

Could Micro-Expressions be Quantified? Electromyography Gives Affirmative Evidence

Jingting Li, *Member, IEEE*, Shaoyuan Lu, Yan Wang, Zizhao Dong, Su-Jing Wang*, *Senior Member, IEEE*,
and Xiaolan Fu, *Member, IEEE*

Abstract—Micro-expressions (MEs) are brief, subtle facial expressions that reveal concealed emotions, offering key behavioral cues for social interaction. Characterized by short duration, low intensity, and spontaneity, MEs have been mostly studied through subjective coding, lacking objective, quantitative indicators. This paper explores ME characteristics using facial electromyography (EMG), analyzing data from 147 macro-expressions (MaEs) and 233 MEs collected from 35 participants. First, regarding external characteristics, we demonstrate that MEs are short in duration and low in intensity. Precisely, we proposed an EMG-based indicator, the percentage of maximum voluntary contraction (MVC%), to measure ME intensity. Moreover, we provided precise interval estimations of ME intensity and duration, with MVC% ranging from 7% to 9.2% and the duration ranging from 307 ms to 327 ms. This research facilitates fine-grained ME quantification. Second, regarding the internal characteristics, we confirm that MEs are less controllable and consciously recognized compared to MaEs, as shown by participants' responses and self-reports. This study provides a theoretical basis for research on ME mechanisms and real-life applications. Third, building on our previous work, we present CASMEMG, the first public ME database including EMG signals, providing a robust foundation for studying micro-expression mechanisms and movement dynamics through physiological signals.

Index Terms—Electromyography (EMG), Micro-expression Characteristics, Low Intensity, Short Duration, Spontaneity

1 INTRODUCTION

RECOGNIZING emotions is essential for effective engagement in social interactions [1]. Among the various ways emotions are expressed, facial expressions are widely regarded as the most intuitive way to recognize emotions and interpret intentions [2]. When individuals intend to suppress or conceal their emotions, micro-expressions (MEs) may leak out on face [3], [4], [5]. Specifically, MEs and macro-expressions (MaEs) are fundamentally distinct in their mechanisms and external appearances. MEs are brief, involuntary facial movements lasting less than 500 milliseconds, often subtle and easily missed, and typically occur in localized facial areas with potential asymmetry [6], [7]. In contrast, MaEs are longer, more intense, consciously observable, involve symmetric movements, and engage a broader range of facial muscles.

As non-contact and imperceptible cues, ME can reflect an individual's true emotions and have potential applications in fields such as healthcare and public safety. However, manually analyzing MEs can be time-consuming and labor-intensive, and the results may be influenced by subjective factors. On the other hand, computer vision-based ME anal-

ysis, which identifies true emotions by extracting motion characteristics of MEs, can be widely, conveniently, and effectively applied in various scenarios.

In 2009, Polikovsky et al. made the first attempt to automatically recognize MEs using 3D-HOG features [8]. Later, with the release of databases such as SMIC [9] and CASME series [10], [11], computer vision-based ME recognition gradually emerged. Then, with the continuous development of deep learning models, ME recognition algorithms have gradually evolved from initial hand-crafted-feature-based machine learning methods to a variety of deep learning algorithms. Meanwhile, since 2016, the release of long-video ME databases such as CAS(ME)² [12] and SAMM-LV [13] has promoted research on ME spotting. As MEs can easily be confused with other types of facial movements or environmental noise, ME spotting is still in its infancy, and its performance is not yet satisfactory [14], [15].

Although the analysis of MEs is rapidly developing, the three ME characteristics mentioned at the beginning still requires further clarification. Previous definitions based on ME samples predominantly relied on visual observation, a method relatively subjective. Specifically, the Facial Action Coding System (FACS) [16], [17], [18], a well-established coding framework, offers effective methods for analyzing both action units (AUs) and emotions. However, the intensity coding in FACS relies solely on subjective ratings by coders [19], which can lead to weak consistency. Consequently, the analysis of ME characteristics lacks objective and replicable quantification method.

For external characteristics of MEs, they can be delineated using objective physiological indicators. For instance, electromyography (EMG), a method known for its high accuracy in physiological signal acquisition [20], would

- * Corresponding author.
- This work is supported, in part, by grants from the National Natural Science Foundation of China (62276252, 62476269, 62106256), and in part, by a grant from the Youth Innovation Promotion Association CAS.
- J.T Li, S.Y Lu, Y Wang, Z.Z Dong and S.J Wang are with the Key Laboratory of Behavior Sciences, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China, and also with the Department of Psychology, University of the Chinese Academy of Sciences, Beijing, 100049, China.
E-mail: {lijt, lusy, wangyan1, dongzz, wangsujiang}@psych.ac.cn
- X.L Fu is with the School of Psychology, Shanghai Jiao Tong University, Shanghai 200030, China E-mail: fuxiaolan@sjtu.edu.cn
E-mail: fuxl@psych.ac.cn

be particularly effective in objectively measuring external aspects of MEs, such as their intensity and duration. Employing objective characterization methods can facilitate the rapid construction of large-scale ME databases, alleviating the small sample size issue prevalent in current intelligent analysis. Moreover, a clear understanding of the external features of MEs can provide essential prior knowledge for ME analysis based on deep learning.

Simultaneously, the internal characteristics are also a crucial component for MEs. **A precise definition of ME characteristics not only enhances the reliability of ME coding but also provides a theoretical foundation for identifying concealed genuine emotions in practical ME-based applications.** While the previous study has established a correlation between self-awareness of facial expressions and external characteristics [21], it remains to be confirmed if the spontaneity of MEs aligns with this conclusion.

Overall, the motivation for this paper comes from two main aspects. First, the theoretical basis for MEs is largely hypothetical, with few studies quantifying their characteristics through physiological signals. Second, the undefined external characteristics of MEs make it difficult for deep learning models to learn and extract effective features. To address these challenges, we propose a quantitative study of MEs using EMG and self-reports as direct indicators. **Specifically, EMG provides a sensitive and objective measure of facial muscle activity with high temporal resolution, enabling precise quantification of ME duration and intensity. By integrating EMG into our study, we enhance the reliability and objectivity of ME observation compared to traditional vision-based annotation methods.** This paper is an extended version of an article we published in the Facial ME workshop at the ACM International Conference on Multimedia [22]. Fig. 1 illustrates the overview of this study. In particular, we estimated intervals for the duration and intensity of MEs. Additionally, we utilized a self-report method to investigate the spontaneity of MEs. Our findings support prior definitions of MEs and provide objective indicators for further analysis. Besides, We organized all the collected data to build an EMG-based ME database called CASMEMG. The contributions of this article can be summarized as follows.

- Firstly, we present objective and dependable findings from facial EMG-based ME studies, indicating that MEs are characterized by shorter duration and lower intensities. Notably, we introduce EMG-based measures as novel indicators of ME intensity. Additionally, we offer precise interval estimations for the duration and intensity of MEs, facilitating a refined quantification of external characteristics.
- Secondly, we conducted an empirical study to investigate the internal characteristics of MEs. We found that individuals exhibit less control and awareness over MEs compared with MaEs. This research provides the theoretical basis for subsequent studies on ME mechanisms and applications in real-life scenarios.
- Thirdly, we have also created a ME database: **CAS-**

MEMG¹, which features EMG signals as objective physiological indicators alongside video recordings. It contains 380 samples of MEs and MaEs in total. Keyframe timings, AUs, and emotion labels are all provided. Additionally, a baseline method and results for EMG-based expression interval detection are included. The release of CASMEMG database aims to enhance research into the underlying mechanisms of ME movements.

2 BACKGROUND

2.1 Duration

Duration is usually used as a criterion to distinguish MEs from MaEs. The different boundaries for ME duration have been given in various research and are systematically presented in the later sub-sections.

Hypothetical Definition: In the early research of MEs, most definitions of MEs came from Ekman’s research team. In particular, the earliest definition was put forward in 1969. Ekman et al. proposed that MEs were subtle and rapid facial movements, lasting less than 1/2 s [23]. Conversely, MaEs were mostly between 1/2 s and 1 s. Subsequently, Ekman et al. proposed that when emotion was hidden or masked, the genuine emotion was expressed as a ME, a transient expression lasting between 1/25 s and 1/5 s [24]. Among these two definitions, the duration of 1/5 s is more agreed upon in the later ME experiments and is considered the classical definition. Besides, Ekman et al. had proposed definitions of ME duration of 1/4 s [25], [26] and 1/3 s [27] in other books, respectively. Nevertheless, the duration definitions of MEs listed above are hypothetical classifications that have rarely been directly verified by empirical research.

Validation Definition: In addition to hypothetical definitions, researchers have attempted to explore the duration definition of MEs through empirical approaches. The first experiment that demonstrated the existence of MEs was done by Porter et al. [28]. The duration of MEs in this study matched Ekman’s earlier definition (1/5 s -1/2 s) [24]. Besides the psychological approach, based on 3D gradient descriptors and FACS coding methods, Polikovskiy et al. indicated that the duration of MEs is between 1/25 s and 1/3 s [8]. Later, Yan et al. explored the distribution of leaked fast facial expressions and gave duration bounds for MEs with a lower limit of about 170 ms and an upper limit of about 500 ms. They also concluded that the duration of the onset phase, i.e., from the onset to the apex, could be a suitable measure for defining MEs with a lower limit of about 65 ms and an upper limit of about 260 ms [5]. However, these precise definitions of ME duration are derived based on FACS. Thus, the conclusions obtained may be inconsistent due to the subjective coding process.

Operational Definition: Due to the unknown nature of the process of ME production, researchers have also explored the operational definitions of MEs and related evidence from the recognition perspective. For instance, Ekman et al. in 1991 validated their proposed ME duration definition based on manual ME recognition. Specifically,

1. The CASMEMG database is available for application by submitting the license agreement through <http://casme.psych.ac.cn/casme/e5>

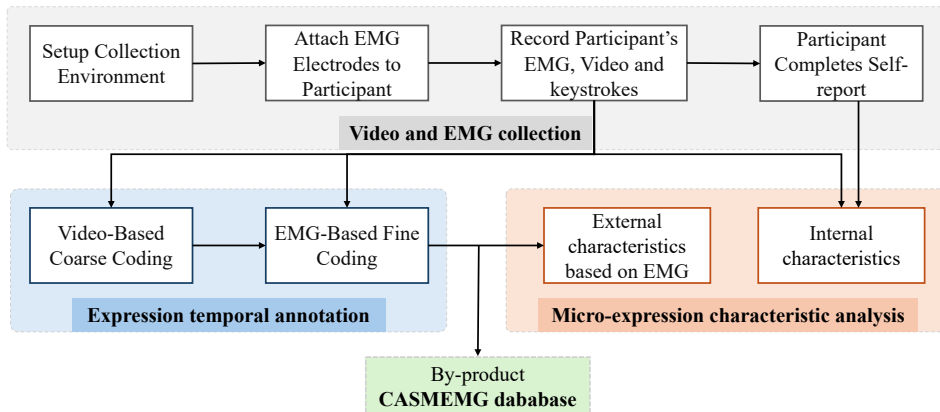


Fig. 1: Overview of EMG-based ME characteristic quantification. Our research process is divided into three main parts: Video and EMG collection, Expression temporal annotation and ME characteristic analysis. During this process, a ME database containing EMG and video, named CASMEMG, was naturally developed.

they used MEs of 1/25 s to explore the correlation between ME recognition rate and the accuracy of deception detection [29]. In addition, Shen et al. found that 200 ms was a turning point for recognition rate in exploring the relationship between ME recognition and duration [30]. When participants observed MEs in pictures with a presentation duration of less than 200 ms, the recognition rate increased with increasing presentation duration. In contrast, the recognition rate remained the same when the presentation duration of the expression pictures was 200 ms or more. In sum, the boundary point between the duration of MEs and MaEs was 200 ms.

