Running head: The STOIC Dynamic Facial Emotional Expressions Database

STOIC: A database of dynamic and static faces expressing highly recognizable emotions

Sylvain Roy[1], Cynthia Roy[1], Catherine Éthier-Majcher[1], Isabelle Fortin[1], Pascal Belin[2], Frédéric Gosselin[1]

1. Department of Psychology, University of Montreal, Montreal, Canada
2. Department of Psychology, University of Glasgow, Glasgow, UK

Corresponding author: Frédéric Gosselin
Département de Psychologie
Université de Montréal
C.P. 6128 Succ. Centre-Ville
Montréal, Québec
H3C 3J7, Canada
(514) 343-7550
Email: frederic.gosselin@umontreal.ca.

Keywords: Dynamic facial expressions, face processing, emotions, database.

Abstract

We recorded about 7,000 short videos of 34 actors expressing facially the six basic emotions (fear, happiness, surprise, anger, sadness, and disgust), pain, and neutrality. The duration of the 1,088 most promising movies was reduced to 500 ms (15 frames). Faces were aligned on points placed on three robust facial landmarks (the pupil centers and the tip of the nose) across frames and videos. The frame containing the peak of the expression was extracted from each of these videos. Participants rated each stimulus with respect to how intensely it expressed happiness, disgust, fear, anger, sadness, surprise, and pain. The STOIC database comprises the 80 movies and corresponding photos most consistently recognized by observers while showcasing five male and five female actors, each expressing facially all basic emotions, pain, and neutrality. It is freely available here.

Facial emotional expressions communicate information from which we can quickly infer the state of mind of our peers, and adjust our behavior accordingly (Darwin, 1872). Most psychophysical studies on facial expressions have been conducted using photos. However, the results from neuroimaging studies suggest that the brain regions involved in processing of facial affect—such as the posterior superior temporal sulcus (pSTS), the amygdala, and insula—respond differently to dynamic—more realistic—than to static emotional expressions (e.g., Haxby, Hoffman, & Gobbini, 2000, 2002; Kilts et al., 2002; LaBar et al., 2003). Furthermore, Humphreys, Donnelly and Riddoch (1993) reported the case of a patient who could accurately recognize emotional expressions from moving points of light, but not from static images; and, reciprocally, Adolphs et al., (2003) reported the case of a patient who could only recognize dynamic emotional expressions. Yet the role played by dynamic features in the *recognition* of facial expression of emotions is still largely unknown (e.g., Ambadar et al., 2005).

Several photo databases of facial expressions are available, such as the popular set developed by Ekman and Friesen (1975; e.g., CAFE, Karolinska Directed Emotional Faces). Likewise, there are a few video databases of facial expressions available (e.g., Battocchi & Pianesi, 2004; Cohen, Sebe, Garg, & Huang, 2002; Douglas-Cowie, Cowie, & Schröder, 2000; Kanade, Cohn, & Tian, 2000; Martinkauppi, Soriano, Huovinen, & Laaksonen, 2002; O'Toole et al., 2005; Pantic, Valstar, Rademaker, & Maat, 2005; Sun, Sebe, Lew, & Gevers, 2004; Wallhoff & [FG-NET]. 2005). None of these databases is perfectly adapted to the experiments that we plan to carry out to examine the role played by dynamic features in the recognition of facial expression of emotions, namely, classification-image and gaze-tracking experiments. Therefore, we developed STOIC, a database of emotions expressed facially conform to our needs. The main characteristics of the database are:

(1) It includes both videos and photos extracted from these videos.

(2) It includes facial expressions of the six basic emotions as well as pain and neutrality. Regardless of whether pain should be considered a basic emotion, its evolutionary significance cannot be denied. It is obvious that the capacity to feel pain (Williams, 2002) and to recognize it in others is just as important as any basic emotions for our survival (Craig, 2004).

(3) The static stimuli as well as every frame of the dynamic stimuli were spatially aligned—and, in the case of the dynamic stimuli, temporally aligned—thus insuring a consistent positioning of facial features on the screen and minimizing head and body movements. This characteristic of the database will greatly simplify the analysis of classification-image and gaze-tracking data.

