ElectroEmotion – A Tool for Producing Emotional Corpora Collaboratively

Lassi A. Liikkanen, Giulio Jacucci, Matti Helin
Helsinki Institute for Information Technology HIIT
Helsinki University of Technology TKK
P.O. Box 9800, FI-02015 TKK, Finland
Firstname.Lastname@hiit.fi
http://www.hiit.fi/UIx

Abstract

Emotion-aware applications supporting natural interaction are currently still a vision. One difficulty in developing these applications is the lack of multimodal corpora suitable for multiple use contexts, such as public spaces. Here, we introduce ElectroEmotion, a research tool prototype for collecting vocal and gestural corpora in novel contexts collaboratively. ElectroEmotion concept includes a public walk-up-and-use interface that allows users to produce multimodal expressions in an interactive environment. We describe the design of this system and report an experimental study, which evaluated the importance of inducting emotion and social influences in corpus acquisition. This preliminary investigation involved 12 users. By performing a video-based interaction analysis, we found that the participants demonstrated spontaneous multimodal activity and more distinctively emotional expressions in response to the emotion induction procedure. Social learning through examples provided by the experimenter influenced the way the subjects interacted. From these observations, we believe that the proposed concept could be developed into a functional system that can help to produce emotional corpora.

1. Introduction

Affective interaction might be divided into the *computational recognition of user emotion* and *communication of emotions back to user*, both relying on the computational modeling of human emotion. To people interested in creating emotion-aware applications or extending existing systems, such as call centre line prioritisation [1], emotion recognition is the crucial topic. Even though some modalities that convey emotion, such as facial expressions, can now be modelled and reproduced quite accurately [2] the same does not apply to aural or haptic senses, for instance.

The present implementations of affective computing rely heavily on machine learning approaches (e.g. [3]). Machine learning methods are used to analyse corpora of tagged emotional expressions and to learn in a bottom-up manner the regularities related to emotion from the input signal. The consequence of this approach is that implementing affective interaction for input purposes becomes largely *corpus dependent*. This severely restricts the usefulness of the corpora. There are several reasons for this. First, emotional expressions are known to be rather *context dependent* indicating that expressions can not easily be generalized across contexts (see [4]). Second, the expressions found in corpora are also *user-dependent* in terms of cultural background, sex, and so forth. Finally, collecting, annotating, and evaluating corpora is a very time consuming task. As a consequence, the producers of corpora have often been reluctant to give away their work to the public. Although there have recently been several advances in the area towards resource sharing, e.g. with speech corpora [5, 6], this fact clearly hinders the deployment of affective intelligence for interactive applications.

In this paper we address this problem. We propose a potential solution for the issue through the production of a collaborative corpus with an interactive environment. Studies regarding embodied interaction with tangible, interactive toys and devices have introduced the concept of affective loop experience [7-10]. This concept is derived from studies in which users interact with a device in a purposefully emotional way. Previous investigations show that this setting tends to recreate real emotions by providing feedback during the embodied interaction. The feedback may be partially generated by the system, but essentially the user is affected by the process of interacting. In other words, emotions are constructed during in interaction [8].

Could these emotional loops be exploited in order to collect a corpus from some context of interest? We believe that a interactive system could be helpful in addressing issues of *context dependency*, *collection, annotation*, and *evaluation efforts*. We propose the idea of creating an interactive tool to create corpora in a user-driven way, a tool that can be transferred to different contexts (physical and social surroundings). A that tool can function as a stand-alone system and attract users to participate in a freely accessible game of producing (collection) and evaluating (evaluation)

978-1-4244-4799-2/09/$25.00 ©2009 IEEE

emotional expressions. With this tool users could collaboratively produce an emotional corpus involving any desired modalities. To the best of our knowledge, this kind of concept has not been discussed before. For this reason, the design of a such system poses a substantial challenge.

