcs** - Actions and movements often follow a circular path. Arcs are very important in reducing the rigidness of movements, helping to achieve natural animations.
4) **Secondary Actions** - Used as a way of reinforcing the main action by adding secondary, subordinate actions (adding movement to expressions, and vice versa). Its use helps reinforce the realism of the animation.

2.4. Expressiveness

Facial Expressions. Facial expressions serve to transmit our emotional state to the person we are communicating with [4] [15] [6], and are one of the briefest emotional signals, usually lasting only mere seconds [7]. Those that last for longer periods of time tend to be associated with more intense feelings. Being universally understood signals, expressions are very useful at portraying each of the basic families of emotion [6], and in addition to enhancing the conversation, can also be used to convey dominance and affiliation.

In order to get a believable character to show facial expressibility, one needs to know the appropriate aspects of the face to transform. One of the more prominent techniques for achieving this is the Facial Action Coding System (FACS) [7]. Developed for measuring facial movement, FACS is able to objectively score each and every 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 [16] and Makarainen et al [17] suggested models that combined the AUs for the six families of basic emotions in a way that would allow for the creation of perceptually valid facial expressions for blending/transition of universal emotions. Using similar methods, one is able to create entirely new, valid expressions.

Back-channeling. When engaging in social interactions with others, it is normal human behavior to use non-verbal signals. These not only help us deliver our communicative intents [18], thus promoting effective communication [19], but also help us establish our personal tendencies, which in turn help define us as truly unique individuals [20]. Non-verbal behaviors give the listener a chance of providing basic communicative functions towards what the speaker is saying [20], and their absence could lead to negative experiences and incorrect interpretations of the situation [19].

Many aspects are important when implementing back-channeling, but two stand out as crucial: *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 [21]. It is worth noting that gaze is not simply relegated to eye movement. Head and even body movement/position can be an active component in gazing. These are not independent from one another and, if treated as such, will appear random and disjointed [22]. **Head Motion** is an integral part of communication. Noding 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 gazes. It also plays an important role with regards 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 [23]

2.5. Believable Virtual Tutoring

Virtual humans have proven to be a valuable method with which to simulate real humans and elicit emotional responses [24]. People often perceive virtual humans in the same way they perceive real humans, and exhibit the same human mannerisms that arise during human-to-human interaction [25]. This would be impossible without the virtual humans' ability of establishing rapport, which builds overtime via nonverbal behaviors and expressions, and is necessary to understand the development of successful personal relationships [25].

The dynamics between virtual and real humans are relevant to us, given that our work will be part of a virtual tutoring application, based on the work of Lima et al. [26]. The application will assist in bridging the gap between students and human tutors, by attempting to create empathic relationships between them and two virtual agents. The application itself is being developed with smartphones in mind, due to convenience offered by mobile platforms, their ability to display convincing virtual humans, and the advantage of facilitating long-term personalized interaction with the students [24].

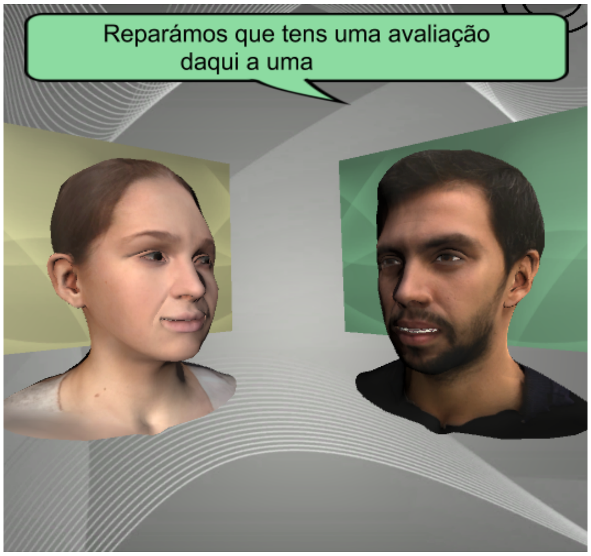


Figure 1: A snapshot of the state of the Virtual Tutoring Application, at the time this document was written

3. Implementation

This work is part of a larger project involving the creation of a Virtual Tutoring application, based on an adaptation of the work brought forth by Lima et al [26], and aims to shorten the gap between students and tutors (see figure 1 for a visual outlook of the application).

