Extended DISFA Dataset: Investigating Posed and Spontaneous Facial Expressions

Mohammad Mavadati, Peyten Sanger, Mohammad H. Mahoor
Department of Electrical and Computer Engineering, University of Denver
2390 S. York Street, Denver, CO, 80208
mavadati.mohammad@gmail.com, peytensanger@gmail.com, mmahoor@du.edu

Abstract

Automatic facial expression recognition (FER) is an important component of affect-aware technologies. Because of the lack of labeled spontaneous data, majority of existing automated FER systems were trained on posed facial expressions; however in real-world applications we deal with (subtle) spontaneous facial expression. This paper introduces an extension of DISFA, a previously released and well-accepted face dataset. Extended DISFA (DISFA+) has the following features: 1) it contains a large set of posed and spontaneous facial expressions data for a same group of individuals, 2) it provides the manually labeled frame-based annotations of 5-level intensity of twelve FACS facial actions, 3) it provides meta data (i.e. facial landmark points in addition to the self-report of each individual regarding every posed facial expression). This paper introduces and employs DISFA+, to analyze and compare temporal patterns and dynamic characteristics of posed and spontaneous facial expressions.

1. Introduction

Majority of the digital technologies that we use on daily bases are computationally complex devices but they lack emotional capabilities. In the last decade, researchers have been developing more intelligent Human-Machine Interaction (HMI) systems capable of understanding humans emotions and affects. This path can lead to more advanced machines that can communicate with human at a more social and primal level. Psychologists have identified a small group of emotions as *basic* emotions that are common among different cultures (e.g., anger, disgust, joy, sadness, fear, and surprise) [4]. Each emotion is defined by physiological and behavioral signatures including facial expressions, head gesture, posture, voice, and pulse rate fluctuation. Most digital devices are currently equipped with a camera and therefore creating smart HMIs that can recognize facial expressions and respond accordingly can be a promising approach.

Developing automated algorithms for computing and recognizing facial expressions has been an active research area in the last decade. Computer scientists with the help of psychologists have been analyzing prototypic facial expressions (e.g. happiness, fear, anger) [3] or facial action units described by Facial Action Coding System (FACS) [8]. Existing literature in facial expression analysis reveals that some of the existing automated systems are capable of detecting posed facial expressions with a high reliably [1, 9]. Posed facial expressions are more exaggerated and oftentimes have different facial muscle activations and dynamics compared to spontaneous ones [5]. In real life we deal with spontaneous facial expressions. Recently with the emergence of publicly available spontaneous facial datasets (e.g. CK+ [13], MMI [17] , DISFA [15], AM-FED [16] and UNBC-McMaster Shoulder Pain Expression Archive [14], BP4D-Spontaneous [21]) researchers can develop automated algorithms for measuring and detecting spontaneous facial actions and expressions. These datasets foster our knowledge and understanding to better model and recognize genuine humans emotions.

In the last few years, there has been a significant needs and interests for investing and studying spontaneous facial expressions. This has inspired scholars to provide a few spontaneous dataset that can be used as benchmark data for developing and evaluating various algorithms. Majority of these datasets annotated the presence of Action Units (AUs). Action units provide principle facial muscle activation on all facial expression [8]). For instance surprise can be describe by combination of AU1 (inner brow raise) and AU2 (outer brow raise) that can occur with AU26 (jaw drop).

The UNBC-McMaster Shoulder Pain Expression Archive [14] is a spontaneous facial expression datasets that was released in 2011 and provides facial expressions associated with pain. This database includes both pain related AUs coded with corresponding pain intensity levels. Belfast [18] used Lab-based emotion induction tasks to capture the emotional responses and of individuals and the self-reported intensity of each emotion. DISFA [15] provided the spontaneous responses of 27 participants watching emotive video clips. DISFA provides the intensity levels of 12 AUs for over 120,000 frames. Affectiva-MIT Facial Expression Dataset (AM-FED) [16] is a naturalistic set of facial expressions collected in-the-wild. AM-FED provides the label for presence of 10 symmetric and 4 asymmetric AUs. Despite the existence of several posed and spontaneous datasets, there is no data for comprehensively compare posed and genuine facial expressions. The question is: *"why and how analyzing posed and spontaneous facial expressions can help researchers?"*

Posed and spontaneous facial expressions can incorporate different facial muscles. Oftentimes, in posed facial expressions, humans intentionally control and move specific set of facial muscles. However in the genuine facial expressions the unconscious facial expressions emerge, which links to emotional states of an individual. Limited studies for some facial actions suggest that posed and spontaneous AUs can exploit different appearances (e.g. different AU co-occurrence and various dynamics and temporal patterns). For instance, [5] shows that spontaneous smiles, in contrast to posed smiles (like a polite smile), are slow in onset, can have multiple AU12 apexes (multiple rises of the mouth corners), and are accompanied by other AUs (e.g. AU6: cheek raiser). Valstar et al. also illustrated that an automated system can differentiate between posed and spontaneous brow activities with about 90% accuracy [20].

