Published in final edited form as:
Curr Opin Psychol. 2017 October ; 17: 27–33. doi:10.1016/j.copsyc.2017.06.009.

Visual Perception of Facial Expressions of Emotion

Aleix M Martinez
The Ohio State University, Columbus, OH 43201

Abstract

Facial expressions of emotion are produced by contracting and relaxing the facial muscles in our face. I hypothesize that the human visual system solves the inverse problem of production, i.e., to interpret emotion, the visual system attempts to identify the underlying muscle activations. I show converging computational, behavioral and imaging evidence in favor of this hypothesis. I detail the computations performed by the human visual system to achieve the decoding of these facial actions and identify a brain region where these computations likely take place. The resulting computational model explains how humans readily classify emotions into categories as well as continuous variables. This model also predicts the existence of a large number of previously unknown facial expressions, including compound emotions, affect attributes and mental states that are regularly used by people. I provide evidence in favor of this prediction.

Introduction

Researchers generally agree that human emotions correspond to the execution of a number of computations by the nervous system. Some of these computations yield facial muscle movements, called Action Units (AUs) [1]. Specific combination of AUs define facial expressions of emotion, which can be visually interpreted by observers.

Here, I hypothesize that the human visual system solves the inverse problem of production, i.e., the goal of the visual system is to identify which AUs are present in a face. Crucially, I show how solving this inverse problem allows human observers to effortlessly infer the expresser’s emotional state.

This hypothesis is in sharp contrast to the categorical model, which assumes that the visual system identifies emotion categories rather than AUs from images of facial expressions, Figure 1. The categorical model propounds that our visual system has an algorithm aimed to categorize facial expressions of emotion into a small number of canonical expressions [2]. This model has, in recent years, included six emotion categories: happiness, surprise, anger, sadness, disgust and fear [3]. The claim is that the visual system knows which image

Conflict of Interest
No conflicts of interest to be declared. Funding agencies have been listed in the acknowledgments.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

features code for each one of these emotion categories, allowing us to interpret the expresser's emotion [4].

A major problem with the categorical model is its inability to provide a fine-grained definition of the expresser's emotion, beyond the six canonical expressions listed above [5]. Also, and crucially, the search for the brain's region of interest (ROI) or ROIs responsible for the decoding of these emotion categories has come up empty [6][7]. This has prompted researchers to propose alternative models [8][9][10]. These models suggest that, rather than emotion categories, facial expressions transmit either continuous variables, such as valence and arousal, or affective attributes and mental states, such as dominance and worry.

Which is the correct model? This paper provides converging computational, behavior and imaging evidence in support of the hypothesis that the visual system is tasked to decode AUs from face images, Figure 1b. I show that once AUs have been successfully decoded from faces, the brain can effortlessly extract high-level information, including canonical and fine-grained emotion categories (e.g., disgusted and happily disgusted), continuous affect variables (e.g., valence and arousal), and affect attributes and mental states (e.g., dominance and worry).

Visual Recognition of Action Units

Which are the computations performed by the human visual system to decode AUs? Facial muscles are hidden under our skin and are, hence, not directly visible to us. The human visual system needs to infer their activation from observable image features.

When we move our facial muscles, the distances between major facial components (chin, mouth, nose, eyes, brows, etc.) change. For example, when people produce a prototypical facial expression of anger, the inner corners of their brows lower (which is labeled AU 4), their lids tightened (AU 7) and their upper and lower lip press against one another (AU 24). If you practice these movements in front of a mirror, you will see that the distance between the inner corners of your brows and mouth decreases and that your face widens. Conversely, when creating a prototypical facial expression of sadness, the combination of AUs (1, 4 and 15) leads to a larger than normal distance between brows and mouth and a thinner face. These second-order statistics (i.e., distance variations) are called configural features.

We have shown that these configural features are extremely accurate when used to visually detect the activation of AUs in images [2,11]. For example, activation of AUs 4 and 24 can be successfully detected with 100% accuracy using a single configural feature--the distance between the inner corners of the brows and mouth (Supplementary Material). But, this algorithm sometimes assumes AUs are active when they are not, i.e., a false positive. This happens when we observe someone who has a brow to mouth distance significantly shorter than the majority of people.