At our disposal we had two synthetic 3D models, alongside some animation clips for the six emotions of Ekman. Most emotions had an idle animation, lasting a couple seconds, and two brief animations that covered both intense and slight expressions of emotion. We will be referring to the intensity of emotions as "High" and "Low" from here on out (ex. "Surprise Low" indicates surprise with low intensity). We also had animations for talking and nodding. These characters and animation clips were obtained via live capture and worked on by 3D animators, courtesy of the Didimo, Inc[4]company. Our contribution to the project provides the visual expressiveness and believability components of these characters, who have been integrated into the application as two virtual tutors. The goal was to bring the characters and animation clips together, via the use of facial blending techniques and changes to the speed, form and frequency of the animations, in order to achieve more interesting and rich interactions, while maintaining a desirable level of believability.

3.1. Architecture

Figure 2 features a finalized version of the architecture of our character animation system. The relevant information flow into our system starts at the "Dialog Manager" and

[4] Didimo, Inc, "Didimo - A digital version of you from a single photo", Software company, http://www.mydidimo.com/

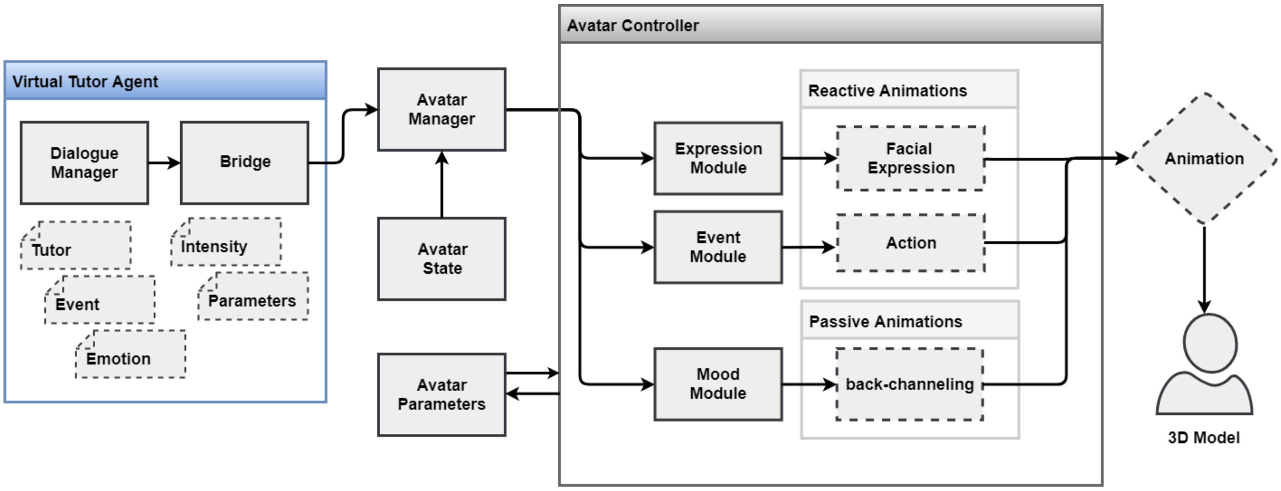


Figure 2: Architecture of our Animation System for the Virtual Tutoring application

"Bridge" modules, which are part of the virtual tutoring application. When a situation arises that requires an animated response from the characters, the manager sends out a request to the bridge module, which refines it to a point where it can be processed by our animation system. The bridge module also features a simple text command parser, which allows the reception of requests in string format (a requirement of the parent VT application, which was added later during development).

