Expressing emotions through animated speech balloons

Ricardo Pereira
Instituto Superior Técnico, Lisboa, Portugal
Email: ricardo.g.pereira@ist.utl.pt

***Abstract*—Nowadays, digital scenes, like games, tend to mostly use model animation, like facial expressions, and spoken dialogue, to convey emotion. However, those approaches are not always possible. This work explores the current approaches to text animation patterns, speech bubbles and background shapes and color, in order to convey emotion. It also aims to create a module that can control these aspects, and integrate it within a Virtual Tutoring application, which also utilizes model animation. We also present experiments, both comparing the existing base version, and final version of the application, and testing the emotional perception of each version, in order to understand if our approaches were correctly implemented. It was possible to observe that the addition of such details led the system to be recognized as better animated, when compared with the earlier version, however it when tested alone, some emotions ended up being less recognized, while others had a better impact on the application. In the end, we observe that such details can help improve the emotional recognition of a scene, however the additional information on one channel can create noise and bias on the other emotional channels, leading to worse results in certain occasions. Nonetheless, they can and should be used in order to reinforce emotional recognition, without creating contrast with the leading emotional channel.**

1. Introduction

Imagine you are interacting with emphatic virtual agents through a smartphone that allows video, sound and text. Now consider that the conversation is taking place in a loud place, in which is impossible to hear properly. Since the conversation is being held through a phone, the expressiveness of such agents may not be fully perceptible. Therefore, in this scenario, text needs to convey sufficient information to the user. Unfortunately, text by itself does not give emotional context, and for that reason, it is helpful to create mechanisms that can help the user understand and convey such emotions in this non-verbal communication channel.

Text was always used as a way of communication, and grew even more with the expansion of the digital era. However, unlike presential and verbal conversation, where gesture, utterance and other subtle aspects play an important role to determine the meaning of the words, it is not as trivial to understand it in written form. It usually underlies in the chosen words, and previous knowledge and trust between the people communicating, and it still might be incorrectly understood. Emotions take a great place in every form of communication, and the fact that there is no correct away to express it through text, makes it a problem worth studying. There are multiple techniques to be explored, like kinetic typography[1], an animation technique that mixes motion and text in order to express ideas, using video animation, and the use of speech bubbles, which are used within comics and can also provide emotional context. Throughout this document, when we mention 2D elements, we are specifically talking about speech bubbles, text, and backgrounds.

Our work will be focused on two aspects: create an independent module that can be integrated in the Virtual Tutoring application (system to help students adapt to the university, by utilizing empathic strategies), allowing the system to easily control the used elements; and add an emotional channel to the scene, which complements the characters expressions, and allows for better emotion recognition. Our hypothesis is that the animation, the speech bubble, and the background can help to convey emotion to the user, and help him to better perceive the world. For this, we will take inspiration from comics and animation, in order to improve the written dialogue between synthetic characters and the user.

In this document, we will start by describing the state of the art, and explain our approach and implementation. We will also present the two experiments performed, and the corresponding results. Finally, we will conclude with the most relevant details and their importance.

2. Related Work

Theoretical and practical work in areas such as Psychology, Human-Computer Interaction, Animation, and Computer Graphics, will be taken into account and studied accordingly. Other, more abstract areas will also be studied, like the art of comic books and art in general, due to their relation to this specific work.

Emotional Intelligence

Emotional intelligence has been defined in multiple ways. The most common are the ability to understand emotions in others and in oneself, and the capacity to use such information to adapt to certain environments [1]. In our work, we are interested in transmitting emotions, and having

1. http://kinetictypography.dreshfield.com/

them being properly recognized. Therefore it is important to understand how emotions are recognized by different people.

Studies found that certain characteristics, like sex and age, influence emotion recognition. When it comes to age, one study by Mill et al. found that the correct recognition of negative valences decreases with age, starting at the age of 30 years old [2]. This decline appears in both vocal and facial modalities. However, this is not true for all emotions. Although being able to perceive emotions, neither younger nor older adults are capable of recognizing emotions with 100% accuracy, most falling between the range of 40% to 90%. This is important, because it means that emotion perception varies between emotions and different people.

There are two most concerning problems that lead to inaccuracies when it comes to emotion recognition. The first issue is the existence of similar movements across multiple emotions. This ambiguity can lead to confusion, resulting in the incorrect perception [3]. Another issue is that different channels and modalities carry different information, which can lead to inconsistencies, if multiple are used. Mower et al. showed evidence that adding a second channel can create channel bias, which can lead to the incorrect perception of emotion [4]. Channel bias appears when one channel is more predominant than another, and therefore captures the receiver's attention, or when two channels transmit contradictory cues. Studies that utilize two channels, tend to focus on different modalities, like facial and auditorial. In our work, we join two visual channels, which can lead to inaccuracies.

Color

Kaya et al. studied the property of color and its association with emotion. The main finding is the fact that the symbolism is dependent on each individual, and which things, objects or physical space they associate each color with [5], [6]. The authors examined how ninety-eight college students viewed several colors. For that, thirteen colors were chosen, divided in three groups - principle hues, intermediate hues, and achromatic - each having less positive emotional responses than the previous ones. Some of the strongest color-emotion associations were green with relaxation and green-yellow with sickness. Unfortunately, these associations change from person to person, and are dependent on their personal experience, and social and cultural background, like age, gender, and nationality.