This effect is illustrated in Figure 2. The left image is consistently perceived as expressing sadness by human subjects. The right image is consistently categorized as expressing anger. But these images correspond to neutral expressions, i.e., a face that does not display any emotion [11,12]. Why then do we perceive emotion in them? Because our visual system assumes that AUs 1 and 15 on the left image and AUs 4 and 24 on the right image are active. The visual system reaches this conclusion because the configural features that define these AU activations are present in the image. This effect overgeneralizes to other species and drawings of facial expressions as shown in Figure S1–S2, i.e., we anthropomorphize.

Of course, very few people have such an uncanny distribution of facial components on their faces and, hence, the number of false positives is small. Furthermore, the brain can use contextual information to correct some, if not most, of them.

Computational Model

The configural features described in the preceding section define the dimensions of the proposed computational model, Figure 3. Note that this model is norm-based. That is, the perception of AU intensity increases with the degree of activation, since this increases/ decreases the value of the corresponding configural feature [11].

But, why use these image features? Are other shape features better determinants of AU activation? To test this, we performed a computational analysis [5]. In this study, the shape of all external and internal facial components was obtained. Then, machine learning algorithms were used to identify the most discriminant shape features of AU active versus inactive. The results demonstrated that the configural changes of our model are indeed the most discriminant image features.

Additional proof of the use of these configural features comes from the perception of AU activation and emotion in face drawings and schematics (Figure S2). Furthermore, a simple inversion eliminates the percept; if you rotate Figure 2 180°, the perception of anger and sadness will disappear [12]. This is a well-known consequence of configural processing [13]. Also, computer vision algorithms that use these features attain extremely accurate recognition of AUs (Figure S3).

These results thus support our hypothesis that the visual system solves the inverse problem of production by identifying which AUs construct an observed facial expression. Yet, if this model is correct, there must be a neural mechanism which implements these computations. Indeed, using multivariate pattern analysis on BOLD (blood-oxygen-level dependent) fMRI (functional Magnetic Resonance Imaging), we have identified a small ROI in posterior Superior Temporal Sulcus (pSTS) consistent with the computations of our model (Figure S4).

Emotion Categories

The computational model summarized in Figure 3 explains how we can detect the presence of AUs in a face. But, how does this model allow us to recognize emotion categories? One hypothesis is that emotion categories are defined by specific sets of AU activations [1,14–18].

In our model, AUs define the dimensions of a face space, Figure 3. Hence, a combination of p AUs corresponds to a p-dimensional orthant of that space. For example, the green quadrant (i.e., orthant of dimension two) in the left image in Figure 3, corresponds to the expression of sadness. This is because facial expressions in this quadrant are recognized as having AUs 4 and 15 active. Similarly, faces in the pink quadrant have AUs 4 and 24 active and, hence, are categorized as expressing anger.[1]

Thus, the positive (in green) and negative (in pink) quadrants of the computational model in Figure 3 describe two distinct emotion categories--sadness and anger. But, what is represented in the other two quadrants (shown in white)? Our model suggests that these are facial expressions described by a combination of AUs employed by distinct emotions. That is, the model hypothesizes that these orthants represent compound emotions (e.g., sadly angry and disgustedly surprised, Figure S4).

Do these compound expressions exist? To test this hypothesis, we took pictures of 230 participants posing 21 of the predicted compounds. No instructions on which facial muscles to move was provided to participants. All images were then manually coded to determine which AUs were used to express each of the 21 emotions (Figure S5). The results [5] demonstrate that AU activation is indeed consistent within and differential between emotion categories, supporting the prediction of the model, i.e., all of us produce these compound emotions using the same AUs.[2]

Our results also show that the AUs used to define a facial expression of a compound emotion are a combination of those employed to express the subordinate categories. For example, a prototypical facial expression of happiness includes AUs 12 and 25, whereas that of surprised is given by AUs 1, 2, 25 and 26. And, as predicted by the model, a prototypical facial expression of happily surprised includes AUs 1, 2, 12, 25 and (typically) 26 (Table S1). AU 12 and 25 come from the expression of happiness, while AUs 1, 2, and 25 express surprise.

But, not all the AUs in the subordinate categories need to be included in the expression of a compound. In some instances, the AUs in the subordinate categories are polar opposite of one another. For example, distinct AUs change the same diagnostic configural features of anger and sadness--AU 1 vs 4 and 15 vs 24, Figures 3 and S6. Is a prototypical facial expression of sadly angry described as an ensemble of AUs 4 and 15? Or AUs 1 and 24? Our results [5] show that, when asked to produce this expression, people use AUs 4 and 15. What is represented by AUs 1 and 24, then? I hypothesize that this facial expression is a yet-to-be-discovered compound. Specifically, this expression is a different type of compound of anger and disgust; possibly, a facial expression of resignation. Note there will also be combinations of AUs that do not define an emotion category; and that some of them may appear strange or funny. And, small deviations of the prototypical AU combinations defined above are common, as demonstrated by our study.

