Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database

Michel F. Valstar, Maja Pantic

Imperial College London / Twente University
Department of Computing / EEMCS
180 Queen's Gate / Drienerlolaan 5
London / Twente
Michel.Valstar@imperial.ac.uk, M.Pantic@imperial.ac.uk

Abstract

We have acquired a set of audio-visual recordings of induced emotions. A collage of comedy clips and clips of disgusting content were shown to a number of participants, who displayed mostly expressions of disgust, happiness, and surprise in response. While displays of induced emotions may differ from those shown in everyday life in aspects such as the frequency with which they occur, they are regarded as highly naturalistic and spontaneous. We recorded 25 participants for approximately 5 minutes each. This collection of recordings has been added to the MMI Facial Expression Database, an online accessible, easily searchable resource that is freely available to the scientific community.

1. Introduction

A key goal in Automatic Human Behaviour Analysis is "really natural language processing" (Cowie and Schröder, 2005), which endows machines with the ability to speak to human users in the same way that a human would speak to another person. Achieving that goal includes finding ways to make machines understand the non-verbal signals that humans send and which are part and parcel of human conversation. Emotion is one signal that Automatic Human Behaviour Analysis scientists have focused on for approximately thirty years already. Automatic detection of emotions has been investigated both from audio, video, and recently also by fusing the audio and video modalities (see (Zeng et al., 2009) for an overview).

Following Darwin, discrete emotion theorists propose the existence of six or more basic emotions that are universally displayed and recognised (Darwin, 1872). These emotions are Anger, Disgust, Fear, Happiness, Sadness, and Surprise. Data from both modern Western and traditional societies suggest that non-verbal communicative signals involved in these basic emotions are displayed and recognised cross-culturally (Keltner and Ekman, 2000). While the basic emotions do not occur very frequently in normal human-human interactions, when they do occur they convey a very strong message to someone's surroundings.

A major issue hindering new developments in the area of Automatic Human Behaviour Analysis in general, and affect recognition in particular, is the lack of databases with natural displays of behaviour and affect. While a number of publicly available benchmark databases with posed displays of the six basic emotions exist, and are well studied (Pantic et al., 2005; Lyons et al., 1998; Kanade et al., 2000), there is no equivalent of this for spontaneous basic emotions.

There do exist a few databases that contain spontaneous emotive content. Table 1 gives an overview of them. In the second to fourth columns of the table, we list how many people were recorded, the total duration of the dataset, and the bandwidth with which the data was recorded. In the 'Availability' column we list whether the database is freely available to anyone (*Public*), freely available to the academic scientific community (*Scientific*), or not available at all (*Private*). In the 'Online' column we list whether the data is an on-line repository, or whether the data is distributed by traditional mail. Finally, in the last column we indicate whether there is an online search option to the database, which would allow researchers to select and download exactly the set of data they require.

Two databases that have been used recently in studies of automatic human behaviour analysis on spontaneous data are the RU-FACS database (Bartlett et al., 2006) and the DS-118 database (Rosenberg et al., 1998). RU-FACS was recorded at Rutgers University. A hundred people participated in false opinion paradigm interviews conducted by retired law enforcement agents. Participants had to either lie or tell the truth about a social/political issue. If participants chose to lie about their opinions, they would gain 50 US Dollars if they convinced the interviewers of their views, and were told that they would have to fill out a long and boring questionnaire if they failed. This raised the stakes of lying, and thus elicited stronger and more natural expressive behaviour.

The DS-118 dataset has been collected to study facial expression in patients with heart disease. Subjects were 85 men and women with a history of transient myocardial ischemia who were interviewed on two occasions at a 4-month interval. Spontaneous facial expressions were recorded during a clinical interview that elicited spontaneous reactions related to disgust, contempt, and other negative emotions as well as smiles.

Unfortunately, these two databases are not freely available to the scientific community. This makes it impossible to reproduce results of automatic behaviour analysis that are tested solely on these databases. Three databases containing displays of the basic emotions that *are* freely available to the academic community are the SAL (Douglas-

Table 1: Overview of databases with spontaneous emotive content.