(4) Over one thousand videos and photos were validated independently. In contrast to what is typically done, we put each stimulus in the perceived emotion category—not necessarily the emotion that the actor intended to express. Only the 80 dynamic and corresponding static stimuli that led to the greatest consensus among observers were kept.

(5) The database is suitable for face identification (10 actors, each expressing facially the seven emotions and neutrality) and gender discrimination (half of the actors are females), in addition to facial expression recognition.

More details about the STOIC database are provided in the following pages.

Stimuli creation

We recorded a total of about 7,000 movies of emotions expressed facially by 34 actors. Four observers selected the best 1,088 movies—those that appeared most genuine and contained the least head movement. Faces in these selected movies were aligned spatially and temporally. One thousand eighty-eight photos were created by extracting the frame that contained the peak expression of every movie.

Actors. A total of 34 actors (16 females) between the ages of 20 and 45 years were recruited among theatrical schools in Montreal. It was reasoned that experienced actors could more easily produce emotions that appear genuine. To insure some uniformity between the visual stimuli, actors were asked not to wear jewelry, or have facial piercing. Powder was used to reduce sweating and reflecting light and a hairnet insured that hair would not get in the way.

Filming. Actors were asked to facially express the six basic emotions (happiness, disgust, fear, anger, sadness, and surprise) as well as pain and neutrality. Filming took place in a semi-anechoic chamber with chroma-key blue background, equipped with two diffuse tungsten lamps. The movie streams were recorded using a Canon XL1S video camera. Data was digitally transferred to a Personal Computer (AMD 1700 processor) and captured using Adobe Premier Pro software. The videos were captured in color at a rate of 29.97 images per second with a resolution of 720 by 480 pixels. The actors were positioned 1.5 meter from the lens of the camera

and centered in the image. We deinterlaced the video track using a blending method. At the beginning of each recording, actors were asked to hold a Kodak colors chart to allow color and luminance calibrations. However, the validation was done only on the achromatic stimuli but remains available for further sudies. Each recording session lasted approximately one hour; actors had to generate multiple exemplars of the eight facial expressions at different intensities (weak, moderate, high). Actors were asked to say "ah" when expressing the emotions. The audio track was removed for the current validation but remains available for further studies (see Belin, Fillion-Bilodeau & Gosselin, 2008).

Movies and photos. The video track was initially segmented into one-second movies, including the full rise of the facial expressions. Four graduate students chose the four best movies for each emotion and actor (i.e., 34 actors * 8 emotions * 4 exemplars = 1,088 movies) based on two criteria: minimum body and head displacements and apparent authenticity of the expressed emotions. For each movie, we isolated facial-muscle movements by aligning three robust facial features using home-brewed Matlab programs. Thus, for every frame of every movie, three points were positioned, by human observers, on the centers of the pupils and on the tip of the nose. Then, we translated, rotated, and scaled the landmark positions of each frame of each movie to minimize the mean square of the difference between them and a template (see Figure 1; e.g., Gonzalez, Woods & Eddins, 2002).

Figure 1. Left: Mean of all frames of a fear movie pre-alignment superimposed with the position of the facial landmarks annotated in red and the average facial landmarks in green. A dynamic version is available here. Right: The same but post-alignment: A significant amount of smear has been removed. A dynamic version is available here.

This template was the average of the landmark positions across all frames and movies scaled so that inter-ocular distance was 100 pixels. A consequence of this spatial alignment is that a featural meaning can now be ascribed to a coordinate. While these transformations worked nicely with clips that contained a lot of head movement, they introduced jitter in those that contained minimal movements. These movie clips were therefore rotated and translated on each frame, and scaled on the mean landmark locations. The frames were cropped at 256 x 256 pixels, centered on the aligned nose landmark. Movies were also aligned temporally by annotating the last neutral frame prior the appearance of the emotional expression and were shortened to 15 frames (500 ms). Our static expressions consisted of the apex of every movie. We've added mid-gray elliptical masks to the movies convolved with a Gaussian filter having a standard deviation of 2 pixels to remove sharp edges. These masks were fitted by emotions and by subjects to reveal all internal and remove all external facial features; when necessary, we fitted individual movies (and photos).

Validations

We proceeded with two separate validations—one for the photos and the other for the movies. This allowed us to derive a measure of stimulus recognizability based on entropy.