Once the bridge module produces a request that our system can process, it is then passed on to the Avatar Manager. This module coordinates between the two tutors, and expedites the requests to the appropriate character. It also has access to all the valid animation states related to the expressions and motions of the characters, and uses them to further refine the requests received from the bridge module. It then passes the relevant information to the appropriate Avatar Controller (or Controllers, since a singular animation request may need to be handled by both controllers). This module, in addition to reading and updating various parameters related to the character animations, features three submodules (discussed further ahead), that handle the three major aspects of our emotional system. The expression module handles the facial expressions of the synthetic characters. The mood module controls the idle behavior of the synthetic characters, as well as the mood the character is experiencing at any given time. Finally, the event module handles any requests for actions and reactions. Once the proper animation response to the request has been established, it is finally transmitted to the animation machine, and the proper sequence of animations is displayed to the user.

Character Animator. Our initial approach to the animator was to create a basic animation machine that would allow us to play the necessary character animations on request and, in-between, play the few idle animations at our disposal. This was a workable first iteration, but it proved lacking in many aspects. Firstly, we were very restricted by the animations we had available to us, since most emotions only had a couple of animation clips that lasted only for a few seconds. This lack of more prolonged animations caused issues when the characters became idle, since having the same animation constantly replaying itself would quickly break any semblance of credibility. Another issue was that the facial expressions did not have any motion to them, and transitioning from a character with its head in motion to a facial expression with a perfectly static head was severely jarring.

We needed a way to keep the characters in motion while the emotional expressions were conveyed to the user, so that the characters did not appear so unnatural. To solve this, we began exploring how to overlay the facial expressions on top of the already present idle animations, with the aid of Unity 3d and the layering capabilities of its Mechanim animation controller. This tool allows one to stack various animations on top of one another via the use of animation layers, and either have them effortlessly playing together, or have the desired animations take the spotlight, when relevant, and relegate the remaining ones to the background. An exposition of the aesthetics and functionality of such an animation controller can be seen in a presentation from Unite 2016[5], and a view of the animation controller we created and assigned to each of our synthetic characters can be seen in figure 3. While not as intricate as the one presented at Unite 2016, it follows some of the same concepts, and it allowed us to achieve our needs of having the characters retain their motion capabilities while still being able to express different emotions, making use of the few animations we had access to.

Expressions. The expression module was designed to fulfill any request for brief emotional expressibility that the

[5] Unite 2016 - Mecanim Bonsai: Lessons from Firewatch and ReCore. http://youtu.be/8VgQ5PpTqjc

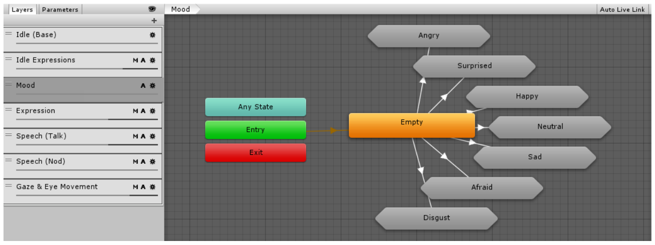


Figure 3: A view of the Mechanim animation controller, assigned to each tutor character.

tutor characters may require. The way we implemented our animation controller allowed us to overcome the limited pool of animations at our disposal. However, as development progressed, it became necessary to support a linear range of emotional expressibility, with decimal values in a scale of 0 to 1, which the early version of the module was unable to do. To solve this, we make use of the blending capabilities of Unity's Mechanim system, and implemented a system of blending trees for each emotion that needed to be expressed at variable, discrete intensities. This allowed us to fine tune the intensity of each expression, by precisely adjusting the weight of each blending variable[6].

Moods. The mood system is in charge of all the back-channeling that occurs during dialog. One of the main problems we had when implementing it was the lack of enough animation clips to cover all the six emotions. We were missing anger and disgust. After exploring the available animations, we decided our high intensity sadness idle would be best suited for the mood of anger, leaving disgust unaccounted for. Based on the work of Makarainen et al [17], we knew that it was plausible to blend two emotional expressions together. With this in mind, we took two of the existing idles and blended them with different weights, until we were left with one that resembled disgust. The next issue we wanted to solve was how to have the mood animations not interfere with the facial expressions module already in place. This is where the Mechanim approach presented in chapter 3.1 proved useful, by allowing us to keep the idle animations and expressions in separate layers, so we could easily combine the two without significant difficulty.

