A Brief Guide: Code for Spontaneous Expressions and Micro-Expressions in Videos

Zizhao Dong
dongzz@psych.ac.cn
CAS Key Laboratory of Behavioral Science, Institute of Psychology
Beijing, China

Gang Wang
just_wg@163.com
School of Computer Science, Jiangsu University of Science and Technology
Zhenjiang, China

Shaoyuan Lu
lusy@psych.ac.cn
CAS Key Laboratory of Behavioral Science, Institute of Psychology
Department of Psychology, University of the Chinese Academy of Sciences
Beijing, China

Wen-Jing Yan
eagan-ywj@foxmail.com
School of Mental Health, Wenzhou Medical University
Wenzhou, China

Su-Jing Wang*
wangsujing@psych.ac.cn
CAS Key Laboratory of Behavioral Science, Institute of Psychology
Department of Psychology, University of the Chinese Academy of Sciences
Beijing, China

ABSTRACT

Facial expressions are an important way for humans to perceive emotions. The advent of facial action coding systems has enabled the quantification of facial expressions. Moreover, a large amount of annotated data facilitates the performance of deep learning for the spotting and recognition of expressions or micro-expressions. However, the study of video-based expressions or micro-expressions requires coders to have expertise while also familiar with action unit (AU) coding. This paper systematically sorts out the relationship between facial muscles and AU to make more people understand AU coding from the principle. For this purpose, we have made a brief guide to get started as quickly as possible for the beginner to code.

CCS CONCEPTS

• **Applied computing → Psychology**.

KEYWORDS

Micro-expression, Expression, Facial muscle, Action unit, Coding

ACM Reference Format:
Zizhao Dong, Gang Wang, Shaoyuan Lu, Wen-Jing Yan, and Su-Jing Wang*. 2021. A Brief Guide: Code for Spontaneous Expressions and Micro-Expressions in Videos. In *Proceedings of the 1st Workshop on Facial Micro-Expression: Advanced Techniques for Facial Expressions Generation and Spotting (FME '21), October 24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3476100.3484464

FME '21, October 24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8683-8/21/10…$15.00
https://doi.org/10.1145/3476100.3484464

1 INTRODUCTION

Emotions are the experience of a person's attitude toward the satisfaction of objective things and are critical to an individual's mental health and social behavior. Emotions consist of three components: subjective experience, external performance, and physiological arousal. The external performance of emotions is often reflected by facial expression, which is an important tool for expressing and recognizing emotions [5]. Expressing and recognizing facial expressions are crucial skills for human social interaction. It has been demonstrated by many research that inferences of emotion from facial expressions are based on facial movement cues, i.e., muscle movements of the face [21].

Based on the knowledge of facial muscle movements, researchers usually described facial muscle movement objectively by creating facial coding systems, including Facial Action Coding System (FACS) [9], Face Animation Parameters [17], Maximally Discriminative Facial Movement Coding System [11], Monadic Phases Coding System [10], and The Facial Expression Coding System [13]. Depending upon the instantaneous changes in facial appearance produced by muscle activity, majority of these facial coding systems divide facial expressions into different action units (AUs), which can be used to perform quantitative analysis on facial expressions.

In addition to facial expression research based on psychology and physiology, artificial intelligence plays a vital role in affective computing. Notably, in recent years, with the rapid development of computing science and technology, the deep learning method begins to be widely adopted to detect and recognize automatically by facial action units and makes automatic expression recognition possible in practical applications, including the field of security [12], clinical [15], etc. The boom in expression recognition is attributed to many labeled expression datasets. For Example, EmotioNet has a sample size of 950,000 [8], which is large enough to fit the tens of millions of learned parameters in deep learning networks. The AU and emotion labels are the foundation for training the supervised deep learning networks and evaluating the algorithm performances.

In addition, many algorithms are developed based on AU because of its importance [16, 20].

However, the researchers found that ordinary facial expressions, i.e., macro-expressions, can not reflect a person's true emotions all the time. By contrast, the emergence of micro-expression has been considered as a significant clue to revealing the real emotion of humans. Studies have demonstrated that people would show micro-expressions in high-stakes situations when they try to hide or suppress their genuine subjective feelings [7]. Micro-expressions are brief, subtle, and involuntary facial expressions. Unlike macro-expression, micro-expression lasts only 1/25s to 1/5s [24].