In this paper we present our first research prototype and insight from the design process of *ElectroEmotion*, a research tool prototype to test this idea. It addresses aspects of both effort involved in collection and context dependency. We also present an evaluation of the prototype to determine the feasibility of the suggested application and discuss how this work may continue.

2. Background

Human communication is essentially multimodal. Typically all modes of human communication, such as prosody, respiration, and hand gestures, are all integrated. It follows that affective computing should also be multimodal if it tries to follow human information processing [11]. However, most approaches to the recognition of emotion so far have either focused on a single modality, mainly facial expressions or speech, or required highly controlled environments in order to function, thereby distancing themselves from human way of handling the matter.

For ElectroEmotion we choose two information streams, voice and hand gesture. *Voice* is a major channel for communication, capable of quickly conveying a large amount of information. This includes emotional content [12, 13] in linguistic and paralinguistic, or nonverbal forms. Different layers of information in the speech signal and their relation to emotional content have been proposed [14, 15]. Typically, analyses concentrate on finding the set of features in the speech signal that are most salient in terms of emotion. These features are then used to analyze a corpus and train algorithms. Compared to vocal expressions, *gestures* have received less attention from emotion researchers [16]. The bodily gestures involved in dance have been investigated [17] but hand gestures are a little known topic despite the fact that they are the basis of sign languages and non-signers have demonstrated the capability to perceive them [18].

To the best of our knowledge, there are no applications that have been explicitly built for the collaborative production of emotional corpora. However, there are a few related systems that may help in drafting a new system. SenToy [7] requires its users to express themselves by interacting with a tangible doll that is equipped with sensors to capture the users' gestures. In a similar way Melder et al. [19] presented a multimodal affective mirror that senses and elicits laughter with vocal and facial affect sensors.

Looking at the state of the art in traditional multimodal corpus acquisition, we observe that typically corpora are elicited from available TV videos or by actors in a laboratory. Closest example to ElectroEmotion comes from Zara et al. [20], present the EmoTaboo protocol for the collection of multimodal emotional behaviors from human-human interactions in a game context. Martin et al. [21] describe how to annotate multimodal behaviors observed during mixed and real emotions from TV interviews. Finally, Bänzinger and Scherer [22] argue that well-designed corpora of acted portrayals of emotion can be useful. They support their arguments by presenting GEMEP corpus as an example of a multimodal corpus.

3. Description of the tool prototype

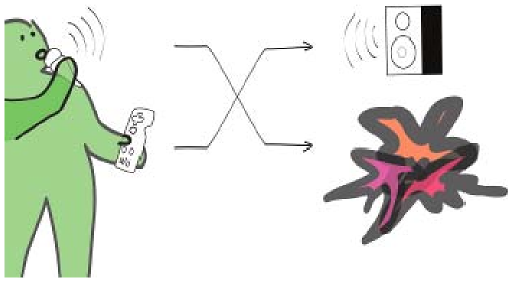

Figure 1: ElectroEmotion concept of multimodality

The ElectroEmotion prototype of was created to investigate how to produce an affective loop experience and use this to produce an audio-gestural corpus collaboratively. The prototype is intended for a single user. It includes portable input devices, a projection screen, and loudspeakers (Fig. 1). Input devices record voice and gesture signals that are used to create realtime audio-visual feedback. The prototype uses standard hardware but the software is specifically programmed in C#. The application is currently straightforward and, once it has been initiated by the experimenter, the user can swiftly interact and create emotional displays (the feedback) in realtime.

