The Computation of Emotion in Facial Expression Using the Jepson & Richards Perception Theory

Brent C.J. Britton

The Media Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

brent@media-lab.media.mit.edu

Abstract

Facial expressions are vital communicators of emotions, and it is in partial response to these expressions that we innately and accurately discern the emotional states of those around us. This paper identifies the activatable facial features that form the language of emotional expression in the face, and the set of emotions that each such feature tends to express. Finally, it is shown how the **fault lattice perception theory** [6] can be used to compute the emotion being registered on a face, given the configuration of the salient features. It is posited that the ability of a computer to make such interpretations would significantly enhance human-computer interaction.

1 Introduction

The ability of the human face to express a wide range of emotions is well supported.[2, 4, 3, 1] The face is stimulus and response in one, a remarkably effective and versatile communicator; human facial muscles are sufficently complex to produce more than a thousand different facial appearances.[2] Our impressions of the facial expressions on those with whom we interact is so seemingly innate, and the communication of emotional information so useful, that a specialized cognitive system may have evolved which is capable of discriminating facial expressions and making inferences about them.[5] Constructing a similar system for use at the computer-human interface seems beneficial.

It is currently possible to communicate with a properly equipped computer system through the use of speech, hand gestures, and eye movements (in addition to, of course, keyboard and mouse input), all of which contribute to a natural, human-like communication environment. Bestowing upon the computer the ability to interpret facial expressions would enrich human-computer communication even more, as it would allow the computer to become sensitive to the emotional state of its human user.

Humans supplement their understanding of facial expressions with contextual knowledge gained through interaction and observation. It is rare that we are expected to interpret a person's emotional state strictly by the examination of his or her face alone. We have the luxury of being able to assimilate other subtle yet important clues such as a person's tone of voice and body movements, as well as any external stimuli that may affect a person's emotional state. The scenario for the computer in this paper, however, is far more simplistic. Here, the computer is assigned the task of determining the user's emotional state (perhaps within the underlying context of ascertaining the user's response to information being provided by the computer) by examining a "snapshot" of the user's face. Adequate vision processing technology is assumed.

The **fault lattice perception theory** [6] serves as the theoretical basis for development. The theory seems particularly well suited to this problem, since, as we shall see, emotional expression is a systematic result of describable facial features (giving rise to a rich set of premises).

2 The Language of Emotions

The manifold facial expressions can be decomposed into gross categories corresponding to a handful of general, primary emotions that people are capable of readily perceiving. The primary emotions can be labeled more or less adequately by the seven terms **happiness, surprise, fear, sadness, anger, disgust**, and **interest**.[1] The face itself can be separated into its constituent parts (brows, eyes, lips, etc.) the relative arrangements of which are likely to be similar and consistent within an emotion category.

2.1 Features

What are the features of a given facial configuration that lead observers to perceive the expression of a particular emotion? To answer this, we must first develop a taxonomy of facial features and their possible states.[1]

[1] The list of features is an adaptation of the **Action Units** of the **Facial Action Coding System** given in [3]. Action Units, however, are anatomical *processes*, whereas the features I list are static *states*.

F1: **brow**[2] can be b1: neutral, b2: raised, b3: lowered

F2: **eyes** can be e1: open (neutral), e2: wide open, e3: slit, e4: closed, e5: tightly closed

F3: **nose** can be n1: neutral, n2: wrinkled, n3: flared (nostrils)

F4: **lips** can be l1: open or wide open, l2: pushed out, l3: curved up, l4: curved down, l5: closed (neutral), l6: tightly closed

F5: **jaw** can be j1: neutral, j2: pushed forward, j3: dropped

The feature knowledge (**F1** through **F5**) serves as a core axiom when the fault lattice perception theory is applied. Note that each of the five features has a neutral state.

The reader may wonder why the teeth are not included as a significant facial feature. As it happens, I originally included them, but soon discovered their state on the face to be just as accurately described by the lips and jaw. Furthermore, it turns out that for nearly every one of the emotions (below), it is acceptable for the teeth to be either open or closed, making them a rather useless indicator!

2.2 Expressions

The aggregate states of the five facial features form the overall facial expression. Let us now see how each of the seven primary emotions is typically expressed by the feature states.[3] It is from these descriptions that useful premises arise for use in the fault lattice perception theory.

happiness: brow raised, **eyes** neutral or slit[4], **nose** neutral, **lips** curved up, **jaw** neutral.

[2] To simplify matters, I use the term **brow** in reference to *both* eyebrows, with the assumption that they are incapable of moving independently of each other.