Participants. Thirthy-five participants (20 females) from Montréal were recruited for the validation dynamic expressions (the mean age and years of schooling were 25 and 16 respectively). Thirthy-five others (19 females) also from Montréal participated in the validation of the static expressions (the mean age and years of schooling were 23 and 16 respectively). All participants had normal or corrected vision.

Procedure. The validations took place in computer rooms at the Université de Montréal. All 1,088 movies (and photos) were presented to all participant using the Internet browser Firefox 2 on Macintosh G5 computers; our website was programmed in PHP/JavaScript. Photos were presented for 500 ms, that is, the same duration as the movies. Movies and photos were preceded and followed by mid-gray frames. Data was automatically saved on a Macintosh server's MySQL database. Participants were told they would see several movies (or photos) of actors expressing facially one of eight possible emotions, i.e., fear, happiness, anger, disgust, pain, sadness, surprise, and neutrality. We added that they could view the movies (or photos) a second time if they felt it was necessary but not more. Participants were instructed to rate each stimulus with respect to how intensely it expresses happiness, disgust, fear, anger, sadness, surprise, and pain,

using seven continuous scroll bars (from leftmost = "not at all" to rightmost = "the most intense possible"). If they perceived an ambiguous facial expression, they were asked to rate the movie on more than one scroll bar. If they perceived neutrality, they were asked to simply set all scroll bars to the leftmost position.

Stimulus entropy. We measured movies (and photos) ambiguity by computing the entropy (E) of their scroll bar ratings:

$$E = -\sum_i \frac{p_i}{\sum_i p_i} \log_2 \left(\frac{p_i}{\sum_i p_i} \right),$$

where p_i is a proportion derived from the scroll bar ratings of emotion i. A stimulus with an entropy of 0 bit was always given a non-zero rating on a single emotion scroll bar—it's as unambiguous as it can be; and a stimulus with an entropy of 2.8074 bits was given equal ratings, on average, on all emotion scroll bars—it's as ambiguous as it can be.

In preparation for the computation of the proportions (p_i), the scroll bar ratings were transformed into z-scores for every participant. This transformation insures that a conservative participant that used only the first third of the scroll bars, for example, is comparable to a blasé participant that used only the second third of the scroll bars and to an ideal participant that used the entire scroll bars; but, importantly, it preserves the relative rating differences between emotions. Then the mean of the z-scores across participants but within emotion (z_i) were transformed into p_i as follows:

$$p_i = \frac{z_i}{2 \times \max(|z_i|)} + 0.5 .$$

We categorized each movie (and photos) as a member of the emotion for which it received the highest p_i. Thus one movie from the final selection was put in the pain category because participants rated the movie highest on the pain dimension even though the actor's intention was to express sadness. Likewise, another movie from the final selection was put in the surprise category even though it was intended to express neutrality. The only exception to this "max" rule was the neutrality category: a movie (or an photo) was categorized as neutral if max(p_i) was smaller than criteria including 1/8 of the movies (or photos).

The database

The STOIC database comprises the 80 movies and corresponding photos associated with the smallest entropy values—most consistently recognized—while showcasing five male and five female actors, each expressing facially all basic emotions, pain, and neutrality. Tables 1 and 2 show the entropy values of every stimulus (see also Figures 2, 3, and 4 for their proportions derived from the scroll bar ratings— p_i).

A 3-way ANOVA (actor gender x stimuli type x emotion) on the entropy values revealed no significant difference between dynamic and static stimuli or male and female actors. A statistically significant effect of emotion was found ($F_{(6,140)} = 30$, $p < .001$). Tukey post-hoc comparisons showed that entropies for fear and pain emotions are significantly larger ($p < .001$) than those for all other emotions—indicating that fear and pain were the most difficult emotions to recognize—but did not differ from one another (*ns*). Happiness and anger were the easiest emotions to recognize and did not differ from one another (*ns*). Moreover, the entropy values for disgust, sadness, and surprise did not differ from one another (*ns*) and constitute moderately difficult emotions to recognize.

Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) awarded to Frédéric Gosselin. We also thank the undergraduate students who have annotated over 28,000 frames.