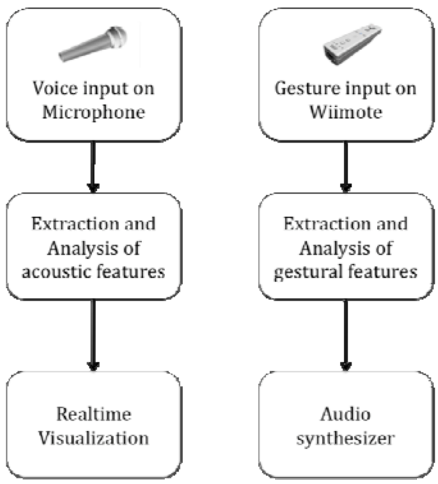

Figure 2: Architecture of the ElectroEmotion prototype in terms of information transformation

In ElectroEmotion gestures are operationalized to hand movements that are recorded using the accelerometers of a Nintendo Wii remote controller. Vocal expressions are captured using either a wireless head-mount portable microphone or a regular hand-held one. These inputs are processed in real time to provide feedback that emphasizes the expressive content of the input by extracting emotion-relevant features (see 3.1). The auditory input controls the visual output (a *visualization* of sound) and gestural input is used to generate sounds (a *sonification* of motion). These modalities are handled independently and the voice picked up by the microphone is not fed back through the speakers. This principle is illustrated in Figure 2.

3.1. Processing input features into outputs

In order to create a meaningful visualization from vocal input, we selected sound features that are likely to be useful for conveying emotional information on the basis of the literature [12, 13]. The candidate features, such as high-frequency energy, were considered with caution, as the application presented an environment, which would address not only speech but all vocal expressions. Thus the hypotheses derived from the literature rather provided a set of working hypotheses for the first prototype. The audio features we selected are indicated in Table 1 along with their visualization features.

Table 1: The association of acoustic and visual output feature

Audio input feature	Visualization feature
Band energy	Block height / visibility
Pitch (averaged)	Row color
HF energy	Spikidness
Intensity	Growth factor

The metaphor for the visualization comes from 3D histograms, so that we draw rows of 3D columns changing in the X and Y dimensions. The time is represented on the Z axis by drawing new sets of columns creating visualization similar to that shown in Figure 3. The relative height of the columns is derived from spectral band power, so the functionality is reminiscent of a spectral analyzer. The other parameters follow the list in Table 1.

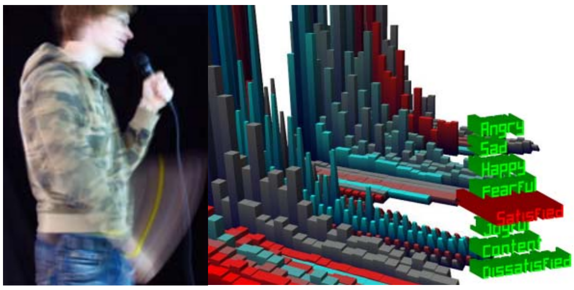

Figure 3. A user interacting with ElectroEmotion and an example of a resulting visualization.

Gestural input is used in ElectroEmotion to trigger and modify sounds generated by a software synthesizer (FM8, Native Instruments GmbH, Germany**)**. The purpose is to provide realtime (40 ms latency) sound responses to hand movements and gestures. This mapping is not intended to produce music, even though initial pitches of the synthesizer sounds are derived from an equal temperament chromatic scale. To prevent musical associations [23], the succeeding sounds triggered by gestures are random, i.e. do not follow any musical key.

Table 2: Mapping gestural input features to synthesizer

Gestural input feature	Auralization feature
Power	Main volume
Spatial extent	Pitch bend
Fluidity	Oscillator detune
Temporal expressivity	Arpeggiator speed
Accelerometers (x,y)	Subpatch morphing
Energy (x,y,z)	Onset velocity

3.2. Interacting with ElectroEmotion

The ElectroEmotion prototype has two operating modes. The *free exploration* mode allows its users to create audio-visual performances freely. The application is in a constant interactive state and ready for user performances. The length of the performances is constrained to 10 + 7 seconds, as will be described below. In the other version, called *prompt,* subjects receive a repeating cycle of practice and perform – prompts related to eight basic emotional categories. We selected the *Excited, Dissatisfied, Happy, Angry, Sad, Joyful, Fearful,* and *Satisfied* categories which correspond to the extreme quadrants of the dimensional models of emotion [e.g. 24, 25]. The prompts first indicate 'Practice X' for 7 seconds, after which the visualization is re-initialized and a prompt 'Perform X' is displayed for ten seconds in which 'X' stands for an emotional category label.