[3] These descriptions are adapted from [4] and [3].

[4] The eyes often appear slit in a happy expression as a result of the intensity of the smile.

surprise: **brow** raised, **eyes** wide open, **nose** neutral, **lips** open or wide open, **jaw** neutral or dropped.

fear: **brow** neutral or lowered, **eyes** closed or tightly closed, **nose** wrinkled, **lips** open or wide open, **jaw** neutral.

sadness: **brow** neutral or lowered, **eyes** neutral or slit, **nose** neutral, **lips** closed, curved down and pushed out, **jaw** neutral or dropped.

anger: **brow** lowered, **eyes** wide open or slit, **nose** wrinkled or flared, **lips** tightly closed, or wide open, **jaw** pushed forward.

disgust: **brow** lowered, **eyes** slit or closed, **nose** wrinkled and flared, **lips** curved down and pushed out, **jaw** neutral or dropped.

interest: **brow** neutral or lowered, **eyes** neutral or slit, **nose** neutral, **lips** closed, **jaw** neutral.

A neutral face is one in which all features are in their neutral states. As it happens, the expression of **interest** is composed almost entirely of neutral facial features, thereby making it a good candidate for the "default" perception, the interpretation chosen in the event that none other is stronger. The practical result is that a neutral face may be perceived as expressing **interest**, and this is likely to have no effect whatsoever on the interaction between human and computer — it should be fair to assume that if the user is expressing no obvious emotion about the information being provided by the computer, then he or she is at the very least *interested.* It is for this reason that the concept of the neutral expression can safely be disregarded in favor of **interest** when ordering interpretations. However, the computer is likely to require an initial "snapshot" of the user's neutral face against which to compare later images to determine the degree of change in the various feature states.

2.2.1 Categories

The seven primary emotions are grouped into two categories, **pleasant** (happiness, interest, surprise[5]) and **unpleasant** (fear, sadness, anger, disgust). Certain feature states are exclusive to pleasant emotions:

b2: raised brow

l3: upwardly curved lips

while certain others are exclusive to unpleasant ones:

e4: closed eyes

e5: tightly closed eyes

n2: wrinkled nose

n3: flared nose

l2: outwardly pushed lips

l4: downwardly curved lips

l6: tightly closed lips

j2: forwardly pushed jaw

Thus it is often possible to narrow down the choices for a given expression by checking for the presence of the exclusive features. Even though more direct mappings (in the form of the feature state descriptions given above) are already at our disposal, they do not always lead to the confident impression of a single primary emotion. The exclusive features, when present, allow us to choose between interpretations from the different categories.

[5] Whether or not **surprise** can be said strictly to be a pleasant emotion is debatable (see, for example, [4], as well as the Conclusions section of this paper). It does, however, seem reasonable in the context of the human-computer interface under discussion here.

2.2.2 Blends

There is some controversy surrounding the concept of legal emotional **blends** in facial expressions.[2, 1, 3] Surely, it is possible to have two emotions blended in the same expression — you can be both happy and surprised, for example — and, it seems, blends may even be more common than single emotions in expressions.[2]

However, it is not clear precisely which emotions can or cannot be blended, owing to the fact that few of the seven primary emotions have, in the remaining six, a direct opposite with which a blend can confidently be called illegal. Furthermore, the descriptions of blends can be fairly complex (is the brow of one emotion blended with the lips of another to be considered identical to the lips of the first blended with the brow of the second?).[2] It does not seem wise, therefore, to attempt to enumerate in detail all the possible emotional blends or the facial features that contribute to these expressions. What *does* seem plausible, though, is that emotional blends across the pleasant/unpleasant boundary are extremely unlikely — a blend of **happiness** and **sadness** is an obvious example — and it is this notion that we shall adopt as a core axiom.

3 Applying the Theory

To invoke the fault lattice perception theory we must first enumerate the infallible core axioms and the fallible premises.

3.1 Core Axioms (not fallible)

Much of the foregoing is embodied in the core axioms. Interpretations at odds with the axioms are regarded as inconsistent.

A1: Accept the feature knowledge **F1** through **F5**.

A2: A neutral expression is interchangeable with an expression of interest.

A3: Accept the pleasant/unpleasant categorization of the emotions, and the exclusive features.

A4: A viewed facial expression is a truthful representation of one of the seven emotions listed above, or a blend.

A5: A pleasant emotion cannot be blended with an unpleasant emotion.