Artists have explored colors to express emotions for years. The main hypothesis described by Melo et al. is that humans find analogies between such properties, and internal and external manifestations when experiencing such emotions. Despite being a complex topic and being difficult to properly use, the richness of their symbolism makes color a great property to manipulate, in order to convey certain emotions [7].

Motion

Motion is the change, in position, of an object over time. It is separated from animation, due to the fact that it only cares about those changes, like speed, amplitude, direction, fluidity, acceleration, shape, angles, and path. It is therefore, a property used by animation. Bartram et al. studied the properties of abstract movement in order to understand what certain types of motions are perceived as. For this, they generated certain common motions, like linear and spiral, and had more complex motions captured while performers enacted certain emotions. After defining and dividing a set of different and abstract motions as small and big motions, they had a group of participants defining them in terms of valence, arousal, dominance, emotion, and abstraction [8].

Their results showed that positive motions are usually larger in amplitude, have few accelerations, and a curvy shape, while negative motions are faster, not smooth, have more obtuse angles, and an angular shape. Finally, calm motions are slow, curvy, and have more decelerations. Depending on whether the motion is big or small, certain positions on screen were also related to certain valences.

Animation

There are twelve known basic principles of animation, and they were introduced by Johnston & Thomas in 1981 in the book The Illusion of Life: Disney Animation [9]. Despite being usually used in character and object animation, some properties are important to this work, specifically in the animation of our speech bubbles. The most relevant principles for this work are: Squash and Stretch, Follow Through and Overlapping Action, and Exaggeration.

There are studies that utilized these principles in order to create better animations and convey emotions. Pires et al. created a tool that uses autonomous synthetic characters to stimulate idea generation during a storytelling activity [10]. To address emotional expression in the creation of their agents, they followed four different means of expression - movement, color, sound, and proxemics. When it comes to movement, they combined the twelve principles of animation, and Darwin's observations regarding movement and posture in emotional expression - stretching the body and mimicking an inflated chest while leaning forward conveys anger, while squashing and tilting down while staying motionless conveys sadness.

Kinetic Typography

Another type of animation is kinetic typography, which can be understood as a communicative medium that adds some of the expressive properties of film to static text [11]. Studies that developed such animations and tested them with users, realized that one of the best ways to convey emotion is by imitating body and sound movements. By doing this, a person can associate specific movements with his own and the particular emotion he had, understanding the text meaning more correctly. Malik et al. started with an

inanimate neutral sentence, and had a design team creating several animations, with varying intensities, and compared how such animations were perceived by the users. Their results were that specific motions, such as shaking, twisting, fading, bouncing, looping, jittery, and flashing, would convey emotion regardless of the content of the sentence, depending on the intensity [12]. Text animation can be used for different things, such as capturing and directing attention, creating characters and expressing emotions.

Out of the two main animations, Gaylord et al. found that gestural animations (mimic the human body) have more consensus than inflectional animations (mimic the modulation of the voice) [13]. Their hypothesis was that gestural animations translate visual gestures into visual movements, while inflectional animations translate auditory qualities, which is fundamentally more subjective. Nonetheless, both are important and should be used, or even combined, in certain situations.

It is important to notice that kinetic typography cannot normally replace or override the intrinsic emotive content of a sentence. Instead, it should be used to reinforce it. In addition, it is also successful in portraying characters, by adding identifiable, distinct and persistent properties and attaching them to the character. Finally, if used properly, it can grab the users' attention and direct it to certain aspects.

Speech Bubbles and Background

The last elements to consider are the balloons and backgrounds. Balloons have been used in comics for a long time to specify which character is speaking and how they are talking. However, each artist creates their own convention in order to convey the physical or emotional state of the character [14]. Despite the diversity, there are a few speech bubbles[2] that are commonly used in the art form - speech, whisper, thought and scream. If properly used, these types of balloons can be created and used as another possibility of conveying specific emotions to the user.

The same statements apply to the use of background. In comics, the background of a scene is also constantly used to specify the state of mind of the character. This comes from the use of color, icons, or even the combination of both. Such artistic details can be seen in multiple comics or any type of visual art, like paintings. Even the most abstract style, such as abstract expressionism, may give the viewer a particular state of mind, or trigger some emotion. However, the more ambiguous and less known the shape is, the less meaning users can read into it.

3. Implementation

Virtual Tutoring Architecture

To summarize the Virtual Tutoring architecture, the application will start by calculating an affective appraisal based on an interaction history, which takes into account the students' preferences and emotional history, interaction choices, and intentions. This will allow the calculation of the student and tutors' affective state. Given the student affective state, an empathic strategy is selected. This strategy will generate intentions, eg. make a study plan, that will be expressed in form of a dialogue and options. From then on, and using the tutors' affective state, the resulting facial and text elements are updated accordingly and shown to the user. All these steps happen in a loop and can change with the user input.