Micro-expression spotting and recognition have played a vital role in defense, suicide intervention, and criminal investigation. The AU-based study has also contributed to micro-expressions recognition. For instance, Davison et al. [3] created an objective micro-expression classification system based on AU combinations; Xie et al. [22] proposed an AU-assisted graph attention convolutional network for micro-expression recognition. Micro-expression has the characteristics of short duration and subtle movement amplitude, which cause that the manual annotation of ME videos requires the data processing personnel to view the video sample frame by frame slowly and attentively. Accordingly, long working hours increase the risk of errors. The current sample size of micro-expressions is still relatively small due to the difficulty of micro-expressions elicitation and annotation.

The prevailing annotation method is to annotate the AU according to the FACS proposed by Ekman et al. [9] as a reference. FACS is the most widely used face coding system, and the manual is over 500 pages long. The manual covers Ekman's detailed explanation of each AU and its meaning, providing schematics and possible combinations of AUs. However, when AU is regarded as one of the criteria for classifying facial expressions (macro-expressions and micro-expressions), a FACS-certified expert is generally required to perform the annotation. The lengthy manual and the certification process have raised the barrier for AU coders.

Therefore, this paper focuses on macro-expression or micro-expression that responds to true emotions, and analyzes the relationship among facial muscles, action units, and expressions. We theoretically deconstruct AU coding based on these analyses, systematically highlighting the specific regions for each emotion. Finally, we provide an annotation framework for the annotator to facilitate the AU coding, expression labeling, and emotion classification.

The following of this article is organized as follows: Section 2 exhibits the process of emotion labeling; Section 3 introduces the relationship between the facial muscles and AUs; Section 4 describes the muscles groups targeting the facial expression; Section 5 presents our conclusion and perspective on code for spontaneous expressions and micro-expressions in videos.

2 EMOTION LABEL

Expressions are generally divided into six basic emotions, happiness, disgust, sadness, fear, anger and surprise. Micro-expressions are usually useful when there is a small negative micro-expression in a positive expression, such as "nasty-nice". For micro-expressions, therefore, there are usually divided into four types, positive, negative, surprise and other. To be specific, positive expression includes happy expressions, which is relatively easy to induce because of some obvious characteristics. Negative expressions like disgust, sadness, fear, anger, etc., are relatively difficult to distinguish, but they are significantly different from positive expressions. Meanwhile, surprise, which expresses unexpected emotions that can only be interpreted according to the context, has no direct relationship with positive or negative expressions. The additional category , "Others", indicates expressions or micro-expressions that have ambiguous emotional meanings can be classified into the six basic emotions.

Emotion labeling requires the consideration of the components of emotions. Generally speaking, we need to take three conditions into account for the emotional facial action: AU label, elicitation material, and the subject's self-report of this video. Meanwhile, the influence of some habitual behaviors should be eliminated, such as frown when blinking or sniffing.

AU label. For AU annotation, the annotator needs to be skilled in the facial coding system and watches the videos containing facial expression frame by frame, the three crucial frames for AU are the start frame (onset), peak frame (apex), and end frame (offset). Then we can get the expression time period for labeling AU. The start frame represents the time where the face changes from neutral expression. The peak frame is the time with the greatest extent of that facial expression. The end frame is the time where the expression ends and returns to neutral expression.

Elicitation material. Spontaneous expressions have high ecological validity compared to posed expressions and are usually elicited with elicitation material. In psychology, researchers usually use different emotional stimuli to induce emotions with different properties and intensities. A stimulus is an important tool for inducing experimental emotions. We use stimuli materials, usually from existing emotional materials databases, to elicit different types of emotions for the subject.

Subject's self-report of this video. After watching the video, the subjects need to evaluate the video according to their subjective feelings. This self-report is an effective means of testing whether emotions have been successfully elicited.

In order to ensure the validity or reliability of data annotation, the process of emotion labeling usually requires the participation of two coders and the calculation of inter-coders confidence must exist in a proper range. The formula is as follows 1:

$$R = \frac{2 \times \left|\bigcap_{i=1}^{N} C_i\right|}{\left|\bigcup_{i=1}^{N} C_i\right|} \tag{1}$$

where C_i represents the set of labeled emotions in the facial expression images by coder i, respectively, and $|\cdot|$ represents the number of labeled emotion in the set after the intersection or merge operations.