The mood of the character, however, is not the only concern of this module. We also need to transmit the full range of back-channeling movements and expressions, such as nods of acknowledge from the characters, eye movement associated with loss/gain of focus and the movements of the head that accompany a natural dialog. To do so, we followed the same logic as before, and added yet another set of overlaying animation layers. By adding the various eye and head animations to them, we were able to have these back-channeling animations play alongside the already existing idles. This was enough to improve how natural the conversation felt.

Events. The events module is in charge of all animations related to actions/events. Once again, we chose to segregate these animations into their own layers, so that we could play them in unison with the already present idle animations. We also implemented a simple action-reaction routine alongside the event/action animation machines, to make the conversation feel more engaged and dynamic. Additionally, to solve some overlapping issues causing the characters having very unnatural faces when talking, we later added a damping mechanism that would trigger while a character talked, and allowed the animations to blend smoothly with the mouth movements[7]. A final hurdle was making these nods and gazes adapt to the emotional state of the character. This was overcame by adding slight differentiation to the speed/frequency of the animations[8]. While these modifiers helped improve the overall feel of the conversation, no tests were conducted to ascertain whether better values could be chosen for any particular mood state. This is something that would be of great value as part of any future work done that focuses on this system.

4. Evaluation

We conducted two experiments to evaluate our implementation. The first was aimed at understanding whether the changes added by our module had improved the animation quality of the system. We enlisted the help fo two separate population samples. One comprised of potential average users of the application, and another containing only users we labeled as "experts". This category includes both PhD students and professors, in the area of intelligent virtual and robotic agents. The intent was to analyze whether the data collected from both populations would show any meaningful discrepancies between the two.

The second experiment focused heavily on emotional recognition, given how important it was for us to understand whether the participants were capable of distinguishing the emotions being transmitted by the application. Akin to the first experiment, we gathered two separate population samples. One group would experience an earlier version of the application, while the other would experience a more finalized version. This would allow us to investigate the positive/negative impact our changes had on the emotional fidelity of the application.

Initial "A" Version. This version represents the application at early stages of development. The system features two 3D synthetic characters, and an animator capable of displaying the animations provided by the Didimo team. Dialog wise, it features simplistic speech balloons with minimal work in terms of shape/color and the same animation being used for every emotion. The environment was also rather simplistic, and comprised solely of two gray background.

[6] Comparison of the result of setting a blend value of 0, 0.5 and 1: https://drive.google.com/file/d/1WXSUDxgQbIdW-HlgrBg4mvcSI8r5cxmL

[7] Difference in talking animations with and without damping: https://drive.google.com/file/d/1BWgsX87uebjtFYmOBkTM8wYyGVHjNQ8V

[8] Difference in Speed/frequency of movements between the moods of happiness and sadness: https://drive.google.com/file/d/1DgTfPPFTKWUae-yKQvDZruXtWkhpRdnD

Final “B” Version. This version supports our full animation system, described in the previous section, capable of smoother transitions and varying intensities. It also gives our characters the ability to express emotions while talking and reacting to events. The speech balloons were completely revised, and feature distinct colors, shapes and animations for different emotions and intensities. The background environments now also assist in conveying the mood of the characters, and in the portrayal of the conversation topic.

4.1. First Experiment

Multiple questionnaires[9] were created, with randomized orders for the presentation of emotions. This was done to minimize bias in the answers. Each question featured two videos; one for version “A” and another for version “B” (their order of appearance was also randomized). Users were asked to evaluate the success of each animation in communicating the specified emotion, on a scale from 1 to 7, with 1 being the lowest rank.

Demographic wise, we collected a total of ten responses for the non-expert population, with five belonging to female participants. The average age was 27 years old, ranging from as early as 23, to as late as 40 years of age. The expert population features equal response totals, and similar distribution of sex and age. The age average this time around was 29, with a range of values going from 22 to 36 years of age. All together, the data for each population totals 130 individual comparisons between both versions of the application (13 questions times 10 participants), and 260 individual visualizations (2 per question).

