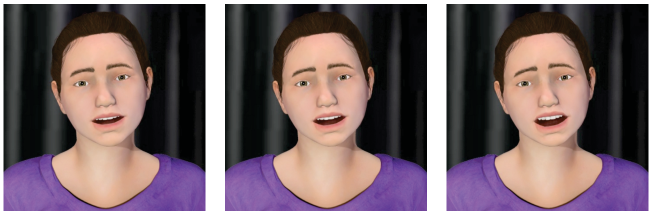
Assessing Naturalness and Emotional Intensity: A Perceptual Study of Animated Facial Motion

Jennifer Hyde[1], Elizabeth J. Carter[2], Sara Kiesler[3], Jessica K. Hodgins[1,2]*
Computer Science Department[1], Robotics Institute[2], HCI Institute[3]
Carnegie Mellon University
Pittsburgh, PA 15213 USA



(a) Damped (b) Normal (c) Exaggerated

Figure 1: *Sample frames from a sad animation with varying levels of facial motion magnitude.*

Abstract

Animated characters appear in applications for entertainment, education, and therapy. When these characters display appropriate emotions for their context, they can be particularly effective. Characters can display emotions by accurately mimicking the facial expressions and vocal cues that people display or by damping or exaggerating the emotionality of the expressions. In this work, we explored which of these strategies would be most effective for animated characters. We investigated the effects of altering the auditory and facial levels of expressiveness on emotion recognition accuracy and ratings of perceived emotional intensity and naturalness. We ran an experiment with emotion (angry, happy, sad), auditory emotion level (low, high), and facial motion magnitude (damped, unaltered, exaggerated) as within-subjects factors. Participants evaluated animations of a character whose facial motion matched that of an actress we tracked using an active appearance model. This method of tracking and animation can capture subtle facial motions in real-time, a necessity for many interactive animated characters. We manipulated auditory emotion level by asking the actress to speak sentences at varying levels, and we manipulated facial motion magnitude by exaggerating and damping the actress's spatial motion. We found that the magnitude of auditory expressiveness was positively related to emotion recognition accuracy and ratings of emotional intensity. The magnitude of facial motion was positively related to ratings of emotional intensity but negatively related to ratings of naturalness.

CR Categories: H.5.1 [Information Interfaces and Presentation (e.g. HCI)]: Multimedia Information Systems—animations

Keywords: emotion, face, audiovisual perception, animation

1 Introduction

Animated characters are used in many domains, including education, therapy, and entertainment. To maximize the efficacy of these characters during human-computer interaction, animators must create characters that emote convincingly [Beale and Creed 2009]. Animated characters can convey emotional information through their facial expressions and paralinguistic vocal cues [Nass and Moon 2000]. Because people are sensitive to these audiovisual cues, even slight variations can result in confusion or an impression of unnaturalness.

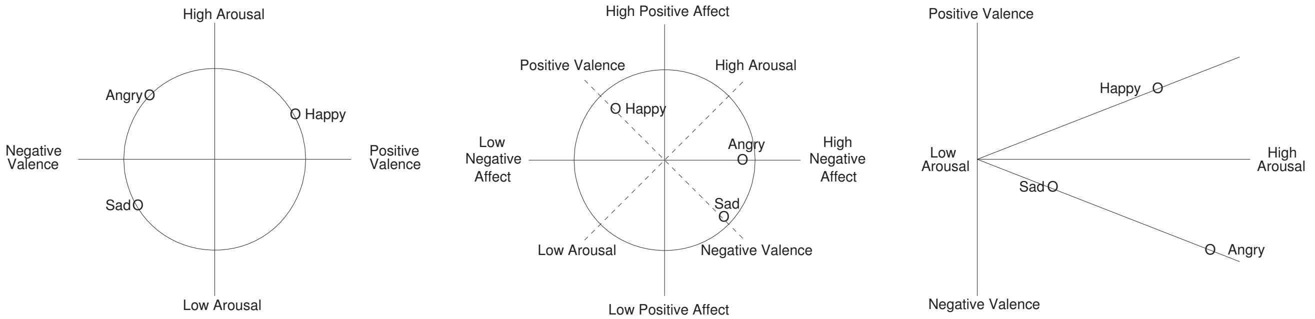
To make animated characters more lifelike, animators have relied on aesthetic precepts, such as exaggerating facial expressions [Thomas and Johnston 1981]. Animators often exaggerate facial expressions to make the emotional content of faces clearer and more intense [Bartneck and Reichenbach 2005]. However, artistic principles such as "exaggeration" and "squash and stretch," which were created in an era of hand-drawn cartoon characters, may no longer be appropriate when used with the more realistic characters [Hodgkinson 2009] often used in interactive applications. When animated characters are realistic rather than cartoonish, viewers may perceive those animated with too much or too little movement as eerie or unnatural [Lasseter 1987; Tinwell et al. 2011].

Researchers have investigated how to enhance realistic faces to improve emotion recognition. Exaggerating the facial expressions depicted in photographs improves emotion recognition accuracy and increases perceptions of emotional intensity [Hess et al. 1997; Calder et al. 1997; Calder et al. 2000]. Exaggeration also negatively affects the perceived naturalness of facial expressions [Calder et al. 2000]. Because these prior studies used still images that lacked motion and audio, we do not know whether those findings generalize to talking animated characters. These characters are less realistic than photographs, which may influence the impact of exaggeration.