The lack of available data for comparing posed and spontaneous expression for same set of participants, encouraged us to capture and annotate a diverse collection of posed and spontaneous facial actions. We called this dataset Extended-DISFA (DISFA+). DISFA+ contains the videos and AU annotations of posed and spontaneous facial expressions of nine participants where the intensity scores of twelve AUs for all of the frames were annotated. DISFA+ provides a ground-truthed data, landmark points, subject-based self report and quantitative and qualitative comparison between posed and genuine facial muscle activations.

The novelties of this article are to present and release a fully annotated subject-based facial expression data in both posed and spontaneous contexts. This dataset presents the 5-level intensity of facial actions in frame-based level. In the following sections, we introduce some initial benchmark comparison between a group of 12 facial actions in posed and genuine settings. In other words, this paper raises this important question: *how the posed facial expressions are different from the spontaneous ones?* To answer this question, we introduced a group of measurements to compare facial expressions and explore their characteristics. The proposed list of metrics includes the temporal characteristics and inherent relation of facial muscle activations. Such knowledge will help researchers to better model and understand facial expressions. In addition the results of our experiments and accessing to DISFA+ will help researchers to build automated algorithms to model and train reliable emotion-aware systems.

The remainder of this paper is organized as follows. Section 2 introduces required terminology for facial expression analysis (e.g. dynamics of facial expressions, event, onset, offset, co-occurrence). Section 3 describes DISFA+ dataset and capturing setting, dataset content and meta-data. Section 4 presents the experimental setting used in our running experiments and discusses the results on comparing posed and spontaneous data. Section 5 concludes the paper.

2. Background

Following seminal efforts by Darwin [6], Duchenne [7] and Hjortsjo [10], Ekman and Friesen [8] developed FACS to describe nearly all possible facial actions. FACS decomposes facial expressions into one or more anatomically based Action Units (AUs). No other systems have such descriptive power. FACS describes facial expressions in terms of 33 anatomically based AUs [8]. The coding of intensity for each of these AUs are in five ordinal levels (A through E) [8]. Using action units empowers researchers to describe facial expressions dynamics, co-occurrence and temporal variations comprehensively.

2.1. Temporal characteristics of AUs

An important intrinsic characteristic of a facial expression is the temporal patterns of AUs. To quantitatively measure important time-related features, bellow some of these features have been defined:

- **Event:** The duration of a facial action that starts and ends with zero intensity and has intensity variations in between.
- **Peak:** The local maximum in the AU intensity curve.
- **Valley:** The local minimum in curves of the intensity of a facial action; (neutral faces and AUs with intensity zero are considered to be a valley)
- **Rising Duration(RD):** The period of time for a facial action to rise from a valley to the next peak.
- **Decaying Duration (DD):** The period of time for a facial action to decline from a peak to the next valley.
- **Peak Length:** The duration that facial action stays in the peak (apex) state.
- **Number of Peaks:** The total number of peaks in one event of a facial action.

Figure 1 illustrates the dynamics of one facial action event for lip corner puller (AU12). This event has three peaks and four valleys and the dashed lines indicate the two middle valleys of the event. One rising time and one decaying duration for AU12 are labeled in the image as well.