Database	Participants	Duration	Video Bandwidth	Audio Bandwidth	Availability	Online	Searchable
DS-118 (Rosenberg et al., 1998)	100	4:10:00	unknown	unknown	Private	No	No
MMI-db Part IV & Part V	25	1:32:00	640x480 pixels @ 29Hz	44.1kHz	Scientific	Yes	Yes
RU-FACS (Bartlett et al., 2006)	100	4:10:00	unknown	unknown	Private	No	No
SAL (Douglas-Cowie et al., 2007)	4	4:11:00	352x288 pixels @ 25 Hz	20kHz	Scientific	Yes	No
SEMAINE (McKeown et al., 2010)	20	6:30:41	580x780 pixels @ 49.979 Hz	48kHz	Scientific	Yes	Yes
Spaghetti db (Douglas-Cowie et al., 2007)	3	1:35	352x288 pixels @ 25Hz	Unknown	Scientific	Yes	No
Vera am Mittag db (Grimm et al., 2008)	20	12:00:00	352x288 pixels @ 25Hz	16kHz	Public	No	No

Cowie et al., 2007), SEMAINE (McKeown et al., 2010), and Spaghetti (Douglas-Cowie et al., 2007) databases. Both the SAL and SEMAINE databases record interactions between a user (the experiment participant) and an operator (someone from the experimenters' team). The operators act out one of four prototypic characters: one happy, one sad, one angry and one neutral. This results in emotionally coloured discourse. As shown in table 1 both databases contain a considerable amount of data. However, except for the emotion 'happiness', emotions are mostly displayed in a very subtle manner. While these databases are very suitable for analysing expressions displayed in natural discourse, they are less suitable for training systems that can identify the basic emotions.

The Spaghetti database on the other hand does contain strong basic emotions; mostly of fear, disgust, surprise, and happiness. The database consists of recordings of an experiment where people were asked to feel inside a box that contained a warm bowl of spaghetti. Because the participants didn't know what's in the box, they reacted strongly when their hands touched the spaghetti. The data was released as part of the HUMAINE database (Douglas-Cowie et al., 2007). Unfortunately, it consists of recordings of only three participants, and the total dataset lasts only one minute and 35 seconds. This makes it very hard to train any automatic Human Behaviour Understanding algorithms on this data.

The MMI-Facial Expression database was conceived in 2002 by Maja Pantic, Michel Valstar and Ioannis Patras as a resource for building and evaluating facial expression recognition algorithms (Pantic et al., 2005). Initially the focus of the database was on collecting a large set of AUs, occurring both on their own and in combination, from either videos or high-quality still images. Later data to distinguish the six basic emotions were added, and in this work the addition of spontaneous data is described.

Recording truly spontaneous instances of basic emotion expressions is extremely difficult, because in everyday life the basic emotions aren't shown frequently. However, when they *are* displayed, they convey a very strong message to someone's surroundings, one which should certainly not be ignored by Automatic Human Behaviour Analysis systems. In order to record a truly spontaneous dataset of sufficient size[1] it would thus be necessary to follow and record the participants for a very long duration. Following the participants for a long time would mean the recording setup would need to be compact and mobile. A side effect of this would be that one loses the ability to control the recording conditions. Instead of waiting for the expressions to occur naturally, we decided to induce them by showing the participants a collage of short video clips. The clips were selected to induce the emotions happiness and disgust.

The remainder of this work is structured as follows: section 2. explains how we recorded the data. The recorded data was manually annotated for the six basic emotions and facial muscle actions (FACS Action Units (Ekman et al., 2002)). The annotation details are presented in section 3.. Section 4. describes the existing MMI Facial Expression Database and the place that the new induced emotional recordings described in this work have in it. We provide a summary and closing remarks in section 5..

2. Data acquisition

The goal of the experiments is to record naturalistic audiovisual expressions of basic emotions. Due to ethical concerns, it is very hard to collect such data for the emotions anger, fear, and sadness. Happiness, surprise and disgust on the other hand are easier to induce. We collected data from these three expressions in two different experiments. In the first experiment we showed the participants short clips of cartoons and comedy shows to induce happiness. Images and videos of surgery on humans and humans effected by various diseases were shown to induce disgust. In the second experiment, only the happiness inducing type of clips were shown. In both experiments, the participants showed expressions of surprise too, often mixed with either happiness or disgust.

The images and videos were shown on a computer screen. The participants were sitting in a chair at approximately 1.5 metres distance to the screen. In both experiments, a JVC GR-D23E Mini-DV video camera with integrated stereo microphones was used to record the reactions of the participants. The camera was placed just above the computer screen, ensuring a near-frontal view of the participants' face as long as they face the screen.

During the recording of the first experiment the experimenters were in the room with the participants. As a result the participants engaged in social interactions with the experimenters about the content of the images shown, and what they thought about this. We regarded this as undesirable components of the recordings, as the experimenters influenced the behaviour of the participants. The recordings were manually cut into 383 segments (called Sessions) that contain distinct displays of affective behaviour. To distinguish this set of data from existing datasets in the database, we will refer to this as Part IV of the MMI Facial Expression database (see section 4.).