Dynamic clips	Fear	S	Happy	S	Anger	S	Disgust	S	Sadness	S	Surprise	S	Pain	S
actor 1	DM1fe	1.74	DM1ha	0.23	DM1an	0.32	DM1di	0.24	DM1sa	0.31	DM1su	1.30	DM1pa	1.80
actor 2	DM2fe	1.76	DM2ha	0.00	DM2an	1.16	DM2di	1.59	DM2sa	0.00	DM2su	0.13	DM2pa	1.47
actor 3	DM3fe	1.66	DM3ha	0.17	DM3an	0.00	DM3di	0.14	DM3sa	0.67	DM3su	0.66	DM3pa	1.82
actor 4	DM4fe	1.58	DM4ha	0.28	DM4an	0.00	DM4di	1.21	DM4sa	0.17	DM4su	0.65	DM4pa	1.83
actor 5	DM5fe	1.57	DM5ha	0.00	DM5an	0.14	DM5di	1.16	DM5sa	0.96	DM5su	0.47	DM5pa	1.39
actrice 1	DF1fe	1.83	DF1ha	0.12	DF1an	0.25	DF1di	0.43	DF1sa	0.00	DF1su	0.53	DF1pa	0.80
actrice 2	DF2fe	1.49	DF2ha	0.17	DF2an	0.16	DF2di	0.25	DF2sa	0.29	DF2su	0.80	DF2pa	1.51
actrice 3	DF3fe	1.07	DF3ha	0.26	DF3an	0.49	DF3di	1.31	DF3sa	0.97	DF3su	0.48	DF3pa	1.70
actrice 4	DF4fe	0.84	DF4ha	0.22	DF4an	0.69	DF4di	0.93	DF4sa	1.60	DF4su	1.19	DF4pa	1.96
actrice 5	DF5fe	1.14	DF5ha	0.78	DF5an	0.81	DF5di	1.32	DF5sa	0.77	DF5su	1.26	DF5pa	1.51
Means		1.47		0.22		0.40		0.86		0.57		0.75		1.58
SEMs		0.33		0.22		0.38		0.54		0.52		0.39		0.33
Static clips	Fear	S	Happy	S	Anger	S	Disgust	S	Sadness	S	Surprise	S	Pain	S
actor 1	SM1fe	1.75	SM1ha	0.28	SM1an	0.32	SM1di	0.52	SM1sa	0.89	SM1su	1.22	SM1pa	1.53
actor 2	SM2fe	1.12	SM2ha	0.04	SM2an	0.67	SM2di	1.29	SM2sa	0.47	SM2su	0.55	SM2pa	1.07
actor 3	SM3fe	1.26	SM3ha	0.00	SM3an	0.39	SM3di	0.92	SM3sa	0.24	SM3su	0.14	SM3pa	1.37
actor 4	SM4fe	1.62	SM4ha	0.29	SM4an	0.13	SM4di	1.22	SM4sa	0.28	SM4su	1.06	SM4pa	1.27
actor 5	SM5fe	1.73	SM5ha	0.09	SM5an	0.88	SM5di	0.58	SM5sa	1.22	SM5su	0.35	SM5pa	1.80
actrice 1	SF1fe	1.99	SF1ha	0.00	SF1an	0.24	SF1di	0.60	SF1sa	0.00	SF1su	0.52	SF1pa	1.18
actrice 2	SF2fe	1.47	SF2ha	0.15	SF2an	0.46	SF2di	0.00	SF2sa	0.17	SF2su	0.49	SF2pa	0.90
actrice 3	SF3fe	0.83	SF3ha	0.14	SF3an	0.46	SF3di	1.32	SF3sa	0.55	SF3su	0.67	SF3pa	1.61
actrice 4	SF4fe	0.76	SF4ha	0.00	SF4an	0.11	SF4di	1.05	SF4sa	1.28	SF4su	0.88	SF4pa	1.45
actrice 5	SF5fe	0.90	SF5ha	0.58	SF5an	0.63	SF5di	1.68	SF5sa	0.83	SF5su	0.97	SF5pa	1.12
Means		1.34		0.16		0.43		0.92		0.59		0.68		1.33
SEMs		0.14		0.06		0.08		0.16		0.14		0.11		0.09

Table 1. Dynamic and static stimuli entropy values. Stimuli names (e.g., "DM1fe") have the following format: Dynamic or static (e.g., "D or S"), gender of the actor (e.g., "M or F"), actor number (e.g., "1"), the first two letters of the expression (e.g., "fe" = "fear").