4. Evaluating the prototype tool

An evaluation was conducted in order to examine the suitability of the designed application for collecting emotional corpora, but not producing a corpus yet. In particular, we needed information about how people would *interact*, approach, and utilize this interactive application. We set out to investigate two crucial interaction aspects that would influence the real-life feasibility of the prototype – the factors of *inducing emotion* and *social influences*.

It is very important to understand the function of constraints on performative interaction. For instance, is it easier to be performative when you have a certain goal to fulfill? Thus we wanted to investigate the effect of suggesting vs. omitting an expressive goal for a performance. To study this factor labeled INDUCTION, we prepared two versions of the software, where only one included explicit emotional prompts. Additionally, we applied an emotional priming procedure as a part of task briefing ([26]) to help the users perform emotionally. In this procedure, the subject was asked to recall a personal episode from their near past associated with a given emotional label, in this case, one for each of the eight emotions. We hypothesized that the prompts would facilitate performative interaction.

The other independent variable concerned the impact of social learning. If we hope to apply the corpus collector across contexts, we need to be aware of how and if social learning influences emotional expressivity. This means we expect that social learning, mimicking, and reflective behavior might play a part in performative interaction. For the present study, we decided to study this by manipulating an example given by the experimenter (a social learning intervention) as an experimental factor called SOCIAL. This was possible as the subjects would use the system alone by themselves. The example included a short vocal expression (less than 5 sec.) using the free-form prototype version and a simultaneous demonstration of sound generation. We assumed that the example would lead subjects to perform in a more uniform fashion.

4.1. Evaluation design and methods

The evaluation utilized a balanced 2 X 2 experimental design with INDUCTION as a within-subject variable and SOCIAL as a between-subjects variable. Subjects were assigned to different factor conditions randomly using a Latin square method. The dependent variables of interest were the number of different interaction activities and the amount of vocal emotional activity, all derived from a video-based interaction analysis. To gather these data, all the test sessions were videotaped digitally in high definition with the participants' written consent. The video observations were supported by the notes of the experimenter. For the purposes of evaluating the usability and comfort of the users we also used qualitative methods, interview and questionnaires. However, these results are reported only partially. All user data (audio and gestures) were logged to a corpus database and an extra video stream was added to simultaneously capture the interaction. The quality of the corpus and the logs will be analyzed in the future.

4.2. Test procedure and space layout

We organized the evaluation in a laboratory at Helsinki University of Technology TKK in Espoo, Finland. The author LL acted as the experimenter and no other staff were present during the tests. Initially, all subjects received a similar kind of briefing about the application, after which some of them received the experimenter example (positive SOCIAL condition). Next they got dedicated briefing for either the free or prompt version and continued to complete the task. This was followed by a short break before the briefing for the next task was provided. This was either prompted or free, depending on which one was completed previously. The data from all users were collected constantly until at least 5 minutes of data for both test versions had been acquired. The sessions concluded with a debriefing about the study.

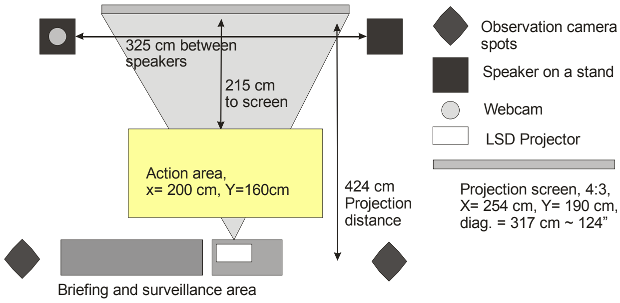

Figure 4: Room layout for the evaluation (not in scale).