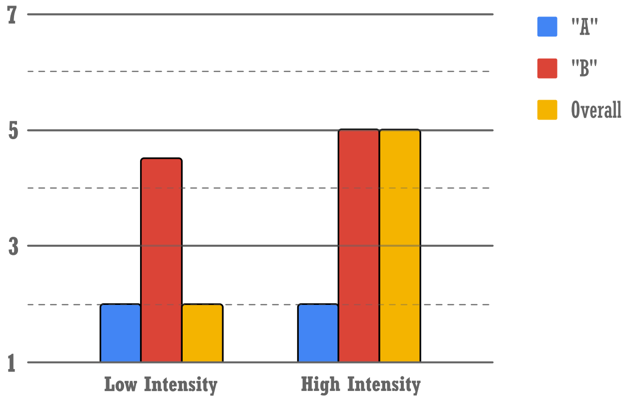


Figure 4: Overall (combined) average scores, of expert and non-expert populations, representing how well each application version conveys emotion

Results. Looking at the median scores we obtained from the non-expert condition, we can observe that the participants found version “B” to be better at conveying the specified emotions. A Wilcoxon signed-rank test backs up this assumption, showing that “B” had statistically better results when compared with “A” (Z = -6.771, p < 0.0005). On a 1 to 7 scale, we have a median value of 3 for version “A” (regardless of emotion intensity), while version “B” in turn features a median value of 5. If one was to take into account emotional intensity, the disparity in median values remains, albeit the gap between the medians of high intensity emotions widens by a small amount.

For the expert population, we can observe a median value of 2 for version “A” (when ignoring emotion intensity). This equates to a small drop of one when compared with the medians of the non-expert sample. Meanwhile, version “B” features an identical median value of 5. If one again takes 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. If we combine the data from both populations (figure 4), it is clear that a preference for the final version of the application exists, irregardless of the participants’ expertise levels. Additionally, we can also see that the scores given to the low intensity emotions were much lower than those given to the high intensity ones. This seems to indicate that lowering the intensity of certain emotions makes it harder for users to perceive the key details that make said emotions distinctive.

4.2. Second Experiment

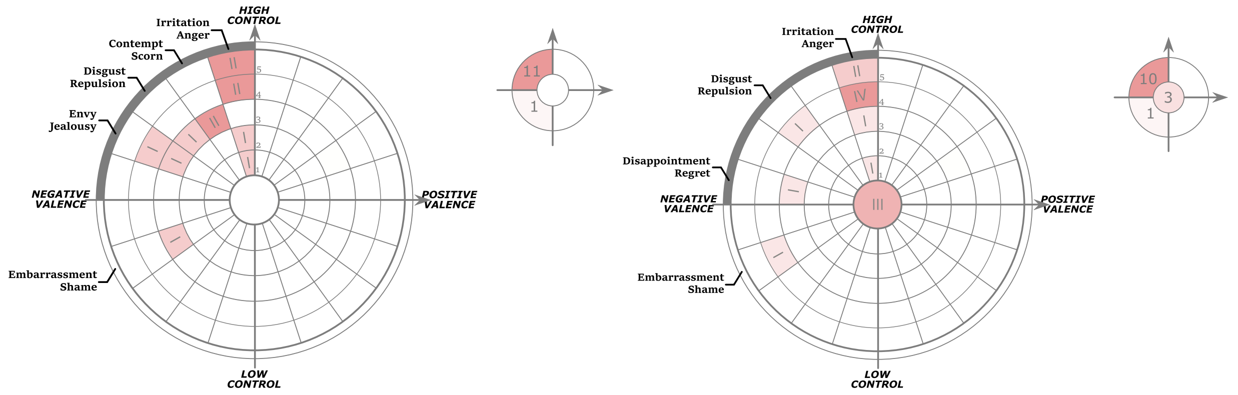
Similarly to the earlier experiment, multiple questionnaires[10] were created, again featuring randomized orders for the presentation of emotions. We made sure to guarantee that each emotion (and intensity) was shown a similar number of times to the participants. For every question, they were asked to identify which emotion they believed was being represented, its intensity and what aspects (characters/ balloons/environment) influenced their decision. Regarding demographic data, we have a total of twenty-six respondents to our survey on version “B”, which results in a total number of videos watched of around 170. For version “A”, we ended up gathering a total of twelve participants, which translates to a sample size of watched videos of around 156.