Yet, the operational definition established from the perspective of ME recognition is different from the definition of ME production. Only MEs occurring for a long enough duration can be self-perceived by the individual. The recognition of MEs by others and the self-perception of MEs are derived from different cognitive processing and thus will yield different results.

Quantitative Definition: Although the definition of the duration of MEs has evolved from approximate estimation to manual frame-by-frame annotation in videos, the consistency of manual coding among coders is challenging. Specifically, researchers have suggested that when the video is running at normal speed, coders are unaware of facial changes occurring within 1/5 s and can only recognize “something is happening” for 2/5 s but cannot identify the specific expression [31]. Therefore, in manual annotation, coders need to observe the video repeatedly at a slow speed to detect the occurrence of MEs. In addition, due to the subtle and transient characteristics of MEs, there are differences in perception ability, subjective definition, and fluctuation of attention among coders when coding the onset and offset frames. These reasons could lead to unreliable and inconsistent manual annotation results.

With the development of high-precision physiological signal detection equipment and artificial intelligence, the duration of MEs can be measured by more objective indicators besides manual coding. Recently, Kim et al. employed electroencephalography (EEG) to detect MEs. However, even with high-resolution EEG, accurately estimating EEG power spectral features at low frequency bands such as theta and alpha requires a relatively long time window (1 s in the cited study). This time window reflects the characteristics of EEG signals over a period, not at a single moment. Consequently, it is not capable of providing information about the onset of MEs (less than 500 ms) [32]. In contrast, the temporal resolution of the EMG signal can reach milliseconds. In 2017, Perusquia-Hernandez et al. detected rapid and subtle smiles through a high-precision wearable EMG device [33]. Regarding the fact that these facial expressions are all expressed by facial muscle action, EMG is seen as an optimal choice for detecting ME and obtaining high precision duration.

2.2 Intensity

Although low intensity is considered one of the characteristics of MEs, there is no compelling empirical approach to demonstrate a significant intensity difference between the ME and MaE. Currently, the popular ME annotation method is to code AUs based on FACS proposed by Ekman et al. [17]. In particular, FACS itself is descriptive, and FACS-based coders code expressions based primarily on observations of facial muscle actions. FACS provides a subjective classification standard for AU intensity, i.e., a descriptive scale of five grades, including distinguishing between trace (A), slight (B), marked or pronounced (C), severe or extreme (D), and maximum (E) intensities. Although FACS specifies intensity coding for many action units, it is difficult to establish and maintain an acceptable level of reliability, and

further research is needed [19]. Furthermore, the FACS-based annotation of facial expressions usually requires a certified expert. In all, this approach is subjective and difficult to replicate, thus adding obstacles to the coding process.

In addition, there has been a lack of attention to intensity measurements for ME. The variation in AU intensity is sometimes treated as a separate event. In subsequent coding of facial expressions based on AU combinations, the intensity of AU is ignored and is not reflected in the coding of expressions. Moreover, in the field of computer science, relatively even less attention has been paid to intensity measurements for ME. Due to the limited number of ME databases with coded facial movement intensity, it is challenging to analyze ME intensities further [34].

Meantime, facial EMG can detect the changes in the electromagnetic field potential during muscle contraction. Particularly, the maximum voluntary contraction (MVC) [35] of facial muscles can be quantified by EMG, and thus the muscle action intensity can be expressed as a percentage of MVC, ranging from resting muscles (0%) to the maximum possible muscle contraction (100%).

2.3 Spontaneity

In the current research, the spontaneity of MEs is also one of the attention-grabbing characteristics. Specifically, the occurrence of MEs is not only difficult for observers to detect but also equally challenging to be controlled by oneself [36]. Accordingly, MEs are often thought not to appear in fabricated emotional expressions and cannot be suppressed or masked, thus providing valuable clues to the genuine emotions [37]. Furthermore, the spontaneity of MEs has become attractive for applications such as deception detection [29], [38], [39], [40].

The spontaneity of MEs is reflected in uncontrollability and unawareness. It can be inferred from the facial feedback hypothesis that individuals can be aware of expressions and emotions only when sensory feedback from facial muscles is strong enough to reach the consciousness [41]. In other words, the subjective experience of facial expressions is correlated with the duration and intensity [42]. Later, research confirmed that intensity is a significant predictor of self-awareness, which means that higher-intensity facial expressions are more likely to be aware by individuals [21]. Therefore, self-awareness of facial muscle actions is more difficult for MEs that are short in duration and lower in intensity. However, previous research has mainly focused on the application that spontaneous MEs can serve as a leakage of genuine emotions. The research on the cognition process behind the spontaneous characteristic of MEs is lacking. Through real-time self-report and subjective experience, differences in the production process of ME and MaE could be further explored.

3 METHODS

3.1 EMG-based External Characteristic Measurement

As mentioned earlier, EMG is effective in capturing and quantifying facial muscle movements through various signal characteristics. In this study, we used EMG to objectively quantify the external characteristics of MEs, specifically

focusing on their short duration and low intensity. We first synchronously collected EMG signals and facial videos. Using manually coded facial expressions from the video, we identified corresponding EMG segments for further analysis. The following subsections detail our EMG acquisition hardware, electrode distribution, signal pre-processing, and the EMG indicators used to characterize the external features of MEs, thereby clarifying our EMG-based methodology.

3.1.1 EMG Acquisition Hardware

To meet our experimental requirements, we designed a custom EMG acquisition module. Subsequent paragraphs will provide an in-depth overview of design of this module, including basic configuration, sampling rate, data transmission, and synchronization with video.

Regarding the EMG signal acquisition device, we needed to capture EMG signals from seven channels to ensure a comprehensive acquisition of facial muscle activities. Meantime, commercial EMG modules are typically limited to recording only two or three channels at a time. To fulfill our acquisition needs, we would have had to use three or four commercial EMG modules, which are both expensive and complex. Additionally, the underlying interfaces are not accessible to end-users. This makes it challenging to flexibly configure commercial EMG modules for experimental purposes. Thus, to reduce costs and customize the device to our needs, we designed our own EMG module based on the Texas Instruments® ADS1299 chip. In detail, it is an eight-channel, low-noise, 24-bit, simultaneous-sampling delta-sigma ($\Delta\Sigma$) analog-to-digital converters (ADCs) with a built-in programmable gain amplifier (PGA), internal reference, and an onboard oscillator [43]. Comparative experiments with commercial modules confirmed the reliability of our system, demonstrating its capability to meet our experimental needs [44].

Regarding the sampling rate, our EMG module allows sampling at 500 Hz, 1000 Hz, or 2000 Hz. For facial EMG signals, predominantly in the 20-500 Hz range [45], we selected a 1000 Hz sampling rate. This rate meets the Nyquist-Shannon theorem [46], ensuring accurate signal

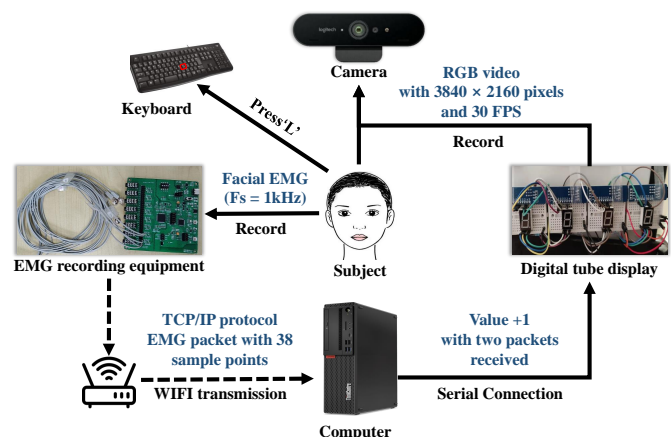


Fig. 2: Equipment layout based on data collection and transmission

capture without aliasing. Sampling at 2000 Hz would introduce unnecessary noise, making 1000 Hz optimal for our experiments.

Regarding the data transmission, Fig. 2 shows the arrangement of acquisition devices and their transmission modes. To prevent interference from utility power, we used Wi-Fi to wirelessly transmit EMG data from our custom module to the computer. Specifically, the data was sent via a 2.4 GHz wireless router using the TCP/IP protocol. TCP, being connection-oriented and reliable, ensures consistent data transmission once a connection is established.

For synchronizing the EMG signal with the video, we integrated a digital tube display into the video background, as illustrated in Fig. 3. The digital tube displayed a changing number, captured alongside the participant's facial muscle movements by a 4K camera, ensuring synchronization between the video frames and the EMG signal. Specifically, when the computer receives two EMG signal packets, the value displayed on the digital tube increments by one. The camera captures this change in real-time, allowing us to align the EMG data packets with the corresponding video frames. Specifically, with an EMG sampling frequency of 1 kHz (1000 samples per second) and each packet containing 38 samples, approximately 26 packets are received per second. Consequently, the digital tube display counts increase by 13 per second, aligning with the video's 30 frames per second rate. The camera simultaneously recorded these changes and the participant's facial movements, marking the continuous video with the digital tube's updates. This precise synchronization ensures that each EMG signal packet corresponds accurately to the video frame sequence, facilitating reliable data analysis.

3.1.2 Electrode Distribution

In our EMG setup, we acquire muscle action potentials by adhering non-intrusive electrode patches to the participant's face. The captured signals are surface EMG, which is widely

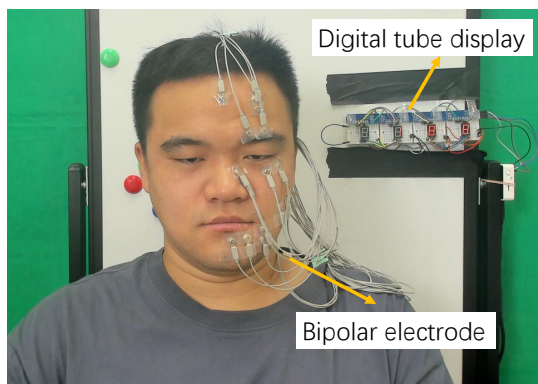


Fig. 3: Data acquisition setup. Participants were recorded on video while watching emotional stimuli. Simultaneously, EMG signals were captured through electrodes attached to the participant's face and EMG module. Furthermore, the digital tube displays helped the coder synchronize the frames in the recorded video with the EMG signals. This example image is selected from recorded videos with the participant's consent.

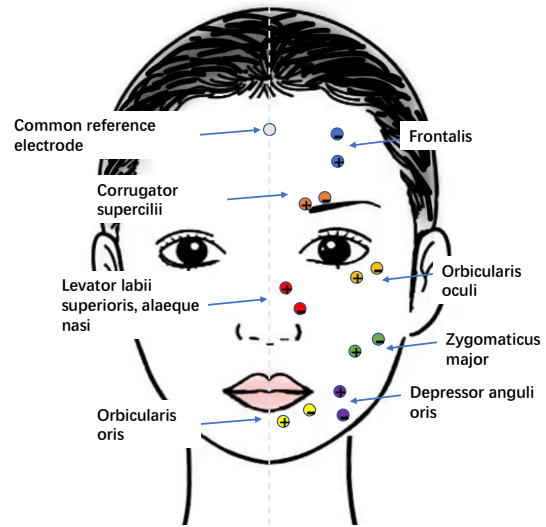


Fig. 4: Electrode position distribution for facial EMG signal acquisition based on seven facial muscles

used in many fields related to muscle movement assessment, such as sports rehabilitation and exercise [47], [48], [49]. As shown in Fig. 4, bipolar electrodes were placed near the targeted muscle tissue. Each bipolar electrode consists of two units, one at the forward input and one at the backward input of the differential amplifier. This setup measures the potential difference between the two electrodes, effectively canceling out electrical signals from distant muscles. Consequently, compared to monopolar electrodes, the bipolar configuration is less prone to interference and crosstalk.