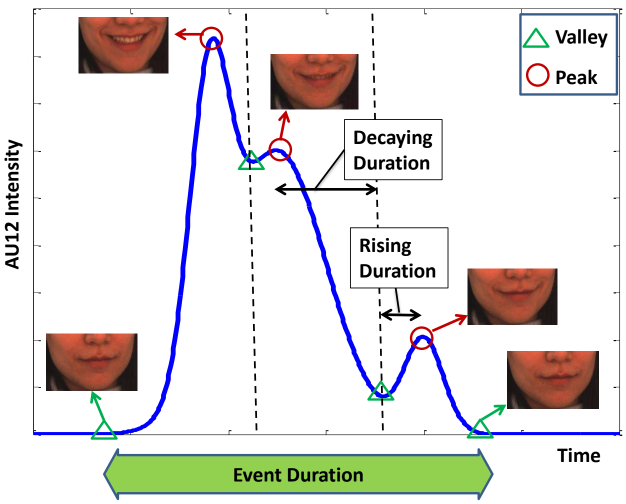
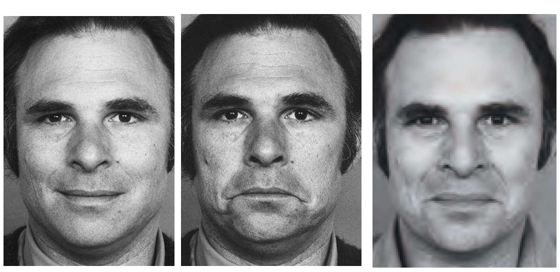


Figure 1: Describing temporal features of facial expression in one event of facial expression

2.1.1 Co-presence and co-absence of AUs

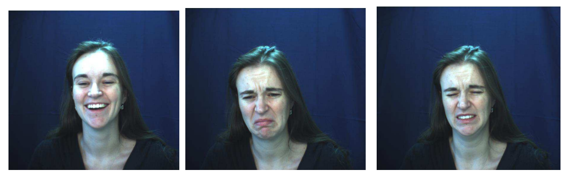
Due to the dynamics of facial actions, oftentimes FACS coders need to measure the intensity of multiple AUs simultaneously, which can be a challenging task; especially, when one AU co-occurs with other AUs in the same part of the face. This can make facial expression annotation a more difficult process as the appearance of an AU in a combination; can be largely different from its stand-alone appearance (i.e. non-additive AU combination). Figure 2 demonstrates an example of the non-additive effect, when AU12 appears alone, the lip corners are pulled up toward the cheekbone (see figure 2(a)); however, if AU15 (lip corner depressor) is also becoming active, then the lip corners are somewhat angled down due to the presence of AU15 (see figure 2(c)) . The non-additive effect increases the difficulty of recognizing AU individually.



(a) AU12 (b) AU15 (c) AU12+AU15

Figure 2: Nonadditive effect in an AU combination (a) AU12 occurs alone., (b) AU15 occurs alone., (c) AU12 and AU15 appear together ([8])

As described in the FACS manual [8], the inherent relationships among AUs can provide useful information to better analyze facial expressions. The inherent relations can be summarized as the co-occurrence relations. The co-occurrence relations characterize the groups of AUs, which oftentimes appear together to show meaningful emotional states. The co-presence of AUs can also indicate the quality and intensity of an emotion. For instance, in Figure 3(a), AU6+AU12+AU25 represents "happiness", and the two types of disgust faces (Figure 3(b) is a combination of sad and disgust emotions (AU9+15+17) and 3(c) is a disgust face with more intensity (AU9+20+25)).



(a) Happy face (b) Disgust face 1 (c) Disgust face 2

Figure 3: Co-occurrence relationships of AUs (a) AU6+AU12+AU25 , (b) AU15 occurs alone., (c) AU12 and AU15 appear together ([8])

For co-presence (co-occurrence) relations among AUs, a matrix A is defined, where each entry $a_{i,j}$ is the probability of $P(AU_i = 1|AU_j = 1)$ and the pairwise co-occurrence dependency between two AUs is computed as follows:

$$P(AU_i = 1|AU_j = 1) = \frac{N_{AU_i+AU_j}}{N_{AU_j}} \quad (1)$$

where $N_{AU_i+AU_j}$ is the total number of positive examples of the AU combination $AU_i + AU_j$ regardless of the presence of other AUs in the database, and N_{AU_j} is the total number of positive examples of AU_j in the database. For instance, $P(AU_1 = 1|AU_2 = 1) = 0.80$ specifies that when $AU2$ is activated, the likelihood of $AU1$ also contracts is 0.80.

2.2. Comparing Spontaneous and Posed Action Units

The next section will introduce the capturing setting, contents of DISFA+. In the following sections, we employ the aforementioned quantitative measures to analyze the AU dynamic relations and temporal patterns in both posed and genuine facial expressions.