The installation was constructed in an office space (layout shown in Figure 4). The test users received instructions, briefings, and other communications from the experimenter in the briefing and surveillance area. After this they performed within the boundaries of the action area marked on the floor. Two different observation camera spots were used depending on the spatial requirements. The experimenter remained constantly in the surveillance area, except in the SOCIAL condition, in which the example was given in the action area. The action area was determined by the limitations of overhead beaming, and the visual angle of the webcam.

4.3. Analyses

Video data were analyzed in ELAN (ELAN 3.6.0 for Windows, Max Planck Institute, Nijmegen, Netherlands) for activity in different modalities. For practical reasons the granularity of the analysis was relatively low. The actions were evaluated in a 17 second window (a single cycle), corresponding to the duration of the practice-perform loop of the prompt version. Four *primary* and two *supporting* categories were used to quantify the different types of activity with the system. The primary categories were verbal and non-verbal vocalizations, hand gestures, and multimodal interactions (synchronized or asynchronous). The two supporting categories were unimodal interactions and whole-body expressions. Emotional vocal activity was additionally scored when this activity was explicit. For each category, it was determined whether the subject exhibited this kind of activity or not. Finally, activity over all cycles was summarized using a four step scale. The scale had the following definitions: 0 = no activity, 1 = seldom (1-2 active cycles), 2 = repeated (> 2 cycles), 3 = continuous (nearly all cycles). These scores were rank ordered for statistical analysis on rank-order average. The time periods in which there were technical problems were excluded and one subject (S09) was excluded because of a software failure. The statistical tests were computed with SPSS 15.0.1.

5. Results

To evaluate "an interactive installation" we recruited 12 subjects (age range 19-45, M=29, SD=9, 4 males, 8 females) who had no information about the real purpose of the study. The subjects were initially contacted by email and asked to participate for an hour; they received two cinema tickets as compensation for their time. To express interest, they filled in an electronic recruitment form with background information. The selected participants were either university students (N=10) or staff and provided their informed consent.

The background questionnaire assessed the subjects' attitudes towards interactive systems, computers and art. We screened for subjects with adequate computer skills and who were familiar with interactive art. The sample of twelve subjects was divided for the analyses. Two subgroups of subjects were also extracted: exploratory pilots (N=2+2) and experimental core (N=8). For the video-based interaction analysis, we had a corpus of 11 subjects (1 excluded due to technical reasons), out of whom 8 subjects with an identical, balanced procedure were analyzed further. The first four subjects served primarily as pilots but their efforts were also analyzed when appropriate. The unit of analysis here is a *test*. A test refers to a single session with a prompt or the free version and implies that that each subject contributed two tests, so that the core group contributed 16 tests and pilots 6.

5.1. Coping with multimodal interaction

Table 3: Classification of different activity types across all tests (N=23, from 13 subjects). Mode classes in bold.

	None	Seldom (1-2 cycles)	Repeated (>2 cycles)	Continuous (~all cycles)
Verbal activity	**14**	5	4	0
Non-verbal activity	3	2	**11**	7
Gestural activity	0	0	7	**16**
Bodily activity	**9**	4	**9**	1
Multimodal activity	3	4	**9**	7
Uni-Modal activity	**8**	5	6	4
No activity	**21**	2	0	0
Emotional vocal act.	**9**	3	4	4

In general the subjects were very active and engaged in multimodal interaction. The video-based activity analysis demonstrated that the majority of the tests were completed with a fair amount of multimodal activity. However there were also several sequences in which people interacted solely in a single domain (unimodal activity). The results collected in Table 3 show how the dominant type of explorative activity was gesturing. It was present in nearly all of the cycles. Non-verbal vocal activity was also quite typical. Gestural and non-verbal activity together produce the multimodal activity score, because there the little verbal vocal activity did not count. An interesting observation was the fact that many users spontaneously produced different bodily gestures even though this was not encouraged. For instance, some people stepped around the action area, some practiced dance moves, some swayed from side to side etc.

In Table 4 below the balance of modalities is broken down showing that although most tests and subjects were balanced, the tendency towards gestural dominance is considerable.

