A Multimodal Database as a Background for Emotional Synthesis, Recognition and Training in E-Learning Systems

Anolli, Luigi; Mantovani, Fabrizia; Mortillaro, Marcello; Vescovo, Antonietta; Agliati, Alessia; Confalonieri, Linda; Realdon, Olivia; Zurloni, Valentino; Sacchi, Alessandro

How to cite

ANOLLI, Luigi et al. A Multimodal Database as a Background for Emotional Synthesis, Recognition and Training in E-Learning Systems. In: Affective Computing and Intelligent Interaction. ACII 2005. Tao, J.; Tan, T.& Picard, R.W. (Ed.). Berlin : Springer, 2005. (Lecture Notes in Computer Science) doi: 10.1007/11573548_73

This publication URL: https://archive-ouverte.unige.ch/unige:97217
Publication DOI: 10.1007/11573548_73

A Multimodal Database as a Background for Emotional Synthesis, Recognition and Training in E-Learning Systems

Luigi Anolli[1], Fabrizia Mantovani[1-2], Marcello Mortillaro[1], Antonietta Vescovo[1], Alessia Agliati[1], Linda Confalonieri[1], Olivia Realdon[1], Valentino Zurloni[1] and Alessandro Sacchi[1]

[1] CESCOM - Centre for Research in Communication Science,
University of Milan – Bicocca, Milan, Italy
{luigi.anolli, fabrizia.mantovani, marcello.mortillaro, .antonietta.vescovo, alessia.agliati, linda.confalonieri, olivia.realdon, valentino.zurloni, alessandro.sacchi}@unimib.it

[2] ATN-P LAB – Applied Technology for Neuro-Psychology Lab,
Istituto Auxologico Italiano, Milan, Italy

Abstract. This paper presents a multimodal database developed within the EU-funded project MYSELF. The project aims at developing an e-learning platform endowed with affective computing capabilities for the training of relational skills through interactive simulations. The database includes data coming from 34 participants and concerning physiological parameters, vocal nonverbal features, facial expression and posture. Ten different emotions were considered (anger, joy, sadness, fear, contempt, shame, guilt, pride, frustration and boredom), ranging from primary to self-conscious emotions of particular relevance in learning process and interpersonal relationships. Preliminary results and analyses are presented, together with directions for future work.

1 Introduction

Designing an automatic system able to express emotions and to detect the emotional state of a user is one of the main aims of the research field defined as affective computing [1]. The acknowledgment of the user's affective state can improve the effectiveness of a number of computer applications in a variety of fields. In particular, e-learning could be deeply empowered in its effectiveness since emotions are recognized to have a deep influence on the learning process. The Learning Companion project, at MIT, was one of the first experiences trying to address this issue in distance learning and computer-assisted training for children. A number of projects are currently being conducted in order to design e-learning platforms endowed with affective computing capabilities. One of these projects, funded by the European Commission and started in September 2004, is called MYSELF "Multimodal elearning System based on Simulations, Role-Playing, Automatic Coaching and Voice Recognition interaction for Affective Profiling" (www.myself-proj.it) and involves 14 partners (among Universities, research labs, educational and IT companies and SMEs) from 6 different EU countries. The target focus of Myself platform will be on training social and relational skills in different professional settings (banking, commerce, health-care, etc.) through interactive simulations. As far as affective computing features are concerned, three main issues are at the moment under investigation: the implementation of a virtual tutor provided with emotional expressive synthesis abilities; a multimodal emotional detection module able to get information on user's state along the learning path; the development of 3D interactive simulations and targeted exercises to improve emotional management, with specific focus on expression and recognition of emotions.

This paper presents preliminary work carried out by the Centre for Research in Communication Science (CESCOM) focusing on building a multimodal emotional database as a background for the development of MySelf platform and simulations.