3. DISFA+: A Posed Facial Expressions Database

DISFA dataset contains the spontaneous facial expressions of 27 young adults who viewed video clips intended to elicit spontaneous facial expressions [15]. To have set of data which allows us to compare the spontaneous and posed facial actions, we captured the posed facial expressions of a subset of DISFA's participants. We recruited nine out of 27 subjects of DISFA and recorded their posed facial actions. These participants have diverse ethnicities (e.g. Asian, African-American, Caucasian) and their facial images are shown in Figure 4. All of the participants signed the informed consent forms for the distribution and use of their video images for research purposes.

Figure 4: Nine subjects of DISFA+ with posed facial expressions.

3.1. Recording Procedure for Posed Expression

To acquire the posed expressions of individuals we designed a software, which instructed and guided users to imitate a set of 42 facial actions during the capturing session (some of these facial actions are shown in Figure 4). (The full list of these facial actions is provided in Table 1. Every subject watched a 3-minute demo to learn how to use the software and record her posed facial expressions for multiple trails. In each trial, the user was asked to mimic a full dynamic of facial actions (i.e. begin with a neutral face, proceed to the maximum intensity of expression and finally end with a neutral face). The user could see her face on an LCD monitor as she mimicked the expression. Figure 5 illustrates a screen-shot of our software while one of the users mimiking *'Surprise'* face. This software has been written in C++ and the OpenCV library [2] was used to record and time-stamp every frame of the video.

Each user was asked to imitate 30 facial actions (i.e. single AU or combinations of AUs) and 12 facial expressions corresponding to the emotional expression (e.g. Surprise, anger, etc.). A list of these facial expressions is provided in the Appendix. Meanwhile an HD camera recorded the facial responses of participants with 1280×720 pixel resolution in 20 frames per second. This frame rate and image resolution was selected to match with the DISFA video acquisition setting and make data comparison and analysis easy. Each individual was instructed to practice each of the facial action first and then begun to record a few trials for each facial action. Users imitated each facial action few times and after finishing each trial, participants rated (in the range of [0-10]) each mimicked facial expression by answering two questions:

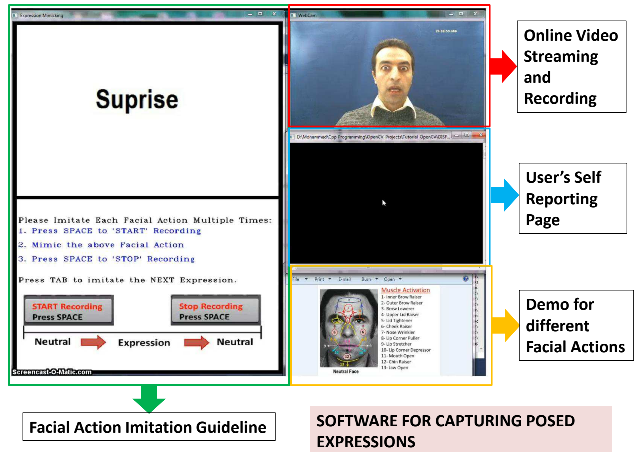


Figure 5: A demo of the designed software for capturing the posed facial expressions of subjects in DISFA+ database.

- How difficult was it for you to make each facial expression?
- How accurate could you mimic/pose the facial expression?

We selected the best trials of each individual with highest score. These trials were then FACS coded and the intensity of each frame was labeled [1].

3.2. Manual FACS Coding

To annotate the ground-truth labels of DISFA+, a certified FACS coder annotated the intensity of 12 AUs. The list of all 12 AUs in DISFA+ is provided in Table 1.

4. Experimental Results: Posed vs. Spontaneous Facial Actions

The semantic and dynamic relations among facial actions are crucial for understanding and analyzing spontaneous and posed expressions. In fact, the coordination and synchronized spatio-temporal interactions between facial actions produce a meaningful facial expression. Tong et al. [19] and Yongqiang et al. [12] employed Dynamic Bayesian Network (DBN) to model the dependencies among AUs for detecting and measuring the intensity of AUs respectively.

[1] DISFA+ metadata, contains the self-report of each individual besides the facial landmark points for every frame.

Table 1: AU description and inter-observer reliability.

AU	Description
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lowerer
5	Upper Lid Raiser
6	Cheek Raiser
9	Nose Wrinkler
12	Lip Corner Puller
15	Lip Corner Depressor
17	Chin Raiser
20	Lip Stretcher
25	Lips Part
26	Jaw Drop

Their proposed approach demonstrates that using the relation among AUs would be helpful information to better model facial expressions.

To quantitatively compare spontaneous and posed facial expressions, we utilized the introduced quantitative measures in Section 2.1. These measurement indices allow us to better explore the AU dynamic relations, temporal patterns and AU co-occurrence for posed and spontaneous facial expressions.