We are interested in understanding how to animate the faces of realistic characters to improve perceptions of emotions and naturalness with the goal of creating more compelling interactive applications. We performed a study in which participants evaluated the emotion, emotional intensity, and naturalness of realistic animated characters that displayed different levels of audiovisual expressiveness. As we will show, exaggerating emotional expressiveness can increase the

*e-mail: {jdtam, ejcarter, kiesler, jkh}@cs.cmu.edu

SAP 2014, August 08 – 09, 2014, Vancouver, British Columbia, Canada.
2014 Copyright held by the Owner/Author. Publication rights licensed to ACM.
ACM 978-1-4503-3009-1/14/08 $15.00



(a) Circumplex Model (b) PANA Model (c) Vector Model

Figure 2: *Location of angry, happy, and sad emotions according to various two-dimensional models of emotion.*

perception of emotional intensity and emotion recognition, but it may do so at the expense of perceived naturalness.

2 Related Work

Emotion communication is not only important for interaction between people, but it is also important for human-computer interaction (HCI). Researchers interested in emotion communication have studied how people use audiovisual cues to understand emotion. Modifying these cues can improve people's emotion comprehension or result in confusion. Understanding how and when people misinterpret these cues can inform HCI designers of the situations when modifications are appropriate. In this section, we summarize the results from emotion communication research, and we review the work on two-dimensional models of emotion to examine how people misinterpret audiovisual cues of emotion.

2.1 Audiovisual Cues for Emotion Recognition

People can accurately recognize emotions from vocal cues alone (e.g., [Wallbott and Scherer 1986]), static facial expressions alone (e.g., [Wallbott and Scherer 1986]), and facial motion alone (e.g., [Bassili 1978; Bassili 1979; Ambadar et al. 2005]). Combining multiple cues improves emotion recognition, such as using dynamic faces that contain both shape and motion (e.g., [Cunningham and Wallraven 2009]) or combining auditory and visual cues [Wallbott and Scherer 1986]. Researchers also have used photographic and synthetic stimuli that contain incongruent emotional cues to understand how vocal cues and facial expressions influence emotion recognition [de Gelder and Vroomen 2000; Massaro and Egan 1996]. Participants rated the anger and happiness depicted in stimuli created by morphing and mixing facial expressions and vocalizations from the angry-happy spectrum. The more expressive cues, be they auditory or visual, exerted more influence on people's decisions than less expressive cues.

2.2 Perceptual Effects of Facial Expression Magnitude

Many researchers have investigated the perceptual effects of spatially exaggerating facial expressions on real and synthetic characters. Hess and colleagues [1997] discovered that exaggerating facial expressions increased emotion recognition accuracy. In the same year, Calder and colleagues [1997] found that exaggerating facial expressions also helped participants recognize emotions faster. Both studies included stimuli that depicted exaggerated and damped facial expressions. These still images were created by morphing photographs of facial expressions. Bartneck and Reichenbach [2005] replicated these experiments using images of a synthetic character. Similar to the studies using photographs, they found that the exaggerated expressions improved emotion recognition and increased ratings of perceived emotional intensity. Together, these studies demonstrate a relationship between perceived emotional intensity and facial expression magnitude as depicted in a still image.

To examine whether similar effects occur with very simple rendering styles, Pollick and colleagues [2003] performed an experiment with exaggerated and damped point-light displays of faces. The point-light displays depicted moving dots on specific areas of the face (e.g., the tip of the nose, mouth, mouth corners, eyebrows, etc.). They found that exaggerating the spatial motion of point-light displays improved emotion recognition and increased ratings of perceived emotional intensity.

In a series of experiments using a silent, 3D, synthetic, human face animated with motion capture data, Hill and colleagues [2005] investigated the effects of exaggerating spatial motion on perceptions of emotional intensity. They found that exaggerating the spatial motion of emotional expressions increased the perceived intensity of the performed emotions. These studies provide more support for the positive effect that spatial exaggeration has on emotion recognition and perceptions of emotional intensity for synthetic characters. Because animations often include audio, we were interested in exploring how changes in spatial motion affect emotion perception for characters that are both expressive in facial motion and in speech.

Researchers also have investigated how exaggerated and damped facial expressions affect perceptions of naturalness. Using a method similar to their previous experiment [Calder et al. 1997], Calder and colleagues [2000] found that participants perceived exaggerated facial expressions from photographs as more emotionally intense but less natural. We wanted to explore whether this effect held for 2D characters in motion.

2.3 Dimensional Models of Emotion

To understand how modifying levels of vocal and facial expressiveness can influence emotion recognition, we use two-dimensional models of emotion as a guide (Figure 2). Although there are other accepted models of emotion, the models with more than two dimensions are useful largely for differentiating between complex emotions and cognitive states. We use simpler 2D models of emotion because we only investigated basic emotions. The circumplex model [Russell 1980], vector model [Bradley et al. 1992], and positive activation-negative activation model (PANA) [Watson and Tellegan 1985] are three prominent 2D models of emotion.

Although all of the models include arousal and valence dimensions, they differ in how they assign emotions to 2D space. The PANA model is a rotated version of the circumplex model. Emotions can span the entire 2D space. In the vector model, emotions must be located on the vectors. This constraint means that emotions with high arousal must also have very negative or positive valence. By understanding how vocal and facial expressiveness relate to the arousal and valence dimensions of emotion, we can predict the types of errors people will make when we modify these cues.