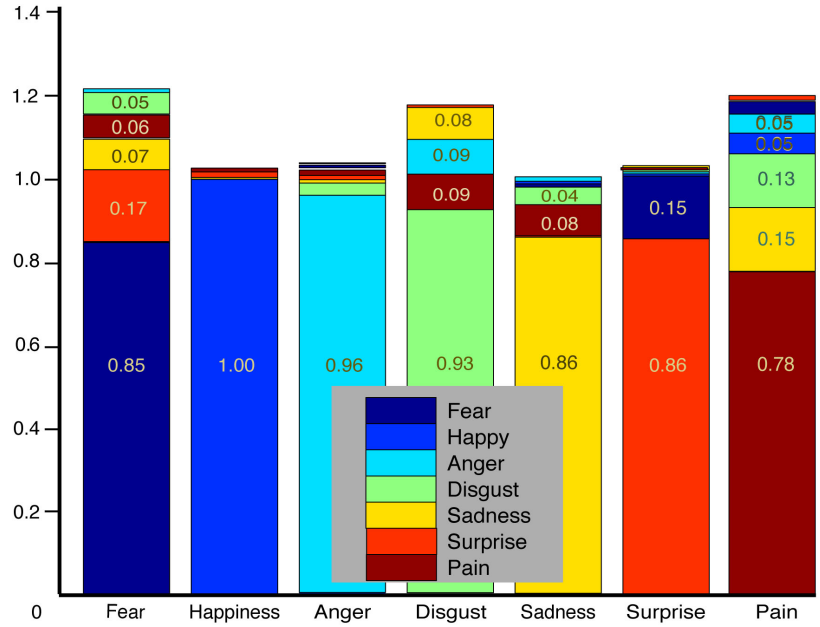

Figure 2. Mean rating proportions (p_i) for photos.

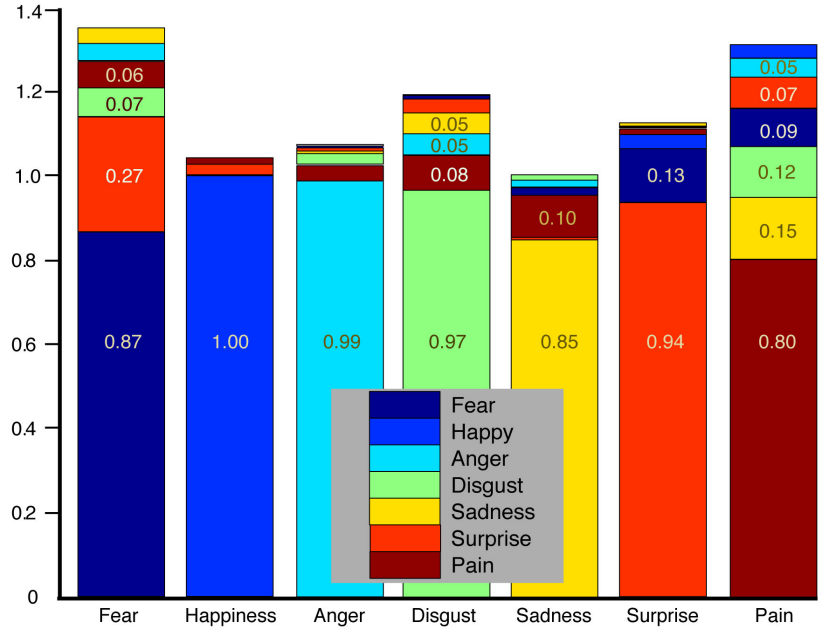

Figure 3. Mean rating proportions (p_i) for movies.

