Recording audio-visual emotional databases from actors: a closer look

Carlos Busso and Shrikanth S. Narayanan

Speech Analysis and Interpretation Laboratory (SAIL)
Viterbi School of Engineering,
University of Southern California, Los Angeles, CA 90089,
busso@usc.edu, shri@sipi.usc.edu

Abstract

Research on human emotional behavior, and the development of automatic emotion recognition and animation systems, rely heavily on appropriate audio-visual databases of expressive human speech, language, gestures and postures. The use of actors to record emotional databases has been a popular approach in the study of emotions. Recently, this method has been criticized since the emotional content expressed by the actors seems to differ from the emotions observed in real-life scenarios. However, a deeper look at the current settings used in the recording of the existing corpora reveals that a key problem may not be the use of actors itself, but the ad-hoc elicitation method used in the recording. This paper discusses the main limitations of the current settings used in collecting acted emotional databases, and suggests guidelines for the design of new corpora recorded from actors that may reduce the gap observed between the laboratory condition and real-life applications. As a case study, the paper discusses the *interactive emotional dyadic motion capture database* (IEMOCAP), recently recorded at the University of Southern California (USC), which inspired the suggested guidelines.

1. Introduction

Humans use intricate orchestrations of vocal and visual modes to encode and convey intent and emotions (Busso et al., 2007b; Cowie and Cornelius, 2003; Ekman and Rosenberg, 1997). The expressive elements in the production and perception of voice, spoken language and non-verbal gestures are central to human communication. Understanding and utilizing these expressive emotional elements, hence, is key to facilitating any creative human experience, whether for learning or entertainment.

One of the major challenges in the study of emotion expression is the lack of databases with genuine interaction that comprise integrated information from the relevant communicative channels (e.g., speech, facial expression, and body posture). Human capabilities in creating expressive emotional experiences through acting provide opportunities to tackle the problem in a systematic and controlled fashion that is impossible or impractical to do with mere observational or post hoc analyses of human interaction data. Unfortunately, the current approaches used to record and post-process the emotional data obtained from actors are less than ideal to generate material that are closer to the emotions observed in real-life scenarios. The use of naïve speakers or inexperienced actors, the lack of contextualization, and the inadequate emotional descriptors are some of the main limitations found in the design of the existing emotional corpora.

The present paper considers the role of acting as a viable research methodology for studying human emotions, noting both the inherent limitations and the advantages the approach provides. This paper discusses some guidelines with the aim of designing emotional databases from actors that will closely represent the emotions observed in real-life scenarios. These guidelines, which are inspired by the lessons learned from our recent experience of recording the *interactive emotional dyadic motion capture database* (IEMOCAP) at USC, emphasizes the importance of using trained actors involved in their roles during interaction, rather than recording monologues or short sentences. Likewise, we highlight the importance of contextualization to collect genuine databases. We hope that the new generation of databases recorded from actors will decrease the discrepancy observed between laboratory and real-life conditions.

The paper is organized as follows. Section 2. presents the related work. It discusses some of the problems found in existing emotional databases. Section 3. provides suggestions that can be used to obtain more genuine emotional databases recorded from actors. As a case study, Section 4. describes the design, collection and evaluation of the IEMOCAP database, in which the suggested guidelines were followed. Finally, Section 5. gives the final remarks and our future directions.

2. Background

Acting and actors have played a key role in the study of emotions. Douglas-Cowie *et al.* reviewed some of the existing emotional databases, and concluded that in most of the corpora the subjects were asked to simulate (“act”) specific emotions (Douglas-Cowie et al., 2003). In most of these cases, naïve speakers or actors without experience were asked to read short utterances or dialogs with few turns, without proper contextualization, which plays a crucial role in how we perceive (Cauldwell, 2000) and express emotions (Douglas-Cowie et al., 2005).

While desirable from the viewpoint of providing controlled elicitation, the use of actors under the current experimental settings has discarded important information observed in real-life scenarios (Douglas-Cowie et al., 2005). As a result, the performance of emotion recognition significantly degrades when automatic recognition models developed using such databases are used in real-life applications (Batliner et al., 2000; Grimm et al., 2007), where a blend of emotions is observed (Douglas-Cowie et al., 2005; Cowie et al., 2005). Differences between spontaneous (“real”) and simulated (“acted”) display of emotions have been studied in previous work. For example, Ekman discussed that there are certain facial action movements that subjects cannot voluntarily display when they are not experiencing certain emotions (e.g., enjoyment, anger, fear and sadness) (Ekman, 1993). Efforts in this direction have focused on analyzing differences between real and acted smiles (Cohn and Schmidt, 2004) and eyebrow actions (Valstar et al., 2006). As a result, the research community has recently shifted to other sources of emotional databases, neglecting acting and creative arts as a viable means for studying emotions.