Related to our work, this system was built into an application to be used by users, and a previewer version, which allowed creators to define and test their dialogues. Our module is integrated in both versions.

Bubble System Architecture

When receiving the data from the Dialogue Manager, the Bubble System Manager can update the background for a specific tutor, and show or hide balloons. When updating the background, it receives the tutor, the emotions and intensities, and the reason for the emotion, and then sends this information to the Background Manager. There, the texture and color will change, if a new reason and emotion is given. When updating balloons, it receives the tutor, the emotions and intensities, the text, and the effects to be used when showing and hiding the text. With this information, The Balloon Manager sets the text and its effects, and the sprite, color and animation of the balloon. The architecture can be found in figure 1.

All this data, like emotion colors and effects, balloon positions, and reason icons, are predefined. A lighter intensity emotion will have its color and text effects attenuated. The text color also changes between black and white, according to the balloon color, to allow readability. Finally, to create a more dynamic scene, the first tutor to speak has its balloon on the top part of the screen. Therefore, the positions swap accordingly. Nonetheless, to allow change, the users creating dialog trees, have multiple commands that can set this information.

Emotional perception approach

Due to the high diversity of emotions, we only focused on the six fundamental emotions of Ekman [15] - Anger, Disgust, Fear, Happiness, Sadness, and Surprise - with the possibility of different intensities.

Balloons

Starting with the balloons, our objective was to create one speech bubble for each one of the emotions. For that, we looked at balloon art, and how they are drawn[3]. Some were trivial to choose, like the neutral balloon (rectangular

2. https://en.wikipedia.org/wiki/Speech_balloon

3. https://www.deviantart.com/youthedesigner/art/30-Hand-Drawn-Speech-Bubble-Photoshop-Brushes-347817364

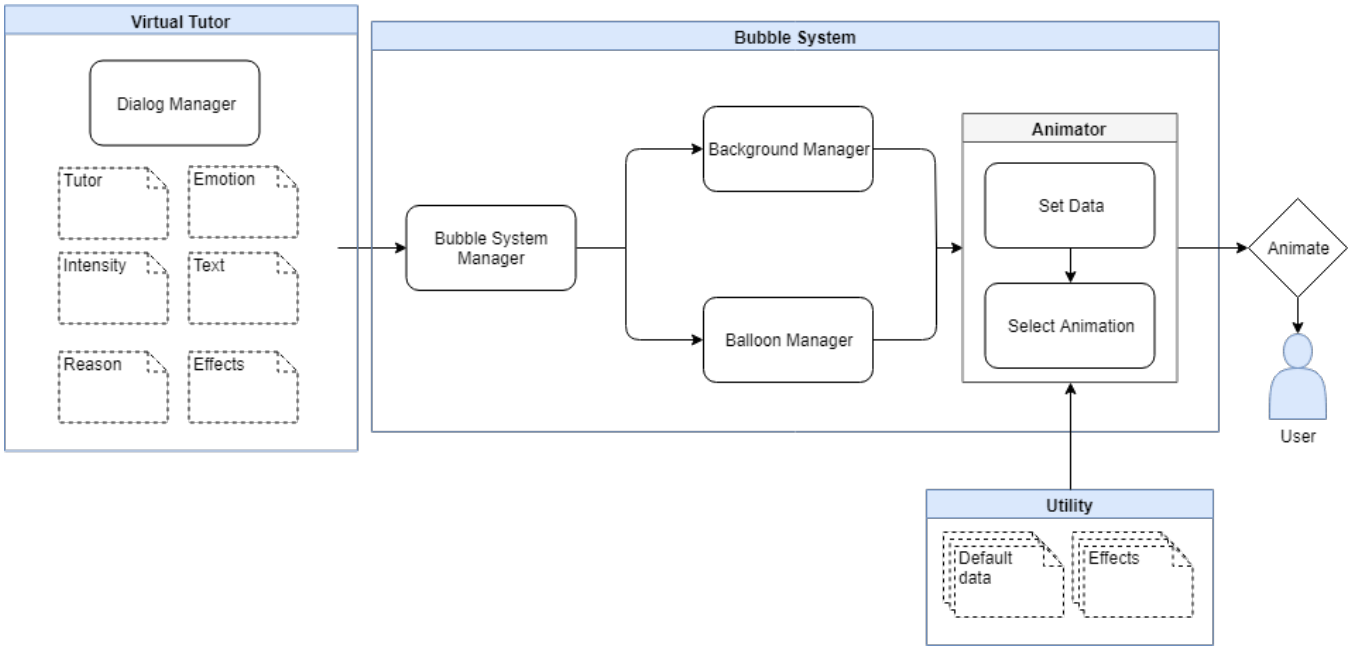


Figure 1: Bubble System Architecture

shape), for non emotional speeches, and the anger balloon (spiked), due to the fact that they are a standard in comic art. However, the rest was a little more abstract, so it came down to a group decision based on the motion they conveyed.