The reason is that in the process of annotation, the coders must make subjective judgments based on their expertise. In order to make these subjective judgments as similar as possible to the perceptions of the majority of people, inter-rater reliability is of paramount importance. Inter-rater reliability is a necessary step for the validity of content analysis (emotion labeling) research. The conclusions of data annotation are questionable or even meaningless without this step.

As mentioned above, it shows that emotion labeling is a complex process, which needs coders to have the expertise with both psychology and statistics, increasing the threshold for being a coder. So we tried to find a direct relationship between emotions and AU to identify specific regions of emotions, as shown in Fig. 1.

3 ACTION UNITS AND EMOTION

Human muscle movements are innervated by nerves, and the majority of facial muscle movements are controlled by the seventh nerve in the brain, the facial nerve (Cranial Nerve VII, CN VII). The facial nerve is divided into five branches, including the *temporal* branch, *zygomatic* branch, *buccal* branch, *marginal mandibular* branch and *cervical* branch [4]. These branches are illustrated in the upper part of Fig. 1.

The *temporal* branch of the CN VII is located in the upper and anterior part of the auricle and innervates the *frontalis*, *corrugator supercilii*, *depressor supercilii*, *orbicularis oculi* muscles. The *zygomatic* branch of the CN VII begins at the *zygomatic bone* and ends at the lateral orbital angle, innervates the *orbicularis oculi* and *zygomaticus* muscles. The *buccal* branch of the CN VII is located in the inferior box area and around the mouth and innervates the *Buccinator*, *orbicularis oris* and other orbicularis muscles. The *marginal mandibular* branch of the CN VII is distributed along the lower edge of the mandible and ends in the descending *depressor anguli oris*, which innervates the lower lip and chin muscles. The *cervical* branch of the CN VII is distributed in the cervical region and innervates the *platysma* muscle.

All facial muscles are controlled by one or two terminal motor branches of the CN VII, as shown in Fig. 1. One or more muscle movements can formation AUs, and different combinations of AUs show a variety of expressions, which ultimately reflect human emotions. Therefore, it is a complex process from muscle movements to emotions. We conclude the relationship between AU and emotion based on the of images in the RAF-AU database [23] and the experience of professional coders.

3.1 The Data-driven Relationship Between AU And Emotion

All the data nearly 5,000 images used to analyze are from RAF-AU [23]. The database consists of face images collected from social networks with varying covering, brightness, and resolution, and annotated through human crowd-sourcing. Seven basic (including neutral) emotion tags were used in the samples. Crowd-sourced annotation is a method, which may be useful for tagging facial expressions in a natural setting, by allowing many observers to heuristically tag a target. Finally, the probability score that the picture belongs to a certain emotion is calculated. The database contains about 200,000 facial expression labeling because that each image was tagged by about 40 independent observers. It should be noted that although the source materials are diverse, the judging group of raters is relatively narrow for the reason that the taggers are all students.

For each image, the corresponding annotation contains both the expert's AU labels and the emotion score obtained from the crowdsourcer's label statistics. We analyzed only the contribution of AUs to the six basic emotions. And we used two methods to analyze the data. One method is to take the highest score as the emotion of the image and then combine it with the labeled AU. In this method, repeated combinations must be removed to avoid the effect on the results due to the predominance of one type of sample, i.e., to mitigate the effect of sample imbalance. Another is to count the weighted sum of the contributions of all AUs to the six emotions without removing repetitions. The pseudocode details of these two methods are shown in Algorithms 1 and 2 below. Table 1 and Table 2 list the top10 AUs contributing to the six basic emotions, respectively. From Table 1, it can be seen that the contribution of AU25 is very high in the six basic emotions, which obviously does not correspond to reality. The reason is that the movement of opening the corners of the mouth in AU25 is caused by the relaxation of the lower lip muscles, the relaxation of the genital muscles and the orbicularis oris muscle. According to our subjective perception, AU25 rarely appears when we have three emotions: happiness, sadness, and anger. The abnormal top statistical data in Table 1 may be caused by the shortcomings of crowdsourced annotations, i.e., the subjective tendency or random labeling of some individuals.