[1] Table S1 lists the AUs defining each of the known facial expressions of emotion. And Table S2 summarizes the configural features most discriminant of several AUs.

[2] Note that, in our model, AUs are probabilistic, i.e., not everyone uses the exact same AUs, as previous authors seem to claim. This is why we talk about prototypical expressions.

Facial Expressions in the Wild

The results summarized in the preceding section show that, as predicted, people can readily and consistently produce facial expressions of compound emotions.

Are the above results observed in the lab also a construct of our methodology? Recall, we did ask participants to produce specific facial expressions of emotion, e.g., “please produce a disgustedly surprised expression.” To verify that our results are not a construct of our (in-lab) approach, we assess the prevalence of compound emotions in spontaneous facial expressions collected outside the lab. These are typically called facial expressions “in the wild.”

Specifically, we downloaded 1 million images of facial expressions of emotion from a variety of Internet sources, including news media, documentaries, and social media [19]. We then used a computer vision algorithm to automatically annotate this image set (Supplementary Material). The results show that the combinations of AUs of prototypical facial expressions of compound emotions are as prevalent (or more) as the previously described six canonical expressions (Figure S7).

Additionally, our computational analysis identified the existence of a large number of categories defining affect attributes and mental states, as suggested by others [9,10]. Indeed, these categories are defined by distinct orthants of the face space, Figure 3. This result suggests that the perception of canonical expressions and other affect attributes and mental states are particular cases of the herein proposed model, Figure 4. These results show that the AU combinations associated to specific emotion categories in the lab are consistently observed in the wild.

The proposed model also explains the perception of valence and arousal. While the axes defining categories (orthants) serves as categorical boundaries, the axes themselves are continuous [2], Figure 4. For example, the brow-to-mouth distance is one such continuous variable. This continuum allows the visual system to distinguish between intensities of AU activation and define variables such as arousal and valence. This, in turn, permits fine-grain interpretations of an expresser’s emotion (e.g., happy, amused and exhilarated).

Discussion and Future Directions

The present paper has introduced a new model of the perception of facial expressions of emotion. This model propounds that the visual system is tasked to identify the AUs that are active in a facial expression. I have delineated the results of several computational, behavioral and imaging studies favoring this model. I have also explained how this model can subsume previous models of the perception of emotion, Figure 4. The results described above also show that the number of emotions (and likely mental states) communicated through facial expressions is much larger than previously thought.

Yet, the studies summarized above have only scratch the surface of the proposed model. Above, I gave an example of the expression given by AUs 1 and 24. An in-depth analysis of the model will identify many more of these expressions. Also, the dynamics of facial expressions [20,21] will need to be incorporated into the model. These remain important open areas of research.

The herein-defined model posits that combinations of AUs (given by orthants in the proposed computational space, Figures 3–4) are innate. For example, AUs 4, 25 and 26 define an orthant of our space. Thus, this facial expression is innate. If this prototypical expression is indeed always and exclusively employed when one is angrily surprised [5], then this emotion would also be innate. But, whether these prototypical expressions are indeed consistent within and differential between emotion categories is still under intense debate. Importantly, as stated above, our studies show that the use of AUs is probabilistic, not binary. That is, not everyone uses exactly the same combination of AUs to express the same emotion, although the differences are small [5,19,22]. It is unclear if these small differences are a consequence of culture, personal experiences or a result of yet-to-be-identified innate mechanisms. It is also unknown if these between-subject differences occur in the production and perception of continuous variables.

It is worth noting that the neural mechanisms of the categorization of emotion are also sketchy. Results in my lab have pinpointed an ROI dedicated to the decoding of AUs. But, what are the neural mechanisms involved in subsequent computations? Also, top-down mechanisms may play an important role [8,23,24] but these are, for the most part, unexplored. For one, there may be top-down mechanisms that modulates the perception of AUs. For example, interacting with someone we really dislike may increase the likelihood of detecting AUs associated with negative valence.