Examples of the most successful efforts to collect natural new emotional databases to date have been based on broadcasted television programs (Belfast naturalistic database, VAM, EmoTV) (Douglas-Cowie et al., 2003; Grimm et al., 2007; Abrilian et al., 2005), recordings in situ (lost luggage) (Scherer and Ceschi, 1997), asking subjects to recall emotional experiences (Amir et al., 2000), inducing emotion with a Wizard of Oz approach (SmartKom) (Schiel et al., 2002), using games specially designed to emotionally engage the users (EmoTaboo) (Zara et al., 2007), and inducing emotion through carefully designed human-machine interaction (SAL) (Cowie et al., 2005; Caridakis et al., 2006). However, these approaches have core limitations such as ethical issues (e.g., inducing emotions), or copyright problems that prevent the wide distribution of the corpora (Cowie et al., 2005). They are also constrained to specific domains. Furthermore, these techniques lack control over the microphone and camera locations, and the lexical and emotional content. In addition, some of the recordings have noisy visual and/or acoustic backgrounds and incomplete information from modalities (only some human communicative channels are recorded). In contrast, recording databases from actors offers the flexibility to control every aforementioned aspect.

We believe that the main problems of existing databases recorded from actors may not be the use of actors itself but the methodologies and materials used to record the existing corpora, which can be made more systematized. For example, different acting styles and methods can be utilized to enable systematic and consistent elicitation (Enos and Hirschberg, 2006). The connection with real-life ("non-acted") scenarios still needs to be clearly established. Furthermore, important aspects of human interaction such as the cognition and multimodal aspects in human interaction, largely ignored thus far, need to be carefully considered in the design of the emotional databases. When some of these aspects are included in the design of a corpus, high quality databases from actors can be recorded (Bänziger and Scherer, 2007). On the other hand, even under these limitations, it is not clear whether (and which of) these differences are clearly perceptively distinguished. For example, Schröder *et al.* reported results on perceptive experiments for induced ("real") and simulated amusement (Schröder et al., 1998). Human evaluators were asked to classify the stimuli between real and acted emotions. The results showed an average accuracy of 58%. Although the performance was over chance, the low accuracy suggests that the task was non-trivial. Human raters were not able to accurately distinguish between real and simulated emotions.

It is in this context that we pose the following question: Can specific acting methods be used to mitigate the limitations of recording emotional data from actors? The creative art forms of human acting exemplify the most enriched forms of expressive human communication. These skills have evolved over centuries, and across cultures, providing insight into how humans use their communication instruments to create and induce specific emotional percepts in others, especially in controlled and deliberate ways. The fields of theater to the contemporary cinematic arts have well-established theories and methods of pedagogy and practice of expressive communication although largely descriptive and non-quantitative offering a fertile ground for allowing research on expressive human behavior in systematic ways.

3. Guidelines to record databases from actors

The guidelines presented in this section were learned from our experience in the design, collection and evaluation of the IEMOCAP database. The main goal for this corpus was to record realistic interaction between subjects in which the emotions were naturally induced in a dialog context. We also needed to acquire detailed visual information to capture the nonverbal behavior by using motion capture technology. Under these requirements, the use of actors was the most suitable choice.

Some of the important requirements that need to be carefully considered in the design of the corpus are the actor selection, material selection, acting styles to be used, modalities to be included, and types of audiovisual sensors. The next subsections discuss some of these guidelines.

3.1. Contextualization and social setting

One of the limitations in many of the existing emotional databases is that they contain only isolated sentences or dialogs (Douglas-Cowie et al., 2003). These settings remove the discourse context, which is known to be an important component for emotion (Cauldwell, 2000). As a result, this approach challenges the actors, who have to face this task, which completely differ from the methodologies and techniques that they have been trained. Furthermore, actors are asked to read the material, which is known to differ from spontaneous speech production. Under these settings, it is not surprising that there is a gap between the expressions in such databases and the emotions observed in real scenarios.