4.1. Temporal Characteristics of AUs

4.1.1 Subject-based analysis: posed vs spontaneous

To compare characteristics of posed and spontaneous facial expressions for each individual, we used paired t-test. Table 2 summarizes the p-value for average of rising and falling duration and also the p-value for average of duration of apex for all 12 AUs. As we have multiple hypotheses for different AUs, we use the Holm-Bonferroni method to control the family-wise error rate [11]. Table 2 report the paired mean p-value for all subjects and it highlights the set of null hypothesis that has been rejected ($\alpha = 0.05$) using Holm-Bonferroni test. The subject-based analysis indicates that the duration apex of few AUs like AU6, AU12, AU15, AU25 are significantly different between posed and spontaneous data. Also the results indicate that mouth open (AU25) has significantly different rising time in pose than spontaneous instances. Next section reports the temporal differences of posed and spontaneous data for the entire data points.

4.2. Entire-dataset: posed vs spontaneous

Table 3 reports the average (# frames) of the described temporal features (e.g. rising duration, decaying duration, peak length, number of peaks in an event, and length of events) for spontaneous and posed AUs. The objective is to quantitatively compare different characteristics of spontaneous and posed facial expressions. We used the unpaired t-test to validate a few hypotheses to differentiate between posed and spontaneous data. The 0.05 significance level was used and the highlighted AUs are those posed and spontaneous AUs that are significantly different.

The results presented in Table 3 confirms that:

- On average, the majority of spontaneous AUs (except AU5 and AU20) have slower rising time than the posed AUs. However using the two-sided unpaired t-test indicates four facial action units (AU6, AU9, AU12 and AU25) in spontaneous context have significantly different rising time than the posed ones.
- On average, all spontaneous AUs (except AU5) have slower decaying (falling slope) than the posed AUs. Moreover, considering (two-sided unpaired) t-test results confirm that the decaying characteristic for majority of AUs (i.e. AU1, AU4, AU5, AU6, AU9, AU12, AU25, AU26) in spontaneous context is significantly different from the posed one.
- Interestingly, the duration of AUs at peak intensity of facial action is largely variant among AUs but specifically the peak duration of 4 action unites (AU5, AU6, AU12 and AU25) is significantly distinct in spontaneous and posed facial expressions. It also worth mentioning that, some of the spontaneous AUs can be elicited in a very subtle way. For instance, AU15, as reported in Table 3, on average stays in the peak intensity for about half a second and then it decays.
- Some of the spontaneous facial actions (e.g.AU12 and AU25) have tendency to have multiple peaks.
- On average the duration of events for posed facial expression about 2.5 seconds. However for the spontaneous facial expressions some AUs (e.g. AU4, AU12, AU15, AU25) have very long duration (over 5 seconds) and they tend to stay on apex for longer time or having multiple peaks.

4.3. Co-presence of AUs

Co-presence of two or more AUs is important for describing and distinguishing between different emotional states of human. In our experiment we compare the co-presence of 12 AUs and came up with the co-occurrence matrix.

For example as shown in Figure 6(a), $P(AU_1 = 1|AU_2 = 1) = 0.71$ means that in the spontaneous facial expression data, when $AU2$ is activated, the likelihood of $AU1$ is also contracted is 0.71. However in posed facial expressions such probability is much higher ($P(AU_1 = 1|AU_2 = 1) = 0.90$). On the contrary, $P(AU_9 = 1|AU_5 = 1) = 0.01$ indicates that the co-occurrence of $AU5$ (eye contraction) and $AU9$ (nose wrinkler) is rarely happening in spontaneous facial expressions. In addition, in posed facial expressions some AUs (like AU5) co-occur with few other AUs (e.g. AU1,AU2, AU4, AU25, and AU26) much more frequently than the spontaneous facial expressions.