3.2 Premises (fallible)

Each of the premises consists of a feature state paired with the emotion or emotions that it seems most likely to indicate in all cases.[6] The premises are constructed so that all seven primary emotions are indicated by at least one of the states of some facial feature. That is, for each facial feature (brow, eyes, etc.), there is at least one state (raised, lowered, etc.) that indicates a given emotion. Although this may seem to be an unwarranted and fairly large reduction in the amount of known information, it must be noted that some of the facial feature states listed in the descriptions of the seven primary emotions in the previous section are somewhat tenuous and do not always apply as typically as others, so their inclusion here could lead to errors. Also, humans seem to find it quite easy to discern the emotions given only a partial view of the face, indicating that only certain facial features are necessary to create the expression.[3] Thus it seems proper simply to choose, for each facial feature state, the emotion or emotions with which the feature state is most closely identified.

B1: A neutral brow indicates interest or sadness.

B2: A raised brow indicates surprise or happiness.

B3: A lowered brow indicates anger, fear, or disgust.

E1: Neutral eyes indicate interest or happiness.

[6] Again, these choices are adapted from [4] and [3], and are formulated in part by my own subjective opinions.

E2: Wide open eyes indicate surprise.

E3: Slit eyes indicate anger or sadness.

E4: Closed eyes indicate disgust.

E5: Tightly closed eyes indicate fear.

N1: A neutral nose indicates happiness, interest, surprise or sadness.

N2: A wrinkled nose indicates anger, disgust or fear.

N3: A flared nose indicates anger.

L1: Open or wide open lips indicate surprise or fear.

L2: Pushed out lips indicate disgust.

L3: Upwardly curved lips indicate happiness.

L4: Downwardly curved lips indicate sadness.

L5: Closed lips indicate interest.

L6: Tightly closed lips indicate anger.

J1: Neutral jaw indicates happiness, interest, or fear.

J2: Forwardly pushed jaw indicates anger.

J3: Dropped jaw indicates surprise, disgust, or sadness.

3.3 Ordering Interpretations

The list of premises may seem formidable, but only a handful of them (five) are in use at a time. Any feature (**b1** - **j3**) present on the face under examination activates the corresponding premise (**B1** - **J3**). Premises based on other feature states are irrelevant. To accept a premise means to assert, obviously enough, that the feature state indicates one of the named emotions, and to fault a premise means to assert that the feature state does not indicate any of the named emotions. The acceptance or faulting of the activated premises in various combinations leads to interpretations of various emotions. The label (emotion) assigned to an interpretation is the emotion or emotions (if any) common to all accepted premises, and not forbidden by the faulted premises.

Figure 1: Expression of surprise.

Consider the photograph in Figure 1. The (somewhat exaggerated) expression on this man's face is, according to its source, universally perceived as **surprise**.[2] Thus, if our premises are sound, the fault lattice perception theory should yield **surprise** as a local maximum in the lattice.

To begin, the face is examined for the presence of key features. Even the most superficial glance at the expression in Figure 1 reveals the following features:

b2: raised brow

e2: wide open eyes

n1: neutral nose

l1: wide open lips

j3: dropped jaw

Consequently, the following premises are activated:

B2: A raised brow indicates surprise or happiness.

E2: Wide open eyes indicate surprise.

N1: A neutral nose indicates happiness, interest, surprise or sadness.

L1: Open or wide open lips indicate surprise or fear.

J3: Dropped jaw indicates surprise, disgust, or sadness.

Let us proceed by attempting to formulate a plausible interpretation while assuming that all of our assumptions about the world are incorrect — that is, when all five of the activated premises are faulted.

Faulting premise **B2** means that a raised brow (or, more specifically, the raised brow in Figure 1) indicates *neither* **surprise** *nor* **happiness**; faulting **E2** means that wide open eyes *do not* indicate **surprise**; and so on for the remaining active premises. Thus, when all of the premises are faulted, the interpretation can be neither **surprise, happiness, interest, sadness, fear**, nor **disgust**. Consequently, invalidating our worldly assumptions leads us to the conclusion that the face in Figure 1 bears an expression of **anger**, the only emotion not explicitly forbidden by the faulted premises!

We are bound, however, by core axioms **A3** and **A5** which tell us infallibly that certain facial features (in this case, b2: raised brow) are exclusive to pleasant emotions, or, conversely, that they can never appear on expressions of unpleasant emotions (in this case, **anger**). The result, then, is that **anger** is an inconsistent intrepretation. This gives us at least a little confidence in the soundness of our premises, so let us now see what happens when we accept just one of them.