3 Hypotheses

Based on the research described earlier, *we hypothesize that (H1) emotion recognition accuracy will increase as auditory emotion level and facial motion magnitude increase*. Because participants may reach ceiling levels in emotion recognition accuracy, we expect that these effects will be more evident when facial motion magnitude and auditory emotion level are mismatched rather than matched. *We also expect that (H2) perceived emotional intensity will increase as auditory emotion level and facial motion magnitude increase. Increased facial motion magnitude may also result in lower ratings of perceived naturalness (H3).*

If we consider that auditory emotion level and facial motion magnitude are related to arousal, then the 2D models of emotion can help predict errors in emotion recognition. The spatial relationship between emotions changes with their arousal level. Therefore, changes to arousal level will alter the likelihood of certain confusion errors. In all three models, decreasing the arousal level of anger makes it more like sadness, but increasing the arousal level of anger makes it less like sadness and happiness. *We predict that increasing the vocal and facial motion levels of angry animations will make anger easier to identify; however, decreasing the vocal and facial motion levels of angry animations will result in more errors where anger is confused with sadness (H4).*

Similarly, if we increase the arousal level of sadness, then it becomes more like anger. In the circumplex and PANA models, decreasing the arousal level of sadness makes it less like anger and happiness. In contrast, the vector model predicts that decreasing the arousal level of sadness makes it more like happiness. *We predict that increasing the vocal and facial motion levels of sad animations will result in more errors where sadness is confused with anger; however, decreasing the vocal and facial motion levels of sad animations will cause more errors where sadness is confused with happiness (H5).*

Happiness has positive valence and a medium level of arousal. In the vector model, happiness is more like sadness than anger, and this relationship holds regardless of how the arousal level of happiness changes. *The vector model predicts that happy animations will be misjudged as depicting sadness (H6a).* In the circumplex and PANA models, happiness is slightly closer to anger. Increasing the arousal level of happiness moves it away from both anger and sadness, but decreasing the arousal level of happiness suggests that it will still be close to anger and move closer to sadness. *The circumplex and PANA models predict that happy animations with increased levels of vocal and facial expression will more clearly exhibit happiness, and that happy animations with lower levels of vocal and facial expression will be misinterpreted as depicting anger and sadness (H6b).*

4 Contributions

Prior research results suggest that people's understanding of emotional facial expressions improve when the expressions are exaggerated. However, it remains unclear whether people's understanding will improve when animated characters are exaggerated because the researchers used stimuli that consisted of morphed photographs or still images [Bartneck and Reichenbach 2005; Calder et al. 1997; Calder et al. 2000; de Gelder and Vroomen 2000; Hess et al. 1997; Massaro and Egan 1996]. Hill and colleagues [2005] used an animated character to examine the effects of exaggerated movement, but their stimuli did not contain audio. The few studies that also included auditory stimuli used emotionally incongruent vocal cues [Hess et al. 1997; Massaro and Egan 1996], but animated characters usually exhibit emotionally congruent audiovisual cues.

Our work extends these previous findings to talking animated characters. We evaluated perceptions of a realistic animated character that used emotional facial expressions and vocal cues. We used an actress's facial motion to animate our character's face, and we systematically altered the motion to test the effects of exaggerated and damped expressions. To investigate whether the size of facial expressions would influence subtler displays of emotion, the actress performed at high and low levels of expressiveness.

5 Method

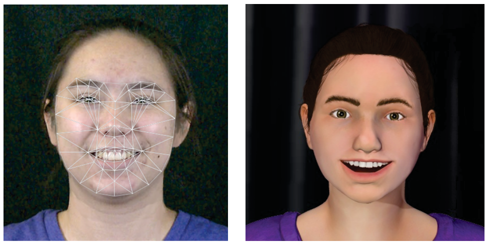
We performed a perceptual study to investigate the relationship between facial and vocal levels of expressiveness on perceptions of emotion. We asked participants to identify the emotion and rate the emotional intensity and naturalness of an animated realistic character that used emotional facial expressions and vocal cues. We hired an actress to provide our animated character with realistic vocal and facial cues. The actress performed each emotion at high and low levels of expressiveness. We used 2D AAMs [Cootes et al. 2001; Cootes et al. 2002; Matthews and Baker 2004] to track the facial motion of the actress. Then we scaled the actress's facial motion up and down to create exaggerated and damped facial expressions on the animated character. Using this method of animation, we created animations that combined different levels of audiovisual expressiveness. For example, to create an animation that mixed auditory and visual levels of expressiveness, we animated our character by damping the facial motion of a highly emotional performance. The character's voice was still highly emotional, but its face was less expressive. In this section, we provide more detail regarding the creation of our stimuli, experimental design, and procedure.

5.1 Animation with Active Appearance Models

We used 2D AAMs to track the actress's facial motion and synthesize the animated face because AAMs can track subtle facial motions, eye gaze, and blinks. Several researchers have found that motions involving the eyes and mouth are important for emotion recognition (e.g., [Ekman and Friesen 1978]).

AAMs require the creation of virtual models for the shape and appearance of the person to be tracked and the character to be animated. Once the models are learned, the person's face can be tracked, and corresponding points on the character's face can be moved (Figure 3).

An AAM consists of two independent models that describe shape and appearance variation. These models define all possible face shapes and appearances for our actress and character. Our face shapes were vectors of 79 coordinates ($\mathbf{s} = (x_1, y_1, ..., x_{79}, y_{79})^T$). We created the shape model with hand-labeled training videos. The shape model is defined in Equation 1 where $\mathbf{s}$ is a new shape, $\mathbf{s_0}$ is the mean shape, and the vectors $\mathbf{s_1}$ through $\mathbf{s_m}$ are the largest basis vectors that span the shape space. The shape parameters, p_i, indicate how much each corresponding basis vector contributes to the overall face shape. A new shape can



(a) Tracking (b) Retargeting

Figure 3: *Example of (a) tracking a person's face and (b) retargeting her motion to a cartoon character.*

then be expressed as the mean shape plus a linear combination of m shape bases.

$$\mathbf{s} = \mathbf{s_0} + \sum_{i=1}^{m} \mathbf{s_i} p_i \tag{1}$$

The appearance model is defined similarly in Equation 2 with appearance, $\mathbf{x} = (x, y)^T$, defined as the pixels that lie within the mean face shape. $A(\mathbf{x})$ is the new appearance, $A_0(\mathbf{x})$ is the mean appearance, $A_1(\mathbf{x})$ through $A_l(\mathbf{x})$ are the largest bases spanning the appearance space, and the λ_i appearance parameters indicate the amount that each appearance base contributes to the new appearance.