Much work has been now carried out in the affective computing domain to perform the detection and inference of emotional state detection from physiological correlates [1-3], facial expressions [4,5], vocal-non-verbal features (such as F0, intensity, etc.) [6,7], verbal speech content, questionnaires or self-report measures and the detection of behavioural events (e.g. mouse-clicking) [8]. Since the integration of multiple sources of information could enhance power to achieve a reliable emotional recognition, building multimodal databases to test recognition algorithms is recognized a very important issue for affective computing research. While this need is clearly perceived within research community very few large multimodal databases are available, in particular including also physiological measures. From a review of the state of the art about emotional databases [9] emerged that among more than 40 emotional databases only 10-12 include more than one measure at the same time. Most of the databases deal only with speech or facial expression (in some cases also gestures) and even when considering few more complete multimodal databases available they mostly combine audio and visual (facial behavior) information; very few [10,11] added physiological features. Furthermore, acquiring ecologically valid emotional data with is very complex. Many of the databases available, in fact, ask subjects to act or pose emotions, in order to extract vocal and facial features. In the last years, this lack of naturalism has been severely criticized, and researchers try to induce emotions (rather than asking subjects to simulate them) or to collect data occuring in everyday life [12-14]. Finally, there is growing awareness that it is important to include a broader range of emotions than the six-seven basic ones. The number of databases that include emotions exceeding traditional primary emotions is constantly increasing [11, 13, 15, 16]. Due to the specific objective of Myself project, the range of emotions of interest include also secondary and self-conscious emotions like pride, shame, frustration, boredom, that do play a key role both in learning process and in interpersonal relationships; their investigation is therefore very relevant when trying to design interactive training simulations dealing with relational skills.

Starting from these premises, the present work focuses on *building a multimodal database* taking into account different sources of data (physiological measures, facial expression and posture, vocal non verbal features) for 10 different emotions (anger, joy, sadness, fear, contempt, shame, guilt, pride, frustration and boredom). We intend to systematically investigate the difference between emotions in the considered modalities and to verify the existence of systematic correlation between the different measures. Furthermore, we are interested in investigating whether these measures are modulated by specific personality characteristics and/or emotional management styles, assessed through dedicated self-report measures. This work, as discussed more into depth in section 4, might hold relevant implications for MySelf project, providing useful guidelines for emotional synthesis, recognition and training.

2 Methodology

2.1 Participants and experimental design

Participants were 34 students (16 male and 18 female) of the University of Milan-Bicocca (M = 22.62; SD = 1.30).
A 10 (emotion) x 2 (sex) mixed factor design (with repeated measures for emotion) was performed. Emotions considered were: anger, joy, sadness, fear, contempt, shame, guilt, pride, frustration and boredom.

Measures considered were: *physiological measures* (Heart Rate, HR; Skin conductance, SC; Respiration Rate, RR; Respiration Amplitude, RA, Finger Blood Amplitude, BA; Electromyography of the extensor muscle of the forearm, EMG); *non-verbal behavior related to facial expression and posture*: in particular, facial expression was coded using FACS (Facial Action Coding System[17]); *vocal acoustic parameters* referring to time (total duration, partial duration, duration of pauses, speech rate and articulation rate), fundamental frequency (F0) (mean, standard deviation, range, minimum and maximum) and intensity (mean, standard deviation, range, minimum and maximum). Besides these measures we administered a battery of paper-and-pencil tests (*Big Five Questionnaire and Emotion Regulation Questionnaire*) to assess some personality and emotional management style characteristics of the participants, as potential variables modulating the emotional expression at different levels.

2.2 Contextualized acting for eliciting emotions

As general framework, we attempted to induce emotions in the laboratory rather than to merely simulate it using actors according to the suggestions by Bachorowski [18]. To this aim, in a preliminary phase we prepared and validated ten prototypical texts eliciting the ten different emotions. The texts, consisting of 200 to 250 words, were characterized by a clear and univocal emotional episode and in each text an identical standard utterance was included on which to carry out subsequently acoustic comparisons. Each text included a first part describing the situation, and a second part where the protagonist speaks in first person. Such a method was previously adopted in the study of vocal correlates of emotions [16, 19, 20]. Procedure was modified in order to acquire other kind of data: participants were asked to read aloud texts written, trying to identify themselves with the main character of every narration (contextualised acting, similarly to Velten procedure, [21], but adding a contextual dimension through narration). To achieve a higher naturalism degree, naïve readers were chosen, and no direct mention of emotional terms was provided by researchers to the participants.