Instead of monologues and short sentences, the database should contain natural dialogues, in which the emotions are suitably and naturally elicited. Furthermore, the average duration of the dialogues should be long enough to contextualize the signs and flow of emotions (e.g., one minute (Douglas-Cowie et al., 2003)). The semantic context of the material should be congruent with the intended emotion, to avoid adding extra difficulties to the actors. The social setting is also important; having interaction between two or more actors is hypothesized to generate more authentic emotions (Douglas-Cowie et al., 2003).

One special experimental case is when the same sentences need to be spoken expressing different emotional categories. This case is important when the goal is to compare neutral versus emotional materials in terms of linguistic units (Busso and Narayanan, 2006). In this case, semantically neutral sentences need to be designed such that they can be adequately contextualized according to the intended

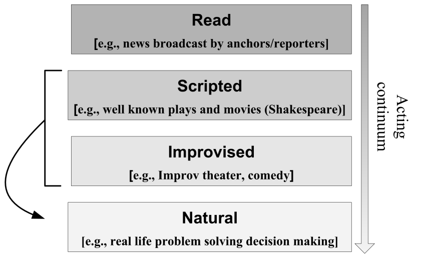

Read
[e.g., news broadcast by anchors/reporters]
Scripted
[e.g., well known plays and movies (Shakespeare)]
Improvised
[e.g., Improv theater, comedy]
Natural
[e.g., real life problem solving decision making]
Acting continuum

Figure 1: Acting continuum – from Fully predetermined to fully undetermined.

emotion. One approach is to record the target sentences embedded in short stories (Martin et al., 2006). For example, for the target sentence "that dress came from Asia", the following contexts could be used: sadness - the subject misses her native land; happiness - the subject receives a gift, anger - the dress was stolen.

3.2. Acting styles

The recording of the database should be as controlled as possible in terms of emotional and linguistic content. The use of specific acting methods and styles can be used to provide a systematic way to control aspects of the expressive communication forms.

Theater theory and practice, over the past several centuries, has been systematized through the work of several scholars resulting in well-known acting systems: notable examples include those of Laban, Delsarte, and Stanislavsky. It is important to investigate creativity at several points along the spectrum between fully pre-specified activity and fully improvised activity. As discussed in Section 4., acting techniques ranging from fully scripted to fully improvised can be used to balance the tradeoff between controllability and naturalness. The two most common genres of theatre are the conventional, scripted approach where dialogue and some instructions to actors are provided a priori, and *improv* in which actors are required to create much of the performance within a set of relaxed constraints (Fig. 1). While the use of scripts provides a way of constraining the semantic and emotional content of the corpus, improvisation gives the actors a considerable amount of freedom in their emotional expression. These two types of acting provide alternatives that the emotional research community should take advantage of when collecting databases (Enos and Hirschberg, 2006).

3.3. Trained actors

Unlike naïve speakers, skilled actors engaged in their role during interpersonal drama may provide a more natural representation of the emotions, avoiding exaggeration or caricature of emotions (Douglas-Cowie et al., 2003). Importantly, as the subjects display facial expressions that are closer to genuine emotions, they may start feeling the emotion, as suggested by Ekman (Ekman, 1993).

Our experiences with actors indicate that rehearsing the material in advance under the supervision of an experienced professional helps to increase the emotional quality of the data. As the actors get familiar with the material (e.g., scripts) and with their colleagues, they will be more confident during the recording.

3.4. Emotional descriptors

One of the most important aspects in the post-processing of the data is defining the emotional description that is conveyed in the data. Defining this ground reference is important since the boundaries between descriptors is usually blurred. The most common techniques to describe the emotional content of a database are discrete (category based) and continuous (primitive based) representation of the emotions. In the discrete emotional representation, categorical labels such as happiness, anger and sadness are used in a time sequence fashion. This approach has been widely used in previous work in describing human emotions. In contrast, the continuous emotional representation is based on primitive attributes. This is an alternative approach to describe the emotional content of an utterance, in which the sentences are described in terms of attributes such as valence, activation (or arousal), and dominance (Cowie and Cornelius, 2003). This approach, which has recently increased popularity in the research community, provides a more general description of the affective states of the subjects in a continuous space. Likewise, it is also useful for analyzing emotion expression variability. Both types of emotional descriptions provide complementary insights about how people display emotions. Adopting either of these schemes will depend on the research questions that will be studied.