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{l} \lambda_i A_i(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathbf{s_0} \tag{2}$$

We followed previous methods [Boker et al. 2009; Hyde et al. 2013; Theobald et al. 2009] to exaggerate and damp the facial motion of the character. By multiplying the face shape variation by values greater than 1, the motion was exaggerated, and by multiplying the face shape variation by values less than 1, the motion was damped. This method of exaggeration and damping affected all facial features. We did not track body motion, so the torso of our character moved rigidly with respect to a pivot point at the mouth. Our character always faced forward as she was created from 2D data. We added rigid points around the top of the character's head to prevent warping, and we damped the face border and nose points by 50% to ensure that the character's face and nose would not appear to be squished or stretched whenever the actress turned her head slightly.

Exaggerating and damping the character's facial expressions did not modify the duration of motion even though our manipulations changed spatial and temporal facial motion. Therefore, we did not need to manipulate the actress's audio. This procedure has been used successfully to manipulate appearance and motion during other studies [Boker et al. 2009; Hyde et al. 2013; Theobald et al. 2009].

5.2 Stimuli

We recorded an actress as she performed three different sentences with angry, happy, and sad emotions. We asked the actress to vary her level of expressiveness for each emotion. To select performances with differing levels of auditory emotion, we calibrated the actress's audio on Amazon's Mechanical Turk (AMT). We first equalized the audio across 63 performances (seven performances for each sentence-emotion pairing) to eliminate obvious volume differences. We used AMT to identify the five most recognizable angry, happy, and sad vocal performances for each sentence. Ninety participants listened to the 63 audio clips and selected the most fitting emotion from the choices: angry, happy, or sad.

We then used AMT again to calibrate these 45 vocal performances for emotion level. Forty-five participants rated each clip's emotional intensity on a scale from 1 to 5. For each sentence-emotion pair, we selected the most and least emotionally intense vocal performances. Using one-way ANalyses Of VAriance (ANOVAs), we determined that all low intensity performances were rated as significantly less intense than the high intensity performances, $p < .0001$.

We used these 18 performances as our low and high emotion audio tracks. We combined the audio tracks with damped, unaltered, and exaggerated animations, resulting in 54 animations that we used as our stimuli (Figure 1). We damped motion by 20% for the damped facial motion condition, and we exaggerated motion by 25% for the exaggerated facial motion condition. We selected these levels of damping and exaggeration based on earlier results that suggested they are perceptually equivalent [Hyde et al. 2013].

We presented all study stimuli and questions on an Apple 27-inch flat panel LED cinema display connected to a MacBook Pro running OSX 10.6, Matlab, and the Psychophysics Toolbox extension [Kleiner et al. 2007]. Participants wore headphones and used a keyboard to respond to questions.

5.3 Experimental Design

We conducted a 3 × 2 × 3 within-subjects, full-factorial experiment with three blocks. The within-subjects factors were performed emotion (angry, happy, sad), auditory emotion level (low, high), and facial motion magnitude (damped, unaltered, exaggerated). We produced three animations from each of the actress's performances, resulting in animations that contained repeated audio tracks. To reduce the effects of memory for a particular audio track, we presented the animations in three blocks. Participants saw 54 different animations across the three blocks. Each block contained 18 animations, and each of these animations had a unique audio track. The animations within each block also spanned every combination of emotion, auditory emotion level, and facial motion magnitude. We selected each participant's trial order pseudo-randomly, taking into account the block constraint.

5.4 Participants

We used a university-based experiment scheduling website to recruit 31 adult participants (16 female, 15 male) with ages ranging from 18 to 72 years ($M = 30$, $SD = 14$). All participants read and signed informed consent forms approved by the Institutional Review Board. We compensated participants for their time.

5.5 Procedure

After obtaining voluntary consent, an experimenter led participants to the study room and seated them in front of a monitor and keyboard. Instructions on the monitor explained that participants would be watching short animations, identifying the emotion, and rating the emotional intensity and naturalness of each animation. Participants began the experiment by hitting the spacebar. A two- to three-second-long animation played automatically, and then participants used a keyboard to answer whether the character had been angry, happy, or sad. After responding, the participants rated the emotional intensity of the character on a scale from 1 to 5 where 1 was *Not Intense* and 5 was *Very Intense*. Participants then rated the naturalness of the character on a scale from 1 to 5 where 1 was *Very*

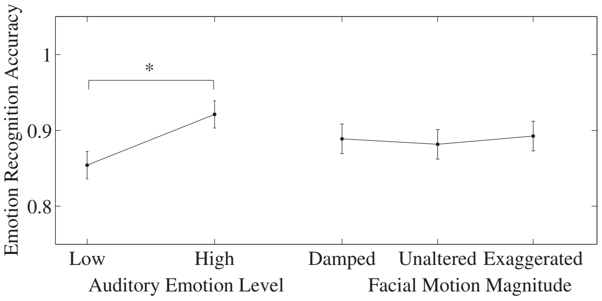


Figure 4: *Influence of auditory emotion level and facial motion magnitude on emotion recognition accuracy. The * indicates significance at $\alpha = .05$.*