Algorithm 1

Initialization: AU's contribution array to emotions $C[6][M] = \{0\}$
M: Max AU number, N: Number of samples, $i = 0$.
repeat
    $i \leftarrow i + 1$
    Split the AU combination into a single AU set
    Take the maximum score of the six emotions as the emotion of the sample, defined as E
    if the combination of AU and emotion E first appears **then**
        Add the emotion score of the sample to the emotion AU
    end if
until $i > N$
Calculate the proportion of AU in each emotion
Sort C in descending order
Output: Contribution array C

Algorithm 2

Initialization: AU's contribution array to emotions $C[6][M] = \{0\}$
M: Max AU number, N: Number of samples, $i = 0$.
repeat
    $i \leftarrow i + 1$
    Split the AU combination into a single AU set
    Add the score of each emotion in the sample to C
until $i > N$
Sort C in descending order
Output: Contribution array C

However, there is room for improvement in the results obtained through the above data-driven approach. The data-driven to results

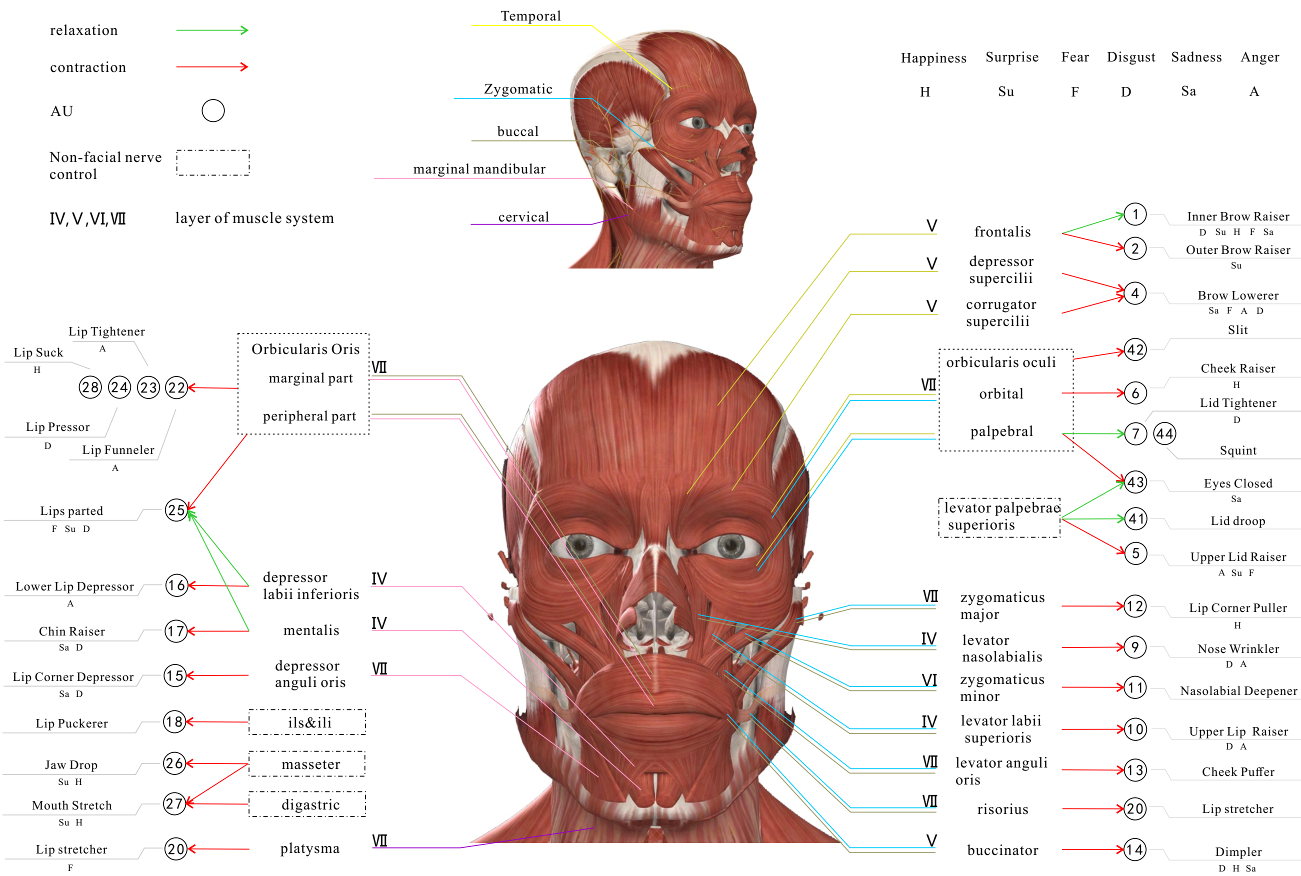

Figure 1: An overview of relationships of facial muscle, AU and emotion based on facial nerve.
Abbreviations ils&ili respectively refer to incisivii labii superioris and incisivii labii inferioris. The source of the human anatomy used in the figure is from https://completeanatomy.cn.