Finally, it is still unclear if the recognition of AUs is featural, holistic or a combination of the two. It is likely that some AUs are detected more holistically than others. For example, behavioral experiments demonstrate that the exposure time and number of pixels needed to analyze an expression varies as a function of its AUs [25]. This suggests the information used to identify distinct AUs might be different, but see [26].

I believe that these and related questions will move us closer to a general understanding of emotion and how it is communicated to observers through facial expressions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research described in this paper was supported by the National Institutes of Health, grants R01 EY 020834 and R01 DC 014498, the Human Frontier Science Program, grant RGP0036/2016, and a Google Faculty Research Award. Special thanks to Fabian Benitez-Quiroz, Ramprakash Srinivasan, Shichuan Du and Qianli Feng.

References

* of special interest

** of outstanding interest

1. Ekman, P., Rosenberg, EL. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). 2012.

2. Martinez AM, Du S. A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. J Mach Learn Res. 2012; 13

3. Ekman P. What Scientists Who Study Emotion Agree About. Perspect Psychol Sci. 2016; 11:31–34. DOI: 10.1177/1745691615596992 [PubMed: 26817724]

4. Ekman P, Cordaro D. What is Meant by Calling Emotions Basic. Emot Rev. 2011; 3:364–370. DOI: 10.1177/1754073911410740

5**. Du S, Tao Y, Martinez AM. Compound facial expressions of emotion. Proc Natl Acad Sci U S A. 2014; 111 The studies in this paper support the view of a much larger number of emotion categories than the typically cited canonical six. doi: 10.1073/pnas.1322355111

6*. Srinivasan R, Golomb JD, Martinez AM. A neural basis of facial action recognition in humans. J Neurosci. 2016; 36 The studies in this paper identify a brain region with computations consistent with those of the herein proposed model. doi: 10.1523/JNEUROSCI.1704-15.2016

7. Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Feldman Barrett L. The Brain Basis of Emotion: A Meta- analytic Review. Behav Brain Sci. 2012; 35:121–202. DOI: 10.1017/S0140525X11000446 [PubMed: 22617651]

8. Barrett LF. The theory of constructed emotion: An active inference account of interoception and categorization. Soc Cogn Affect Neurosci. 2016; :nsw154.doi: 10.1093/scan/nsw154

9. Tamir DI, Thornton MA, Contreras JM, Mitchell JP. Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. Proc Natl Acad Sci U S A. 2015; 113:194–199. DOI: 10.1073/pnas.1511905112 [PubMed: 26621704]

10. Skerry AE, Saxe R. Neural Representations of Emotion Are Organized around Abstract Event Features. Curr Biol. 2015; 25:1945–1954. DOI: 10.1016/j.cub.2015.06.009 [PubMed: 26212878]

11**. Neth D, Martinez AM. Emotion perception in emotionless face images suggests a norm-based representation. J Vis. 2009; 9 This paper describes the original studies that identified configural changes as major image features for the recognition of emotion in faces. doi: 10.1167/9.1.5

12. Neth D, Martinez AM. A computational shape-based model of anger and sadness justifies a configural representation of faces. Vision Res. 2010; 50doi: 10.1016/j.visres.2010.05.024

13. Maurer D, Le Grand R, Mondloch CJ. The many faces of configural processing. Trends Cogn Sci. 2002; 6:255–260. DOI: 10.1016/S1364-6613(02)01903-4 [PubMed: 12039607]

14. Aristotle. Aristotle: Minor Works. Loeb Classical Library; 1939.

15. Hume, D. Four Dissertations. Nabu Press; 2010. 1141285118

16. Descartes, R. The Passions of the Soul and Other Late Philosophical Writings. Oxford University Press; 2016.

17*. Duchenne, C-B. The Mechanism of Human Facial Expression. Cambridge University Press; 2006. This classical text was the first to provide physiological proof of the use of facial muscle actions to produce consistent and differential facial expressions of emotion

18. Darwin, C. The Expression of the Emotions in Man and Animals. Penguin Classics; 2009.

19*. Benitez-Quiroz CF, Srinivasan R, Martinez AM. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016 The most extensive study of facial expressions of emotion outside the lab. The studies in this paper evaluated one million facial expressions in the wild. This paper demonstrates that expressions typically seen in the real world are consistent with (and more extensive than) those identified in the lab.