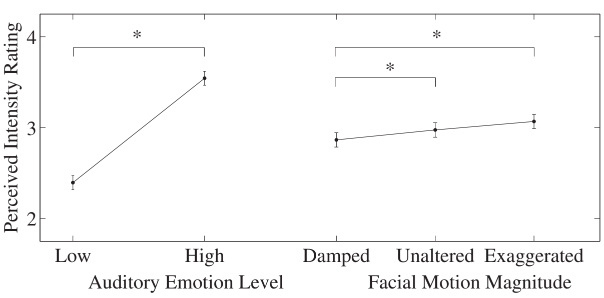


Figure 5: *Influence of auditory emotion level and facial motion magnitude on ratings of perceived emotional intensity. The * indicates significance at $\alpha = .05$.*

Unnatural and 5 was *Very Natural*. After answering the questions, participants could take a break or continue to the next animation by pressing the spacebar. The experiment lasted 10 to 15 minutes.

6 Results

To explore the effects of auditory emotion level and facial motion magnitude on perceptions of emotion in animated characters, we conducted a repeated measures ANOVA with emotion recognition accuracy, ratings of perceived intensity, and ratings of perceived naturalness as dependent variables. Performed emotion, auditory emotion level, and facial motion magnitude were within-subjects independent variables.

6.1 Preliminary Analysis

Although participants completed the experiment in only 10 to 15 minutes, it was possible for participants to get bored and lose focus on the task partway through. To ensure that participants paid attention during the entire study, we checked their emotion recognition accuracy across blocks. We found that emotion recognition improved slightly over the blocks, with mean accuracy rates of 84%, 89%, and 89% for the three blocks. All accuracy rates were well above the 33% chance rate. Because the participants demonstrated good attention during the entire study and the learning effect was small, we used data from all three blocks in our analysis.

6.2 Emotion Recognition Accuracy

We expected emotion recognition accuracy to be better for animations with high auditory emotion than for animations with low auditory emotion (H1). As expected, we found a main effect of auditory emotion level on emotion recognition accuracy, $F(1, 1626) = 20.45$, $p < .0001$. On average, participants were 85% accurate on the animations with low auditory emotion and 92% accurate on the animations with high auditory emotion (Figure 4).

We also hypothesized that participants would be more accurate with exaggerated facial motion (H1), but we found no main effect of facial motion magnitude, $F(2, 1626) = .18$, $p = .8332$. Participants recognized emotions with 88-89% accuracy, regardless of magnitude (Figure 4). The near-ceiling accuracy of our participants indicates that damping facial motion did not reduce recognition. Because we used vocal performances that clearly evoked happy, sad, or angry affect, participants may have relied more on auditory cues than visual cues to determine emotion. We also found no significant interaction between auditory emotion level and facial motion magnitude. The lack of significant interaction also suggests that the auditory cues may have been clearer and more influential than the visual cues in our stimuli.

6.3 Perceived Emotional Intensity

We expected auditory emotion level and facial motion level to affect perceptions of emotional intensity (H2). We found a significant main effect of auditory emotion level on perceived emotional intensity, $F(1, 1626) = 672.97, p < .0001$. As anticipated, participants found animations with low auditory emotion less emotionally intense than animations with high auditory emotion (Figure 5). We also hypothesized that animations with exaggerated facial motion would be perceived as more emotionally intense than animations with damped facial motion. We found a significant main effect of facial motion magnitude, $F(2, 1626) = 6.99, p = .0009$, and confirmed our hypothesis by contrasting emotional intensity perceptions for animations with exaggerated and damped facial motion, $F(1, 1626) = 13.95$, $p = .0002$. Participants found animations with damped motion less emotionally intense than animations with normal motion, $F(1, 1626) = 4.07$, $p = .0439$ (Figure 5). We found no significant interaction between auditory emotion level and facial motion magnitude on perceived emotional intensity.

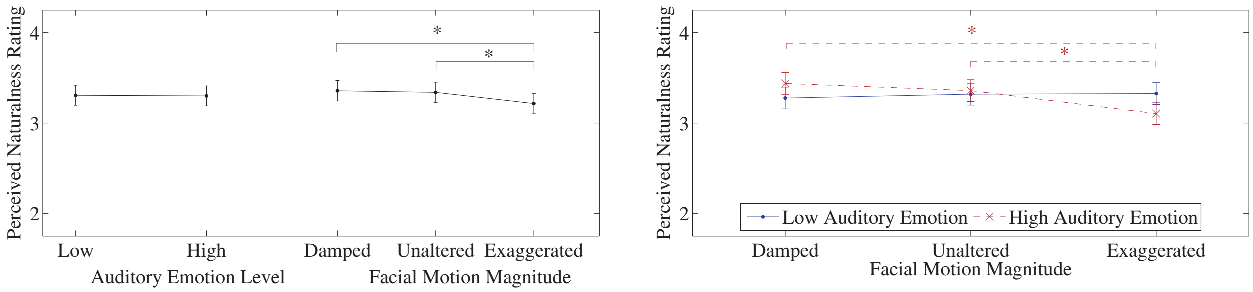
6.4 Perceived Naturalness

We hypothesized that ratings of naturalness would decrease as facial motion increased (H3). As expected, we found a main effect of facial motion on perceived naturalness, $F(2, 1626) = 3.41$, $p = .0334$. Animations with exaggerated facial motion were significantly less natural than the animations with normal motion, $F(1, 1626) = 4.38$, $p = .0365$, and even more unnatural than the damped animations, $F(1, 1626) = 5.74$, $p = .0167$ (Figure 6(a)).