Regarding the electrodes' distribution on the face, since we needed to balance ME coding and EMG signal acquisition, we chose the left half of the face to acquire EMG signals. Previous studies have shown that the left facial region tends to express stronger emotions than the right [50]. In particular, we selected seven muscles on the left side of the face as the central muscle regions for EMG signal acquisition, as shown in Fig. 4, namely frontalis (c1), corrugator supercilii (c2), orbicularis oculi (c3), levator labii superioris alaeque nasi (c4), zygomaticus major (c5), orbicularis oris (c6), and depressor anguli oris (c7). Specifically, as presented in Appendix A, the seven facial muscles selected in this study largely encompass the range of AUs associated with the six basic emotions. The right side of the face was left without occlusion to facilitate the coder's observation of facial muscle movements during the ME coding.

3.1.3 EMG Signal Pre-Processing

The raw EMG signal was pre-processed before the meaningful interpretation of the EMG data because the raw EMG signals usually carry a lot of noise. In our study, the EMG signal pre-processing involved removing Direct Current (DC) offset, denoise, full-wave rectification, and linear envelope, as shown in Fig. 5. All of the processes were implemented on the MATLAB platform.

Removing DC offset: The DC component represents the mean value of the signals. By subtracting the mean value

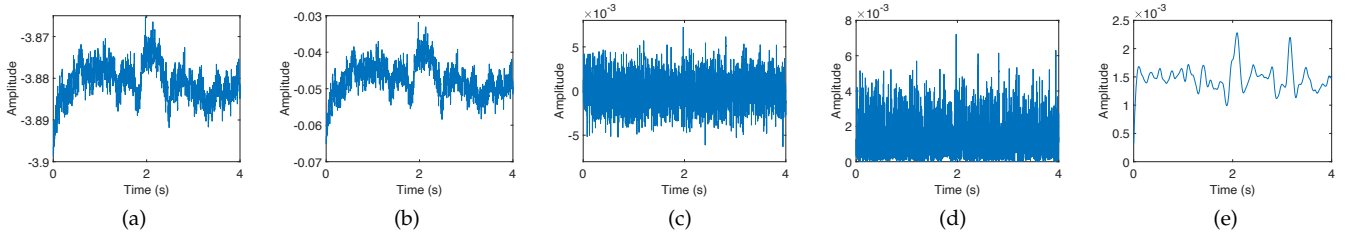


Fig. 5: Four steps of EMG signal pre-processing: removing DC offset, denoise, full-wave rectification, and linear envelope. (a) Raw EMG data; (b) EMG with DC Offset Removed; (c) EMG with Noise Removed; (d) EMG with Full-Wave Rectification; (e) Linear Envelope of EMG.

from the raw EMG signal E_r , the EMG signal with DC offset removed E_d was obtained (Fig. 5b).

$$E_d = E_r - \text{mean}(E_r) \quad (1)$$

where $\text{mean}(\cdot)$ denotes calculating the average value of the variable.

Denoise: The predominant frequency range of facial EMG signals is 20-500 Hz [45]. Thereby, to minimize noise within the EMG signals, we applied a 2nd-order band-pass Butterworth filter (BF_{bp}) from roughly 20 Hz to 450 Hz on E_d and obtained E_b (Fig. 5c).

$$E_b = \text{BF}_{\text{bp}}(E_d, 2, [Wn_1, Wn_2]) \quad (2)$$

where Wn_1 and Wn_2 represent the lower and the higher cutoff frequencies, respectively, $Wn_1 = 20 \times 2/1000$, $Wn_2 = 450 \times 2/1000$.

Full-wave rectification: The EMG signal fluctuates around the 0-value baseline. Thus, the EMG signal E_b was rectified by absolute value to achieve full-wave rectification of the EMG signals [51] (Fig. 5d).

$$E_f = \text{abs}(E_b) \quad (3)$$

where $\text{abs}(\cdot)$ denotes calculating the absolute value of the variable.

Linear envelope: A more intuitive data representation could facilitate researchers to compare and analyze the amplitude and wavelength of the EMG signals. The linear envelope of the EMG signals E_l is implemented through a low-pass filter which is a 2nd-order Butterworth filter (BF_{lp}) with a low-pass cutoff frequency of 6 Hz (Fig. 5e).

$$E_l = \text{BF}_{\text{lp}}(E_f, 2, Wn) \quad (4)$$

where $Wn = 6 \times 2/1000$ denotes the normalized cutoff frequency. All the following data analysis is based on the envelope EMG signal E_l .

3.1.4 EMG Indicators

Two indicators of EMG signals, $\text{MVC}\%$ and iEMG , are chosen for the quantification of ME external characteristics.

MVC%: For each facial expression, the EMG signal of the channel with the most distinct signal variation at the appearance of the expression is conserved as the specific EMG signal for this facial expression.

$$A_{i,j}^k = \max(\text{Amp}(E_{t,i,j}^k)) \quad (5)$$

$$E_{i,j} = E_{t,i,j}^k \quad \text{where } k = \arg \max_{k \in \{1, \dots, 7\}} (A_{i,j}^k)$$

where i , j , and k denote the index of participants, the expression, and the EMG channels, respectively. Then, as shown in Eq. 6, the amplitude peak value $A_{i,j}^k$ of the EMG signal $E_{i,j}$ of each facial expression was divided by the MVC of the corresponding channel to obtain the amplitude ratio. This ratio, denoted as $\text{MVC}\%$, represents the degree of muscle activation. This relative comparison could minimize between-participant variability [35].

$$\text{MVC}\%_{i,j}^k = \frac{A_{i,j}^k}{\text{MVC}_i^k} \quad (6)$$

iEMG: The integrated ENG (iEMG) represents the mathematical integration of the absolute value of the EMG signal $E_{i,j}$. In other words, this indicator measures the sum of muscle fibers' temporal and spatial electrical activity in the recorded electrode areas [52]. Furthermore, as shown in Eq. 7, the EMG signals were normalized, enabling the iEMG result comparison across participants.

$$\text{iEMG}_{i,j} = \int_{t_1}^{t_2} |\text{Norm}(E_{i,j}(t))| dt \quad (7)$$

where t_1 and t_2 represent the begin and ending time of facial expression, respectively. Norm denotes the normalization calculation.

3.2 Self-report-based Internal Characteristic Measurement

In our study, we employed the second-generation ME elicitation paradigm [53], using intense emotional stimuli to evoke participants' emotions, as described in Section 4.3. Participants were instructed to maintain a neutral facial expression while viewing these stimuli. Video stimuli often trigger sensations, perceptions, and emotions, leading to facial expressions. We assessed participants' awareness and control over these expressions using two self-report methods. First, while viewing the stimuli, the participants were required to press the 'L' on the keyboard when they were aware of an expression leakage. The keystroke response responds to the rate of self-awareness. Subsequently, after viewing each video clip, participants were asked to assess their ability to control their facial expressions using a 9-point Likert scale. The scale ranges from 0 (no control over facial expressions) to 8 (complete control), allowing participants to select the number that best represents their level of self-control during the video [54]. Using the two self-report methods, we could analyze whether there is a significant difference between the occurrence of MEs and MaEs.

4 EXPERIMENT

4.1 Participants

Thirty-five participants (17 females; Mean age = 24.7 years, Standard Deviation = 1.6 years) participated in the experiment. They all had normal or corrected-to-normal vision. In addition, none of them had a history of neurological, psychiatric, or other serious illnesses that could affect the experiment. Each participant was given a detailed description, signed an informed consent form before the experiment, and paid after finishing the experiment. Our study followed the Declaration of Helsinki and was approved by Institute of Psychology, Chinese Academy of Sciences.

In addressing individual differences, our study primarily investigates the common patterns of MEs. Therefore, we recruited participants aged between 18 and 30 years, a range associated with peak crystallized intelligence. This selection ensures optimal experimental outcomes, facilitating the identification of underlying patterns. Besides, we examined the relationship between individual traits and ME spontaneity within our sample of 35 participants. The analysis revealed no significant correlations between ME frequency and several individual traits (See more detail in Appendix C). Furthermore, extensive research, including studies by Ekman and colleagues [55], has demonstrated high consistency and universality in the recognition of six basic facial emotions: happiness, anger, sadness, fear, surprise, and disgust. Although our participant group consists of Chinese university and graduate students, these six basic emotions are widely recognized as universal across different cultures.

4.2 Experiment Scene

All equipment was arranged in a quiet, soundproof lab with stable, soft lighting provided by two LED lights with reflective umbrellas. A monitor with 2564×1440 pixels resolution and speakers was placed in front of the participants to play the stimuli. A Logitech C1000e camera positioned on the monitor recorded the participants' facial responses at a resolution of 3840×2160 pixels and 30 FPS. The captured facial areas had a resolution of approximately 700×900 pixels. As shown in Table 1, the SAMM database [56] and MMEW [57] have the highest facial region resolution, around 400×400 pixels, among published ME databases. Our facial region resolution is four times higher, enabling more accurate coding of facial expressions. The camera's 30 FPS frame rate effectively captures facial movements at intervals above 33 milliseconds, which suffices for our purposes. In our study, video recordings of MEs corroborate EMG data, helping us identify and match the timing of expressions with corresponding EMG signals. EMG signals sampled at 1000 Hz are then used to precisely quantify the temporal characteristics of MEs.

4.3 Elicitation Materials

Numerous methods are commonly used to elicit participants' emotions, with video elicitation offering advantages such as dynamism and non-deceptiveness [59]. We selected 18 storytelling video clips from two databases that are frequently used to elicit expressions and MEs. Specifically,

12 clips were ME elicitation stimuli from the CASME database [10], [11], and six were chosen from the Chinese Positive Emotion Database (CPED) [60]. To enhance the emotional variety, six additional open-source clips from films were added. In total, as listed in Table 2, there are 24 video clips, with four clips representing each of the six emotion categories: happiness, sadness, disgust, anger, fear, and surprise [61].

4.4 Procedure

The experimental procedure, illustrated in Fig. 6, began with a detailed explanation of the study's requirements to the participants, followed by the signing of an informed consent form. During the preparation phase, electrodes were attached to the participants' faces for facial EMG acquisition in parts A and B.

In part A, participants were guided to perform seven different facial muscle movements corresponding to the seven EMG channels. Each movement was repeated three times at maximum intensity to capture the MVC on the seven facial muscles, which served as a baseline for further EMG-based ME intensity analysis. However, this baseline MVC data is not presented to the coders during the manual coding process, as it is unrelated to their task, i.e., ME identification and localization without coding the intensity.

In part B, we elicited MEs using 24 emotional video stimuli, presented in a randomized, balanced order to minimize stimulus interaction. These clips, lasting 30 to 120 seconds each, were evenly distributed across six emotion categories. Participants were instructed to maintain a neutral expression while watching the stimuli to encourage the emergence of MEs when trying to hide genuine expressions. Additionally, participants were asked to press the 'L' key if they perceived an expression leak. After viewing each video, they completed a 9-point Likert scale to rate their self-control over their emotions.

TABLE 1: Facial region resolution summary of public spontaneous ME databases.

Databases	Resolution	Databases	Resolution
CASME [10]	150×190	SMIC [9]	190×230
CASME II [11]	280×340	CAS(ME) ² [12]	200×340
SAMM [56]	400×400	MMEW [57]	400×400
CAS(ME) ³ [53]	250×300	4DME [58]	160×200

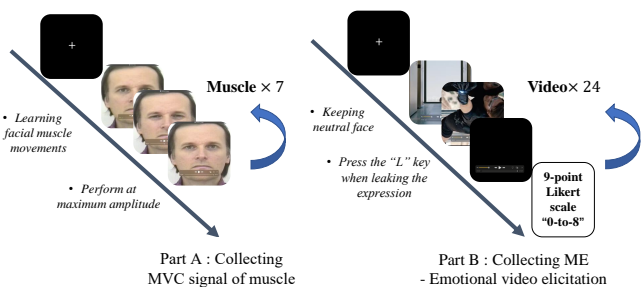


Fig. 6: The experimental procedure consists of two parts, A and B. Part A collected the MVC of each muscle, and part B used video stimuli to evoke expressions.