20. Jack RE, Garrod OGB, Schyns PG. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Curr Biol. 2014; 24:187–192. DOI: 10.1016/j.cub.2013.11.064 [PubMed: 24388852]

21. de la Rosa S, Giese M, Bülthoff HH, Curio C. The contribution of different cues of facial movement to the emotional facial expression adaptation aftereffect. J Vis. 2013; 13doi: 10.1167/13.1.23

22. Du S, Martinez AM. Compound facial expressions of emotion: From basic research to clinical applications. Dialogues Clin Neurosci. 2015; 17

23. Gilbert CD, Li W. Top-down influences on visual processing. Nat Rev Neurosci. 2013; 14:350–363. DOI: 10.1038/nrn3476 [PubMed: 23595013]

24. Cox D, Meyers E, Sinha P. Contextually evoked object-specific responses in human visual cortex. Science (80-). 2004; 304:115–117. DOI: 10.1126/science.1093110

25. Du S, Martinez AM. Wait, are you sad or angry? Large exposure time differences required for the categorization of facial expressions of emotion. J Vis. 2013; 13doi: 10.1167/13.4.13

26. Yan X, Young AW, Andrews TJ. Differences in holistic processing do not explain cultural differences in the recognition of facial expression. Q J Exp Psychol. 2016; 0:1–15. DOI: 10.1080/17470218.2016.1240816

Highlights

- To interpret facial expressions of emotion, the human visual system solves the inverse problem of production.
- The task of the visual system is to identify muscle actions, i.e., Action Units (AUs).
- Second-order configural features are a primary mechanism to identify AUs in images.
- A brain region with computations consistent with those of the proposed model is identified.

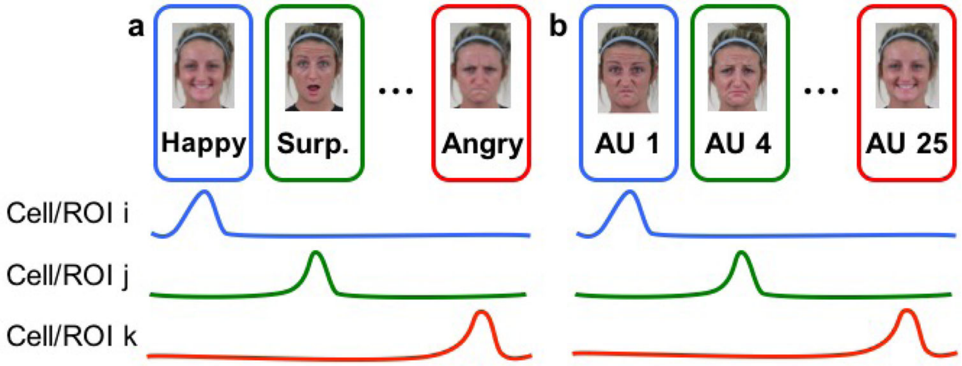

Figure 1.

a. The categorical model posits there must be a group of cells, region of interest (ROI), ROIs or brain networks that differentially respond to specific emotion categories. **b.** The model proposed in the present paper postulates the existence of an ROI dedicated to the decoding of Action Units (AUs) instead. That is, cells in this ROI decode the presence of AUs, not emotion category.

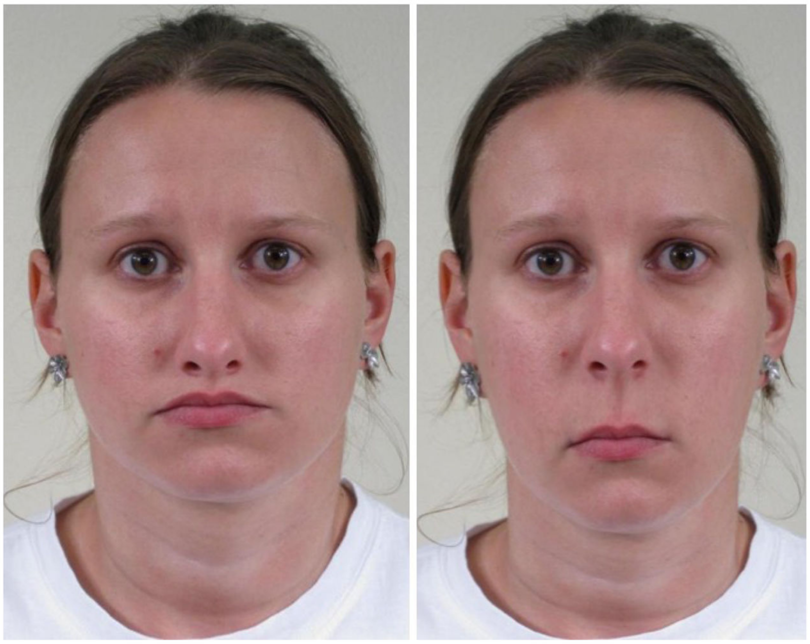

Figure 2.