In the second experiment, the participants would hear the

[1] Sufficient meaning enough recordings to be able to perform studies that can have statistically significant results.

Figure 1: A selection of expressions of disgust and happiness added to the MMI-Facial Expression Database. The first row is taken from Part V of the database and shows expressions of happiness and disgust. The second row shows expressions of happiness and disgust taken from Part IV of the database. The third row shows four frames of a single sequence in which the participant showed an expression of disgust.

sound of the stimuli over headphones instead of over computer speakers, as was the case in experiment 1. This resulted in less noise in the audio signal. Another refinement was that the participants were left in a room on their own while the stimuli were provided. The idea behind this is that the participants would be less socially inhibited to show their emotions if there were no other people around. Also, without interruptions caused by interactions between participants and experimenters, the data can now be used to analyse the changes in behaviour over time. To allow the latter, we chose not to cut the recordings of the second experiment into smaller clips. We will refer to this data as Part V of the MMI Facial Expression database.
In total, 25 participants aged between 20 and 32 years took part in the experiments, 16 in experiment 1 and 9 in experiment 2. Of these, 12 were female and 13 male. Of the female participants, three were European, one was South American, and eight were Asian. Of the men, seven were European, two were South American and four were of Asian background.

3. Annotation of affect, laughter and facial muscle actions

Part IV of the database has been annotated for the six basic emotions and facial muscle actions. Part V of the database has been annotated for voiced and unvoiced laughters.

3.1. Emotion and Action Unit annotation

All Sessions of Part IV were annotated to indicate which of the six basic emotions occurred in each clip. This annotation is valuable for researchers who wish to build automatic basic emotion detection systems. Not all affective states can be categorised into one of the six basic emotions. Unfortunately it seems that all other affective states are culture-dependent.

Instead of directly classifying facial displays of affective states into a finite number of culture-dependent affective states, we could also try to recognise the underlying facial muscle activities. These muscle activations are objective measures of facial expression. These can then be interpreted in terms of (possibly culture-dependent) affective categories such as emotions, attitudes or moods. The Facial Action Coding System (FACS) (Ekman et al., 2002) is the best known and the most commonly used system developed for human observers to describe facial activity in terms of visually observable facial muscle actions (i.e., Action Units, AUs). Using FACS, human observers uniquely decompose a facial expression into one or more of in total 31 AUs that produced the expression in question.

Note that all possible facial expressions can be uniquely described by this 31-dimensional space. That is a rather low-dimensional space, which means that learning a mapping from AUs to affective states requires significantly less space than a mapping directly from images to affective states would need.

All Sessions belonging to Part IV have been FACS AU-coded by a single FACS coding expert. For each Session, the annotation indicates which AUs were present at some time during the clip.

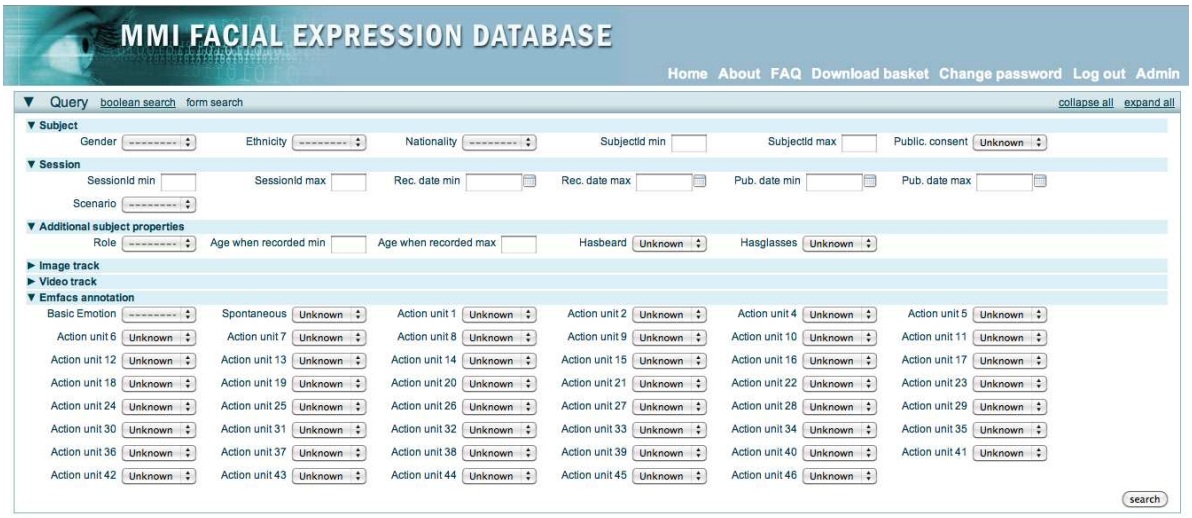

Figure 2: The online search form of the MMI Facial Expression Database.