Table 4: Classification of tests and subjects by activity.

Tests by modal preference	Tests	Percent
Slightly vocal	2	8.7 %
Balanced	10	43.5 %
Slightly gestural	6	26.1 %
Gestural	2	8.7 %
Dominantly gestural	3	13.0 %
Total	23	

Subjects by modal preference	Subjects	Percent
Slightly vocal	1	7.7 %
Balanced	6	46.2 %
Slightly gestural	4	30.8 %
Dominantly gestural	2	15.4 %
Total	13	

There are several explanations for the gestural preference. The two dominantly gestural subjects both stated in the interview that they did not feel comfortable about expressing themselves vocally. This shows in the way they interacted with the system relying completely on a single modality and highlights the individual differences in daring to explore the expressive potential of voice. The amount of control over the sounds produced might have facilitated gestural interaction, presenting it as a tangible toy, where as vocal performances might have been too personal to feel comfortable with, as indicated by the interviews.

5.2. Experimental outcomes

Four subjects (half of the identical procedure group) were presented an example by the experimenter before starting the first task. This SOCIAL factor changed the interaction as hypothesized. The subjects, who had received the example, had higher activity scores in non-verbal and multimodal activity and less unimodal activity (see Table 5; negatively related to multimodality). Gestural activity was not affected. This maybe explained by the general lack of variation in gestural activity– nearly all the participants performed at the ceiling level of the scale used and differences were thus unlikely to occur. Verbal vocal activity was not affected either, probably because the example was non-verbal and gestural and non-verbal activity dominated.

Table 5: Effect of experiment example on activity scores.

	Mean rank		Kruskal-Wallis	
Factor	**Example**	**No-example**	**Chi^2**	**p**
Uni-Modal activity	4.938	12.063	9.578	**0.002**
Multimodal activity	11.688	5.313	7.668	**0.006**
Non-verbal activity	11.063	5.938	5.175	**0.023**
Verbal activity	9.938	7.063	1.817	0.178
Gestural activity	9.500	7.500	1.250	0.264
Bodily activity	9.313	7.688	0.551	0.458

Note : 1 degree of freedom for all factors' Chi^2 scores

We also compared two conditions of INDUCTION, where only one within-subjects group was prompted to perform emotionally. In the interaction activity analysis, there were no differences between the prompt and free versions. However, vocal emotional activity was significantly increased by emotional prompts (Kruskal-Wallis $\chi2=4.846$, $p=.028$, $N=10$). From the qualitative data, it seemed that the prompt had highly individual effects. Some participants who tried and managed to follow the task instructions seemed to be activated by the prompt. For others, the existence of the prompt seemed to increas the lethargy of already passive subjects, and comparable to social pressure. feedback from users indicated that this may have been due to inadequate time to orient to varying affective prompts.

6. Discussion and conclusions

In this paper we have addressed the challenge of creating multimodal emotional corpora in a novel, collaboriate way. This was motivated by the needs of machine learning approaches used to implement emotion recognition in various contexts. Recent work on the creation of multimodal corpora for affective computing relies on naturally elicited or enacted corpora [19-21]. We advocate producing corpora in interactive environments for two reasons. First, it enables expressive features in different modalities to be investigated by eliciting them in realtime. This may support the development of alternatives to machine learning approaches. Second, the emergence of affective loop experiences in interaction may give rise to realistic emotions and enable to store these displays to a corpus. For these reasons we presented a research tool concept to meet the challenge based on the idea of an expressive, affective loop experience. We also introduced a tool prototype called ElectroEmotion to investigate how people could interact successfully with such system.

In our initial study, ElectroEmotion elicited expressions in multiple modalities. It demonstrated that the dimensions of social interaction and inducing emotion are influential factors in this kind of application, even though we could not yet fully control them. From an activity analysis we found that unimodal emotional expressions predominated over coordinated multimodal ones. In particular, the subjects preferred gestural interaction and explored the sounds produced by gestures. Exclusively vocal performances were rare. Interestingly, most auditory expressions were non-verbal hissing, grunting, whistling, or humming. The application and the context clearly did not particularly encourage people to talk or sing.