2.3 Procedure

Participants were introduced in a laboratory setting (individually), they were shortly briefed about the sensors and gave their consent to being video and audio recorded.

First of all, a baseline (3 minutes) for physiological parameters was measured for all participants. Then, the eliciting texts were presented on a single PC screenshot, in randomized order. Participants were asked to read each text first time silently and imagining the situations described, then reading them aloud in a natural and spontaneous way trying to identify themselves with the protagonist. At the end participants were administered the two questionnaires (Big Five and ERQ). The overall recording session took about 30 minutes.

2.4 Materials and instruments

Physiological data were collected through Procomp Thought Technology, able to detect real-time measure of physiological indices. One desktop PC (NEC, Powermate) was connected to the Procomp, and had the Biograph Infinity software for feature extraction installed. *Video data* were acquired using a webcam (Logitech Quickcam) placed on the top of the screen in front of the participant. Video files were analyzed through Theme software, a tool to code and analyse behavioural events, as well as to detect regularities within behavior, that are graphically translated in time patterns. Time patterns (T-patterns) are repeated sequences of behavioural units occurring in the same order and with approximately the same time distance during an observation period; according to recent studies [22], the use of this tool might provide very interesting insights in the dynamic and hierarchical features of nonverbal behavior organization. *Voice* was acquired through a unidirectional dynamic microphone (Shure, SM 48); vocal features were then extracted and analyzed through Computerized Speech Lab (CSL 4500 by Kay Elemetrics). A laptop PC (ASUS M3000N), ran the PowerPoint text slideshow that participants read.

3 Results

Due to space constraints and to the early stage of analysis, here are presented some preliminary data. Next steps for analysis will include: investigating the correlations between measures taken from different modalities; training and testing computational algorithms for automatic clustering and recognition; identifying individual profiles according to personality and emotional management characteristics.

Physiological data. We considered the mean value detected during the standard utterance period. Repeated-measure parametric and non parametric tests were performed on all the measures (according to the normality of their distributions). A significant main effect for emotion was acknowledged in ANOVA statistics for mean values of BA ($F_{9, 288} = 3.44$; $p< .01$), RR ($F_{9, 288} = 8.981$; $p< .01$), RA ($F_{9, 288} = 2.753$; $p< .01$). For EMG Friedman's non-parametric test showed a significant main effect for emotion ($Chi^2 = 27.915$; $p< .05$). A number of post hoc analyses showed significance patterns allowing to defining preliminary groupings of variables (Due to their number it is not possible to report all of them here). Results are generally consistent with previous studies: for example, emotions like joy, anger, fear are clearly differentiated from sadness, registering a higher level in skin conductance and EMG and a lower in finger blood amplitude, while other measures like heart rate fails to discriminate. For what concerns self-conscious emotions and emotions connected to learning, Pride and Shame showed similar means for every measure with the exception of RR, while Frustration and Guilt differed in respiration measures (RA and RR) and SC. From a whole consideration of these data subsequent analysis will be directed to the identification of possible physiological patterns for each emotion.

Table 1. Mean and sd of the physiological indices for the 10 emotions considered

	HR		BA		*RR*		RA		*SC*		EMG	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Anger	91,03	20,3	4,99	4,5	14,32	2,7	3,19	2,1	11,38	4,1	12,03	11,8
Boredom	91,69	17,6	4,53	3,7	14,55	3,2	3,23	2,1	11,35	4,7	7,60	8,1
Contempt	92,43	17,0	4,75	3,8	10,55	3,1	4,06	2,1	11,86	4,7	11,43	12,0
Fear	91,27	20,6	5,13	3,8	12,83	2,9	3,48	1,9	11,56	4,9	10,63	11,6
Frustration	87,35	19,7	6,00	4,8	10,19	2,9	4,14	2,6	11,38	4,9	8,65	10,5
Guilt	87,52	18,5	5,92	4,4	13,47	3,8	2,79	1,7	11,00	4,3	9,61	9,0
Joy	89,22	19,2	4,40	3,5	11,42	3,2	3,59	1,4	11,66	4,9	11,71	11,1
Pride	90,96	16,8	5,44	5,3	12,32	2,8	3,10	1,7	11,26	4,5	11,84	9,1
Sadness	89,74	20,2	5,85	4,9	11,36	3,4	3,65	1,9	11,22	4,4	9,31	11,0
Shame	91,75	18,2	5,06	4,2	13,89	2,6	3,39	2,3	11,32	4,3	11,05	10,5