In regards to demographics information from the participants, akin to the first experiment, the expected low sample size would not have allowed us to perform any tests that would yield statistically significant results, and thus we decided to forgo the collection of said data altogether for this second experiment.

Results. By comparing the new data with the results of the previous experiment, we are able to identify emotions that were perceived as better animated, but also suffer from poor emotional perception. The emotion of sadness was perceived as better animated in version “B”, but exhibits lower perception values. Specifically, the low intensity expression of sadness had poor accuracy in both versions, with version

[9] https://goo.gl/PmUztE

[10] https://goo.gl/8ZaZT6



(a) Heatmap for version "A"

(b) Heatmap for version "B"

Figure 5: Heatmap for the emotion of anger (low intensity), for the listed application versions.
For the bigger wheel, the roman numerals indicate the amount 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 mirrors the larger one, but instead indicates the total number of emotions recognized in each quadrant

TABLE 1: Overlap (blue background) of Bassili's findings and the emotional recognition mistakes of version "B"

Could Be confused with ...

	Happiness	Surprise	Sadness	Fear	Disgust	Anger	Happiness	Surprise	Sadness	Fear	Disgust	Anger
Happiness	■	1		1			■					
Surprise		■		1		1		■				1
Sadness			■		-1	1	1		■	-1		
Fear	-2	3		■	-2	3		-2	2	■	-1	-1
Disgust		1		1	■	-3				2	■	-4
Anger						■			1	2	-2	■
	Low Intensity						High Intensity					

"B" having 0% of correct answers. Surprise low was also more accurately perceived in version "A", 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 (figure 5), fear high, and surprise high. The remaining emotions do not show significant signs of either better or worse emotional perception.

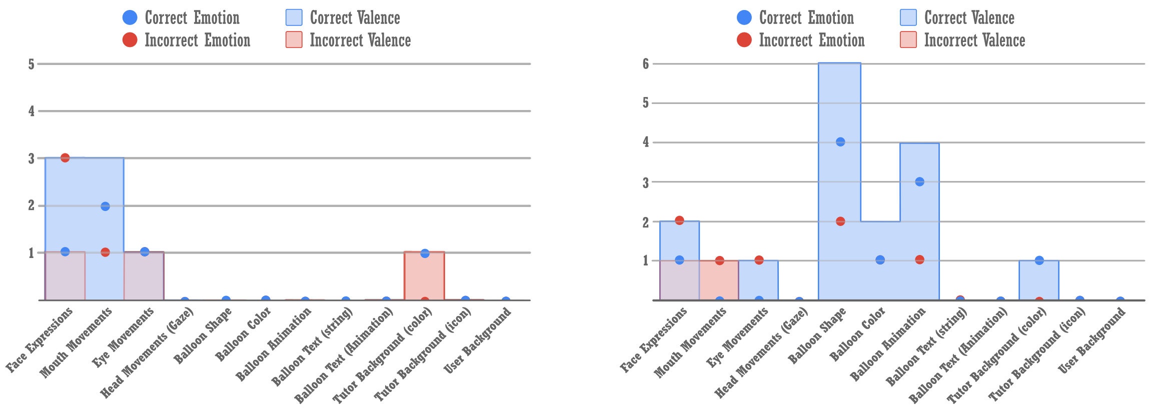
If we crosscheck our results with the findings of Bassili on emotional confusion (table 1) we can infer that most confusions falling outside said findings were corrected in version "B", and the remaining cases exhibit only singular occurrences. The exception here is low intensity fear, which began being mistakenly recognized as anger.

In terms of the what aspects influenced these perceptions, we asked the participants to indicate which of the primary emotional channels (characters, balloons or environment) most influenced their choice of which emotion they recognized from the videos. We can observe a relative split between the characters and the 2D elements (balloons and environment). Characters are 1.45 times more influential than the balloons, and 5.42 times more than the environment. Balloons themselves are 3.55 times more influential than the environment. The more prominent specific aspects (in regards to 3D and 2D elements) are, respectively, the facial expressions (29.02%), followed by the shape of the speech balloons (18.13%). It is worth noting that aspects such as the character gaze, balloon strings and background icons, despite having lower relevance percentually, end up mostly just contributing to the existing inaccuracy, when it comes to the correct recognition of emotion.