TABLE 2: Video clips selected as elicitation stimuli for this experiment.

Clip Name	Source	Emotion	Time	Clip Name	Source	Emotion	Time
Massacre	CASME	Anger	144s	School bullying	CPED	Anger	220s
Dog abuse	CASME	Anger	96s	Attack	Film	Anger	144s
Accident	CASME	Sadness	117s	CJ7	CASME	Sadness	140s
Farewell	CPED	Sadness	159s	Olympic Games	CASME	Sadness	91s
Eat Cockroach	CASME	Disgust	42s	Young emperor	CPED	Disgust	145s
Eat Worm	CASME	Disgust	34s	Myopia operation	CASME	Disgust	116s
Spider	Film	Fear	25s	Toilet psychic incident	CPED	Fear	68s
Elevator thriller	CASME	Fear	77s	Teiko	CASME	Fear	161s
Climb the window	Film	Surprise	67s	War	Film	Surprise	51s
Magic	Film	Surprise	53s	Car chase	Film	Surprise	35s
Honey-trap	CASME	Happiness	67s	Cartoon	CASME	Happiness	92s
Vocal concert	CPED	Happiness	172s	Won the game	CPED	Happiness	90s

4.5 Facial Expression Coding

To identify the EMG signals corresponding to facial expressions, we first coded the expressions in the recorded videos frame by frame using the Facial Action Coding System (FACS). We then searched for distinct EMG signal amplitudes within the time intervals corresponding to the onset and offset frames of the coded expressions. Specifically, one trained coder manually annotated the videos, marking the onset, apex, and offset of all emotional expressions while excluding facial movements caused by mechanical actions like swallowing and blinking. After coding, we synchronized the timing of these facial expressions with the EMG signals using the digits displayed on the digital tube. As illustrated in Fig. 7, if the expression’s occurrence coincided with a variation in the EMG signal, the channel exhibiting the most significant change was selected for further analysis as the EMG signal representing that expression.

5 RESULT AND DISCUSSION

In the experiment, 406 facial expressions were initially coded, but 26 were excluded during data processing. Specifically, 21 expressions were removed due to the absence of corresponding EMG signals, likely because only seven facial muscles on the left side were monitored, leaving other facial regions uncaptured. Additionally, 5 expressions were excluded due to the failure in capturing the maximal voluntary contraction in some participants, preventing cross-participant and cross-channel comparisons. Consequently,

380 expressions with corresponding EMG signals were included in the analysis, comprising 233 MEs and 147 MaEs. **Specifically, we did not control the interval between the appearances of MEs and MaEs. The entire data collection process was spontaneous, meaning that the timing of expression displays varied naturally among participants. Besides, our study employed an emotion suppression paradigm, requiring participants to maintain a neutral facial expression as much as possible. This condition resulted in a significantly higher number of MEs compared to MaEs, as MaEs are more easily detected and consciously controlled.**

5.1 External Characteristics

We analyzed the data of EMG signals to explore the two external characteristics of MEs: low intensity and short duration.

5.1.1 Low Intensity

MVC% was used as an EMG indicator to measure the intensity characteristics of expressions. The basic statistics of the MVC% are listed in Table 3. We analyzed each facial expression using Independent Samples T-test analysis to explore whether there were differences in intensity under the ME and MaE. The results showed that the intensity differed significantly between the two groups, reflected in the left side of Fig. 8. **MVC% was significantly lower for MEs ($M = 8.11, SD = 8.54$) than for MaEs ($M = 23.09, SD = 21.27$), $t(378) = 9.6, p < 0.05, d = 1.01$.** MVC% of MEs was lower than MaEs by 1.01 times the standard deviation. According to Cohen’s criterion, the effect size could be considered to be at a high level.

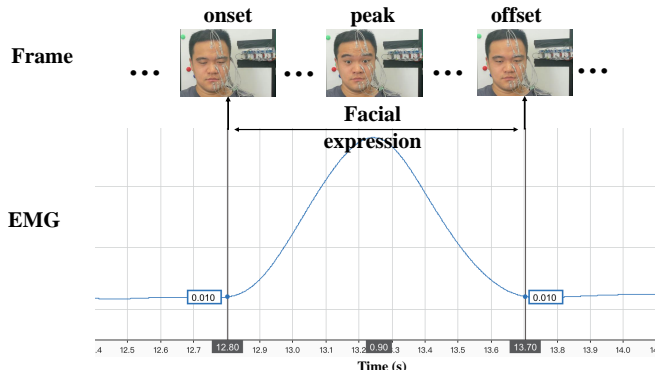


Fig. 7: EMG signal extraction based on the frame-based annotation of facial expression

TABLE 3: The numerical distribution of the two indicators of ME and MaE. N , $Mean$ and SD represent the number of facial expressions, the average value, and the standard deviation, respectively. In addition, in the results of the T-test, df , t , p and d represents the degree of freedom, the test quantity, the error probability that the results are representative of the population, and the standard difference between the mean values, respectively.

Indicators	Type	N	$Mean$	SD	df	t	p	d
MVC%	MaE	147	23.09	21.27	378	9.6	0.00	1.01
	ME	233	8.11	8.54				
iEMG	MaE	147	440.89	459.97	378	11.89	0.00	1.25
	ME	233	78.8	55.6				

MVC% provides feasibility for directly comparing muscle activities across channels and participants. The average MVC% apparently demonstrates the characteristics of low intensity. Furthermore, the intensity significantly differs between ME and MaEs. Hence, the minor MVC% value of ME could be considered as a preliminary empirical basis for ME characteristics of low intensity.

In addition to the above qualitative observations, we also inferred EMG-based interval estimates. First, the histogram visualizes the intensity distribution (Fig. 9). Then, we used the interval estimation method in statistics to infer the intensity intervals of the population parameters of MEs. As listed in Table 4, the results of the interval estimation at the 95% confidence interval showed that the lower confidence limit of the intensity indicator MVC% of MEs was 7.01%, and the upper confidence limit was 9.21%, which means that the intensity of MEs was roughly between 7% and 9.2% of the MVC.

5.1.2 Correlation between Duration and Intensity

iEMG, as an EMG indicator with two dimensions of time and amplitude, represents the normalized integrated EMG value. **Independent Sample T-test results showed that iEMG of ME ($M = 78.8, SD = 55.6$) was significantly lower than that of MaE ($M = 440.89, SD = 459.97$), $t(378) = 11.89, p < 0.05, d = 1.25$.** iEMG for MEs was 1.25 times lower than that for MaEs in standard deviation. According to the Cohen criterion, the effect size could be considered to be at a high level. As shown on the right side of Fig. 8, the intensity and duration in the EMG signal of MEs were significantly lower than that of MaEs. Since the iEMG reflects a composite character of the amplitude and time of EMG signal, the low intensity and short duration of MEs can be tentatively verified.

Moreover, we explored the relationship between two characteristics of expressions: intensity and duration. Statistical analysis were conducted with intensity as the dependent variable and duration as the predictor variable. Linear

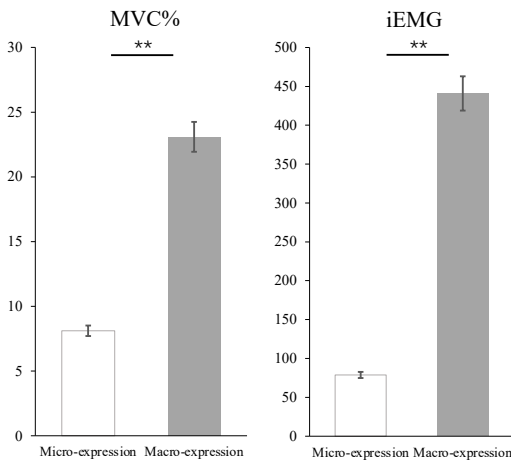


Fig. 8: Statistical analysis on MVC% and iEMG between MEs and MaEs. The EMG data of MEs and MaEs were significantly different by independent sample T-test with a 95% confidence interval. The notation ** represents $p < 0.01$.

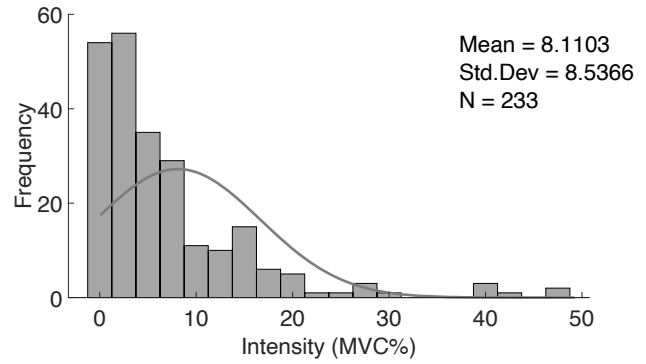


Fig. 9: Distribution of ME intensity

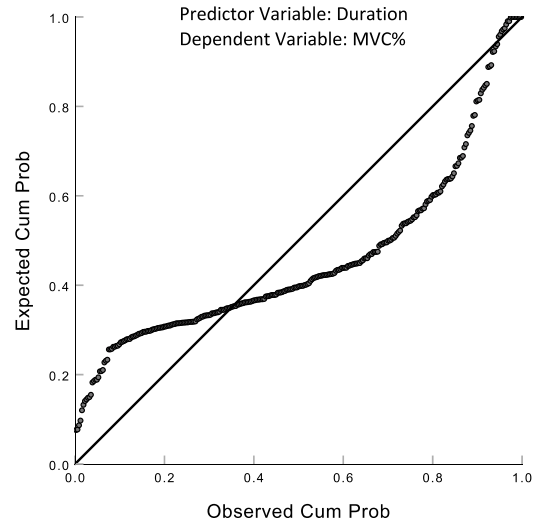


Fig. 10: Normal P-P Plot of Regression Standardized Residual.

regression analysis shown in Fig. 10 revealed that duration was a significant predictor variable of intensity, $\beta = 0.492, t(378) = 10.98, p < 0.05$. The experimental results indicate a significant linear correlation between the duration and intensity of MEs; specifically, a shorter duration corresponds to a lower intensity. This result highly coincides with the results of the iEMG indicator.

5.1.3 Short Duration

As previously described, our analysis of the correlation between intensity and duration, along with the iEMG indicator that reflects both temporal and intensity characteristics, leads us to conclude that the ME duration is brief, significantly shorter than that of MaE. This forms the basis for our further investigation into the ME duration.

TABLE 4: Interval range of duration and intensity of ME and MaE obtained by Interval Estimation

Characteristics	Type	Lower limit	Upper limit
Intensity (MVC%)	MaE	19.63	26.56
	ME	7.01	9.21
Duration (Second)	MaE	1.027	1.295
	ME	0.307	0.327

The duration distribution histogram is illustrated Fig. 11. Similar to the intensity estimation, we also performed an interval estimation of the duration of MEs. The results showed that the lower confidence limit for the ME duration was 0.307 s, and the upper confidence limit was 0.327 s, with a confidence probability of 95%. This interval estimation could be representative of the population parameters behind the ME samples used for duration analysis. Our results for ME duration interval estimation are consistent with the most commonly used definition of 1/2 s by current researchers [8], [27]. As listed in Table 4, we also obtain interval estimates of the duration and intensity of MaEs. Overall, by taking advantage of EMG, we inferred the main distribution intervals of two external characteristics for MEs, which lays the foundation for future research on the physiological mechanisms.

In order to further verify the reliability of the duration and intensity characteristic intervals, we used *k*-means cluster for an unsupervised analysis on 380 expression samples, i.e., the expression samples were not classified as ME and MaE based on the 500 ms threshold in clustering analysis. Specifically, *k*-means [62] algorithm divided 380 indicator combinations of duration and intensity into two clusters. The positions of the centroids in two clusters were optimized by the iterations calculating the indicator-point-to-cluster-centroid distance. As shown in Fig. 12, the clustering result showed that the intensity and duration of one of the categories were within the above range of the psychological-based interval estimation of ME shown in Fig. 9 and Fig. 11. This consistency indicates that MEs and MaEs can be distinguished just as well when departing from the 1/2 s division criterion. Thus, through statistical analysis, psychological analysis, and clustering calculations, our study shows that MEs can be effectively distinguished from MaEs by intensity and duration. This conclusion enriches the previous view that duration is the only difference between MEs and MaEs [30].