We also found a significant interaction between auditory emotion level and facial motion magnitude, $F(2, 1537) = 5.74, p = .0033$. A post-hoc analysis ($\alpha = .05$) revealed that facial motion magnitude only influenced ratings of naturalness for the animations with high auditory emotion level (Figure 6(b)). Participants perceived the animations with high auditory emotion level as less natural when they were exaggerated than when they had normal or damped motion, $F(1, 1626) = 9.28$, $p = .0024$ and $F(1, 1626) = 15.92$, $p < .0001$, respectively. Participants did not find the animations with unmatched levels of auditory and facial expressiveness less natural than unaltered animations.

6.5 Emotion-Specific Results

We performed additional analyses to better understand the influence that performed emotion had on emotion recognition accu-



(a) Main effects of auditory emotion level and facial motion magnitude

(b) Interaction of auditory emotion level and facial motion magnitude

Figure 6: *Influence of auditory emotion level and facial motion magnitude on ratings of perceived naturalness. The * indicates significance at* $\alpha = .05$.

racy, perceived emotion intensity, and perceived naturalness (Figure 7). As predicted in H1, participants recognized angry animations with high auditory emotion level better than angry animations with low auditory emotion level. An analysis of misclassifications (Table 1) illustrates that participants mislabeled angry animations with low auditory emotion level as sadness more often than happiness, $\chi^2(1, N = 53) = 18.13, p < .0001$, supporting H4.

Also following H1, participants recognized sad animations with high auditory emotion level better than sad animations with low auditory emotion level. The analysis of participants' misclassifications of sad animations also supported H5. Participants misjudged sad animations with high auditory emotion level as depicting anger more often than happiness, $\chi^2(1, N = 17) = 13.24, p = .0002$. Participants misidentified sad animations with low auditory emotion as happy and angry equally, $\chi^2(1, N = 37) = .68, p = .4111$.

Auditory emotion level did not affect emotion recognition accuracy for happy animations. In support of H6a and the vector model of emotion, we found that happy animations, regardless of auditory emotion level, were mislabeled as sad, $\chi^2(1, N = 32) = 6.13, p = .0133$ and $\chi^2(1, N = 38) = 8.53, p = .0035$ for low and high auditory emotion level, respectively. Hypothesis H6b was not supported as greater auditory emotion level did not improve emotion recognition accuracy of happy animations, and decreased auditory emotion level did not cause equal mislabeling of happy animations.

We then analyzed how each performed emotion influenced perceptions of emotional intensity (Figure 7(b)). We found a significant main effect of performed emotion on perceived emotional intensity, $F(2, 1626) = 15.82, p < .0001$, with angry animations perceived as the most intense compared to happy and sad animations ($F(1, 1626) = 17.35$, $p < .0001$ for angry vs. happy and $F(1, 1626) = 28.69$, $p < .0001$ for angry vs. sad). These results indicate a possible relationship between the arousal dimension of emotion and emotional intensity.

Lastly, we examined how each performed emotion influenced perceptions of naturalness (Figure 7(c)). We found a main effect of performed emotion on perceived naturalness, $F(2, 1626) = 40.61, p < .0001$. Participants considered sad animations the most natural, followed by angry and then happy animations ($\alpha = .05$). There was also a significant interaction between performed emotion and auditory emotion level, $F(2, 1626) = 18.09$, $p < .0001$. Paired contrasts ($\alpha = .05$) revealed that participants found the sad animations with high auditory emotion level the most natural, followed by the sad animations with low auditory emotion level, angry animations with both auditory emotion levels, and happy animations with low auditory emotion level. Participants found happy animations with high auditory emotion level the least natural. Although the interaction between performed emotion and facial motion level was not quite significant, $F(4, 1626) = 2.29, p = .0575$, we can see that exaggeration did not affect the perceived naturalness of all emotions equally. Exaggeration negatively impacted the perceived naturalness of happy animations the most, $F(1, 1626) = 46.58, p < .0001$. These results indicate that something other than auditory emotion level and facial motion magnitude influenced perceptions of naturalness.

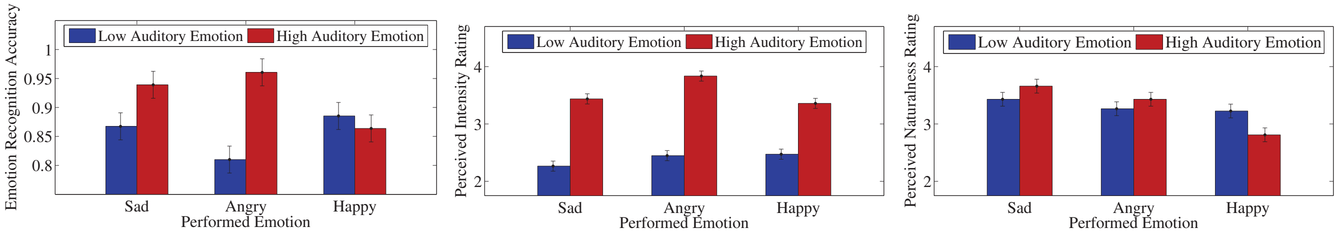
Table 1: *Confusion matrix of participants' emotion recognition responses. Each cell in the table contains the number of trials that participants labeled with the perceived emotion.*

Perceived Emotion	Performed Emotion by Auditory Emotion Level					
	Sad		Angry		Happy	
	Low	High	Low	High	Low	High
Sad	242	262	42	8	23	28
Angry	16	16	226	268	9	10
Happy	21	1	11	3	247	241