Table 1: Top10 of AU's contribution to the six basic emotions (Method 1)

Emotion	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score
Happiness	25	0.1577	12	0.1424	10	0.0725	1	0.0649	6	0.0616	2	0.0565	26	0.0506	9	0.0498	4	0.0472	27	0.0447
Surprise	25	0.1760	1	0.1093	5	0.1088	2	0.0962	26	0.0772	12	0.0683	10	0.0499	27	0.0499	16	0.0473	4	0.0452
Anger	25	0.1454	10	0.1118	4	0.1034	9	0.0997	16	0.0701	12	0.0506	27	0.0502	5	0.0480	26	0.0402	7	0.0365
Fear	25	0.1669	12	0.0997	1	0.0873	27	0.0866	5	0.0835	4	0.0742	10	0.0734	16	0.0703	2	0.0580	26	0.0479
Disgust	4	0.1303	25	0.1226	10	0.1199	9	0.0782	17	0.0611	1	0.0488	12	0.0488	7	0.0470	26	0.0448	6	0.0398
Sadness	4	0.1526	25	0.1249	10	0.0745	1	0.0689	17	0.0578	12	0.0566	26	0.0505	15	0.0455	9	0.0455	16	0.0375

Table 2: Top10 of AU's contribution to the six basic emotions (Method 2)

Emotion	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score	AU	score
Happiness	12	0.2312	35	0.2059	19	0.2015	2	0.1746	6	0.1635	28	0.1598	30	0.1543	27	0.1513	26	0.1470	14	0.1470
Surprise	2	0.3742	5	0.3625	35	0.3186	1	0.3051	26	0.2859	27	0.2832	21	0.2783	34	0.2721	28	0.2588	25	0.2398
Anger	9	0.3387	33	0.3333	16	0.2726	23	0.2662	7	0.2604	10	0.2549	30	0.2521	24	0.2346	29	0.2332	32	0.2328
Fear	20	0.2500	27	0.2054	33	0.1944	5	0.1854	16	0.1813	2	0.1674	1	0.1631	12	0.1444	30	0.1398	21	0.1307
Disgust	24	0.3112	32	0.3026	17	0.2942	15	0.2789	19	0.2642	14	0.2622	7	0.2484	4	0.2387	18	0.2377	9	0.2325
Sadness	39	0.4294	15	0.2810	43	0.2543	17	0.2355	4	0.1957	14	0.1686	6	0.1621	28	0.1598	7	0.1490	24	0.1467

can be affected by many aspects, first of all by the data source, such as the possible homogeneity of the RAF-AU database (number of subjects, gender, race, age, etc.), the uneven distribution of the sample, and the subjective labels based on human perception resulting from crowdsourcing annotation. Furthermore, the analysis method we used is based on a maximum value and probability weighting. Although straightforward, such analytical approaches represent the contribution of AU to the six basic emotions, are less comprehensive. More analysis is also needed in dealing with unbalanced data. In response to the challenges posed by data and analytical methods to data-driven methods, we could combine data-driven and experience-driven research methods. In this way, we could draw on the objectivity of data-driven and the robustness of experience-driven to realize the construction of the AU coding system for expressions/micro-expressions.

3.2 The Experience-driven Relationship Between AU And Emotion

It usually exists difficulties for the data-driven method to analyze with a theoretical basis. For example, the typical "black box" characteristic brings the problem of poor interpretability. Meanwhile, the results by data-driven are highly dependent on the quality (noiseless) and quantity (wide and massive) of the database. By comparison, the experience-driven method, based on coding and common sense knowledge, is a way to label emotion. Three advantages are listed below: (1) The experience-driven method can help reduce the noise by using coding and common sense knowledge. (2) Experience-driven method has a reliable theory as supported, making the results convincing. (3) Experience-driven can often solve most universal laws with just a few simple formulas. Therefore, we combine experience-driven and data-driven methods to get the final AU and emotional relationship summary table, as shown in Table 3, by using their respective advantages.