The nature of the task significantly influenced the nature of expressions produced with ElectroEmotion. The induced emotion condition produced very consistent results. The finding that the subjects responded to emotional prompts emotionally suggests that feedback may have created something similar to an affective loop experiment, promoting emotional expression. With free exploration this did not occur.

Social influences will be a design challenge for the future. We found that even simple manipulation of the task by giving or omitting an experimenter example influenced the outcomes. This finding should influence several design decisions, e.g. how many users are allowed at a time and how people are instructed or encouraged to participate and interact. All these findings should be considered in the design of the next generation interactive corpus production tool.

6.1. Issues with the approach

We set out to propose a rather radical solution to a known conundrum. We must acknowledge that there are several issues related to the idea we propose. The nature of the corpora that might be produced by a system similar to ElectroEmotion is debatable. We assumed that if an affective loop can be created, this would take us a long step from acted towards real ones. However, it raises the issues of how genuine are these emotions and how much they are affected by the particular realtime feedback provided? Constraints on the application and input modalities exist and one could question whether the corpus might become *application specific*. This method of corpus production has both its pros and cons in terms of corpus quality. These should be empirically tested to see whether a corpus collaboratively produced with a future version of ElectroEmotion is better, worse, or equal to corpora gathered in traditional, labor intensive ways. We believe that ElectroEmotion could still be used to acquire corpora useful for specific typology of interactive applications and social settings.

There are also ethical concerns in collaborative corpus acquisition. If we envision that this tool would be accepted and could be used to sample, for instance, emotional expressions in public spaces, the following issues rise: How would one retrieve informed consent, and should one deal with minors? Can we ethically acquire data that is difficult to anonymize data without losing the main content from modalities such as facial video and speech?

6.2. Conclusions and future work

The research tool we presented here provides an example of how to organize corpus collection in a collaborative and interactive way. Relying on the idea of an affective loop experience we suggested that experiences evoked by an interactive system might be utilized in corpus creation. Our results show that people can adopt this way of working and create emotional displays, but several constraints exist. However, we think that the breakthrough of social media, crowd sourcing, and similar phenomena relating to users empowered by new technology, will be exploited in the production of emotional corpora in the future. We have proposed a way to leverage this resource.

However, the ElectroEmotion prototype is just a start and future systems must also address corpus annotation and evaluation. Evaluation can easily be coupled to ElectroEmotion, for instance, by having users to compete in guessing emotional labels for other people's performances. This could be implemented by something akin to Feeltrace [27]. The production of a future version will also require experimental studies on different modalities and effective feedback strategies.

Acknowledgements

Supported by European Community's Sixth Framework research project CALLAS. ElectroEmotion was designed and implemented with the help of Rodolfo Samperio, Umair Khan, and Jari Kleimola, thank you!

References

[1] V. A. Petrushin: Emotion Recognition In Speech Signal: Experimental Study, Development, And Application. Proc. Sixth International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China: 4, 2000.

[2] F. Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. D. Carolis: From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent, International Journal of Human-Computer Studies, 59, (1-2): 81-118, 2003.

[3] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias: Emotion Analysis in Man-Machine Interaction Systems, in Emotion Analysis in Man-Machine Interaction Systems: 318-328, 2005.

[4] E. Douglas-Cowie, L. Devillers, J. C. Martin, R. Cowie, S. Savvidou, S. Abrilian, and C. Cox: Multimodal databases of everyday emotion: Facing up to complexity. Proc. INTERSPEECH 2005, Lisbon, Portugal2005.

[5] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach: Emotional speech: Towards a new generation of databases, Speech Communication, 40, (1-2): 33-60, 2003.