We measured non-rigid facial motion by calculating the average displacement of tracked AAM vertices from their neutral positions. We aligned the tracked vertices in each video frame to the vertices of the actress's average face using a similarity transform, which removed the effects of head motion (i.e., translation, in-plane rotation, and scaling of the face). We then averaged across all vertices and scaled our results by interpupillary distance. We discovered that happy animations had the most non-rigid facial motion, followed by angry and then sad animations (2.66, 2.17, and 1.90 mm, respectively). This discovery suggests that the amount of non-rigid motion is negatively related to perceived naturalness. To investigate this relationship further, we compared the amounts of non-rigid facial motion in animations grouped by auditory emotion level and emotion. Sad animations with low auditory emotion had the least amount of non-rigid facial motion, followed by angry animations with low auditory emotion, sad animations with high auditory emotion, happy animations with low auditory emotion, and angry animations high auditory emotion. Happy animations with high auditory emotion had the most non-rigid facial motion.

7 Discussion

We conducted a study to gain preliminary insight into how the expressiveness of vocal and facial cues affect perceptions of emotion. Our results suggest that people depend more heavily on vocal cues than on facial expressions when decoding emotion. Changes in facial motion magnitude did not affect emotion recognition, but increasing levels of auditory emotion improved emotion recognition. Our results also support previous findings on the positive relation-



(a) Emotion Recognition

(b) Perceived Emotional Intensity

(c) Perceived Naturalness

Figure 7: *Influence of auditory emotion level and emotion on (a) emotion recognition accuracy, (b) ratings of perceived emotional intensity, and (c) ratings of perceived naturalness.*

ship between perceived emotional intensity and facial motion magnitude, and the negative relationship between perceived naturalness and facial motion magnitude.

We also investigated participants' incorrect responses to better understand how participants misclassified emotions. This analysis revealed a possible relationship between the amount of non-rigid facial motion and the arousal dimension of emotion. When we aligned non-rigid motion with the arousal dimension of emotion, we could use established emotion models to predict participants' errors. For example, when we used exaggeration to increase the non-rigid facial motion of sad animations, they were mislabeled as angry more often than happy relative to when they were not exaggerated. Similarly, angry animations with damped facial motion were mislabeled as sad. These results suggest a relationship between the amount of non-rigid facial motion and the arousal dimension of emotion. Our results are similar to those of Hill and colleagues [2005], whose participants made similar recognition errors for exaggerated sad and damped angry expressions.

Our work highlights potential trends that should be investigated further in future work. Although we found a difference between the influence of vocal and facial cues on emotion recognition, this result may have been due to our stimulus selection process. We selected our actress's most recognizable high and low intensity vocal performances. We did not validate the emotionality of the corresponding physical performances, which may have resulted in animations that had stronger vocal cues than facial cues. We did find a positive relationship between our actress's auditory emotion level and the amount of her non-rigid facial motion, suggesting that our actress's facial and vocal expressiveness were relatively matched. Moreover, we found that facial motion magnitude affected perceptions of emotional intensity. Our results suggest that when vocal cues are stronger than facial cues, people will identify basic emotions based on vocal cues, but still interpret nuances of emotion from facial motion magnitude. To explore this idea further, we suggest expanding our study to include subtler motion levels and more complex emotions.

We focused our investigation on participants' perceptions when confronted with audiovisual cues that had mixed levels of expressiveness but the same emotion. For this reason, we specifically chose emotions that were easy to recognize (angry, happy, and sad) and performances that were easily recognizable. Now that we have evidence that facial motion magnitude affects perceptions of emotional intensity for recognizable emotions, it would be interesting to explore the effect of facial motion magnitude on more subtle and complex emotions. For example, a shy smile might become a confident smile when exaggerated.

In addition, future studies should use more performers and non-professional actors. We created our animations using the motion of a professional actress who was trained to be as expressive as possible. Therefore, her expressions may have appeared closer to an average person's expressiveness when we damped her motion or become physically impossible when we exaggerated her motion. These characteristics could have affected participants' ratings of naturalness. In the future, it would be interesting to explore whether the effect of facial motion magnitude on perceptions of naturalness is similar for non-actors.

We found evidence that our animation technique may have affected participants' perceptions of naturalness. Although we tried to map our actress's facial motion to the animated character accurately, 2D AAMs have limitations that affect tracking and retargeting. The biggest limitation is that 2D AAMs cannot track head turns and nods accurately; therefore, our actress was asked to limit her head motion. The lack of head turns and nods may have been perceived as unnatural, and exaggerating facial motion without adding head rotations may have caused the character to look more unnatural. There are many other animation techniques that could have been used that would have kept head rotations; however, we used 2D AAMs because they are more capable of capturing smaller motions around the mouth, eye gaze, and blinks.

Moreover, any visual artifacts introduced by the AAMs would have been amplified by exaggeration. For example, we did not animate our characters with the typical textural information that one would expect to see on a person, such as wrinkles around the mouth, eyes, and brow. It is possible that this lack of textural information was emphasized in the animations with more non-rigid facial motion, with the result that participants found the animations with more non-rigid facial motion to be more unnatural. The character's lack of bottom teeth and tongue may have also been more apparent in the exaggerated condition when the character opened her mouth wider.

Our results have implications for the design of affective human characters. Animators of realistic characters cannot rely on the traditional animation principles used with more cartoonish characters to have their intended effect (e.g., exaggeration to convey emotion). In our study, exaggeration did not improve emotion recognition, but it may have created confusion. Participants misinterpreted low arousal emotions performed by exaggerated characters as high arousal emotions. To avoid confusion, animators may need to avoid exaggerating expressions of low arousal.

