

\*\*\*\*\*

Third Call for Papers

**International Workshop on EMOTION:  
CORPORA FOR RESEARCH ON EMOTION AND AFFECT**

23 May 2006  
Half day workshop: afternoon session

In Association with  
5th INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION  
LREC2006 <http://www.lrec-conf.org/lrec2006/>

Main Conference  
24-25-26 May 2006

Magazzini del Cotone Conference Center  
Genoa - Italy

\*\*\*\*\*

-----  
Summary of the call for participation  
-----

Papers are invited in the area of corpora for research on emotion and affect. They may cover one or more of the following topics. What kind of theory of emotion is needed to guide the area? What are appropriate sources? Which modalities should be considered, in which combinations? What are the realistic constraints on recording quality? How can the emotional content of episodes be described within a corpus? Which emotion-related features should a corpus describe, and how? How should access to corpora be provided? What level of standardisation is appropriate? How can quality be assessed? Ethical issues in database development and access. The organisers of this workshop proposal are members of the Humaine NoE, and have a main role in its databases WP. They will be able to reach researchers in databases of emotions both through the resulting contacts and through the HUMAINE portal.

-----  
MOTIVATIONS  
-----

This decade has seen an upsurge of interest in systems that register emotion (in a broad sense) and react appropriately to it. Emotion corpora are fundamental both to developing sound conceptual analyses and to training these 'emotion-oriented systems' at all levels - to recognise user emotion, to express appropriate emotions, to anticipate how a user in one state might respond to a possible kind of reaction from the machine, etc. Corpora have only begun to grow with the area, and much work is needed before they provide a sound foundation.

The HUMAINE network of excellence (<http://emotion-research.net/>) has brought together several groups working on the development of databases, and the workshop aims to broaden the interaction that has developed in that context.

Papers are expected to address some of the following areas of concern.

Many models of emotion are common enough to affect the way teams go about collecting and describing emotion-related data. Some which are familiar and intuitively appealing are known to be problematic, either because they are theoretically dated or because they do not transfer to practical contexts. To evaluate the resources that are already available, and to construct valid new corpora, research teams need some sense of the models that are relevant to the area.

What are appropriate sources?

In the area of emotion, some of the hardest problems involve acquiring basic data. Four main types of source are commonly used. Their potential contributions and limitations need to be understood.

Acted:

Many widely used emotion databases consist of acted representations of emotion (which may or may not be generated by actors). The method is extremely convenient, but it is known that systems trained on acted material may not transfer to natural emotion. It has to be established what kind of acted material is useful for what purposes.

Application-driven:

A growing range of databases are derived from specific applications (eg call centres). These are ideal for some purposes, but access is often restricted for commercial reasons, and it is highly desirable to have more generic material that could underpin work on a wide range of applications.

General naturalistic:

Data that is representative of everyday life is an attractive ideal, but very difficult to collect. Making special-purpose recordings of everyday life is a massive task, with the risk that recording changes behaviour. Several teams have used material from broadcasts, radio & TV (talk shows, current affairs). That raises issues of access, signal quality, and genuineness.

Induction:

A natural ideal is to induce emotion of appropriate kinds under appropriate circumstances. Satisfying induction is an elusive ideal, but new techniques are gradually emerging.

Which modalities should be considered, in which combinations?

Emotion is reflected in multiple channels - linguistic content, paralinguistic expression, facial expression, eye movement, gesture, gross body movement, manner of action, visceral changes (heart rate, etc), brain states (eeg activity, etc). The obvious ideal is to cover all simultaneously, but that is impractical - and it is not clear how often all the channels are actually active. The community needs to clarify the relative usefulness of the channels, and of strategies for sampling combinations.

What are the realistic constraints on recording quality?

Naturalism tends to be at odds with ease of signal processing. Understanding of the relevant tradeoffs needs to be reached. That includes awareness of different applications (high quality may not be crucial for defining the expressive behaviours a virtual agent should show) and of timescale for solving particular signal processing issues (eg recovering features from images of heads in arbitrary poses).

How can the emotional content of episodes be described within a corpus?

Several broad approaches exist to transcribing the emotional content of an excerpt - using everyday emotion words; using dimensional descriptions rooted in psychological theory (intensity, evaluation, activation, power); using concepts from appraisal theory (perceived goal-conduciveness of a development, potential for coping, etc). These are being developed in specific ways driven by goals such as elegance, inter-rater reliability, faithfulness to the subtlety of everyday emotion, relevance to agent decisions, etc. There seems to be a real prospect of achieving an agreed synthesis of the main schemes.

Which emotion-related features should a corpus describe, and how?

Corresponding to each emotion-related channel is one or more sets of signs relevant to conveying emotion. For instance, paralinguistic signs exist at the level of basic contours - F0, intensity, formant-related properties, and so on; at the level of linguistic features of prosody (such as 'tones and breaks' in TOBI); and at more global levels (tune shapes, repetitions, etc). Even for speech, inventories of relevant signs need to be developed, and for channels such as idle body movements, few descriptive systems have been proposed. Few teams have the expertise to annotate many types of sign competently, and so it is important to establish ways of allowing teams that do have the expertise to make their annotations available as part of a database. Mainly for lower level features, automatic transcription methods exist, and their role needs to be clarified. In particular, tests of their reliability are needed, and that depends on data that can serve as a reference.

How should access to corpora be provided?