Specifically, firstly, based on the analysis results listed in Tables 1 and 2 (data-driven), the preliminary selection is made by comparing the description and legend of each AU in FACS, and combining with the meaning of emotion. We obtained a preliminary AU system for emotion. Then, with large amounts of facial expression images on search engines such as Google and Baidu, the preliminary AU system for emotion was screened by eliminating non-compliant AU in these images. In this way, the ultimate relationship is shown in Fig. 1 and Table 3.

Table 3: The relationship between AU and emotion

Emotion	AU
Happiness	1, **6**, **12**, 14, 26, 27, **28**
Surprise	1, **2**, 5, 25, 26, 27
Anger	4, 5, 9, 10, **16**, **22**, **23**
Fear	1, 4, 5, **20**, 25
Disgust	1, 4, **7**, 9, 10, 14, 15, 17, **24**, 25
Sadness	1, 4, 14, 15, 17, **43**

Based on Tabel 3, we assume that the set of six basic emotions containing AU is S_1, S_2, S_3, S_4, S_5 and S_6. Let $S = \{S_1, S_2, S_3, S_4, S_5, S_6\}$, then

$$Q_i = S_i \setminus \bigcap_j S_j \tag{2}$$

where $i = 1, ..., 6$, and $j = \{1, ..., i-1, i+1, ..., 6\}$. $\bigcap$ is the intersection operation of the set. $\setminus$ represents the set of symmetric difference, for example, we assume that $A = \{3, 9, 14\}$, $B = \{1, 2, 3\}$, then $A \setminus B = \{9, 14\}$. Q_i denotes the AU set that is exclusive to the S_i emotion.

According to Table 3, we can infer which AU appears in a certain emotion but does not appear in other emotions. See Table 3 in bold for details. Therefore, we can conclude that the appearance of certain AU represents related emotion. For instance, if AU20 appears, we assume that fearful emotion emerges.

4 THE SPECIFICITY OF THE RELATIONSHIP BETWEEN FACIAL MUSCLE AND EMOTIONS

According to Fig. 1 and Table 3, we make further analysis of facial muscle and emotions to guarantee that each emotion can be targeted at a specific AU.

4.1 The Muscle That Classifies Positive And Negative Emotions

The basic dimensions for emotions are the two main categories, positive and negative emotion. Positive emotions are associated with the satisfaction of demand and are usually accompanied by a pleasurable subjective experience, which can enhance motivation and activity. By comparison, negative emotions represent a negative or aversive emotion such as sadness, disgust, etc., by an individual.

The *zygomaticus* is controlled by the *zygomatic* branch of the CN VII. The *zygomatic* branch of the CN VII begins at the *zygomatic bone* and ends at the lateral orbital angle, innervates the *orbicularis oculi* and *zygomaticus*. The *zygomaticus* includes the *zygomaticus major* and the *zygomaticus minor*. The *zygomaticus major* begins in the *zygomatic bone*, and ends at the *angulus oris*. The responsibility of *zygomaticus* is to pull the corners of the mouth back or up to smile. The *zygomaticus minor* begins in the lateral profile of *zygomatic bone*, and ends at the *angulus oris*. The function is to raise the upper lip, such as grinning.

The *corrugator supercilii* begins in the medial end of the arch of the eyebrow and ends at the skin of the eyebrow, which is located at the *frontalis* and *orbicularis oculi* muscles back. It is innervated by the *temporal* branch of the CN VII. The contraction of *corrugator supercilii* depresses the brow and generates a vertical frown.

It has been found that the *corrugator supercilii* induced by unpleasant stimuli is more intense than that induced by pleasant stimuli, and the *zygomaticus* is more intense by pleasant stimuli [1]. In a word, pleasant stimuli usually leads to greater electromyography(EMG) activity in the *zygomaticus*, whereas unpleasant stimuli leads to greater EMG activity in the frowning muscle [14].

In the AU encoding process, *zygomaticus* activity and *corrugator supercilii* activity can reliably recognize positive emotion and negative emotion, respectively. This conclusion also supports the discrete emotion theory [2]. For example, oblique lip-corner contraction (AU12), together with cheek raising (AU6) can reliably signals enjoyment [6], while brow furrowing (AU4) tends to signal negative emotion [1]. The correlation between emotion and facial muscle activity can be summarized as follows: (1) The main muscle area of the zygomatic is a reliable discriminating area for positive emotion; (2) The corrugator muscle area is a reliable identification area for negative emotion.