Table 2: Subject-based comparison: spontaneous and posed temporal characteristics

paired p-value(posed,spont.)	**AU1**	**AU2**	**AU4**	**AU5**	**AU6**	**AU9**	**AU12**	**AU15**	**AU17**	**AU20**	**AU25**	**AU26**
Rising Duration (P-Val)	0.49	0.16	0.48	0.16	0.09	0.11	0.09	0.63	0.73	0.18	0.08	0.65
Peak Duration (P-Val)	0.93	0.19	0.06	0.78	**0.001**	0.015	**0.003**	**0.003**	0.07	0.97	**0.003**	0.33
Decaying Duration (P-Val)	0.57	0.74	0.12	0.01	0.03	0.39	0.03	0.37	0.73	0.44	**0.003**	0.11

Table 3: Comparing spontaneous and posed temporal characteristics (Duration in frames): (RD: Rising Duration, DD: Decaying Duration, PD: Peak Duration, ED: Event Duration)

	AU1	**AU2**	**AU4**	**AU5**	**AU6**	**AU9**	**AU12**	**AU15**	**AU17**	**AU20**	**AU25**	**AU26**
RD (Spontaneous)	9.33	12.74	12.50	5.90	17.43	14.04	16.64	14.35	10.59	5.80	22.43	11.70
RD (Posed)	7.04	6.09	7.71	6.25	7.24	6.17	7.54	9.52	7.15	6.78	8.06	7.55
Unpaired T-Test (P-Val)	0.41	0.22	0.11	0.64	**0.02**	**0.04**	**0.01**	0.69	0.59	0.32	**0.00**	0.13
DD (Spontaneous)	19.92	20.05	46.19	8.39	24.34	29.83	28.97	21.26	13.67	12.90	39.03	23.19
DD (Posed)	10.39	11.20	11.60	12.51	7.17	8.34	7.69	13.44	9.59	6.66	13.70	10.47
Unpaired T-Test (P-Val)	**0.02**	0.06	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	0.38	0.53	0.08	**0.01**	**0.03**
PD (Spontaneous)	25.42	34.10	29.36	14.67	68.04	45.95	53.50	41.78	24.10	30.61	55.70	37.23
PD (Posed)	25.43	24.52	28.74	23.32	32.91	28.47	28.05	32.56	29.22	26.52	23.44	23.28
Unpaired T-Test (P-Val)	1.00	0.18	0.91	**0.00**	**0.00**	0.06	**0.00**	0.30	0.37	0.46	**0.00**	0.06
# Peaks in Event (Spont.)	1.07	1.07	1.22	1.13	1.10	1.12	1.29	1.24	1.07	1.04	1.37	1.14
# Peaks in Event (Posed)	1.08	1.08	1.04	1.09	1.07	1.08	1.06	1.06	1.08	1.05	1.04	1.05
Unpaired T-Test (P-Val)	0.88	0.87	**0.01**	0.29	0.50	0.29	**0.00**	**0.01**	0.85	0.78	**0.00**	0.07
ED (Spont.)	56.52	64.19	104.9	55.59	103.98	94.70	124.70	110.90	49.71	46.82	157.1	79.81
ED (Posed)	44.09	47.72	47.69	46.64	51.10	51.27	49.84	52.62	51.88	49.25	47.85	48.65
Unpaired T-Test (P-Val)	0.20	0.13	**0.01**	0.30	**0.00**	**0.00**	**0.00**	**0.00**	0.75	0.66	**0.00**	**0.04**

Such comparison along with the temporal information would be a benchmark for exploring posed and spontaneous facial expressions and more analyses will be advised.

5. Conclusion

We introduce a new publicly available database (DISFA+) of posed and spontaneous facial expressions, which annotates anatomically based descriptions (or AUs). An expert FACS coder labeled the intensity of each AU on a six-point intensity scale [0-5]. Using DISFA+ allowed us to investigate the temporal and dynamic characteristics of facial expressions in both posed and spontaneous domains. Comparison between the rising and decaying times of AUs revealed that the majority of spontaneous AUs has slower rising and decaying time than the posed ones. We believe accessing to a well-annotated data for same set of participants, serve as a great resource to scholars to better explore posed and unposed facial expression data. It helps researchers to develop and evaluate novel methods for spontaneous facial expression recognition that eventually will be used in affect-aware intelligent systems.

6. Appendix

To emote the posed facial expression we designed a software which loop through 42 facial expressions (30 AUs, 12 textual expressions) shown in Figure 7. The software asked each individual to pose each expression for a full dynamics (i.e. from neutral to apex back to neutral). To have a control over how people can pose prototypic facial expressions, we ask them to pose 6 prototypic emotional states using two approaches: 1) providing ”the description of emotion” by listing the set of facial muscles contributing to the emotions 2) Only using the ”emotion names”. Every individual posed each expression for multiple trials and grade herself on *how well she could pose the expression*. Some of the facial actions such as inner-brow raiser and upper lid raiser appeared to be challenging to imitate. In our analysis we selected only one trial per facial expression with highest imitation score that the participant provided.