Apart from analyzing ME duration in our experiment, we also explored the definition of ME duration in previous studies. While many definitions exist, our interval estimation aligns closely with Ekman and Polikovsky [8], [27]. We addressed the inconsistencies in other results, noting that the process of ME production differs fundamentally from ME recognition. For instance, definitions like 1/25 s and 1/5 s [24], derived from recognition methods, are not suitable for defining the actual duration of ME production.

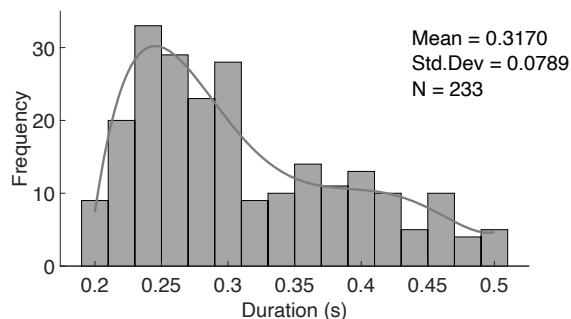


Fig. 11: Distribution of ME duration

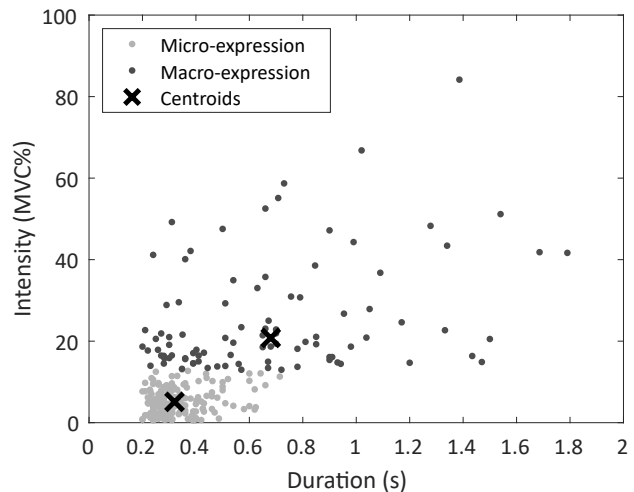


Fig. 12: *k*-means clustering for intensity (MVC%) and duration. The 380 facial expressions were divided into two clusters. The crossed symbols in the figure represent the centroids of these clusters, aligning with the distribution and interval estimates of duration and MVC%. For better visualization of sample distribution, facial expressions lasting longer than the mean plus variance in each category were excluded from the figure.

MEs, as theoretically defined, are brief, involuntary leaks of genuine emotion, reflecting the cognitive state at the moment of occurrence. Meanwhile, the extremely small value of 1/25 s, likely reflects the minimum threshold for visual recognition by the human eye [63], rather than a meaningful duration of ME occurrence. The significance of our precise interval is to define ME duration with a 95% confidence level of high occurrence probability, unlike the vague temporal definitions used before.

Furthermore, we discuss the commonly used “less than 1/2 s” definition in computer science, highlighting the lack of empirical evidence from cognitive neurology to support it. Without concrete neural or physiological markers, the boundary for ME duration is behaviorally derived: the shorter the duration, the more likely an expression is to be a ME. However, the broad 1/2 s boundary is more practical for computer-based ME analysis, which requires large sample sizes.

In summary, previous studies relied solely on duration to define MEs, without considering intensity [23], [24], [25], [27]. We demonstrated significant differences between MEs and MaEs in both duration and intensity. Our study proposes intensity as an additional criterion, offering a more precise definition of MEs. This two-dimensional approach of duration and intensity serves as a foundation for intelligent ME analysis, enabling more targeted machine learning methods and improving the distinction between MEs and other facial movements in practical applications.

5.2 Internal Characteristic

The external appearance of MEs occurs only after the individual experiences emotional reactions at the cognitive level. Therefore, after analyzing the external characteristics

TABLE 5: The numerical distribution of self-awareness and facial expression frequency. The response with a keystroke is represented as aware, and the response without a keystroke is represented as unaware. ME and MaE represent micro- and MaEs, respectively.

Keystroke response	ME	MaE
Aware(Yes)	17 (7%)	39 (27%)
Unaware(No)	216 (93%)	108 (73%)
Total (Frequency)	233 (100%)	147 (100%)

of MEs, we attempted to explain the internal characteristics through the rate of awareness and control, via keystroke response and self-reports. In particular, Chi-square tests were conducted separately for the frequencies of keystroke response and control rate reported in the self-reports.

Besides, we would like to make a distinction between the two terms in order to avoid confusion. In the context of this paper, unawareness represents the ability of the individual’s senses to be unable to perceive, while unconsciousness is the state of the individual’s inability to mental percipience.

5.2.1 Self-Awareness

The types of expression (ME or MaE) and the keystroke response (yes or no) were analyzed for correlation. The basic statistics of the keystroke frequency are shown in Table 5. There was a significant correlation between the type of expression and the ability to be aware of the expression, $\chi^2(1, N = 380) = 26.54, p < 0.05, Cramer’s V = 0.26$. More expressions reported by individuals were MaEs (70%). On the contrary, most of the MEs were not reported (93%). This result indicated that individuals were more likely to be aware and report MaEs by keystrokes and less likely to be aware when MEs appear.

This finding suggests that ME leakage is neither consciously aware nor consciously expressed [36]. Furthermore, based on our experimental results and previous research, it can be assumed that duration can serve as a predictor of intensity, while intensity can predict self-awareness [21]. Thereby, as a facial movement with short-duration and low-intensity characteristics, MEs occur below the threshold of self-awareness.

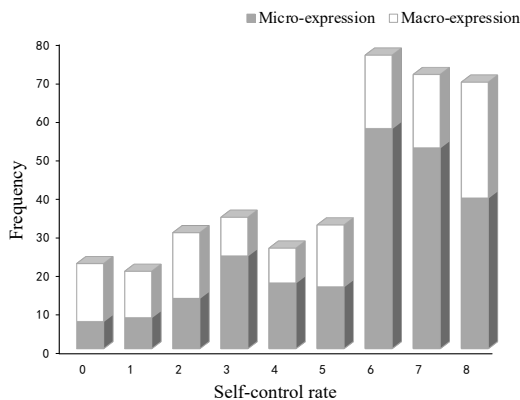


Fig. 13: The facial expression frequencies under different self-control rates

5.2.2 Self-Control

The correlation of expression type (ME or MaE) and the emotional self-control rate (0-8) were also explored. In the chi-square test, $\chi^2(8, N = 380) = 30.06, p < 0.05, Cramer’s V = 0.28$. The results showed a significant correlation. Precisely, the higher the self-control rate, the more MEs appear, as shown in Fig. 13. Individuals leaked plenty of MEs when they perceived they had higher self-control over their emotions. Therefore, self-control over emotions could only be reflected in the restriction of MaEs. MEs were challenging to control and were spontaneous.

This finding is corroborated by our statistical analysis, which showed that MaEs are less likely to leak and are more easily suppressed compared to MEs. This result aligns with earlier hypotheses suggesting that individuals’ genuine emotions manifest uncontrollably as MEs, appearing briefly and unnoticed on their faces [3], [25], [31]. In other words, the leakage of MEs is unconscious and beyond the individual’s control. Moreover, the results support the idea that MEs can reveal an individual’s true emotions. This conclusion justifies the use of MEs as a valid indicator in practical applications and enhances the reliability of ME analysis in real-world settings.

6 CASMEMG DATASET

We organized the data from our previous study to develop a ME database, CASMEMG. This database contains 147 MaE and 233 ME samples, each accompanied by corresponding EMG signals and videos. The original raw data from which these samples were segmented is also provided. In this section, we introduce the annotation of emotion labels and provide a baseline method and results for EMG-based facial expression interval detection. The CASMEMG is open for application, with the aim of promoting the study of ME mechanisms and the dynamics of movement changes grounded in EMG signals.

6.1 Emotion Annotation

We have coded the samples in the CASMEMG dataset for Action Units (AUs) and emotions. All samples have been coded for AUs and emotions by two trained coders, exhibiting high coding consistency. Initially, the AU coding followed the requirements and guidelines of the FACS system. AUs that have emotional significance in the facial expressions were coded, covering 27 AUs related to facial expressions. Based on the FACS coding method, the onset frame, apex frame, and offset frame of the 380 expressions, along with their corresponding AU combinations, were determined.

After the two coders independently completed the coding, the consistency R_C of the AU coding was calculated. The method referred to the reliability calculation formula used in CAS(ME)² [12]:

$$R_C = \frac{N_{C_1 \cap C_2} \times 2}{N_{C_1 \cup C_2}} \quad (8)$$

where $N_{C_1 \cap C_2}$ and $N_{C_1 \cup C_2}$ represent the number of samples in the intersection and union of the coding results from two coders, respectively. In the database created for

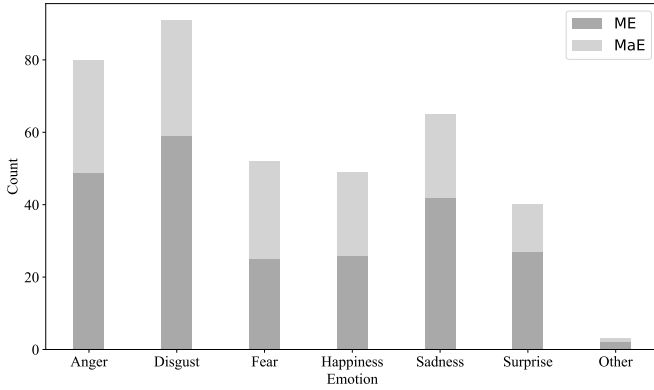


Fig. 14: Graphical presentation of data distribution with respect to each emotion categories

this study, the consistency of AU coding by the two coders was 83%, indicating high reliability. Additionally, each AU was coded for whether it was unilateral/bilateral, and it was found that only one expression had inconsistent coding, with a consistency as high as 99.7%.

Secondly, for emotional annotation, the coders need to make judgments by integrating the AU labels and the emotional type of the eliciting material. The two coders separately annotated the emotions of 380 expressions with coded AUs. Under a seven-category standard of six basic emotions and others (emotions ambiguous or difficult to classify into the six basic categories), the consistency of emotional coding was as high as 99%. The final consensus coding results, along with the coding statistics, are shown in Table. 6 and Fig. 14.

During the data collection process, participants were instructed to maintain a neutral expression while watching emotionally stimulating materials, resulting in a slightly higher number of MEs compared to MaEs in CASMEMG. Additionally, we used 24 elicitation videos, more than other datasets and almost twice as many as CAS(ME)³ (13 elicitation videos), to balance the six basic emotions as much as possible. This ensures a relatively even distribution of emotional samples in our database. Furthermore, Table. 6 shows the correlation between AUs and emotions, such as AU4 and AU7 with negative emotions, and AU12 with happiness. Additionally, anger and disgust can be differentiated by (AU16, AU23) and (AU9, AU17) respectively. Fear, another challenging negative emotion to distinguish, can be identified by focusing on AU5. Sadness and surprise each have their own distinctive AUs. Utilizing these patterns can further enhance the performance of intelligent ME analysis based on expert knowledge.

6.2 Baseline method on EMG-based ME interval spotting

We proposed a facial expression interval detection method based on EMG peak and trough detection, and compared it with manually annotated results to demonstrate the potential of semi-automatic coding based on EMG. (Code is available on <https://github.com/MELABIPCAS/CASMEMG.git>).