As shown in Fig. 1, AU4, which is controlled by contraction of the depressor supercilii and corrugator supercilii, is present in all negative emotions. Most of the AU associated with happiness is controlled by the zygomatic branch, which mainly innervates the zygomatic muscle. Therefore, the coder should focus more on the cheekbones, i.e., the middle of the face and the mouth if they want to catch the expressions or micro-expressions elicited by positive stimuli. For those elicited by negative stimuli, the coder should focus more on the forehead, i.e., the eyebrows and the upper part of the face.

4.2 Further Specific Classification of The Muscles of Negative Emotions

In the six basic emotions, the negative emotions usually manifested as sadness, disgust, anger and fear, which are all highly associated with the *corrugator supercilii*, the brow and upper region. Therefore, in combination with the lower face, launching a further distinguishing of these four emotions from facial muscles is crucial for emotional classification.

Muscle group specific for sadness. The *depressor anguli oris* begins at the genital tubercle and the oblique line of the mandible, ends at the *angulus oris*. It is innervated by the *buccal* branch of the CN VII and the *marginal mandibular* branch. It serves to depress the *angulus oris*. The study found that when the participants produced happy or sad emotions by recalling, the facial EMG of the frowning muscle in the sadness was significantly higher than that in the happiness [18]. This suggests that the combination of *corrugator supercilii* and textitdepressor anguli oris may be effective in classifying sad emotions.

Muscle group specific for fear. The *frontalis* begins in the *epicranial aponeurosis*, extends to terminates in the skin of the brow and nasal root, and into the *orbicularis oculi* and *corrugator supercilii*. It is innervated by the *auricular posterior* nerve and the *temporal* branch of the CN VII. The *frontalis* is a vertical movement that serves to raise the eyebrows and increase the wrinkles at the level of the forehead, often seen in expressions of surprise. In expression coding, the action of raising the inner brow is coded as AU1. The *orbicularis oculi* begin in the pars nasalis ossis frontalis, the frontal eminence of the upper skeleton and the medial palpebral ligament, surrounds the orbit and ends at the adjacent muscles. Anatomically it is divided into the orbital and palpebral portions. It is innervated by the *temporal* and *zygomatic* branches of the CN VII. The function is to close the eyelid. In the study of the positive intersection of facial expressions and emotional stimuli, the researchers asked the subjects to maintain the fear feature of facial muscles, involving *corrugator supercilii*, *frontalis*, *orbicularis oculi* and *depressor anguli oris* [19].

5 CONCLUSION

In this article, first of all, we sort out the whole process of emotion label. Afterward, with the help of statistical analysis, a data-driven approach is used to obtain a quantifiable system between AU and emotion. In the next part, we further obtain a robust correspondence system between AU and emotion by combining with an empirically driven comparison with actual data (from the web). Finally, the physiological theoretical support for AU labeling of emotions was obtained by adding facial muscle movements. The main manifestations are listed below:

Based on the Fig. 1, Table 3 and the theoretical of section 4, there is a brief guide for coding:

(1) When corners of lips pulled up (AU12) appears, it can be coded as a positive emotion, i.e., happy; In addition, cheek rise (AU6), lip suck (AU28) are both happy specific action units and can also be coded as positive emotions;
(2) When brow rise (AU2) is present, it can be coded as surprise;
(3) When frown (AU4) is present, it can be coded as a negative emotion;
(4) It can be coded as anger when gnashing (AU16, AU22 or AU23), which only occur in the specific action units, appear;
(5) When movements of the eyebrows (AU1 and AU4), eyes (AU5) and mouth (AU25) are present simultaneously, they can be coded as fear;
(6) It can be coded as disgusted when the specific action unit of disgust, lower eyelid rise (AU7), mouth tightly closed (AU24), is present;
(7) It can be coded as sad when frown (AU4) and eyes wide open (AU5) are present at the same time; eyes closed (AU43) is the specific action unit for sadness and can also be coded as sadness.

ACKNOWLEDGMENTS

This paper is supported by grants from the National Natural Science Foundation of China (61772511, U19B2032), and the China Postdoctoral Science Foundation (2020M680738). The authors would like to thank Shouyun Mao and Xudong Lei for linguistic assistance during preparation of this manuscript.