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 straight lines could potentially create the idea of aggressiveness. Since surprise can either be a positive or negative feeling, we chose a mixture between balloons that had such valences. Therefore, it became similar to the anger balloon, but with less and more curved spikes. Disgust, as it is usually caused by sickness, 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. All the balloons tails, followed the same ideas, with ones being more curved, and others more spiked, to complement the valence that we wanted to transmit. All the balloons created can be found in figure 2.

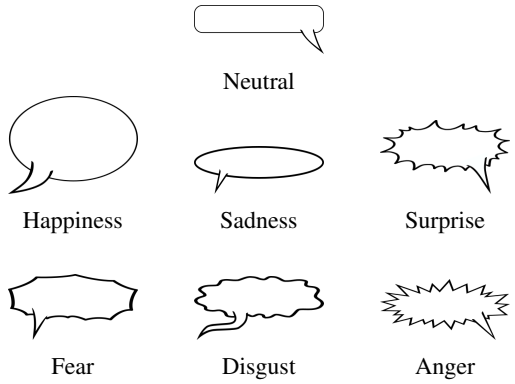


Figure 2: Palette of speech bubbles

Color

Since the association between colors and emotions changes depending on the social and cultural background of each person, it was not an easy decision to make. For that, we decided to use the same approach as Pires et al. [10], and get inspiration from Disney™ movie Inside Out. Since Ekman was one of the scientific consultants for this movie, and each character is identified with a color, representing an emotion, we decided to use the same idea. Anger was represented with red, happiness with yellow, disgust with yellow-green, sadness with blue, and fear with purple. For surprise, we thought upon the idea of the emotion itself. Surprise is an emotion that is usually positive, despite being possibly considered as negative. Moreover it is usually similar to fear, with surprise being positive. Since fear is a mix between red and blue, which in our case is anger and sadness, we decided to mix anger and happiness, to have the aggression of red and the positive valence of yellow. Therefore, surprise ended up being defined as orange. Finally, to be shown as neutral, we decided upon a mix between green and blue, since it is usually seen as relaxed. It is important to note that the movie colors were only used as a guideline, which means that the exact same values were not used.

Throughout the development, there was an important focus on having emotional intensity. In order to achieve this, all details that can be animated had a parameter that could

control its strength. In simple terms, balloon animations had a speed parameter that could control its speed; text animations, depending on the effect, had parameters that allowed to control the speed, or even frequency of such animation; and finally, colors were also given the possibility to be attenuated based on the intensity value.

Balloon Animation

When it comes to animation, there are numerous properties that can be changed over time and in multiple ways. For instance, text has a position and can have capitalized characters, while the balloon and background are images that have a position and offset. Both can be transformed geometrically - scaled, translated, rotated - and have colors. All of these attributes - position, offset, color and transformations - can be animated in time, with varying speed, duration and rate, creating distinct effects like linear, pulse, and jitter. We defined two different type of animations: hand made balloon animations, using the Unity 3D[4] animator system which allows the creation of certain behaviours more easily, and also allows blending between animations if needed; and mathematical animations based on animation curves.

For the unity system, we defined fourteen animations, two for each emotion (one to show the balloon, and one to hide it), plus two for the neutral state. The neutral animation was already in the project as a simple scale over time. For that reason, we defined the rest based on this one. In order to differentiate valences, we ended up adding different motions, at different speeds, to each emotion. Happiness was achieved with bounces, due to its use in animation; for sadness we lowered the speed of the scale, since it is also use in animation; to 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 sea sickness; and finally anger was defined with a chaotic path, to show disorder.

Text Animation

For text animations, we went with a mathematical approach over animation curves (curves that define functions) for two reasons: the text plug-in used[5] allows us to control characters vertices, which gives us more flexibility on the effects we can achieve; and we can generalize effects and apply different curves, giving different effects. All effects created can be combined together to create more complex animations, however not all combinations will work. For instance, a combination of fade out and fade in is incoherent, therefore only the latter will be used.

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 jitter; disgust was created as a warp over a curve, which is just a displacement; and finally anger was defined with a shake. Options balloons, on the other hand, do not have an inherent emotion, therefore do not use any effects.

Icons

Finally, we had to create icons to be used by the background, and animate the corresponding planes. For abstract backgrounds, we wanted any image with any unrecognizable pattern, as long as we could define its color. For icons, which needed to give some sort of context, we had to be more careful with the design. The Virtual Tutoring application applies six metrics, in order to decide how applied the student is, and how everything is going. Three of such metrics are subjective (qualitative) and provided by the user, whilst the others are objective (quantitative), and depend on the student's results. For challenge, we decided upon a figure trying to reach the summit of a mountain; for enjoyment we used the sock and buskin (symbols for comedy and tragedy); importance was defined as a podium with an exclamation mark, to transmit the idea of a hierarchy towards a goal; performance was designed as a speedometer; effort as Sisyphus (figure pushing a rock over a slope); and finally engagement, since it is how regularly tasks are performed, was designed as a task list being filled. All icons can be found in figure 3. The backgrounds, being images were also given animations, mostly to change between them, and their respective color.