In most of the previous emotional corpus collections, the subjects were asked to read a sentence, expressing a given emotion, which is later used as the emotional label. A drawback of this approach is that it is not guaranteed that the recorded utterances reflect the target emotions. Additionally, a given display can elicit different emotional percepts. To avoid these problems, the emotional description should rely on perceptual human evaluation collected from as many evaluators as possible ($\geq$3).

Subjective evaluations are expensive and, therefore, need to be suitably designed in advance (running pilot tests is highly suggested). When categorical description is adopted, a critical aspect is the emotional classes included in the evaluation. On the one hand, if the number of emotions is too extensive, the agreement between evaluators will be low (which is an inherent problem of emotional subjectivity). Ad-hoc solutions such as clustering *similar* emotional categories after the perceptive evaluation should be avoided, since the emotional partition will most probably differ from the one obtained with the new labels. On the other hand, if the list of emotions is limited, the emotional description of the utterances will be poor and likely less accurate.

Another important aspect is the order in which the material will be presented. We suggest presenting the material in order (i.e., not in isolated fashion), so that the evaluators can judge the emotional content based on the sequential development of the dialogs. Likewise, all available modalities should be presented, so the evaluators can use all their senses to assess the emotion. Finally, long and tedious subjective evaluations should be avoided.

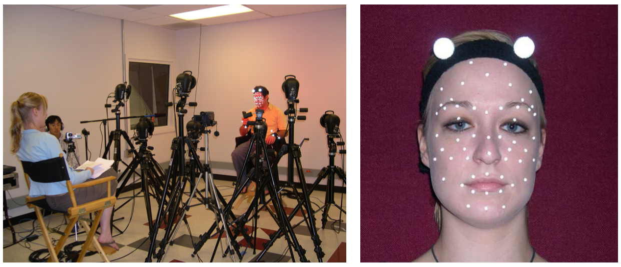


Figure 2: VICON motion capture system with 8 cameras, and an actress showing the markers on the face and headband.

4. A case study: The USC IEMOCAP corpus

As a case study, we recently collected an audiovisual database, we refer to as the *interactive emotional dyadic motion capture database* (IEMOCAP) (Busso et al., 2007a). This is an extensive corpus with over twelve hours of data comprising multimodal information. This corpus was designed, collected and evaluated following the guidelines suggested in Section 3.. This section briefly describes this multimodal corpus.

4.1. Designing the corpus

This database was collected from seven professional actors and three senior students (5 female and 5 male) from the Drama Department at USC. These experienced actors were recorded in dyadic sessions to facilitate a social setting suitable for natural interaction (Sec. 3.1.).

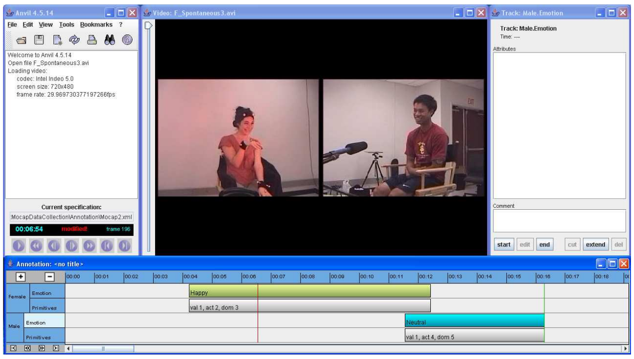
In contrast to providing material that an actor reads under the target emotional condition, the two approaches mentioned in Section 3.2. were selected: the use of plays (scripted sessions), and improvisation-based hypothetical scenarios (spontaneous sessions). The first approach is based on a set of scripts that the subjects were asked to memorize and rehearse. In the second approach, the subjects were asked to improvise based on hypothetical scenarios that were designed to elicit specific emotions (happiness, anger, sadness, frustration and neutral states). These recording settings are familiar to the actors since they were trained to memorize and improvise scripts (Sec. 3.3.).

4.2. Data collection

To capture non-verbal behavior of the subjects, markers were attached to their face, head and hands, which provide detailed information about their facial expression and hand movements. To track those markers, a VICON motion capture system with 8 cameras was used (Fig. 2). Due to equipment constraints, only one of the subjects' movements were captured at a time. Then, the markers were placed on the other subject and the material was recorded again.