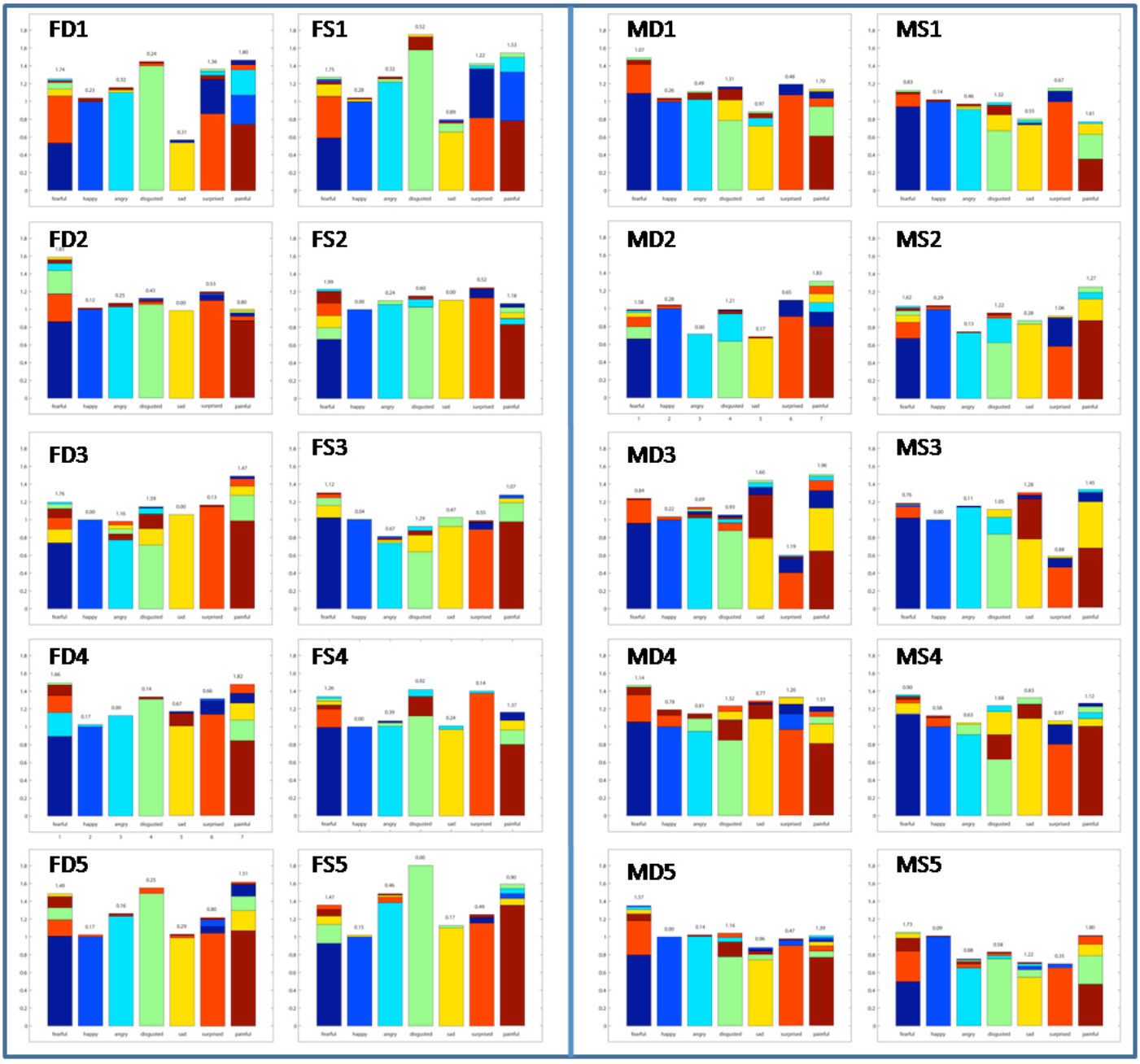

Figure 4. Mean rating proportions (p_i) for all photos (left half) and movies (right half). F and M stand for Female and Male respectively. S and D stand for static and Dynamic respectively. Finally 1 through five is the female or male actor number.

Reference

Adolphs, R., Tranel, D., & Damasio, A. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition, 52*, 61-69.

Ambadar, S., Schooler, J., & Cohn, J. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychological science, 16*(5), 403-410.

Bassili, J. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 373-379.

Battocchi, A., & Pianesi, F. (2004). *DaFEx: Un Database di Espressioni Facciali Dinamiche.* Paper presented at the SLI-GSCP Workshop of the Comunicazione Parlata e Manifestazione delle Emozioni, Padova (Italy).