We also found that exaggeration made our realistic character appear unnatural. Participants rated characters with high auditory emotion as less natural when they exhibited exaggerated motion compared to when they had unaltered or damped motion. In contrast, exaggerating the characters with low auditory emotion did not affect participants' perceptions of the character's naturalness. Animators should refrain from exaggerating the facial expressions of highly emotional characters to prevent lowering perceived naturalness.

Although animators of realistic characters should use exaggeration sparingly, they can damp the facial motion of their characters to decrease perceived emotional intensity. This method may be useful if a character's voice actor is unavailable to rerecord a less emotionally intense performance. Damping facial motion did not effect emotion recognition accuracy or perceptions of naturalness.

8 Acknowledgments

Disney Research funded this work in part, and award R03HD068816 from the Eunice Kennedy Shriver Institute of Child Health and Human Development of the NIH supported E.J. Carter. Thanks to our participants, R. Browne, B. Kelly, T. Simon, P. Carr, I. Matthews, M. Mahler, J. Krahe, C. Heiner, and M. Danver.

References

AMBADAR, Z., SCHOOLER, J. W., AND COHEN, J. F. 2005. Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science 16*, 5, 403–410.

BARTNECK, C., AND REICHENBACH, J. 2005. Subtle emotional expressions of synthetic characters. *International Journal of Human-Computer Studies 62*, 2, 179–192.

BASSILI, J. N. 1978. Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance 4*, 3, 373–379.

BASSILI, J. N. 1979. 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.

BEALE, R., AND CREED, C. 2009. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies 67*, 9, 755–776.

BOKER, S. M., COHN, J. F., THEOBALD, B.-J., MATTHEWS, I., BRICK, T. R., AND SPIES, J. R. 2009. Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars. *Philosophical Transactions of the Royal Society B 364*, 1535, 3485–3495.

BRADLEY, M. M., GREENWALD, M. K., PETRY, M. C., AND LANG, P. J. 1992. Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition 18*, 2, 379–390.

CALDER, A. J., YOUNG, A. W., ROWLAND, D., AND PERRETT, D. I. 1997. Computer-enhanced emotion in facial expressions. *Philosophical Transactions of the Royal Society B 264*, 919–925.

CALDER, A. J., ROWLAND, D., YOUNG, A. W., NIMMO-SMITH, I., KEANE, J., AND PERRETT, D. I. 2000. Caricaturing facial expressions. *Cognition 76*, 2, 105–146.

COOTES, T. F., EDWARDS, G. J., AND TAYLOR, C. J. 2001. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 6, 681–685.

COOTES, T. F., WHEELER, G., WALKER, K., AND TAYLOR, C. J. 2002. View-based active appearance models. *Image and Vision Computing 20*, 9-10, 657 – 664.

CUNNINGHAM, D. W., AND WALLRAVEN, C. 2009. Dynamic information for the recognition of conversational expressions. *Journal of Vision 9*, 13, 1–17.

DE GELDER, B., AND VROOMEN, J. 2000. The perception of emotions by ear and by eye. *Cognition and Emotion 14*, 3, 289–311.

EKMAN, P., AND FRIESEN, W. V. 1978. *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, Palo Alto, CA.

HESS, U., BLAIRY, S., AND KLECK, R. E. 1997. The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior 21*, 4, 241–257.

HILL, H. C., TROJE, N. F., AND JOHNSTON, A. 2005. Range- and domain-specific exaggeration of facial speech. *Journal of Vision 5*, 10, 793–807.

HODGKINSON, G. 2009. The seduction of realism. In *Proc. of ACM SIGGRAPH ASIA 2009 Educators Program*, 1–4.

HYDE, J., CARTER, E. J., KIESLER, S., AND HODGINS, J. K. 2013. Perceptual effects of damped and exaggerated facial motion in animated characters. In *Proc. of IEEE Automatic Face and Gesture Recognition 2013*, 1–6.

KLEINER, M., BRAINARD, D., AND PELLI, D. 2007. What's new in Psychtoolbox-3? In *Perception*, vol. 36, ECVP Abstract Supplement.

LASSETER, J. 1987. Principles of traditional animation applied to 3D computer animation. In *Proc. of ACM SIGGRAPH 1987*, 35–44.

MASSARO, D. W., AND EGAN, P. B. 1996. Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review 3*, 2, 215–221.

MATTHEWS, I., AND BAKER, S. 2004. Active appearance models revisited. *International Journal of Computer Vision 60*, 135–164.

NASS, C., AND MOON, Y. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues 56*, 1, 81–103.

POLLICK, F. E., HILL, H., CALDER, A., AND PATERSON, H. 2003. Recognising facial expression from spatially and temporally modified movements. *Perception 32*, 813–826.

RUSSELL, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology 39*, 6, 1161–1178.

THEOBALD, B.-J., MATTHEWS, I., MANGINI, M., SPIES, J. R., BRICK, T. R., COHN, J. F., AND BOKER, S. M. 2009. Mapping and manipulating facial expression. *Language and Speech 52*, 369–386.

THOMAS, F., AND JOHNSTON, O. 1981. *Disney Animation: The Illusion of Life*. Abbeville Press.

TINWELL, A., GRIMSHAW, M., NABI, D. A., AND WILLIAMS, A. 2011. Facial expression of emotion and perceptions of the Uncanny Valley in virtual characters. *Computers in Human Behavior*, 2, 741–749.

WALLBOTT, H. G., AND SCHERER, K. R. 1986. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology 51*, 4, 690–699.

WATSON, D., AND TELLEGAN, A. 1985. Toward a consensual structure of mood. *Psychological Bulletin 98*, 219–235.