Accepting **B2** means that the raised brow indicates **surprise** or **happiness**. Neither of these is yet a plausible interpretation, however; each is ruled out by one or more of the remaining (faulted) premises. More precisely, **happiness** is inconsistent with **N1** (which, being faulted, forbids the **happiness** interpretation, among others), and **surprise** is inconsistent with *all* the remaining premises (which, being faulted, summarily forbid the **surprise** interpretation). Thus, accepting only **B2** leads to an inconsistent interpretation to which we cannot assign a label. It should also be obvious that accepting only **E2** leads to the very same sort of inconsistent interpretation.

But what happens when we accept only **N1**? This means that the neutral nose indicates **surprise, happiness, interest** or **sadness**, one of which, namely **interest**, is not forbidden by any of the remaining (faulted) premises, nor is it inconsistent with any of the core axioms. Thus we are able to assign a label of **interest** to the interpretation in which **N1** is the only accepted premise.

Though continuing to enumerate the rest of the 32 possible combinations of accepted and faulted premises would be an instructive exercise, it soon becomes clear that the process is unnecessarily tedious. It is straightforward enough simply to examine the sets of emotions indicated by the activated premises and decide by inspection which combinations are interesting and lead to plausible interpretations. For example, it is easy to see that accepting **B2** and **N1** together leads to the interpretation of **happiness**, since this emotion is indicated by both premises and forbidden by none of the others. The fault lattice for the expression in Figure 1 which we have been studying appears in Figure 2. Many of the unlabeled nodes are not shown. Note that, as was hoped, **surprise** does indeed turn out to be a maximal node.

The expression in Figure 1 was nice to us; it was exaggerated enough to trigger the feature states beyond any doubt. However, in the majority of

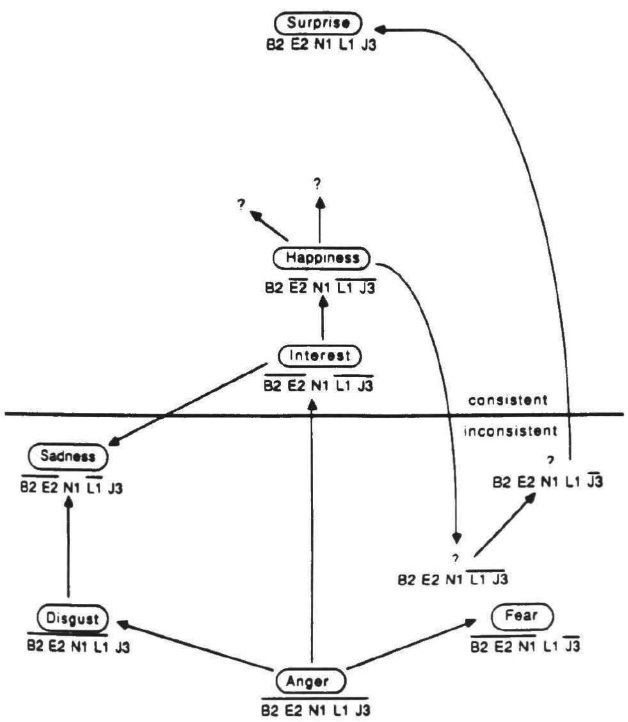

Figure 2: Fault lattice for the expression in Figure 1.

Figure 3: Expression of intense anger.

the faces we see (unlike those typically used in subjective studies) the facial features are quite subtle. We have assumed that our computer has on hand a "snapshot" of the user's neutral face against which to compare later "snapshots" of the same user to determine the amount of change in the features, and we can further assume that this comparison process is carried out at the resolution necessary to detect such subtleties. For the purposes of this paper, however, all we have to go on are our own judgements, because we are not always provided with neutral "snapshots" for comparison. We must therefore take care not to let our judgements of the facial features be biased by our expectations.

Again, there was little doubt about our choices for the expression in Figure 1, but consider the expression in Figure 3, described by its source as intense anger.[2] The features on this face are quite subtle, but it seems safe to say that the brow is neutral, the eyes are slit, the lips are tightly closed, and the jaw is neutral. The nose, however, could be neutral, but it *could* be flared, too. It's just not possible to tell from this one photograph. Carrying out the analysis as if the nose is neutral (using premise **N1**) results in a lattice (not shown) in which **anger** is not the maximal node, but is a *local* maximum. The lattice in which the nose is taken to be flared is shown in Figure 5, and in it, **anger** is the maximal consistent node. The case in which no premises are faulted is unlabeled, since there is no emotion common to all premises. Furthermore, the lips are taken to be tightly closed, and since this feature is exclusive to unpleasant emotions, then interpretations of pleasant emotions are inconsistent by core axioms **A3** and **A5**. For example, in the case where **J1** is the only accepted premise, the interpretation is a *don't know*, a toss up between **fear** and **happiness**. Since **happiness** is a pleasant emotion, however, it is inconsistent, so this node can be labeled **fear**. The same situation exists in the five-fault case with **disgust** and **surprise**.