Belin, P., Fillion-Bilodeau, S. & Gosselin, F. (2008). The "Montreal Affective Voices": a validated set of nonverbal affect bursts for research on auditory affective processing. Behavior Research Methods, 40, 531-539.

Cohen, I., Sebe, N., Garg, A., & Huang, T. (2002). *Facial Expression Recognition from Video Sequences.* Paper presented at the International conference on Multimedia and Expo.

Craig, K. D. (2004). Social communication of pain enhances protective functions: a comment on Deyo, Prkachin and Mercer (2004). *Pain, 107*(1-2), 5-6.

Douglas-Cowie, E., Cowie, R., & Schröder, M. (2000). *A new emotion database: Considerations, sources and scope.* Paper presented at the ISCA Workshop on Speech and Emotion, Northern Ireland.

Ekman, P., & Friesen, W. (1975). *Unmasking the face. A guide to recognizing emotions from facial clues*. New Jersey: Prentice-Hall.

Ekman, P., & Friesen, W. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1*(1), 56-75.

Gonzalez, R. C., Woods, R. E. & Eddins, S. L. (2002). *Digital Image Processing Using MATLAB.* New Jersey: Prentice Hall. pp. 624.

Gosselin, F., & Schyns, P. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research, 41*, 2261-2271.

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science, 4*(6), 223-233.

Haxby, J., Hoffman, E., & Gobbini, M. (2002). Human Neural Systems for Face Recognition and Social Communication. *Biological Psychiatry, 51*(1), 59-67.

Humphreys, G., Donnelly, N., & Riddoch, M. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces neuropsychological evidence. *Neuropsychologia, 31*(173-181).

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perceptioni, 30*, 875-887.

Kanade, T., Cohn, J., & Tian, Y. (2000). *Comprehensive Database for Facial Expression Analysis.* Paper presented at the Automatic Face and Gesture Recognition Proceedings. Fourth IEEE International Conference.

Kilts, CD., Egan, G., Gideon, DA., Ely, TD., Hoffman, JM. (2003). Dissociable Neural Pathways are involved in the recognition of emotion in static and dynamic facial expressions. NeuroImage, 18, 156-168.

LaBar, K., Crupain, M., Voyvodic, J., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex, 13*(10), 1023-1033.

Lundqvist, D., Esteves, F., & Ohman, A. (2004). The face of wrath: The role of features and configurations in conveying social threat. *Cognition & Emotion, 18*(2), 161-182.

Martinkauppi, B., Soriano, M., Huovinen, S., & Laaksonen, M. (2002). *Face video database.* Paper presented at the First European Conference on Color in Graphics Imaging and Vision, Poitiers, France.

Mehrabian, A. (1968). Communication Without Words. *Psychology Today, 2*(4), 53-56.

O'Toole, A., Harms, J., Snow, S., Hurst, D., Pappas, M., Ayyad, J., et al. (2005). *A Video Database of Moving Faces and People.* Paper presented at the IEEE Transactions on pattern analysis and machine intelligence.

Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). *Web-based database for facial expression analysis.* Paper presented at the IEEE Int'l Conf. on Multimedia and Expo, Amsterdam, The Netherlands.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research, 20*(1), 81-91.

Smith, M., Cottrell, G., Gosselin, F., & Schynss, P. (2005). Transmitting and decoding facial expressions. *Psychological science, 16*(3), 184-189.

Sun, Y., Sebe, N., Lew, M., & Gevers, T. (2004). Authentic Emotion Detection in Real-Time Video. In *Computer Vision in Human-Computer Interaction* (Vol. 3058, pp. 94-104): Springer Berlin / Heidelberg.

Wallhoff, F., & [FG-NET]. (2005). Facial Expressions and Emotions Database from Technical University of Munich. from http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. (2000). Studying the Dynamics of Emotional Expression Using Synthesized Facial Muscle Movements. *Journal of Personality and Social Psychology, 78*(1), 105-119.

Williams, A. C. (2002). Facial expression of pain: an evolutionary account. *Behav Brain Sci, 25*(4), 439-455; discussion 455-488.