Vocal features. A significant main effect of emotion was found in ANOVA statistics for every index considered with the exception of Pause: Speech Rate ($F_{9,\ 270}$ = 5.132; p< .01), Articulation Rate ($F_{9,\ 270}$ = 12.017; p< .01), Total Duration ($F_{9,\ 270}$ = 3.907; p< .01), Partial Duration ($F_{9,\ 270}$ = 10.796; p< .01), Intensity mean ($F_{9,\ 270}$ = 32.138; p< .01), Intensity SD ($F_{9,\ 270}$ = 4.423; p< .01), F0 mean ($F_{9,\ 270}$ = 9.222; p< .01), F0 SD ($F_{9,\ 27}$ = 3.628; p< .01). Also for vocal correlates a high number of post hoc analysis resulted significant and allowed a preliminary grouping of emotions according to their vocal nonverbal features profiles. In general, our results are consistent with scientific literature for the primary emotions included in our study. For example emotions like anger, joy and fear showed a significant higher Intensity mean and F0 mean than sadness. For secondary emotions, it emerged that Boredom and Frustration were quite similar (except for Intensity SD), as resulted for Pride and Shame, (except for F0 mean).

Non Verbal behaviour (facial expression and posture). Video sequences were analyzed and coded frame by frame using Theme Coder and then coded were entered in Theme Software, that generates both the frequency of every behavioural unit observed and Time patterns. A multivariate ANOVA with repeated measures showed a significant main effect of emotion ($F_{33,\ 2340}$ = 2.289; p< .01). At univariate tests a significant effect emerged for several nonverbal behaviour units: AU1, inner brow raising ($F_{9,\ 288}$ = 2.709; p< .05); AU2, outer brow raising ($F_{9,\ 288}$ = 3.469; p< .01); AU4, eyebrow lowering ($F_{9,\ 288}$ = 3.514; p< .01); AU6, cheek raising ($F_{9,\ 288}$ = 8.393; p< .01), AU7 ($F_{9,\ 288}$ = 4.459; p< .05); AU9, nose wrinkling ($F_{9,\ 288}$ = 3.098; p< .05); AU10, upper lip raising ($F_{9,\ 288}$ = 3.115; p< .05); AU12, lip corners raising ($F_{9,\ 288}$ = 22.081;

$p< .00$); AU26, jaw drop ($F_{9,\ 288} = 3.832$; $p< .05$); AU33, cheek blowing ($F_{9,\ 288} = 31.703$; $p< .00$); AU37, lips wiping ($F_{9,\ 288} = 2.646$; $p< .05$); hand on table ($F_{9,\ 288} = 4.203$; $p< .05$); trunk on chair ($F_{9,\ 288} = 3.539$; $p< .05$). Some nonverbal units mean frequencies were significant to identify a typical expressive nonverbal pattern for different emotions. In particular AU1 and AU2 were significantly more manifested in pride, AU4 and "hand on chair" in fear; AU6 and AU12 in happiness; AU33 in boredom and "trunk on chair" in guilt. By way of example, figure 1 represents a T-pattern detected during an emotional sequence, specifically during boredom: a sequence (recurring twice) of cheek blowing (AU33), eyebrow lowering (AU4) and head left inclination (AU55), that seems well represent the typical boredom-related facial expression.