After selecting the best posed videos for all 42 facial actions, a trained FACS labeler annotated the intensity of 12 AUs on frame-based level. Extended DISFA contains the list of videos for 9 subjects and each of them has 42 facial videos with manually annotation of the intensity of all 12 AUs. The main use of Extended-DISFA dataset is for research purposes

	AU1	AU2	AU4	AU5	AU6	AU9	AU12	AU15	AU17	AU20	AU25	AU26
AU1	1.00	0.71	0.16	0.35	0.05	0.12	0.05	0.12	0.10	0.09	0.08	0.10
AU2	0.60	1.00	0.10	0.39	0.04	0.03	0.04	0.08	0.06	0.11	0.06	0.07
AU4	0.45	0.33	1.00	0.17	0.23	0.76	0.09	0.50	0.46	0.32	0.20	0.22
AU5	0.11	0.14	0.02	1.00	0.00	0.00	0.01	0.02	0.01	0.04	0.01	0.02
AU6	0.12	0.10	0.18	0.03	1.00	0.46	0.45	0.20	0.24	0.22	0.29	0.22
AU9	0.10	0.03	0.22	0.01	0.17	1.00	0.04	0.19	0.18	0.11	0.09	0.08
AU12	0.18	0.15	0.12	0.08	0.71	0.18	1.00	0.09	0.22	0.18	0.44	0.37
AU15	0.11	0.09	0.16	0.04	0.08	0.21	0.02	1.00	0.29	0.29	0.08	0.10
AU17	0.14	0.10	0.24	0.04	0.16	0.32	0.09	0.47	1.00	0.52	0.07	0.12
AU20	0.04	0.07	0.06	0.06	0.05	0.07	0.03	0.17	0.18	1.00	0.04	0.06
AU25	0.43	0.35	0.37	0.24	0.68	0.57	0.65	0.45	0.24	0.36	1.00	0.88
AU26	0.30	0.25	0.22	0.16	0.28	0.29	0.30	0.32	0.23	0.36	0.48	1.00

(a) Spontaneous AU co-occurrence

	AU1	AU2	AU4	AU5	AU6	AU9	AU12	AU15	AU17	AU20	AU25	AU26
AU1	1.00	0.90	0.43	0.67	0.09	0.14	0.14	0.28	0.27	0.16	0.37	0.49
AU2	0.73	1.00	0.29	0.71	0.02	0.18	0.10	0.14	0.13	0.04	0.34	0.49
AU4	0.58	0.47	1.00	0.51	0.47	0.84	0.19	0.66	0.67	0.45	0.37	0.34
AU5	0.63	0.82	0.35	1.00	0.06	0.25	0.10	0.24	0.18	0.08	0.37	0.53
AU6	0.06	0.02	0.22	0.04	1.00	0.64	0.54	0.34	0.42	0.38	0.38	0.27
AU9	0.06	0.10	0.27	0.12	0.44	1.00	0.08	0.36	0.42	0.09	0.23	0.21
AU12	0.09	0.09	0.10	0.07	0.57	0.12	1.00	0.05	0.10	0.69	0.29	0.21
AU15	0.08	0.05	0.13	0.07	0.15	0.22	0.02	1.00	0.57	0.21	0.04	0.07
AU17	0.10	0.06	0.18	0.07	0.24	0.35	0.05	0.76	1.00	0.24	0.04	0.11
AU20	0.06	0.02	0.13	0.03	0.22	0.08	0.39	0.29	0.25	1.00	0.04	0.04
AU25	0.39	0.44	0.28	0.41	0.61	0.54	0.44	0.15	0.13	0.11	1.00	0.88
AU26	0.35	0.42	0.18	0.40	0.29	0.34	0.22	0.19	0.21	0.08	0.60	1.00

(b) Posed AU co-occurence

Figure 6: Comparing the co-occurrence of AUs in spontaneous and posed data: (a) spontaneous data (Entire DISFA database) (b) posed data (12 basic expression of DISFA+ database))