[6] D. Ververidis, and C. Kotropoulos: Emotional speech recognition: Resources, features, and methods, Speech Communication, 48, (9): 1162-1181, 2006.

[7] A. Paiva, R. Prada, R. Chaves, M. Vala, A. Bullock, G. Andersson, and K. Höök: Towards tangibility in gameplay: building a tangible affective interface for a computer game. Proc. 5th international conference on Multimodal interfaces, Vancouver, BC: 60-67, 2003.

[8] K. Höök: Affective Loop Experiences – What Are They? Proc. PERSUASIVE 2008, Oulu, Finland: 1-12, 2008.

[9] P. Sundstrom, A. Stahl, and K. Hook: In situ informants exploring an emotional mobile messaging system in their everyday practice, International Journal of Human-Computer Studies, 65, (4): 388-403, 2007.

[10] P. Sundström, and K. Höök: eMoto: affectively involving both body and mind2005.

[11] Z. Zeng, M. Pantic, and T. S. Huang: Emotion Recognition Based on Multimodal Information, Affective Information Processing: 241, 2008.

[12] R. Cowie, and R. R. Cornelius: Describing the emotional states that are expressed in speech, Speech Communication, 40, (1-2): 5-32, 2003.

[13] K. R. Scherer: Vocal communication of emotion: A review of research paradigms, Speech Communication, 40, (1-2): 227-256, 2003.

[14] L. Devillers, L. Lamel, I. Vasilescu, C. Limsi, and F. Orsay: Emotion detection in task-oriented spoken dialogues2003.

[15] T. Vogt, and E. Andre: Comparing Feature Sets for Acted and Spontaneous Speech in View of Automatic Emotion Recognition. Proc. IEEE International Conference on Multimedia and Expo (ICME 2005) Amsterdam, Netherlands: 474-477, 2005.

[16] J. Montepare, E. Koff, D. Zaitchik, and M. Albert: The Use of Body Movements and Gestures as Cues to Emotions in Younger and Older Adults, Journal of Nonverbal Behavior, 23, (2): 133-152, 1999.

[17] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe: EyesWeb: Toward gesture and affect recognition in interactive dance and music systems, Computer Music Journal, 24, (1): 57-69, 2000.

[18] J. K. Hietanen, J. M. Leppanen, and U. Lehtonen: Perception of emotions in the hand movement quality of Finnish sign language, Journal of Nonverbal Behavior, 28, (1): 53-64, 2004.

[19] W. A. Melder, K. P. Truong, M. Den Uyl, D. A. Van Leeuwen, M. A. Neerincx, L. R. Loos, and B. S. Plum: Affective multimodal mirror: sensing and eliciting laughter. Proc. HCM'07, Augsburg, Germany: 31-40, 2007.

[20] A. Zara, V. Maffiolo, J. Martin, and L. Devillers: Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics, in Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics: 464-475, 2007.

[21] J. C. Martin, G. Caridakis, L. Devillers, K. Karpouzis, and S. Abrilian: Manual annotation and automatic image processing of multimodal emotional behaviors: validating the annotation of TV interviews, Personal and Ubiquitous Computing, 13, (1): 69-76, 2009.

[22] T. Bänziger, and K. Scherer: Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus, in Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus: 476-487, 2007.

[23] L. L. Balkwill, and W. F. Thompson: A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues, Music Perception, 17, (1): 43-64, 1999.

[24] A. Mehrabian: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament, Current Psychology, 14, (4): 261-292, 1996.

[25] J. A. Russell: A circumplex model of affect, J. Pers. Soc. Psychol., 39, (6): 1161-1178, 1980.

[26] N. Ravaja, T. Saari, K. Kallinen, and J. Laarni: The role of mood in the processing of media messages from a small screen: Effects on subjective and physiological responses, Media Psychology, 8, (3): 239-265, 2006.

[27] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor: Emotion recognition in human-computer interaction, IEEE Signal Processing Magazine, 18, (1): 32-80, 2001.