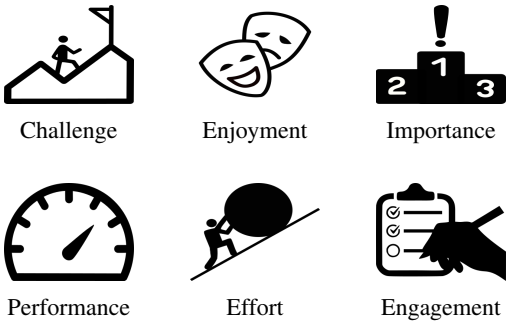


Figure 3: Palette of icons

4. Evaluation

To analyze the effects of our work, we organized two separate experiments. The objective of the first experiment was figuring out if the new system was perceived as better animated than the initial one, regardless of emotional recognition. We performed the experiment with two separate population samples, in order to study the difference in answers between experts and non expert participants. In the expert category, we included PhD students and professors who worked in the area of intelligent virtual and

4. The application is implemented with Unity 3D Engine: https://unity3d.com

5. https://assetstore.unity.com/packages/essentials/beta-projects/textmesh-pro-84126

robotic agents. In the non expert category, we included potential (non-expert) users of the final application. It is worth mentioning that the results of this experiment only explain if we managed to improve the emotional display of the application, when compared with the earlier version, and not whether participants would be able to distinguish the emotions being expressed in a less restrictive context.

Given the importance of emotional recognition, it was essential to also conduct another experiment, in order to understand if the participants were actually capable of distinguishing the emotions being transmitted by both the initial and final versions of the application. Akin to the first experiment, two conditions were introduced that lead to two separate population samples. One group would view an earlier version of the application, while the other would view a more finalized version. With this separation, we hoped to understand if the application was giving the correct emotional perception, as well as understand if there were improvements between the two versions.

Initial "A" Version. Version "A" represents the initial prototype of the application. When it comes to the bubble system, this version had a very simplistic module working, with the same balloons, animations and colors being used for every emotion. For the background, there was only a gray layout behind the tutors. The avatar system featured two tutors as 3D synthetic characters, with a small amount of facial animations, representing emotions.

Final "B" Version. This version supports a much more complex animation system, capable of smoother transitions and with the capacity for dealing with emotional expressions of various intensities. The improved animation system also gives the tutors the ability to express emotions while talking and reacting to events, something which was not previously possible. The speech balloons were completely revised, and the application now features balloons with distinct colors, shapes and animations for different emotions, as well as animated text and the capability of supporting various emotional intensities. Environment wise, a background display was added for each tutor, to further help in conveying their current emotional state, as well as assist in portraying the conversation topic to the user. A snapshot of the representation of anger, with high intensity, can be found in figure 4.

Figure 4: "B" anger high emotion

First Experiment

Questionnaire. Multiple questionnaires[6] were created, and the order by which each emotion and video were presented to the participants was randomized, in order to minimize bias. For every question, each participant was shown two videos, depicting the same emotion for both versions, and was asked to rate the success of each animation in communicating the specified emotion, on a scale from 1 to 7, going from lowest to highest.

Demographic Data. For the first condition (non-expert participants), we had ten responses, with half being from male participants. The age of the participants ranged from 23 to 40 years old, giving an average of 27 years old. For the second condition (expert participants), after collecting all the data, we had a total of ten responses, with an even split between male and female participants. The average of ages was 29 years old, with a range of values that went from 22 to 36 years of age.

All together, this equates to a total of 260 watched videos, and 130 comparisons between different versions of the application, for each condition.

Second Experiment

Questionnaire. Multiple questionnaires[7] were randomized, in both the order the videos were shown and which emotions were selected to be part of the questionnaire itself, in an attempt to minimize bias. This was done in a way that assured us that every emotion was visualized about the same number of times. For every question, each user was asked to identify which emotion they believed was being represented; how intense that emotion was; and what aspects of the scene influenced their decision.

Demographic Data. For version "B", we ended up with twenty-six answers, which translates to 170 watched videos. 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, we expected a low number of respondents, which would now allow us to perform any significant statistical tests. Due to this reason, we decided not to collect data for this experiment.

Results

First Experiment. Looking at the median of the scores we obtained from the non-expert condition, we can verify 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" ($Z = -6.771$, $p < 0.0005$).

6. https://goo.gl/PmUztE
7. https://goo.gl/8ZaZT6

On a scale of 1 to 7, we have a median value of 3 for version "A", when considering both low and high intensity emotions together, while version "B", in turn, features a median value of 5. If one was to take into account the intensity of the emotion, the disparity in median values remains, albeit the gap between the median scores of the emotions with high intensity does widen by a small amount.

For the expert condition, if we join intensities, we can see that, on a scale from 1 to 7, version "A" had an overall median score of 2, which is less than the previous condition, while version "B" remained with a median score of 5. By separating intensities, we can observe that the tendency remains, with high intensities having a slight accentuation.

From this analysis, we can observe the preference for the final version of the application, when compared with the initial version.

Second Experiment. If we compare the results we obtained from both experiments, there are emotions that were perceived as better animated in the first experiment, and had low emotional perception in the second experiment. If we join experts and non experts, we can see that the emotion sadness, disregarding intensity, fell in this category. Despite the 0% accuracy, it was perceived better in version "B". Another emotion that followed this tendency was surprise low. Both versions heatmaps can be seen in figure 5.