Practically, it is clearly important to find ways of establishing a sustainable and easily expandable multi-modal database for any sorts of emotion-related data; to develop tools for easily importing and exporting data; to develop analysis tools and

application programmers interfaces to work on the stored data and meta-data; and to provide ready access to existing data from previous projects. Approaches to those goals need to be defined.

What level of standardisation is appropriate?

Standardisation is clearly desirable in the long term, but with so many basic issues unresolved, it is not clear where real consensus can be achieved and where it is better to encourage competition among different options.

How can quality be assessed?

It is clear that some existing corpora should not be used for serious research. The problem is to develop quality assurance procedures that can direct potential users toward those which can.

Ethical issues in database development and access

Corpora that show people behaving emotionally are very likely to raise ethical issues - not simply about signed release forms, but about the impact of appearing in a public forum talking (for instance) about topics that distress or excite them. Adequate guidelines need to be developed.

All of the questions above will be studied during the workshop and will contribute to the study of practical, methodological and technical issues central to developing emotional corpora (such as the methodologies to be used for emotional database creation, the coding schemes to be defined, the technical settings to be used for the collection, the selection of appropriate coders).

-----  
IMPORTANT DATES  
-----

|   |             |
|---|-------------|
| 1st call for paper                          | 16 December |
| 2nd call for paper                          | 8 January   |
| Deadline for 1000 words abstract submission | 24 February |
| Notification of acceptance                  | 10 March    |
| Final version of accepted paper             | 10 April    |
| Workshop half-day                           | 23 May      |

-----  
SUBMISSIONS  
-----

The workshop will consist of paper and poster presentations.

Final submissions should be 4 pages long, must be in English, and follow the submission guidelines at <http://www.chi2006.org/docs/chi2006pubsformat.doc>

The preferred format is MS word.

The .doc file should be submitted via email to [lrec-emotion@limsi.fr](mailto:lrec-emotion@limsi.fr)

-----

As soon as possible, authors are encouraged to send to [lrec-emotion@limsi.fr](mailto:lrec-emotion@limsi.fr) a brief email indicating their intention to participate, including their contact information and the topic they intend to address in their submissions.

Proceedings of the workshop will be printed by the LREC Local Organising Committee.

Submitted papers will be blind reviewed.

-----  
TIME SCHEDULE AND REGISTRATION FEE  
-----

The workshop will consist of an afternoon session. There will be time for collective discussions. For this half-day Workshop, the registration fee will be specified on <http://www.lrec-conf.org/lrec2006/>

-----  
THE ORGANISING COMMITTEE  
-----

Laurence Devillers / Jean-Claude Martin  
Spoken Language Processing group/ Architectures and Models for Interaction,  
LIMSI-CNRS,  
BP 133, 91403 Orsay Cedex, France  
(+33) 1 69 85 80 62 / (+33) 1 69 85 81 04 (phone)  
(+33) 1 69 85 80 88 / (+33) 1 69 85 80 88 (fax)  
[devil@limsi.fr](mailto:devil@limsi.fr) / [martin@limsi.fr](mailto:martin@limsi.fr)  
<http://www.limsi.fr/Individu/devil/>  
<http://www.limsi.fr/Individu/martin/>

Roddy Cowie / School of Psychology  
Ellen Douglas-Cowie / Dean of Arts, Humanities and Social Sciences  
Queen's University, Belfast BT7 1NN, UK  
+44 2890 974354 / +44 2890 975348 (phone)  
+44 2890 664144 / +44 2890 \*\*\*\*\* (fax)  
<http://www.psych.qub.ac.uk/staff/teaching/cowie/index.aspx>  
<http://www.qub.ac.uk/en/staff/douglas-cowie/>  
[r.cowie@qub.ac.uk](mailto:r.cowie@qub.ac.uk) / [e.douglas-Cowie@qub.ac.uk](mailto:e.douglas-Cowie@qub.ac.uk)

Anton Batliner - Lehrstuhl fuer Mustererkennung (Informatik 5)  
Universitaet Erlangen-Nuernberg - Martensstrasse 3  
91058 Erlangen - F.R. of Germany  
Tel.: +49 9131 85 27823 - Fax.: +49 9131 303811  
[batliner@informatik.uni-erlangen.de](mailto:batliner@informatik.uni-erlangen.de)  
<http://www5.informatik.uni-erlangen.de/Personen/batliner/>

-----

## PROGRAM COMMITTEE

-----  
Roddy Cowie, QUB, UK  
Ellen Douglas-Cowie, QUB, UK  
Laurence Devillers, LIMSI-CNRS, FR  
Jean-Claude Martin, LIMSI-CNRS, FR  
Anton Batliner, Univ. Erlangen, D  
Nick Campbell, ATR, J  
Elisabeth André, Univ. Augsburg, D  
Stephanos Kollias, ICCS, G  
Marc Schröder, DFKI Saarbrücken, D  
Catherine Pelachaud, Univ. Paris VIII, FR  
Elisabeth Shriberg, SRI and ICSI, USA  
Izhak Shafran, Univ. Johns Hopkins, CSLP, USA  
Ioana Vasilescu, ENST, FR  
Fiorella de Rosis, Univ. Bari, I  
Isabella Poggi, Univ. Roma Tre, I  
Nadia Bianchi-Berthouze, Univ. Aizu, J  
Susanne Kaiser, UNIGE, S  
Valérie Maffiolo, FranceTelecom, FR  
Véronique Aubergé, CNRS-STIC, FR  
Shrikanth Narayanan, USC Viterbi School of Engineering, USA  
John Hansen, Univ. of Texas at Dallas, USA  
Christine Lisetti, EURECOM, FR  
-----