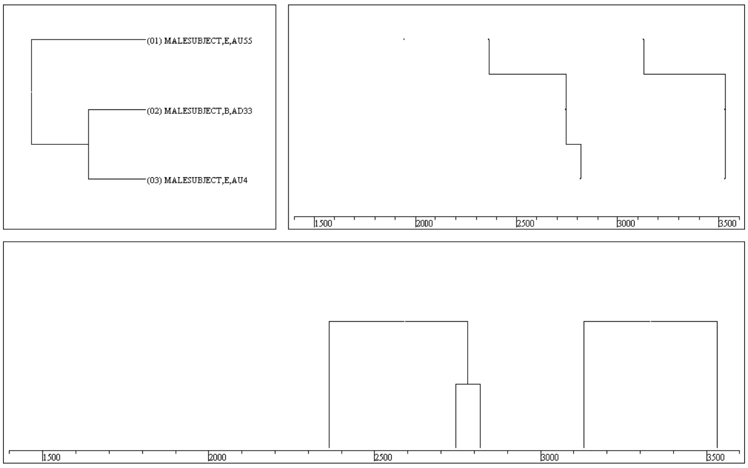

Fig. 1. T-pattern detected in a male subject during boredom sequence. (Legend: Male subject, ends, AU55; male subject, begins, AD33; male subject, ends, AU4).

4 Discussion and conclusions

In this paper we presented a multimodal emotional database which included vocal non verbal features, physiological signals, facial expression and posture for ten different emotions (anger, joy, sadness, fear, contempt, shame, guilt, pride, frustration and boredom).

Preliminary data presented in this paper supported the hypotheses of the existence of significant differences among emotions in the different modalities investigated; results for the single modalities seem to be, in general, quite consistent with scientific literature. Also, the investigation of complex, self-conscious emotions seemed to show specific multimodal profiles.

This work might have relevant implications for the ongoing MySelf project, at different levels.

First of all, for *emotional expression synthesis:* identifying prototypical nonverbal configurations for the different emotions might provide a useful basis to animate the 3D virtual tutor and the virtual characters of the interactive simulations. Of particular interest are data on facial expression and postural behavior: the use of Theme software allowed to investigate non only differences among emotions in terms of relative frequencies of single action units, but also in terms of emergence of higher order, temporal patterns characteristics of different emotions. We will try to use these prototypical configurations for 3D character animation (virtual tutor and simulation characters) and assess their effectiveness in terms of recognizability and representativeness. Issues related to temporal synchrony across the different modalities are also worth deeper investigation, especially for their inputs to animating virtual characters.

Second, the work might provide useful indications for the implementation of the *user's emotion recognition system:* work for training and testing multimodal recognition algorithms (in line with works like [3, 5, 23]) will be performed in the near future. These analyses should support efforts in the investigation of criteria for feature and algorithm selection, in order to identify the best combination of features and algorithm according to different clustering problems. Furthermore, within the project we will focus on how to integrate in the recognition modules information coming also from verbal speech content, self-report measures and specific events detectable by the e-learning system (such as failing/passing a test, etc.) in order to enhance consistency and reliability.

Third, material gathered within this study could be extremely useful for the implementation of *training exercises* specifically aimed at improving emotional expression and recognition skills.

Future work within the project will be also aiming at building a second database, more context-dependent (given the importance of restricting use domain in order to increase reliability). This work, which will start after the release of the beta version of the platform, will be hopefully able to enhance ecological validity of data gathered and allow a useful comparison with data coming from the first database presented in this paper.

Acknowledgments

The present work is part of the European Commission 6FP research project MYSELF (SME-2003-1-508258). Web site: www.myself-proj.it . The authors would also like to thank Dr. MariaChiara Gritti, Sara Manieri and Simona Raspelli for their active collaboration in this study.