However, there are emotions that were better recognized in version "B", while also being considered better at conveying emotion when compared with version "A", such as: fear high, surprise high, and anger low.

Looking at anger with low intensity, we can observe that version "B" had better emotional accuracy than version "A". Most incorrect answers were perceived as neutral or within the right quadrant in version "B", while version "A", on the other hand, had the responses more dispersed in the same quadrant, despite having less emotional precision. Anger low heatmaps can be seen in figure 6. Finally, happiness high, and happiness low were recognized similarly between versions, despite being recognized as better animated, when compared directly.

Using the work of Bassili [3], we can also create a confusion table, depicted in table 1, in order to compare each emotion and intensity, between versions, with their corresponding mistakes. For this analysis, we will only focus on the six basic emotions, discarding the rest. The numbers signify the difference between mistakes from version "A" to version "B". This means that negative numbers show the amount of mistakes reduced in version "B", when compared to version "A", while positive numbers represent the opposite. For this analysis, we will discard the cases with only one mistake.

If we focus on high intensities, we can see that we reduced or even eliminated most confusions. Some confusions still occur, like disgust and anger being similar to fear, and fear being recognized as sadness. Nonetheless, these mistakes are, according to Bassili [3], common (represented in a blue background). Low intensities also follow this tendency, with almost all mistakes being reduced or eliminated. However, fear was also recognized as surprise and anger, with the latter being uncommon. The majority of mistakes that were not resolved had at most one occurrence, which is not significant. Therefore, overall, we were able to reduce or remove most confusions that are considered abnormal.

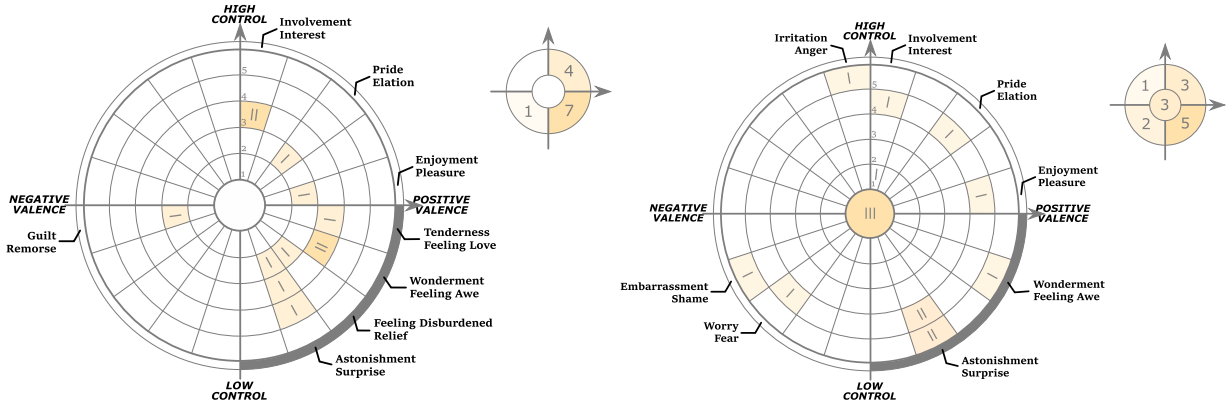
Factors influencing emotion recognition. Focusing on the influences that led to this perception, when it comes to version "B", we can observe that the expressions, mouth and eyes of characters, and the shape of balloons contributed the most, while in version "A" the focus was mainly on the characters. However, other details also make an appearance in the final version, like the gaze of the characters, the background icons, and the text shown by the balloons. This could signify that these aspects need to be handled with care, since despite contributing less than other aspects, they seem to be a source of misdirection when it comes to reaching the desired answer. Furthermore, the string displayed by the balloons, which was not supposed to influence the results, appears to have had the biggest influence out of these three aspects. This confirms the importance of context and appropriately defined dialogues, due to the cultural and emotional meaning behind certain sentences.

There are emotions, in which 2D elements tend to be negligent, while in others they end up being important for the perception. For instance, in surprise with low intensity, the importance of the characters remained between both versions, with balloons only having little and insignificant impact on version "B". The characters had an identical weight for all details, with the only difference being the absence of gaze and the misleading of the face expressions in the final version.

If one compares the fear low intensity charts, depicted in figure 7a, we can see that the balloons had more positive impact in the final version of the application, which can be seen in the shape, color, and animations. We can also observe that, despite being irrelevant on version "A", on version "B", the shape is the most influential aspect for the correct emotion, being even more predominant than the characters themselves. It is worth mentioning that this tendency flips when considering the high version of the emotion, with characters regaining their predominance.