Regarding the methodology, the EMG signal from the channel with the largest amplitude among all channels was selected and normalized. The ratio of the mean value of the normalized EMG data E_n is used to define the threshold scale, T_r . This threshold scale is given by:

$$T_r = k \times \text{mean}(E_n) \quad (9)$$

where k is a scaling factor, and E_n represents the normalized EMG data.

A sliding window approach is used to evaluate the EMG data E_n . The number of sampling points in each window is denoted as W_L . The step size, or the number of sampling points by which the window moves forward or backward, is denoted as S_L . For each window position, the EMG energy is calculated. Windows that exceed the threshold scale T_r are marked with an indicator. This indicates that the average energy within the window is above the threshold. If the number of consecutive indicators exceeds the threshold S_n , it suggests a peak in the EMG data. This implies a potential facial movement within this segment composed of these consecutive windows. This can be mathematically expressed as:

$$\sum_{i=n}^{n+S_L-1} \mathbf{1}(x_i > T_r) \geq S_n \quad (10)$$

where $\mathbf{1}(\cdot)$ is the indicator function that equals 1 if the condition inside is true and 0 otherwise.

Once an EMG peak indicating the presence of a facial movement is identified, the onset and offset times of the facial movement are determined. The algorithm defines the troughs before and after the EMG peak as the start and end times of the movement, respectively.

For the **Onset Time**:

$$\text{Onset Time} = \arg \min_{t \in [p-W_f, p]} EMG(t) \quad (11)$$

where W_f represents the number of windows moved backward, and p is the position of the peak.

In contrast, for the **Offset Time**:

$$\text{Offset Time} = \arg \min_{t \in [p, p+W_b]} EMG(t) \quad (12)$$

where W_b represents the number of windows moved forward. The identified minimum values (troughs) within these ranges are recorded to provide the onset and offset times of the facial expressions.

6.3 Experimental result and analysis

During the algorithm validation process, the EMG segments were derived from annotated MEs. Specifically, we extended the annotated expression times by 0.5 seconds before and after within the complete EMG signal. The validation of the EMG detection algorithm involves pinpointing the start and end times of MEs within these segments. This serves two purposes: demonstrating that EMG can effectively help coders accurately locate key time points of MEs within roughly identified regions, and providing a baseline result for an EMG-based expression interval detection algorithm in the EMG database.

A total of 380 EMG facial expression samples were used in the experiment. In cases where multiple values appear

TABLE 6: The statistical classification of seven emotional categories

Emotion	ME	MaE	Total #	Dominant AUs	Main EMG channels
Anger	49	31	80	AU4, AU7, AU16, AU23	c2, c3, c6
Disgust	59	32	91	AU4, AU7, AU9, AU17	c2, c3, c4, c6
Fear	25	27	52	AU4, AU5, AU7	c2, c3
Happiness	26	23	49	AU6, AU12, AU14	c3, c5, c7
Sadness	42	23	65	AU14, AU15, AU24, AU43	c3, c6, c7
Surprise	27	13	40	AU2, AU5	c1, c3
Other	2	1	3	AU45	c3

in the onset and offset detection results, the final onset and offset values used to determine the key moments of MEs are the maximum and minimum values, respectively.

We conducted 900 iterations to explore the parameters involved in the algorithm. This process identified the optimal parameter combination: $W_L = 60$, $S_L = 30$, $k = 1$, $S_n = 5$, $W_f = 2$, $W_b = 6$.

As listed in Table 7, we analyzed the results at both the interval and moment levels. At the interval level, there were only 8 segments where the EMG detection algorithm failed to detect the corresponding interval. Additionally, we compared the overlap (IoU) rates for all detected intervals. We found that 95.3% of the intervals had an overlap rate greater than 0.5 (among 380 EMG segments). The average overlap rate for all intervals was 84.7%.

We then evaluated the performance of onset and offset position estimation. We calculated the differences between the predicted onset and offset positions and the ground truth. The table lists the mean, Standard Error (SE) and Root Mean Square Error (RSME) values. The results show that EMG can provide accurate predictions for the start and end times of expressions. It is also expected that the SE and RSME values of the onset moment are better than those of the offset moment. This is because the start of an action is usually regular, typically transitioning from weak to strong. However, at the end of an action, it does not always return to a neutral state, making accurate detection challenging as it may not form a clear trough. This phenomenon is consistent with the findings from the previous research [64].

In addition, we conducted a comparative analysis between our EMG-based ME spotting method and a top-performing vision-based method from the MEGC2022 challenge [65] under identical experimental settings. The results are summarized in the table below. Due to the limitation of small sample sizes and the very subtle motion characteristics of MEs, current ME spotting methods often exhibit relatively poor performance, with many missed detections and false positives. Since our focus is on accurately localizing ME intervals to assist annotators in efficiently labeling key frames (onset and offset), we have chosen to ignore false positives and analyze only those detection segments that successfully correspond to ground truth MEs or expressive intervals. As a result, our EMG-based interval localization method demonstrates more accurate results both in terms of interval overlap rate and differences with ground-truth key frames. Despite the superior performance of the EMG-based method in accurately localizing MEs, it is important to acknowledge the limitations associated with EMG technology, as highlighted in your second comment. Specifically, the accessibility and usability of EMG devices in real-world applications remain constrained. Developing

vision-based ME spotting methods is therefore essential for practical deployment. We expect that our EMG-based semi-automatic annotation system could facilitate the creation of a substantial and objectively labeled ME database, which is indispensable for training and enhancing vision-based ME spotting models. By providing a rich dataset, our approach supports the improvement of computer vision techniques, ultimately contributing to more effective and accessible ME spotting in real-world settings.

TABLE 7: EMG-based facial expression interval detection performance, compared with computer vision-based ME interval spotting method. N_nan represents the number of EMG segments in which expression intervals were not detected. Mean_IoU and P(IoU)>0.5 indicate the average overlap rate and the proportion of overlap rates greater than 0.5.

Metric		EMG	Vision	
Interval level	N_nan ↓	8	172	
	Mean_IoU ↑	0.847	0.268	
	P(IoU > 0.5) ↑	0.953	0.148	
Moment-level	ΔOnset	Mean ↓	0.206s	0.417s
		SE ↓	0.005s	0.089s
		RMSE ↓	0.230s	1.015s
	ΔOffset	Mean ↓	0.203s	-0.432s
		SE ↓	0.012s	0.084s
		RMSE ↓	0.309s	0.970s

7 CONCLUSION AND PERSPECTIVE

To address the lack of empirical studies on the physiological characteristics, we characterized MEs via EMG signals. The study showed that MEs could be characterized as facial expressions that are short in duration, low in intensity, and spontaneous (unaware and uncontrollable). First, this study accurately estimated the ME duration interval, with the main duration ranging from 307 to 327 ms. This duration range is a more precise time interval estimation of ME duration compared with the duration of less than 1/2 s. Second, we demonstrated that the intensity of MEs was significantly smaller than that of MaEs and then provided an estimate of the intensity interval of MEs. The maximum EMG signal amplitude of MEs ranged from 7.01% to 9.21% of the MVC. In other words, the amplitude of muscle movements during ME production is measured through a specific quantitative indicator, i.e., MVC%, which is a preliminary attempt to investigate the physiological mechanisms behind ME production in the future. Third, we also explored the internal characteristic, i.e., the cognitive dimension of the ME expression process. The results of the keystroke response and self-report studies suggest that MEs are unaware and uncontrollable.

Overall, this is the first in-depth and objective exploration of ME external characteristics by EMG. Along with a further internal characteristic analysis, this study lays the foundation for subsequent studies of ME physiological mechanisms. Furthermore, this work has guiding importance for the algorithm design of automated ME analysis. These characteristics can be used as the valuable priori knowledge for feature learning, thus improving the network’s ability to analyze MEs. Additionally, we have released CASMEMG, the first ME database that includes EMG signals, facilitating deeper exploration of ME movement mechanisms.

In addition to providing an overview of our work, this section also aims to discuss the limitations encountered in our research and the perspective.

Asymmetrical MEs: As mentioned at the beginning of Section 5, twenty-one expression samples were excluded due to the absence of corresponding EMG data. If the ME occurs only on the right side of the face, we can code the expression but cannot capture the corresponding EMG signal. Although we aimed to balance facial EMG signal acquisition and expression coding, our method still has room for improvement. Future work could explore the design of flexible transparent electrodes [66] tailored for facial EMG acquisition and optimized electrode placement along the facial contour.

High ecological validity elicitation paradigm: Due to the “expressionless” elicitation paradigm, facial movements with large intensities in our experiment were suppressed, the intensity of our collected MaEs reached only 23% of MVC. There may be differences in intensity between passively evoked expressions in the laboratory and actively flowing expressions in interpersonal interactions [67], [68]. In future studies, we will collect facial expressions in interaction situations for comparative analysis with current data.

Culture Difference: We acknowledge that different cultural background influence how emotions are expressed. Different cultures may have varying tendencies to express negative emotions either more directly or more subtly. This can affect the frequency and manifestation of MEs. In future research, we plan to recruit a more diverse participant pool, encompassing a wider age range and various nationalities and ethnicities. We will also enhance the control and analysis of cultural difference factors during the experimental process. By expanding our sample to include a balanced representation of different cultures and designing culturally tailored emotional stimuli, we aim to directly and thoroughly investigate the similarities and differences in MEs across cultures. We believe that these steps will make our conclusions more universal and reliable, and better illustrate the impact of cultural factors on MEs.

Semi-automatic EMG-based ME coding: Capturing all muscle actions in an ME with the naked eye is challenging, and traditional manual coding is prone to self-report and coder biases. In contrast, facial EMG can detect subtle muscle activations more accurately and objectively, reducing these biases (see the user study in Appendix C). However, EMG signals also capture non-emotional movements, making pre-processing difficult. Combining automatic EMG-based interval detection with manual annotation could enhance the accuracy of ME coding. This kind of semi-

automatic coding could accelerate the creation of large ME databases, facilitating improved intelligent analysis and exploration of physiological mechanisms. **Furthermore, a large-scale ME database with rich and objectively labeled samples is essential for developing high-performance ME recognition models based on the ample visual data collected. In real-world applications, the accessibility and usability of EMG devices are limited. To overcome this, we could leverage models trained on EMG-collected data and apply techniques such as model transfer and fine-tuning to achieve effective ME recognition in real-world scenarios, even in contexts with small sample sizes.**

Multi-modal models integrating EMG: Multi-modal model with computer vision techniques hold significant potential. Future research could focus on combining EMG data with video samples to develop multi-modal ME recognition models. This integration will involve addressing challenges such as modality heterogeneity, inconsistent information distribution, multi-modal data correlation, dynamics, and adaptability. By learning deep associations between emotional features from different modalities, these hybrid models could achieve comprehensive cross-modal emotional understanding and expression.

Practical Applications: EMG-based ME analysis holds significant application value in both social interactions and lie detection. In social settings, EMG-defined MEs offer valuable insights for emotion recognition without the need for wearable devices, enhancing communication in fields such as counseling and customer service. For lie detection, EMG devices effectively identify concealed genuine emotions, thereby improving the accuracy and reliability of polygraph tests and reducing the dependence on highly trained professionals. Additionally, integrating EMG-based systems with traditional polygraphs could further enhance lie detection accuracy. Future research will focus on developing smaller, more comfortable EMG sensors for practical use in social environments, combining EMG with polygraph systems to increase efficiency, and ensuring data privacy and ethical standards in sensitive applications like security and lie detection.

APPENDIX A

CORRESPONDENCE BETWEEN EMG CHANNELS, AUs, AND EMOTIONS

Table 8 lists the relationships between facial muscle channels and AUs, while Table 9 examines the primary muscle movements corresponding to different emotions.