4.3. Evaluation

After the data were recorded, the dialogs were manually segmented at the dialog turn level. The emotional categorical labels in this corpus were assigned based on agreements derived from subjective emotional evaluations (3 evaluators per sentence). For the aforementioned purpose, the *annotation of video and spoken language* tool ANVIL (Kipp, 2001) was used (Fig. 3). This tool is particularly useful to jointly annotate verbal and nonverbal behaviors observed from the actors. The evaluators were asked to sequentially assess the turns, after watching the videos. Thus, the acoustic and visual channels, and the previous turns in the dialog were available for the emotional assessment, so that the evaluators could judge the emotional content based on the sequential development of the dialogs (Sec. 3.4.). For the evaluation, the emotional categories surprise, fear, disgust, excited, and other were also included. While categorical description is already evaluated, we are currently assessing the database in terms of the primitive attributes to have a more complete emotional description.



Welcome to Anvil 4.5.14
Open file F_Spontaneous3.avi
Loading video:
codec: Intel Indeo 5.0
screen size: 720x480
frame rate: 29.969730377197266fps

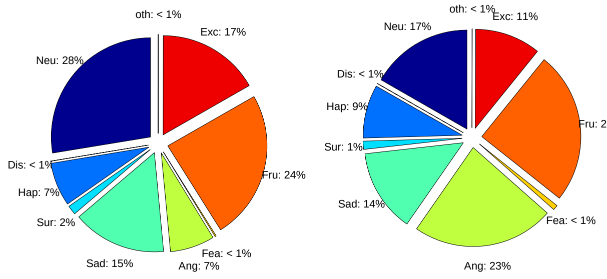
Current specification:
:MocapDataCollection\Annotation\Mocap2.xml
00:06:54 modified! frame 196

Track: Male.Emotion
Time: ---

Figure 3: ANVIL annotation tool used for emotion evaluation. The tool is set for categorical and primitive based evaluation.

Majority voting was used to tag the sentences with the emotional categories. Under this criterion, 74.6% of the sentences were assigned one emotional category (spontaneous sessions: 83.1%; scripted session: 66.9%). For the sentences in which the evaluators reached agreement, the resulting Fleiss kappa statistic was κ= 0.40 (spontaneous sessions: $\kappa = 0.43$; scripted session: $\kappa = 0.36$). These levels of agreement, which are considered as fair/moderate agreement, are expected since people have different perception and interpretation of the emotions. They also show the difficulties in the assignment of emotional labels. Interestingly, the results reveal that for the spontaneous sessions the levels of inter-evaluator agreement are higher than in the scripted sessions. While spontaneous sessions were designed to target specific emotions, portions of the scripted sessions elicited a wider range of emotion categories, increasing the confusion between evaluators.

Figure 4 shows the emotional content of the database. For the target emotions (happiness, anger, sadness, frustration and neutral states), the figure indicates that a balanced emotional content was obtained which validates the proposed design. Notice that for scripted sessions, the emotional labels are less balanced than for spontaneous sessions. From a design viewpoint, these results suggest that improvisation may be preferred if a balanced corpus with higher agreement level is required.



(a) Scripted session

(b) Spontaneous sessions

Figure 4: Distribution of the data for each emotional category.

5. Conclusions

This paper described guidelines with the aim of designing controlled emotional databases from actors that are closer to the emotions observed in real-life scenarios. Based on the limitations of current settings used in the recording of existing corpora, we discussed the importance of contextualization, the use of skilled actors, the use of different acting styles and suitable emotional descriptors.

As a case study, the paper presented the IEMOCAP database. Based on the settings used to elicit the emotions and the achieved results, we consider that the emotional quality of this database is closer to natural than those from prior elicitation settings. The quality of this database suggests that genuine realization of the emotions can be recorded from actors when the settings are carefully designed.

While this methodology was a significant step forward in the use of actors in emotions in research, it did not exploit or control for the nature of the expressive behavior such as through specific acting styles or the nature of improvisation. Further analysis is needed to identify the recording methodologies that will aid emotional recording from actors that resemble real emotions observed in daily human interaction. In fact, acting methods such as the one proposed by Stanislavsky (Carnicke, 1998), in which the actors are encouraged to feel their characters, could be exploited to capture realistic realization of the emotions. These are some of the goals of our future work.

Acknowledgements

This research was supported in part by funds from the NSF, and the Department of the Army. The authors wish to thanks the anonymous reviewers for their thoughtful and insightful comments. Thank also go to the colleagues in the emotion research group for their valuable comments.

6. References

S. Abrilian, L. Devillers, S. Buisine, and J.C.Martin. 2005. EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *11th International Conference on Human-Computer Interaction (HCI 2005)*, pages 195–200, Las Vegas, Nevada, USA, July.