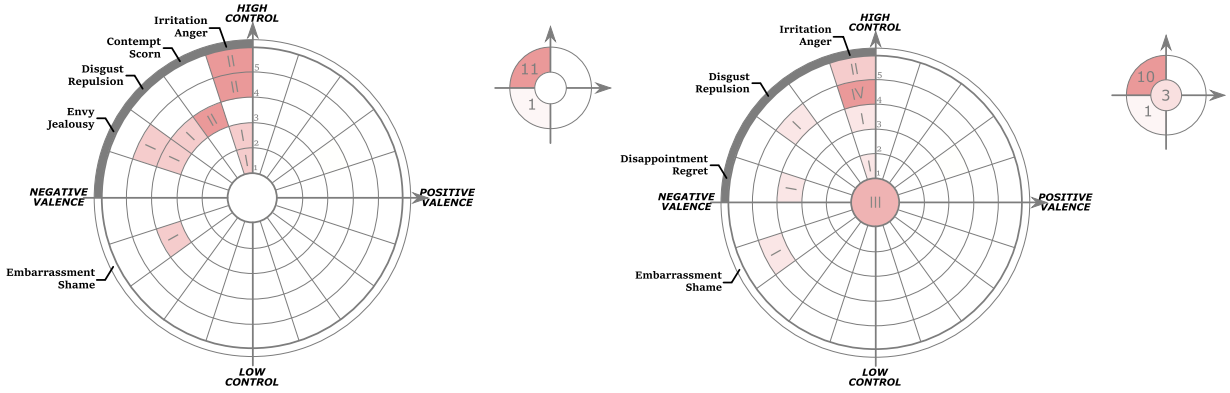
Disgust low was as interesting emotion, since it was recognized as being worse animated and less perceptible than the initial version. Nonetheless, some balloons' aspects helped transmitting the right emotion, while also misleading some participants in terms of valence.

Finally, in anger low, we can see that the majority of aspects, irregardless of the category, that influenced the right perception, also led to the incorrect recognition in other participants. Nonetheless, we can verify that the balloons, more specifically the shape and color, were the most referenced details, with more mentions than the characters. Anger low influences can be seen in figure 7b.



(a) Version "A" (b) Version "B"

Figure 5: Surprise low heatmaps (Roman numerals show the number of answers per emotion and intensity, with intensity growing outwards from 1 to 5, and colors getting darker with the number of answers. On the side, there is a small quadrant heatmap, showing the responses per quadrant.)



(a) Version "A" (b) Version "B"

Figure 6: Anger low heatmaps

Relevance

The first essential aspect to address is that both balloons and environment are less relevant than the characters themselves. This is not true for every emotion, but holds true for the majority of emotions. This idea is in line with our hypothesis that such 2D elements should not be used on their own, but instead be used alongside characters to reinforce the emotion transmitted by them. However, their usage needs to be carefully considered, in order to minimize noise and to not overshadow the characters, which may cause some confusing to the users of the application. Within the realm of 2D elements, balloons tend to be more easily recognizable than backgrounds, since they are common in certain types of artistic works, more specifically cartoons and comics.

Looking specifically at the balloons, we now understand that shape is the most influential characteristic, followed closely by the color, and finally the animation. This aligns with the idea that some shapes are prone to being associated with particular emotions. The color-emotion association is present in a similar fashion, albeit being somewhat ambiguous, given the differences in ones culture and personal experiences. Finally we have the animation aspect, achieved by mimicking movements and/or voice modulation, that tends to be even more abstract, thus having severely less impact than the others. A downside was that, due to the relevance of context, the sentence presented inside the balloons was picked up as being meaningful by some of the users. This sentence needs to be carefully defined for future experimental purposes, in order to minimize the bias it introduces. On the other hand, for practical uses, the sentence may assist in reinforcing the emotion or even add contrast, depending on the intent, and thus should not be disregarded when the objective is passing emotions that are appropriate to the situation.

Finally, the environment once again needs to be tested with an already established context, if one hopes to understand if it can be a positive addition to the experience. In our case, the color helped in some cases, but the misuse of icons without context, and the addition of a user background that was left unused during the experiments, led to the wrong perception of emotion by our participants. Without context,

TABLE 1: Overlap, via a blue background, of Bassili's findings with the emotional recognition mistakes found in our study.

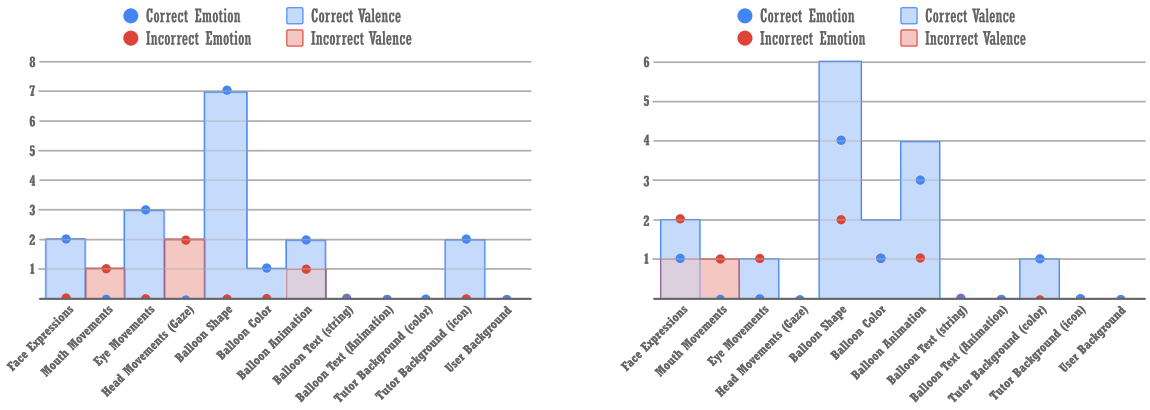
Could Be confused with ...