4 Conclusions

The premises the we have adopted for use in this exercise seem to be fairly sound, and have generated lattices in which the expected interpretations are at least local maxima. A more robust analysis might consider more complex premises that account for the meanings of certain *combinations* of facial feature states, or that use more granularity in the descriptions of the feature states themselves.

Although core axioms **A3** and **A5** seem to be exceptionally useful in preventing obviously incorrect interpretations from becoming consistent perceptions, the notions of exclusive features and pleasant/unpleasant categorizations are in some ways too restrictive. More precisely, there are situations in which it does not seem wise to relegate **surprise** and **interest** strictly to the pleasant category, thus making them unable to be blended with unpleasant emotions. Indeed, these two emotions can almost be thought of as *measures* for the

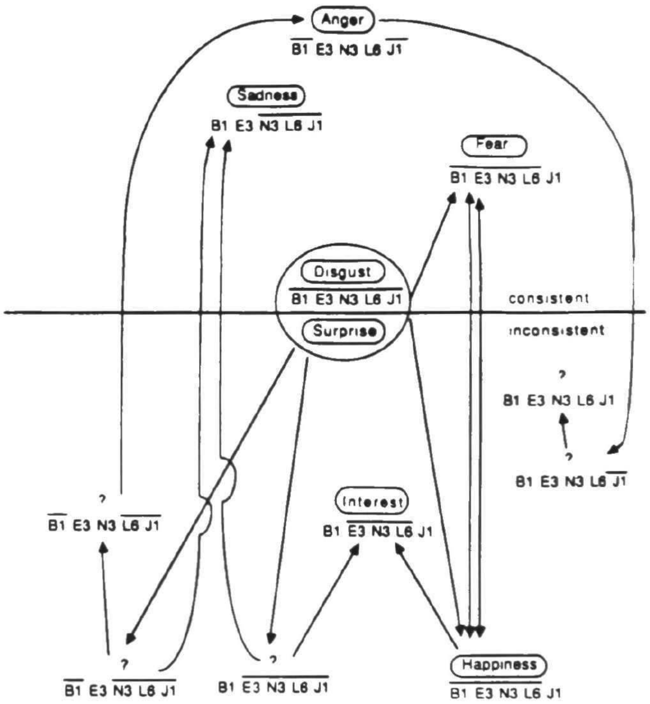

Figure 4: Fault lattice for the expression in Figure 3 when the nose is taken to be flared.

remaining five, levels of saturation along a continuum, always present to some degree in every face. In this respect, greater precision may be required in the categorization of emotions, and the determination of legal blends.

It has been assumed throughout this paper that the computer system attempting to read the expression of its user is doing so without regard for any other forms of input, and without the use of any higher-level reasoning. It does seem entirely plausible, however, to program a system to make use of contextual clues such as what the user says, how the rest of the user's body moves, and the sort of reaction that the information being provided by the computer is *likely* to illicit based on general expectations, or a knowledge of how this particular user has reacted in similar situations in the past.

In summary, the analysis of facial expressions is certainly not an exact science, but the results shown here indicate that it is possible to correctly interpret emotion by examining the configuration of certain facial features. The decisive task is to compose a rich set of premises that accurately describe the relationships between the features and the emotions they indicate.

References

[1] Graham Davies et al., editors. *Perceiving and Remembering Faces.* Academic Press, London, 1981.

[2] Paul Ekman, editor. *Darwin and Facial Expression.* Academic Press, New York, 1973.

[3] Paul Ekman, editor. *Emotion in the Human Face.* Cambridge University Press, Cambridge, 1982.

[4] Hadyn Ellis et al., editors. *Aspects of Face Processing.* Martinus Nijoff, Dordrecht, 1986.

[5] Nancy L. Etcoff. The neuropsychology of emotional expression. MIT Center for Cognitive Science Occasional Paper number 31.

[6] A. Jepson and W. Richards. What is a perception? AI Version, 1989.

[7] Whitman Richards, editor. *Natural Computation.* MIT Press, Cambridge, Massachusetts, 1988.

[8] Clea Theresa Waite. The facial action control editor, face. Master's thesis, MIT Media Arts and Sciences Section, 1989.