3.2. Laughter annotation

Laughter is a an event in which a person smiles and produces a typical sound. It is therefore an audio-visual social signal. Laughters have been further divided into two categories: voiced and unvoiced laughters (Bachorowski et al., 2001). To allow studies on the automatic audio-visual analysis of laughter (e.g. (Petridis and Pantic, 2009)), we annotated in all recordings of Part V of the database exactly when voiced and unvoiced laughters events occurred. In our annotation rules, a laughter event was coded as voiced if any part of that laughter event had a voiced component.
Part V consists of nine recordings. In total we, annotated 109 unvoiced and 55 voiced laughters. The average duration of an unvoiced laughter was 1.97 seconds, while a voiced laughter lasted 3.94 seconds on average (exactly twice as long). The laughter annotation was performed using the ELAN annotation tool (Wittenburg et al., 2006).

4. MMI Facial Expression Database

The MMI Facial Expression Database is a continually growing online resource for AU and basic emotion recognition from face video. In the following sections we will describe the database's structure, and how to use its web-interface.

4.1. Database organisation

Within the database, the data is organised in units that we call a *Session*. A Session is part of a *Recording*, which is a single experiment, i.e. all data of a single participant watching all the stimuli. Each Session has one or more single sensor data files associated with it. For the data presented in this paper, this is the audio-visual data stream recorded by the camera. We call these database entries *Tracks*. There are low-quality previews of all tracks on the web-based interface of the database so that users can get an idea of what content each track has, before choosing to download it. The fourth component that makes up the database are the *annotations*. In the case of the MMI Facial Expression database, FACS and basic emotion annotation data is available in so-called EMFACS annotations, and laughter annotations are available in the form of ELAN files.
The first data recorded for the MMI Facial Expression Database were mostly displays of individual AUs. In total, 1767 clips of 20 participants were recorded. Each participant was asked to display all 31 AUs and a number of extra Action Descriptors (ADs, also part of FACS). After all AUs were recorded, the participants were asked to perform two or three affective states (e.g. sleepy, happy, bored). The participants were asked to display every facial action twice, to increase the variability of the dataset. During recording a mirror was placed on a table next to the participant at a 45 degree angle with respect to the camera. This way, a completely synchronised profile view was recorded together with the frontal view. We will refer to this part of the data as Part I of the database. It consists of Sessions 1 to 1767, and is video-only.
The second set of data recorded were posed displays of the six basic emotions. In total, 238 clips of 28 subjects were recorded. Again, all expressions were recorded twice. People who wear glasses were recorded once while wearing their glasses, and once without. This time we focused on obtaining a higher spatial resolution of the face. Therefore we did not use a mirror to obtain a profile view, and we tilted the camera to record the faces in portrait-orientation. We will refer to this part of the data as Part II of the database. It consists of Sessions 1767 to 2004, and is video-only.
Part III, the third set of data recorded consists of high-quality still images. Similar to Part I, the participants were asked to display all AUs and the six basic emotions. In total 484 images of 5 subjects were recorded. Part III consists of Sessions 2401-2884. The acquisition of Parts I-III are described in detail in (Pantic et al., 2005).
0Part IV and Part V are described in detail in section 2. . Part IV consists of Sessions 2005-2388, while Part V consists of Sessions 2895 to 2903.

4.2. Search

The web-based interface of the database has search form functionality that allows the user to collect and preview the exact set of data that he or she needs. Search options include searching on basic emotion, FACS coding, whether the expression was posed or spontaneous, ethnographic variables of the subjects and technical details of the recorded media. Figure 2 shows the search form layout.

One particularly interesting search criterium is the recording scenario type. To describe the different possible scenarios in which emotionally coloured data can be acquired, we divide the possible scenarios in three groups: *acting*, *reacting*, and *interacting*. In this taxonomy, all posed expression databases fall in the category *acting*, while all scenarios in which an emotion is induced by providing the experiment participants with a certain stimulus while measuring their reactions fall in the *reacting* category. The additional datasets described in this work (i.e. Part IV and Part V) are classic examples of *reacting* scenarios. Finally, scenarios in which participants are interacting freely fall in the *interacting* category. Good examples of this are the SAL and SEMAINE databases (Douglas-Cowie et al., 2007; McKeown et al., 2010).