5 References

[1] Picard, R. W.: Affective computing. The MIT Press, Cambridge MA (1997)

[2] Prendinger, H., Dohi, H., Wang, H., Mayer, S., Ishizuka, M.:. Empathic embodied interfaces: Addressing users' affective state Tutorial and Research. Workshop on Affective Dialogue Systems (ADS-04). Springer LNAI 3068, Kloster Irsee, Germany: (2004) 53-64

[3] Scheirer, J., Fernandez, R., Klein, J., Picard, R. W.: Frustrating the user on purpose: A step toward building an affective computer. Interacting with computers 14 (2002) 93-118

[4] Chen, L.S., Huang, T.S., Miyasato, T., Nakatsu, R.: Multimodal human emotion/expression recognition. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, Nara, Japan, IEEE Computer Society (1998)

[5] Kapoor, A., Picard, R.W., Ivanov, Y.: Probabilistic combination of multiple modalities to detect interest. In: Proceedings of International Conference on Pattern Recognition, Cambridge, England (2004)
[6] Oudeyer, P-Y.: The production and recognition of emotions in speech: features and algorithms. International Journal of Human Computer Interaction 59 (2003) 157--183
[7] Batliner, A., Fischer, K., Huber, R., Spilker, J., Nöth, E.: How to find trouble in communication. Speech Communication, 40 (2003) 117-143
[8] Pantic, M., Rothkrantz, L.J.M.: Toward an Affect Sensitive Multimodal Human-Computer Interaction. In: Proceedings of the IEEE Vol. 91, 9 (2003) 1370-1390
[9] Douglas-Cowie, E. et al.: Deliverable D5c-Preliminary plans for exemplars: Databases. Project Deliverable of Humaine Network of Excellence (Available online at: http://emotion-research.net/deliverables/) (2004)
[10] Amir, N., Ron, S., Laor, N.: Analysis of an emotional speech corpus in Hebrew based on objective criteria. In: Proceedings of ISCA ITRW on speech and emotion, Newcastle, Belfast, Textflow (2000) 29-33
[11] McMahon, E., Cowie, R., Kasderidis, S., Taylor, J., Kollias, S.: What chance that a DC could recognise hazardous mental states from sensor outputs? In: Tales of the disappearing computer, Santorini (2003)
[12] Campbell, W. N.: Recording Techniques for capturing natural everyday speech. In: Proceedings of LREC, Las Palmas (2002)
[13] Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P.: Emotional speech: towards a new generation of databases. Speech Communication, 40 (2003) 33-60
[14] Devillers, L., Vasilescu, I., Vidrascu, L.: . Anger versus Fear detection in recorded conversations. Speech Prosody (2004)
[15] Cowie R., Cornelius R.: Describing the emotional states that are expressed in speech. Speech Communication, 40 (2003) 5-32.
[16] Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. Journal of personality and social psychology, 70 (3), 614-636.
[17] Ekman, P., Friesen, W. V.: The Facial Action Coding System: A technique for measurement of facial movement. Palo Alto, CA: Consulting Psychologists Press (1978)
[18] Bachorowski, J. A.: Vocal expression and perception of emotion. Current Directions in Psychological Science,. 8 (1999) 53-57
[19] Douglas-Cowie, E., Cowie, R., Schroeder, M. (2000). A new emotion database: considerations, sources and scope. Proceedings of ISCA ITRW on speech and emotion, Newcastle, Belfast, Textflow (2000) 39-44
[20] Iriondo, I., Guaus, R., Rodriguez, A., Lazaro, P., Montoya, N., Blanco, J., Beradas, D., Oliver, J., Tena, D., Longhi, L.: Validation of an acoustical modelling of emotional expression in Spanish using speech synthesis techniques. In: Proceedings of ISCA ITRW on speech and emotion, Newcastle, 5-7 September 2000, Belfast, Textflow (2000) 161-166
[21] Velten, E.: A laboratory task for induction of mood states. Behaviour research and therapy, 6 (1968) 473-482
[22] Anolli, L., Duncan, S.Jr., Magnusson, M., Riva, G.: The hidden structure of interaction: From neurons to culture patterns. Amsterdam. IOS Press (2005)
[23] D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R. W., Graesser, A. C.: Integrating affect sensors in an intelligent tutoring system. In: Affective Interactions: The Computer in the Affective Loop, Workshop at 2005 International conference on Intelligent User Interfaces. New York: AMC Press (2005) 7-13