TABLE 8: Correspondence between EMG channels and AUs

ID	Muscle	AUs
c1	Frontalis	AU1, AU2
c2	Corrugator supercilii supercilii	AU4
c3	Orbicularis oris	AU5, AU6, AU7, AU4, AU41, AU42, AU43, AU44, AU45
c4	Levator labii superioris alaeque nasi	AU9, AU10
c5	Zygomaticus major	AU11, AU12
c6	Orbicularis oris	AU15, AU16, AU17, AU18, AU22, AU23, AU24, AU28
c7	Depressor anguli oris	AU14, AU25

TABLE 9: Relationship between emotions, AUs and EMG channels

Emotion	Action Units (AUs)	Channels	Manifestation
Happy	AU6, AU12	c3, c5	Smile: Pulls the corner of the mouth back or upward.
Sad	AU45, AU15	c3, c7	Grinning; Eyes closing.
Fear	AU4, AU7	c2, c3	Frowning and eye contraction.
Surprise	AU2, AU5	c1, c3	Widening eyes and raising eyebrows.
Disgust	AU4, AU9, AU17	c2, c4, c6	Frowning, nose twist, and curling lips.
Angry	AU4, AU7	c2, c3	Frowning and eye widening.

APPENDIX B USER STUDY ON EMG-BASED CODING PROCESS

We conducted a user study comparing purely manual coding with and without the aid of EMG. This means that no algorithms were involved in detecting changes in EMG or visual signals; the process relied entirely on the observations of the coders. Specifically, **the study involves four participants: two experienced FACS-certified coder and two novice coder who received brief training (presented with 10 ME samples), separated in two groups.** Regarding the video to be coded, **we selected the 12 videos with an average duration of 124 seconds through a randomized process. Specifically, we randomly chose several videos that contained only one ME and several videos that featured multiple MEs.** This approach was primarily aimed at identifying the presence of MEs without considering the emotional categories during the selection process. These videos had previously been coded for the occurrence times (onset and offset) of MaEs/MEs by two experienced coders based on EMG and video data, serving as the ground truth for this experiment. Ten of these videos contained only one expression each, while the other two videos each contained four expressions. Detailed information about these videos is provided in Table 10. **Regarding the study setup, since the primary objective was to demonstrate that EMG assists coders in quickly locating MEs, the annotation process was timed based on when coders successfully identified the accurate positions of MEs.**

During the preliminary experiment, we discovered that novice coders found it challenging to complete the coding task within 20 minutes without any hints. Therefore, without compromising the experimental goal (i.e., proving the effectiveness of EMG in assisting ME coding), we reduced the difficulty of the coding task. The number of expressions in each video was explicitly informed to the participants, and they were only required to accurately code the onset and offset times of the expressions they identified within 20 minutes. Table 10 presents the time cost comparison of the two participants' encoding performance with and without the aid of EMG. Fig. 15 shows the average encoding times for different types of coders under various conditions. It is evident that with the aid of EMG, both experienced and novice coders experienced a significant increase in coding efficiency. **Specifically, for videos containing a single expression, the improvements for experienced and novice coders were 64.00% and 61.51% , respectively. For videos with multiple expressions, the improvements were 66.46% for experienced coders and 39.61% for novice coders. Noteworthy, without EMG, one novice coder failed to detect any MEs in two of the videos.** Besides, in post-study interviews, both experienced and novice coders reported that the EMG system

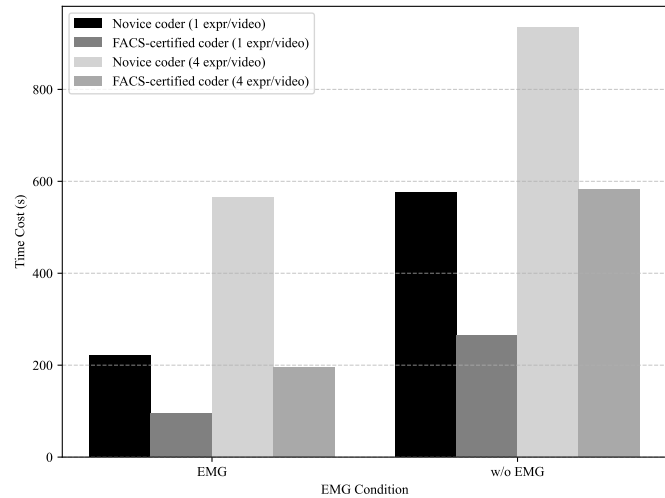


Fig. 15: Average encoding times under various conditions

greatly assisted their coding efforts. It not only enhanced their coding efficiency but also boosted their confidence in coding accuracy, making the system user-friendly.

This improvement can be attributed to the following reasons: Without the aid of EMG, coders needed to meticulously watch the entire video, often in slow motion. However, with EMG, coders could quickly pinpoint the locations of expressions based on changes in EMG signals. Additionally, it is noteworthy that the improvement was more pronounced for experienced coders. While EMG provides potential time frames for facial movements, a certain level of experience is required to filter these candidate segments and accurately time-code the facial movements. This highlights the importance of familiarity with facial action coding.

It is important to note that conducting large-scale user studies for ME coding is very challenging. On one hand, there are very few experienced coders with FACS certification, which requires extensive knowledge of facial muscle movement patterns and often a background in psychology for accurate coding. On the other hand, ME coding, especially without EMG assistance, is extremely time-consuming. For novice coders, ME coding is particularly difficult. In our table, we only recorded the time taken by coders to encode a single video. In reality, the novice coder took a total of 4 hours to complete the coding. This was partly due to the time required to thoroughly study the ME sample examples. Additionally, the novice coder found that coding MEs in a video demanded intense concentration on the screen, necessitating breaks after coding each video before proceeding to the next. These factors limited our ability to conduct large-scale user studies. However,

TABLE 10: Coding efficiency comparison with and without EMG. n_e denotes the number of facial expressions in one video to be coded. t_N and t_F represent time costs for novice and FACS-certified coders, respectively. NaN means that the novice coder failed to detect any MEs in the video.

EMG	n_e	video index	duration (s)	Group 1		Group 2	
				t_N (s)	t_F (s)	t_N (s)	t_F (s)
with	1	7-4	83	377.023	104.029	167.025	93.098
		8-2	78	345.012	72.009	105.079	66.068
		10-16	229	298.028	176.019	126.064	109.037
		11-1	153	193.017	94.005	60.093	63.065
	13-1	153	200.023	76.015	343.009	96.011	
	4	5-11	149	559.016	191.048	569.033	200.003
w/o	1	7-6	78	735.031	225.017	NaN	183.088
		8-3	52	846.097	182.025	218.096	157.028
		10-13	85	486.035	252.056	341.062	259.014
		11-12	167	904.071	469.031	287.043	216.065
	13-17	104	884.075	381.003	476.084	313.057	
	4	5-5	154	934.02	577.012	NaN	589.056

we believe that the coding experiments conducted with these two participants are representative and demonstrate the effectiveness and user-friendliness of the EMG-assisted coding system.

APPENDIX C

RELATIONSHIP BETWEEN INDIVIDUAL TRAITS AND ME SPONTANEITY

We examined the relationship between individual traits and ME spontaneity within our sample of 35 participants. The analysis revealed no significant correlations between ME frequency or detection rates and measures of emotion suppression, anxiety, depression, or personality traits. Specifically:

- Emotion Suppression: No correlation with ME frequency ($r(33) = 0.094, p = 0.59$) or detection rate ($r(33) = 0.193, p = 0.268$).
- Symptom Self-Assessment: No correlation with ME frequency ($r(33) = 0.071, p = 0.685$) or detection rate ($r(33) = 0.038, p = 0.827$).
- Beck Depression Inventory: No correlation with ME frequency ($r(33) = 0.008, p = 0.965$) or detection rate ($r(33) = 0.074, p = 0.672$).
- Big Five Personality Traits: No significant correlations across all five dimensions (Neuroticism, Agreeableness, Openness, Conscientiousness, Extraversion) with ME frequency or detection rates (see the following table).

APPENDIX D

EMOTION-SPECIFIC DURATION AND INTENSITY OF MES

We conducted a statistical analysis of the MEs associated with each of the six basic emotions. The average intensity and duration distributions are presented in Fig. 16. Our analysis revealed significant differences in both the duration and intensity of MEs across the six basic emotions. Fear exhibited the highest average intensity (10.830% MVC) and the longest duration (0.338s), indicating strong and sustained muscle activity during fearful expressions. Conversely, happiness showed the lowest intensity (6.540% MVC) with a moderate duration (0.317s), reflecting more subtle emotional

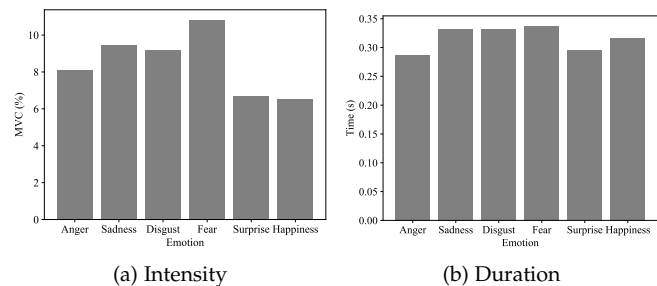


Fig. 16: Average emotion-specific intensity and duration distributions of ME

expressions. Anger had an intensity of 8.095% MVC and a slightly shorter duration (0.287s), suggesting more abrupt muscle movements. Sadness and disgust displayed intermediate intensities (9.485% and 9.170% MVC) and durations (0.332s each), while surprise had an intensity of 6.715% MVC and a duration of 0.295s. These results highlight that different emotions vary in both their intensity and temporal expression, enhancing the precision of ME detection models through emotion-specific calibration of intensity and duration parameters.

APPENDIX E

ADDITIONAL SUBSET OF CASMEMG

To enrich our EMG database, we would include an additional set of ME samples collected using a mock crime paradigm from 31 participants (13 of whom overlapped with the primary study). This resulted in 173 expressions (59 MEs and 104 MaEs). However, due to differences in the paradigms—active suppression (intentional deception) versus passive suppression (emotion regulation while viewing stimuli), the two paradigms affect the intensity and duration of MaEs differently, making combined analysis inappropriate. Nevertheless, this supplementary data enhances our database and allows future researchers to explore paradigm-specific differences.

REFERENCES

[1] C. Darwin, *Der Ausdruck der Gemüthsbewegung bei dem Menschen und den Thieren*. E. Schweizer, 1872.