	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



(a) Fear Low

(b) Anger Low

Figure 7: Version "B" influences

these backgrounds end up as too ambiguous, and sadly, too open to interpretation.

To transmit the idea of different intensities, in our case, we used only variations in animation and color. These changes may have ended up being too exaggerated. However, by reducing the color saturation, or by smoothing and dampening the animations, we may end up making the variations so subtle that users may not be able to recognize any emotion at all. This means that all aspects need to be carefully defined and tested further. In the end, tests with a defined context should exist, since context could easily give different results.

To conclude, 2D elements can work within a scene, to transmit a certain valence or a specific emotion, but tests should be performed, both with and without context, to understand if the choices work within the scene being created. Irregardless, we maintain that they should be used primarily as a complement to the characters, in order to reinforce the expressed emotions, rather than ending up competing with them for the attention of the user. After all, characters were shown as being more influential than flatter elements like balloons and backgrounds. Finally, it is important to mention that all aspects should be tested both individually, as well as in combination with each other, since results could differ. Certain aspects may work just fine on their own, but end up having a negative impact when grouped with other elements.

5. Conclusions

In this thesis, we created a prototype that allowed us to control and use all 2D elements in the scene. At the same time, we created a palette of balloons and backgrounds to be used by the application, chose a color scheme to associate with emotions, and developed a range of effects that could be used, both for the balloons and the text. Later, we also developed commands that could be used to define all these properties, by the dialog writer, creating a more flexible interface. With a functional system, we decided to test the final version of the application with users.

We created a set of two experiments, in order to understand how the application was being perceived by the participants. With the first, we observed that generally, the final version of our application is perceived as better animated than the original version, when compared directly, both by experts and non-experts. However, there were certain emotions that did not follow this tendency. With this in mind, we performed the other experiment, which allowed us to test each version separately and understand the inner problems.It was possible to observe that some emotions were better recognized in the final version, whilst others were better perceived in the initial version. Therefore, despite being perceived as better animated, the final version has lost some emotional identity, misleading the participants, when it comes to some emotions.

Nonetheless, it was possible to observe that, overall, the use of 2D elements does have an impact in the perception of users, even if less than the characters, as hypothesized. Some of these elements, due to their common use in comic art, led to better results than others. A good example of this, is the anger emotion, which utilized the anger speech bubble, commonly used in comics. The use of red, which in our culture is also recognized as aggressive, also helped to transmit the right valence. Unfortunately the opposite also happened, with the emotion sadness not even being recognized as a negative valence. This means that if utilized correctly, such elements can help reinforce the emotional perception of a scene. However, since their use is not standardized, they should be used carefully, and only as complementary to the characters, since their can create unwanted contrast, misleading users.

There is still more work to be done, with certain aspects yet to be utilized, while others need to be reiterated upon. Nonetheless, we believe that these elements, despite being less influential, have the potential to transmit valence and help with the emotional recognition of a scene.

References

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

[2] Aire Mill, Jüri Allik, Anu Realo, and Raivo Valk. Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, 9(5):619–630, 2009.

[3] 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.

[4] 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, aug 2009.

[5] Naz Kaya and Helen 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*, 2004.

[6] Naz Kaya and Helen H. Epps. Relationship between Color and Emotion: A Study of College Students. 2004.

[7] Celso M. de Melo and Jonathan 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*, pages 1–7. IEEE, 9 2009.

[8] Lyn Bartram and Ai Nakatani. What Makes Motion Meaningful? Affective Properties of Abstract Motion. In *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*, pages 468–474. IEEE, 11 2010.

[9] Ollie Johnston and Frank Thomas. *Disney Animation: The Illusion of Life*. Abbeville Press, 1981.

[10] André Pires, Patrícia Alves-Oliveira, Patrícia Arriaga, and Carlos Martinho. Cubus: Autonomous Embodied Characters to Stimulate Creative Idea Generation in Groups of Children. pages 360–373. 2017.

[11] Jodi Forlizzi, Johnny Lee, and Scott Hudson. The kinedit system. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, page 377, New York, New York, USA, 2003. ACM Press.

[12] Sabrina Malik, Jonathan Aitken, and Judith Kelly Waalen. Communicating emotion with animated text. *Visual Communication*, 8(4):469–479, 11 2009.

[13] Weston Gaylord, Vivian Hare, and Ashley Ngu. Adding Body Motion and Intonation to Instant Messaging with Animation. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15 Adjunct*, pages 105–106, New York, New York, USA, 2015. ACM Press.

[14] Scott McCloud. *Understanding Comics: The Invisible Art*. 1993.

[15] Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the Human Face*. 3rd edition, 1972.