Regardless of the overall influence, the weight of importance for each aspect seems to vary depending on what particular emotion is being observed. An example of this trend can be seen in figure 6. Both fear (high) and anger (low) are emotions that were better perceived in version "B" of the application. Fear was strongly impacted by the various aspect of the characters (gaze is absent). Balloons are completely absent, and the environment only shows up via mention of the backgrounds. On the other hand, anger is heavily influenced by the speech balloons, even if there are still some mentions of various aspects pertaining to the characters. It is worth noting that these mentions (be they character or balloon related) served to both help and mislead the users, but given that the emotional perception for these two particular cases improved, we hypothesize that their overall contribution was more positive than negative.

5. Conclusions

The intent of this work was to create an animation system, capable of being integrated into the context of a virtual tutoring application, and able to support the animation requirements necessary to achieve believable interactions between synthetic characters and a human user. To accomplish this goal, we focused on aspects such as ways to properly convey emotional expressiveness, how to present the feelings and moods of the two characters in a way that



(a) Area map for the Fear (High) emotion

(b) Area map for the Anger (Low) emotion

Figure 6: Area maps indicating what influenced the listed emotions, for version "B" of the application.
Each dot indicates the number of times each aspect contributed to the correct (blue) and incorrect (red) recognition of the emotion.
The areas indicate the number of times each aspect contributed to the correct (blue) and incorrect (red) recognition of the emotion's valence.

did not break the natural flow of a dialog and how to have the characters exhibit suitable levels of back-channeling.

We conducted two experiences to validate our solution. One aimed at studying whether any animation improvements had occurred by comparison with an earlier, pre-development application version. The results showed our finalized system was recognized as better animated than in the previous version, when compared directly. This holds true after polling both regular users, and users who are familiar with the relevant field of expertise. The following experiment aimed to ascertain the quality of our solution, in regards to emotion recognition. Results showed that, despite our best efforts, some emotions were not being recognized as frequently as we would have hoped, while the recognition levels of others showed improvement. Regarding intensity, we observed that more intense expressions of emotion were more recognizable than subtler ones. At the same time, our changes caused all emotions to be seen as more intense even when presented with relatively low intensity values, indicating that further work is required to regain the distinction between slight and strong displays of emotion, in a way that would not compromise the current levels of emotional recognition accuracy.

Regarding our approach, the idea of blending animations via layering, while functional, is not ideal for animations that affect the same region of a character model. Combining a smile and an open mouth, given that both affect the mouth area, proved to be tricky. We hypothesize that minimizing the amount of collision by curating the animation clips that are used by the controller could, to a certain extent provide a more refined outcome. The attenuation mechanisms added to the animations during conversation proved beneficial, but they too could be further polished. Current hypotheses is that having separate damping values for different emotional states would be beneficial in treating aggressive animations differently than softer, less prominent ones. The role of gaze and head movements in conversations is another aspect to refine, given their current negative impact on emotion recognition. While important for achieving a natural dialog, these should not sabotage the expressiveness of other emotional channels. Lastly, the current use of variable parameters to tweak the intensity of animations and movements may not be the ideal solution since, as pointed out, it blurs the differences between higher and lower intensity displays of emotion. When adjusting aspects such as frequency and speed, one should test the use of different values that adapt to particular ranges of emotional expressions, thus hopefully finding those that work best for a particular mood and intensity state.

Overall, we feel synthetic characters play a major role in conveying emotion to the audience, even when paired with other channels, and how much their contribute seems to vary depending on the situation. In stronger emotions, such as anger, the iconic colors and blunt shapes of the accompanying speech balloons allowed them to become the primary channel. Regardless, this partnership of emotional avenues shows potential as a way to have a system complement the shortcomings of another, but still has issues with channel bias, perhaps more so than those that occur when one mixes visual and audio channels. Further improvements need to be considered for the system as a whole (tutoring application combining 3D and 2D animated elements), in order to truly achieve the end result: Two virtual characters having a believable dialog with a human counterpart.