- [2] P. A. Stewart, B. M. Waller, and J. N. Schubert, "Presidential speechmaking style: Emotional response to micro-expressions of facial affect," *Motivation and Emotion*, vol. 33, no. 2, pp. 125–135, 2009.
- [3] P. Ekman and M. O'Sullivan, "From flawed self-assessment to blatant whoppers: The utility of voluntary and involuntary behavior in detecting deception," *Behavioral Sciences & the Law*, vol. 24, no. 5, pp. 673–686, 2006.
- [4] P. Ekman, "Lie catching and microexpressions," in *The Philosophy of Deception*. Oxford University Press, 2009, vol. 1, pp. 118–133.
- [5] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013.
- [6] K.-H. Liu, Q.-S. Jin, H.-C. Xu, Y.-S. Gan, and S.-T. Liong, "Micro-expression recognition using advanced genetic algorithm," *Signal Processing: Image Communication*, vol. 93, p. 116153, 2021.
- [7] X. Hong, W. Peng, M. Harandi, Z. Zhou, M. Pietikäinen, and G. Zhao, "Characterizing subtle facial movements via riemannian manifold," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 3s, pp. 1–24, 2019.
- [8] S. Polikovskiy, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*. IET, 2009.
- [9] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013, pp. 1–6.
- [10] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, "CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *IEEE International Conference & Workshops on Automatic Face & Gesture Recognition*. IEEE, 2013, p. 1–7.
- [11] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "CASME II: an improved spontaneous micro-expression database and the baseline evaluation," *PLoS One*, vol. 9, no. 1, p. e86041, 2014.
- [12] F. Qu, S.-J. Wang, W.-J. Yan, H. Li, S. Wu, and X. Fu, "CAS(ME)²: a database for spontaneous macro-expression and micro-expression spotting and recognition," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 424–436, 2018.
- [13] C. H. Yap, C. Kendrick, and M. H. Yap, "SAMM long videos: A spontaneous facial micro-and macro-expressions dataset," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020, pp. 771–776.
- [14] J. Li, M. H. Yap, W.-H. Cheng, J. See, X. Hong, X. Li, S.-J. Wang, A. K. Davison, Y. Li, and Z. Dong, "MEGC2022: ACM multimedia 2022 micro-expression grand challenge," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 7170–7174.
- [15] A. K. Davison, J. Li, M. H. Yap, J. See, W.-H. Cheng, X. Li, X. Hong, and S.-J. Wang, "MEGC2023: ACM multimedia 2023 me grand challenge," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 9625–9629.
- [16] P. Ekman, W. V. Friesen, and S. S. Tomkins, "Facial affect scoring technique: A first validity study," 1971.
- [17] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.
- [18] —, "Felt, false, and miserable smiles," *Journal of Nonverbal Behavior*, vol. 6, no. 4, pp. 238–252, 1982.
- [19] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," *The Handbook of Emotion Elicitation and Assessment*, vol. 1, no. 3, pp. 203–221, 2007.
- [20] U. Dimberg, M. Thunberg, and S. Grunedal, "Facial reactions to emotional stimuli: Automatically controlled emotional responses," *Cognition & Emotion*, vol. 16, no. 4, pp. 449–471, 2002.
- [21] F. Qu, W.-J. Yan, Y.-H. Chen, K. Li, H. Zhang, and X. Fu, "you should have seen the look on your face...": self-awareness of facial expressions," *Frontiers in Psychology*, vol. 8, p. 832, 2017.
- [22] S. Lu, J. Li, Y. Wang, Z. Dong, S.-J. Wang, and X. Fu, "A more objective quantification of micro-expression intensity through facial electromyography," in *Proceedings of the 2nd Workshop on Facial Micro-Expression: Advanced Techniques for Multi-Modal Facial Expression Analysis*, 2022, pp. 11–17.
- [23] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.
- [24] —, "Unmasking the face: A guide to recognizing emotions from facial clues," 1975.
- [25] P. Ekman, "Telling lies—a how-to-guide for all those who want to detect lies," 1985.
- [26] —, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2003.
- [27] P. Ekman and E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Encoding System (FACS)*. Oxford University Press, 2005.
- [28] S. Porter and L. Ten Brinke, "Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions," *Psychological Science*, vol. 19, no. 5, pp. 508–514, 2008.
- [29] P. Ekman and M. O'Sullivan, "Who can catch a liar?" *American Psychologist*, vol. 46, no. 9, p. 913, 1991.
- [30] X.-b. Shen, Q. Wu, and X.-l. Fu, "Effects of the duration of expressions on the recognition of microexpressions," *Journal of Zhejiang University Science B*, vol. 13, no. 3, pp. 221–230, 2012.
- [31] E. A. Haggard and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of Research in Psychotherapy*. Springer, 1966, pp. 154–165.
- [32] H. Kim, D. Zhang, L. Kim, and C.-H. Im, "Classification of individual's discrete emotions reflected in facial microexpressions using electroencephalogram and facial electromyogram," *Expert Systems with Applications*, vol. 188, p. 116101, 2022.
- [33] M. Perusquia-Hernandez, M. Hirokawa, and K. Suzuki, "A wearable device for fast and subtle spontaneous smile recognition," *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 522–533, 2017.
- [34] S. Wang, L. Hao, and Q. Ji, "Facial action unit recognition and intensity estimation enhanced through label dependencies," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1428–1442, 2018.
- [35] G. J. Lehman and S. M. McGill, "The importance of normalization in the interpretation of surface electromyography: a proof of principle," *Journal of Manipulative and Physiological Therapeutics*, vol. 22, no. 7, pp. 444–446, 1999.
- [36] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.
- [37] H. G. Johnson, P. Ekman, and W. V. Friesen, "Communicative body movements: American emblems," *Semiotica*, vol. 15, no. 4, pp. 335–354, 1975.
- [38] M. G. Frank and P. Ekman, "The ability to detect deceit generalizes across different types of high-stake lies," *Journal of Personality and Social Psychology*, vol. 72, no. 6, p. 1429, 1997.
- [39] T. E. Slowe and V. Govindaraju, "Automatic deceit indication through reliable facial expressions," in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*. IEEE, 2007, pp. 87–92.
- [40] M. H. Yap, H. Ugail, and R. Zwigelaar, "Facial behavioral analysis: A case study in deception detection," *British Journal of Applied Science and Technology*, vol. 4(10), pp. 1485–1496, 2014.
- [41] S. S. Tomkins, "Affect, imagery, consciousness: The positive affects (Vol. 1)," 1962.
- [42] P. K. Adelman and R. B. Zajonc, "Facial efference and the experience of emotion," *Annual Review of Psychology*, vol. 40, no. 1, pp. 249–280, 1989.
- [43] Texas Instruments, "ADS1299-x low-noise, 4-, 6-, 8-channel, 24-bit, analog-to-digital converter for EEG and biopotential measurements," [Online], 2017, <http://www.ti.com/lit/ds/symlink/ads1299.pdf>.
- [44] J. Zhang, S. Huang, J. Li, Y. Wang, Z. Dong, and S.-J. Wang, "A perifacial emg acquisition system for facial-muscle-movement recognition," *Sensors*, vol. 23, no. 21, p. 8758, 2023.
- [45] A. Van Boxtel, "Facial emg as a tool for inferring affective states," in *Proceedings of Measuring Behavior*, vol. 7, 2010, pp. 104–108.
- [46] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.
- [47] I. Campanini, C. Disselhorst-Klug, W. Z. Rymer, and R. Merletti, "Surface EMG in clinical assessment and neurorehabilitation: barriers limiting its use," *Frontiers in neurology*, vol. 11, p. 934, 2020.
- [48] L. McManus, G. De Vito, and M. M. Lowery, "Analysis and biophysics of surface EMG for physiotherapists and kinesiologists: Toward a common language with rehabilitation engineers," *Frontiers in neurology*, vol. 11, p. 576729, 2020.
- [49] H. Türker and H. Sze, "Surface electromyography in sports and exercise," *Electrodiagnosis in new frontiers of clinical research*, vol. 2013, pp. 175–194, 2013.

- [50] W. G. Dopsion, B. E. Beckwith, D. M. Tucker, and P. C. Bullard-Bates, "Asymmetry of facial expression in spontaneous emotion," *Cortex*, vol. 20, no. 2, pp. 243–251, 1984.
- [51] U. Hess, "Facial EMG," *Methods in Social Neuroscience*, pp. 70–91, 2009.
- [52] T. Moritani, A. Nagata, and M. Muro, "Electromyographic manifestations of muscular fatigue," *Medicine and Science in Sports and Exercise*, vol. 14, no. 3, pp. 198–202, 1982.
- [53] J. Li, Z. Dong, S. Lu, S.-J. Wang, W.-J. Yan, Y. Ma, Y. Liu, C. Huang, and X. Fu, "CAS(ME)³: A third generation facial spontaneous micro-expression database with depth information and high ecological validity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 2782 – 2800, 2023.
- [54] J. P. Tangney, A. L. Boone, and R. F. Baumeister, "High self-control predicts good adjustment, less pathology, better grades, and interpersonal success," in *Self-regulation and self-control*. Routledge, 2018, pp. 173–212.
- [55] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.
- [56] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "SAMM: A spontaneous micro-facial movement dataset," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 116–129, 2018.
- [57] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng, and Y.-J. Liu, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5826 – 5846, 2022.
- [58] X. Li, S. Cheng, Y. Li, M. Behzad, J. Shen, S. Zafeiriou, M. Pantic, and G. Zhao, "4DME: A spontaneous 4D micro-expression dataset with multimodalities," *IEEE Transactions on Affective Computing*, vol. Early Access, pp. 1–18, 2022.
- [59] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition & Emotion*, vol. 9, no. 1, pp. 87–108, 1995.
- [60] Y. Zhang, G. Zhao, Y. Shu, Y. Ge, D. Zhang, Y.-J. Liu, and X. Sun, "CPED: a Chinese positive emotion database for emotion elicitation and analysis," *IEEE Transactions on Affective Computing*, vol. Early Access, pp. 1–1, 2021.
- [61] M. M. Donia, A. A. Youssif, and A. Hashad, "Spontaneous facial expression recognition based on histogram of oriented gradients descriptor," *Computer and Information Science*, vol. 7, no. 3, pp. 31–37, 2014.
- [62] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [63] F. Becattini, F. Palai, and A. Del Bimbo, "Understanding human reactions looking at facial micro-expressions with an event camera," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 9112 – 9121, 2022.
- [64] W.-J. Yan, S.-J. Wang, Y.-H. Chen, G. Zhao, and X. Fu, "Quantifying micro-expressions with constraint local model and local binary pattern," in *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13*. Springer, 2015, pp. 296–305.
- [65] J. Yu, Z. Cai, Z. Liu, G. Xie, and P. He, "Facial expression spotting based on optical flow features," in *Proceedings of the 30th ACM International Conference on Multimedia, 2022*, pp. 7205–7209.
- [66] L. Inzelberg, M. D. Pur, S. Schlisske, T. Rödlmeier, O. Granoviter, D. Rand, S. Steinberg, G. Hernandez-Sosa, and Y. Hanein, "Printed facial skin electrodes as sensors of emotional affect," *Flexible and Printed Electronics*, vol. 3, no. 4, p. 045001, 2018.
- [67] U. Hess, R. Banse, and A. Kappas, "The intensity of facial expression is determined by underlying affective state and social situation," *Journal of Personality and Social Psychology*, vol. 69, no. 2, p. 280, 1995.
- [68] S. Porter, L. Ten Brinke, and B. Wallace, "Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity," *Journal of Nonverbal Behavior*, vol. 36, no. 1, pp. 23–37, 2012.



Jingting Li is currently an associate researcher at the Institute of Psychology, Chinese Academy of Sciences (CAS). She received the PhD degree in Signal, Image, Vision from Centrale-Supélec in 2019. She served as the chair of the ACM'21, 22 and 23 FME workshop and MEGC Grand challenge, organized and hosted several China Society of Image and Graphics (CSIG) online ME workshop sessions. Her current research interests include image processing, computer vision and pattern recognition.



Shaoyuan Lu received the B.S. degree in Psychology major from Anhui Normal University, China, in 2020. She is currently pursuing the M.Psy. degree in the Institute of Psychology, Chinese Academy of Sciences. Her current research interests include facial ME, visual psychophysics and neurophysiology.



Yan Wang received the B.S. degree in Psychology major from Anhui Normal University, China, in 2020. He is currently pursuing the M.Psy. degree in the Institute of Psychology, Chinese Academy of Sciences. Her current research interests include facial ME, visual psychophysics and neurophysiology.



Zizhao Dong received the B.S.Ed degree in applied psychology major from China Women's University, China, in 2018. She is currently pursuing the M.Psy. degree in the Institute of Psychology, Chinese Academy of Sciences. Her current research interests include facial ME, visual psychophysics and neurophysiology.



Su-Jing Wang (M'12-SM'19) is an Associate Researcher, PhD supervisor at the Institute of Psychology, CAS. He received the Ph.D degree from the College of Computer Science and Technology of Jilin University in 2012. His current research interests include pattern recognition and machine learning. He won the first prize of the 8th Wu Wenjun Artificial Intelligence Science and Technology Award in 2018. He was selected as one of the top 2% of scientists in the world in 2020 for "Impact of the Year".



Xiaolan Fu received her Ph. D. degree in 1990 from Institute of Psychology, Chinese Academy of Sciences. Currently, she is a Senior Researcher at Cognitive Psychology. Her research interests include visual and computational cognition: (1) attention and perception, (2) learning and memory, and (3) affective computing. At present, she is the dean of School of Psychology, Shanghai Jiao Tong University.