REFERENCES

[1] Serena-Lynn Brown and Gary E Schwartz. 1980. Relationships between facial electromyography and subjective experience during affective imagery. *Biological psychology* 11, 1 (1980), 49–62.
[2] JT Cacioppo, GG Berntson, JT Larsen, KM Poehlmann, TA Ito, M Lewis, and JM Haviland-Jones. 2000. *Handbook of emotions.* The Guilford Press, New York.
[3] Adrian K Davison, Walied Merghani, and Moi Hoon Yap. 2018. Objective classes for micro-facial expression recognition. *Journal of imaging* 4, 10 (2018), 119.
[4] Richard Drake, A Wayne Vogl, and Adam WM Mitchell. 2009. *Gray's anatomy for students E-book.* Elsevier Health Sciences.
[5] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384.
[6] Paul Ekman, Richard J Davidson, and Wallace V Friesen. 1990. The Duchenne smile: emotional expression and brain physiology: II. *Journal of personality and social psychology* 58, 2 (1990), 342.
[7] Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA.
[8] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. 2016. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 5562–5570.
[9] E Friesen and Paul Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3, 2 (1978), 5.
[10] Carroll E Izard, Robin R Huebner, D Risser, and LM Dougherty. 1980. The young infant's ability to produce discrete emotion expressions. *Developmental psychology* 16, 2 (1980), 132.
[11] Carroll Ellis Izard and Maya Weiss. 1979. *Maximally discriminative facial movement coding system.* University of Delaware, instructional resources Center.
[12] Qiang Ji, Peilin Lan, and Carl Looney. 2006. A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on systems,*

man, and cybernetics-Part A: Systems and humans 36, 5 (2006), 862–875.

[13] Ann M Kring and Denise Sloan. 1991. The facial expression coding system (FACES): A users guide. *Unpublished manuscript* (1991).

[14] Jeff T Larsen, Catherine J Norris, and John T Cacioppo. 2003. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* 40, 5 (2003), 776–785.

[15] Patrick Lucey, Jeffrey F Cohn, Iain Matthews, Simon Lucey, Sridha Sridharan, Jessica Howlett, and Kenneth M Prkachin. 2010. Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41, 3 (2010), 664–674.

[16] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. 2019. Multi-label Co-regularization for Semi-supervised Facial Action Unit Recognition. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/cf67355a3333e6e143439161adc2d82e-Paper.pdf

[17] Igor S Pandzic and Robert Forchheimer. 2003. *MPEG-4 facial animation: the standard, implementation and applications.* John Wiley & Sons.

[18] Gary E Schwartz, Paul L Fair, Patricia Salt, Michel R Mandel, and Gerald L Klerman. 1976. Facial expression and imagery in depression: an electromyographic study. *Psychosomatic medicine* (1976).

[19] Roger Tourangeau and Phoebe C Ellsworth. 1979. The role of facial response in the experience of emotion. *Journal of Personality and Social Psychology* 37, 9 (1979), 1519.

[20] S. Wang, G. Peng, and Q. Ji. 2020. Exploring Domain Knowledge for Facial Expression-Assisted Action Unit Activation Recognition. *IEEE Transactions on Affective Computing* 11, 4 (Oct 2020), 640–652. https://doi.org/10.1109/TAFFC.2018.2822303

[21] Thomas Wehrle, Susanne Kaiser, Susanne Schmidt, and Klaus R Scherer. 2000. Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of personality and social psychology* 78, 1 (2000), 105.

[22] Hong-Xia Xie, Ling Lo, Hong-Han Shuai, and Wen-Huang Cheng. 2020. Au-assisted graph attention convolutional network for micro-expression recognition. In *Proceedings of the 28th ACM International Conference on Multimedia*. 2871–2880.

[23] Wen-Jing Yan, Shan Li, Chengtao Que, Jiquan Pei, and Weihong Deng. 2020. RAF-AU Database: In-the-Wild Facial Expressions with Subjective Emotion Judgement and Objective AU Annotations. In *Proceedings of the Asian Conference on Computer Vision*.

[24] Wen-Jing Yan, Qi Wu, Jing Liang, Yu-Hsin Chen, and Xiaolan Fu. 2013. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior* 37, 4 (2013), 217–230.