N. Amir, S. Ron, and N. Laor. 2000. Analysis of an emotional speech corpus in Hebrew based on objective criteria. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 29–33, Newcastle, Northern Ireland, UK, September.

T. Bänziger and K.R. Scherer. 2007. Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In A. Paiva, R. Prada, and R.W. Picard, editors, *Affective Computing and Intelligent Interaction (ACII 2007), Lecture Notes in Artificial Intelligence 4738*, pages 476–487. Springer-Verlag Press, Berlin, Germany, September.

A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. 2000. Desperately seeking emotions or: actors, wizards and human beings. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 195–200, Newcastle, Northern Ireland, UK, September.

C. Busso and S.S. Narayanan. 2006. Interplay between linguistic and affective goals in facial expression during emotional utterances. In *7th International Seminar on Speech Production (ISSP 2006)*, pages 549–556, Ubatuba-SP, Brazil, December.

C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, and S.S. Narayanan. 2007a. IEMOCAP: Interactive emotional dyadic motion capture database. *Submitted to Journal of Language Resources and Evaluation.*

C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan. 2007b. Rigid head motion in expressive speech animation: Analysis and synthesis. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3):1075–1086, March.

G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou, and K. Karpouzis. 2006. Modeling naturalistic affective states via facial and vocal expressions recognition. In *Proceedings of the 8th international conference on Multimodal interfaces (ICMI 2006)*, pages 146–154, Banff, Alberta, Canada, November.

S.M. Carnicke. 1998. *Stanislavsky in focus*. Routledge, Taylor & Francis Group, Oxford, UK.

R. Cauldwell. 2000. Where did the anger go? the role of context in interpreting emotion in speech. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 127–131, Newcastle, Northern Ireland, UK, September.

J. Cohn and K. Schmidt. 2004. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2:1–12, March.

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

R. Cowie, E. Douglas-Cowie, and C. Cox. 2005. Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural Networks*, 18(4):371–388, May.

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

E. Douglas-Cowie, L. Devillers, J.C. Martin, R. Cowie, S. Savvidou, S. Abrilian, and C. Cox. 2005. Multimodal databases of everyday emotion: Facing up to complexity. In *9th European Conference on Speech Communication and Technology (Interspeech'2005)*, pages 813–816, Lisbon, Portugal, September.

P. Ekman and E.L. Rosenberg. 1997. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)*. Oxford University Press, New York, NY, USA.

P. Ekman. 1993. Facial expression and emotion. *American Psychologist*, 48(4):384–392, April.

F. Enos and J. Hirschberg. 2006. A framework for eliciting emotional speech: Capitalizing on the actors process. In *First International Workshop on Emotion: Corpora for Research on Emotion and Affect (International conference on Language Resources and Evaluation (LREC 2006))*, pages 6–10, Genoa,Italy, May.

M. Grimm, K. Kroschel, E. Mower, and S. Narayanan. 2007. Primitives-based evaluation and estimation of emotions in speech. *Speech Communication*, 49(10-11):787–800, October-November.

M. Kipp. 2001. ANVIL - a generic annotation tool for multimodal dialogue. In *European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370, Aalborg, Denmark, September.

O. Martin, I. Kotsia, B. Macq, and I. Pitas. 2006. The eNTERFACE'05 audio-visual emotion database. In *Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW 2006)*, Atlanta, GA, USA, April.

K.R. Scherer and G. Ceschi. 1997. Lost luggage: A field study of emotionantecedent appraisal. *Motivation and Emotion*, 21(3):211–235, September.

F. Schiel, S. Steininger, and U. Türk. 2002. The SmartKom multimodal corpus at BAS. In *Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain, May.

M. Schröder, V. Aubergé, and M.A. Cathiard. 1998. Can we hear smiles? In *5th International Conference on Spoken Language Processing(ICSLP 1998)*, pages 559–562, Sydney,Australia, November-December.

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

A. Zara, V. Maffiolo, J.C. Martin, and L. Devillers. 2007. Collection and annotation of a corpus of human-human multimodal interactions: Emotion and others anthropomorphic characteristics. In A. Paiva, R. Prada, and R.W. Picard, editors, *Affective Computing and Intelligent Interaction (ACII 2007), Lecture Notes in Artificial Intelligence 4738*, pages 464–475. Springer-Verlag Press, Berlin, Germany, September.