So, to find parts IV and V within the MMI Facial Expression database, one possibility would be to search for the scenario option *reacting*, in the EMFACS section of the search form. Alternatively one could search directly using the Session numbers described above.

4.3. Database availability

The database is freely available to the academic scientific community, and is easily accessible through a web-interface. The url of the database is http://www.mmifacedb.com. Prospective users have to register with that website to create an account. To activate this account, the prospective users need to sign and return an End User License Agreement (EULA). Among other things, the EULA prohibits using the data to train or test commercial products, and it prohibits all military use.

The EULA allows the user to use imagery contained in the database for academic publications and presentations, provided that the participants shown in that particular imagery have specifically allowed this. One can search for this in the search form.

5. Conclusion

We have presented an addition to the MMI-Facial Expression corpus consisting of spontaneous data. Emotions of happiness, disgust, and surprise were induced in 25 participants and their displays/outbursts of affect were recorded on video and audio. This resulted in 1 hour and 32 minutes of data, which is made available in 392 segments called Sessions. Part IV of the data was manually annotated for the six basic emotions, and for FACS Action Units. Part V was annotated for voiced/unvoiced laughters. We believe this dataset can be of great benefit to researchers in the field of Automatic Human Behaviour Understanding.

6. Acknowledgments

This work has been funded in part by the European Community's 7th Framework Programme [FP7/20072013] under the grant agreement no. 231287 (SSPNet). The work of Michel Valstar is further funded in part by the European Community's 7th Framework Programme [FP7/2007-2013] under grant agreement no. 211486 (SEMAINE). The work of Maja Pantic is also funded in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

7. References

J. A. Bachorowski, M. J. Smoski, and M. J. Owren. 2001. The acoustic features of human laughter. *Journal of the Acoustical Society of America*, 110(1):1581–1597.

M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I.R. Fasel, and J.R. Movellan. 2006. Automatic recognition of facial actions in spontaneous expressions. *Journal of Mutlimedia*, pages 1–14, Oct.

R. Cowie and M. Schröder. 2005. Piecing together the emotion jigsaw. *Machine Learning for Multimodal Interaction*, pages 305–317, Jan.

C. Darwin. 1872. *The Expression of the Emotions in Man and Animals*. John Murray, London.

E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, Lowry O., M. McRorie, J. Martin, L. Devillers, S. Abrilian, A. Batliner, N. Amir, and K. Karpouzis. 2007. The humaine database: Addressing the collection and annotation of naturalistic and induced emotional data. *Lecture Notes in Computer Science*, 4738:488–501, Jan.

P. Ekman, W.V. Friesen, and J.C. Hager. 2002. *Facial Action Coding System*. A Human Face.

M. Grimm, K. Kroschel, and S. Narayanan. 2008. The vera am mittag german audio-visual emotional speech database. *IEEE International Conference on Multimedia and Expo*, pages 865–868.

T. Kanade, J.F. Cohn, and Y. Tian. 2000. Comprehensive database for facial expression analysis. *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53.

D. Keltner and P. Ekman. 2000. Facial expression of emotion. In M Lewis and J.M. Haviland-Jones, editors, *Handbook of emotions*, pages 236–249. Guilford Press.

M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. 1998. Coding facial expressions with gabor wavelets. *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205.

G. McKeown, M.F. Valstar, M. Pantic, and R. Cowie. 2010. The semaine corpus of emotionally coloured character interactions. *Submitted to ICME*, pages 1–6, Jan.

M. Pantic, M.F. Valstar, R. Rademaker, and L. Maat. 2005. Web-based database for facial expression analysis. In *IEEE International Conference on Multimedia and Expo*, pages 317–321.

S. Petridis and M. Pantic. 2009. Is this joke really funny? judging the mirth by audiovisual laughter analysis. In *IEEE International Conference on Multimedia and Expo*, pages 1444 –1447, 28 2009-july 3.

E.L Rosenberg, P. Ekman, and J.A. Blumenthal. 1998. Facial expression and the affective component of cynical hostility in male coronary heart disease patients. *Health Psychology*, 17(4):376–380, Aug.

P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. 2006. Elan: a professional framework for multimodality research. *Proceedings of Language Resources and Evaluation Conference*.

Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58.