Acknowledgments

The author would like to thank his thesis advisor professor Carlos António Roque Martinho, for his guidance during the course of this thesis, as well as his colleges Ricardo Pereira and Ricardo Rodrigues, for their help and assistance.

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with ref. UID/CEC/50021/2013, and FCT grant from project Tutoria Virtual with ref. TDC/IVCPEC/3963/2014.

References

[1] James Cameron. Avatar: Motion capture mirrors emotions.

[2] W Scott Reilly. Believable social and emotional agents. Technical report, Carnegie Mellon University. Pittsburgh, PA. Department of Computer Science, 1996.

[3] Ana Paiva, Joao Dias, Daniel Sobral, Ruth Aylett, Sarah Woods, Lynne Hall, and Carsten Zoll. Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence*, 19(3-4):235–266, 2005.

[4] Ollie Johnston and Frank Thomas. *The illusion of life: Disney animation*. Disney Editions, 1981.

[5] Aaron B Loyall. Believable agents: Building interactive personalities. Technical report, Carnegie Mellon University. Pittsburgh, PA. Department of Computer Science, 1997.

[6] Paul Ekman and Wallace V Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk, 2003.

[7] Paul Ekman. *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Macmillan, 2007.

[8] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[9] Peter Salovey and John D. Mayer. Emotional Intelligence. *Imagination, Cognition and Personality*, 9(3):185–211, 03 1990.

[10] John 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*, 37(11):2049–2058, 1979.

[11] 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*, 11(5):843–855, 08 2009.

[12] Jun'ichiro Seyama and Ruth S. Nagayama. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16(4):337–351, 2007.

[13] Andrew Ortony. On making believable emotional agents believable. *Trappl et al.(Eds.)*, pages 189–211, 2002.

[14] ACM. *Principles of traditional animation applied to 3D computer animation*, volume 21, 1987.

[15] Michael Lewis, Jeannette M Haviland-Jones, and Lisa Feldman Barrett. *Handbook of emotions*. Guilford Press, 2010.

[16] Ali Arya, Steve DiPaola, and Avi Parush. Perceptually valid facial expressions for character-based applications. *International Journal of Computer Games Technology*, 2009, 2009.

[17] Meeri Mäkäräinen and Tapio Takala. An approach for creating and blending synthetic facial expressions of emotion. In *Intelligent Virtual Agents*, pages 243–249. Springer, 2009.

[18] Jina Lee, Zhiyang Wang, and Stacy Marsella. Evaluating models of speaker head nods for virtual agents. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 1257–1264. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[19] RM Maatman, Jonathan Gratch, and Stacy Marsella. Natural behavior of a listening agent. In *International Workshop on Intelligent Virtual Agents*, pages 25–36. Springer, 2005.

[20] Elisabetta Bevacqua, Maurizio Mancini, and Catherine Pelachaud. A listening agent exhibiting variable behaviour. In *Intelligent Virtual Agents*, pages 262–269. Springer, 2008.

[21] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. Designing effective gaze mechanisms for virtual agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 705–714. ACM, 2012.

[22] Brent Lance and Stacy C Marsella. Emotionally expressive head and body movement during gaze shifts. In *International Workshop on Intelligent Virtual Agents*, pages 72–85. Springer, 2007.

[23] Carlos Busso, Zhigang Deng, Michael Grimm, Ulrich Neumann, and Shrikanth Narayanan. Rigid head motion in expressive speech animation: Analysis and synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1075–1086, 2007.

[24] Sin-Hwa Kang, Andrew W Feng, Anton Leuski, Dan Casas, and Ari Shapiro. The effect of an animated virtual character on mobile chat interactions. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, pages 105–112. ACM, 2015.

[25] Jonathan Gratch, Ning Wang, Anna Okhmatovskaia, Francois Lamothe, Mathieu Morales, Rick J van der Werf, and Louis-Philippe Morency. Can virtual humans be more engaging than real ones? In *International Conference on Human-Computer Interaction*, pages 286–297. Springer, 2007.

[26] André Lima. Virtual tutor - virtual tutoring agent using empathy and rapport techniques. Master's thesis, Instituto Superior Técnico, 2017.