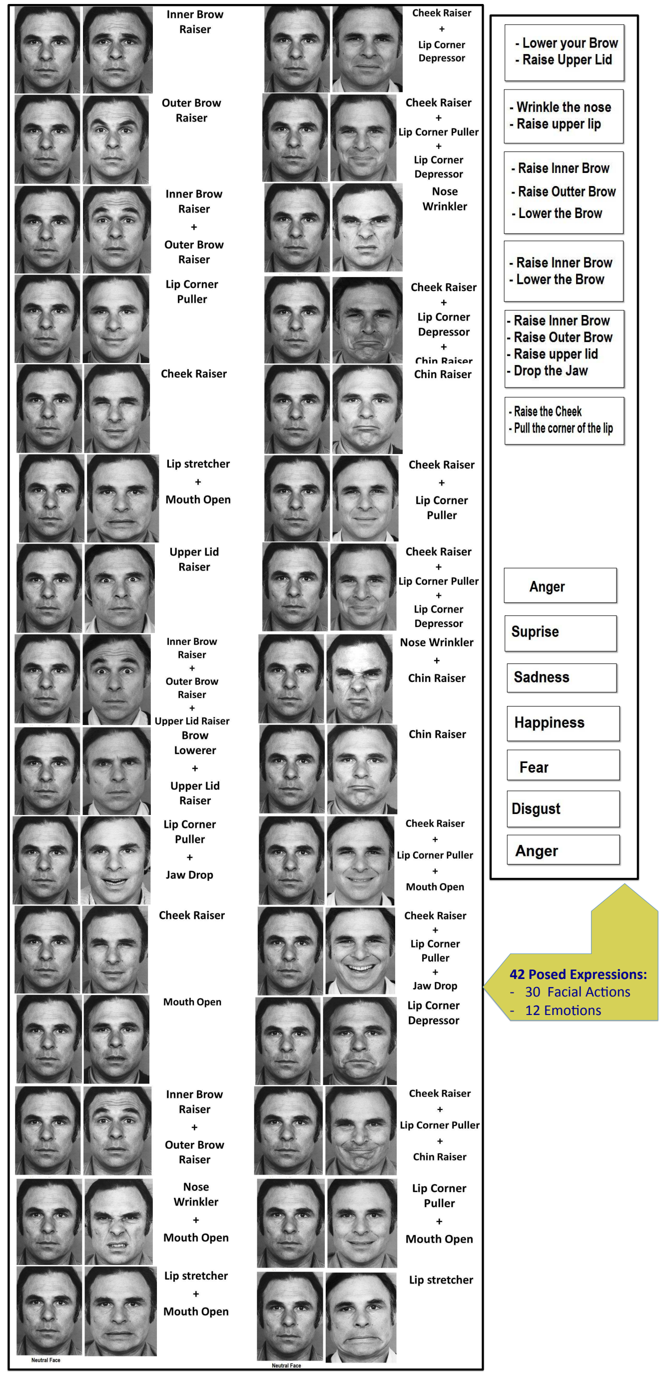


Figure 7: List of acquired 42 posed facial expressions

References

[1] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, volume 5, pages 53–53. IEEE, 2003.

[2] G. Bradski. *Dr. Dobb's Journal of Software Tools*.

[3] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91(1):160–187, 2003.

[4] 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*, pages 203–221, 2007.

[5] J. F. Cohn and K. L. Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(02):121–132, 2004.

[6] C. Darwin. *The expression of the emotions in man and animals*. Oxford University Press, 1998.

[7] D. G. B. Duchenne. *Mécanisme de la physionomie humaine...* typ. F. Malteste, 1862.

[8] P. Ekman, W. V. Friesen, and J. C. Hager. *Facial action coding system*. A Human Face Salt Lake City, 2002.

[9] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.

[10] C.-H. Hjortsjö. *Man's face and mimic language*. Studentlitteratur, 1969.

[11] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[12] Y. Li, S. M. Mavadati, M. H. Mahoor, and Q. Ji. A unified probabilistic framework for measuring the intensity of spontaneous facial action units. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–7. IEEE, 2013.

[13] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE, 2010.

[14] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 57–64. IEEE, 2011.

[15] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, pages 151–160, 2013.

[16] D. McDuff, R. El Kaliouby, T. Senechal, M. Amr, J. Cohn, and R. Picard. Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected in-the-wild. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 881–888, 2013.

[17] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 5–pp. IEEE, 2005.

[18] I. Sneddon, M. McRorie, G. McKeown, and J. Hanratty. The belfast induced natural emotion database. *Affective Computing, IEEE Transactions on*, 3(1):32–41, 2012.

[19] Y. Tong, J. Chen, and Q. Ji. A unified probabilistic framework for spontaneous facial action modeling and understanding. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(2):258–273, 2010.

[20] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn. Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 162–170. ACM, 2006.

[21] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10):692–706, 2014.