The left image appears to express sadness, even though no AU is active and, hence, the true expression is neutral. Compare this with the image to its right, where we have reduced the distance between brows and mouth and increased the width of his face. The right image is consistently categorized as expressing anger by human subjects.

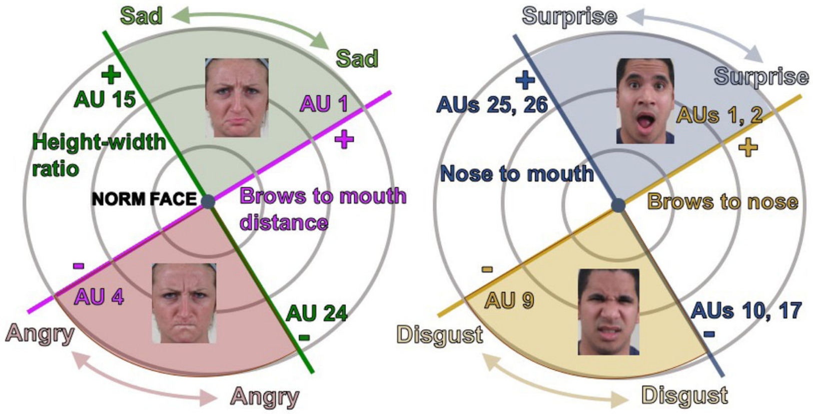

Figure 3.

The dimensions of the face space of the proposed model define AUs. Since AUs are not directly visible in faces, the visual system needs to estimate their presence from image features. Converging evidence supports the view that configural features are used to make this inference. Shown here are four dimensions of this computational model; the dimensions of these two 2-dimensional spaces are orthogonal to one another. One of these dimensions defines the distance between brows and mouth. This distance is increased (indicated with a + sign) by activating AU 1. The same distance can be decreased (−) with AU 4. Other AUs are used to increase or decrease additional configural features of the model, as shown above. These increases/decreases of the distance between facial components are with respect to the norm face. The norm face is the average value of these configural features in the faces we see in our daily lives. Thus, the norm face will vary depending on where you grow up and currently live. This causes the so-called other-race effect, i.e., we make additional mistakes when classifying emotion in faces of other cultures [26]. For example, Asian faces tend to be wider than Caucasian faces. Asian faces also generally have a smaller distance between brows and mouth. Hence, Asian faces are typically perceived as angrier by Caucasians [12]. Each orthant in this computational space defines an emotion category, given by the perception of a set of AUs.

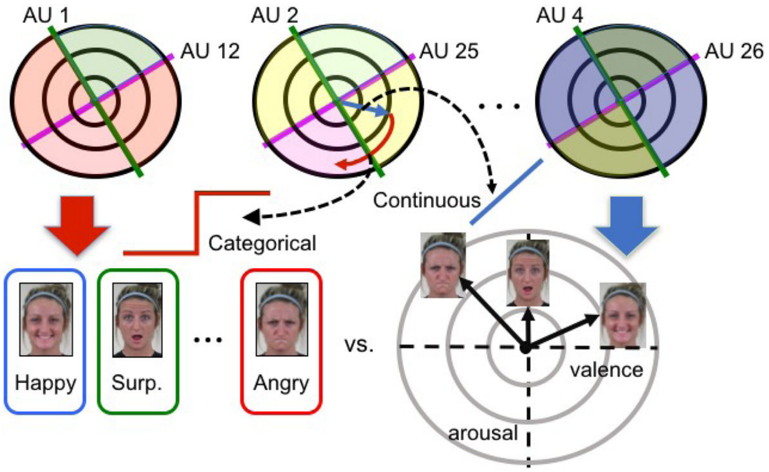

Figure 4.

The dimensions of the computational model derived in the present paper define AUs. This yields a hard (categorical) boundary between orthants of the resulting spaces (shown in different colors in the figure). This result explains how we categorize emotion in faces. But, AU activation is computed using configural image features. This results in continuous variables that can be used to estimate intensity of AU activation. These computations can also be employed to define continuous spaces of emotion, e.g., one given by valence and arousal.