

HUMAINE Workshop « Theories and Models of Emotion », June 17-19, 2004

Work Group 3 – Synthesis Chair: Tanja Bänziger (tanja.banziger@pse.unige.ch)

This document follows the WP3 workshop held in Geneva on the 17-19th of June 2004. One of the aims of this workshop was to initiate several working groups. The directions emerging from the discussions that took place during the first meeting of **working group 3** are presented in this document.

Working group 3 will be focusing on an issue taken from the larger field of individual differences in emotional communication, namely: the **assessment of individual differences in the decoding (recognition) of emotional expressions**. The specific HUMAINE approach to this field will take the form of a set of propositions for the **assessment of the performance of automatic emotion recognition systems**.

Psychologists have, since long, been concerned with the issue of assessing individual sensitivity to nonverbal communication (e.g. see the work of Rosenthal et al., 1979). Furthermore, the recent surge of interest in "Emotional Intelligence" has raised a renewed interest in the assessment of sensitivity to *emotional* communication. Nevertheless, there are currently no widely accepted tests at hand for assessing sensitivity to emotional expressions. Worse, there appears to be a general lack of concern regarding the specific features that such tests ought to include. It is, for instance, not infrequent to find tests in which emotional sensitivity is assessed by asking a participant to select an emotion label among 4-5 given alternatives for relatively few still pictures displaying somewhat ambiguous expressions (e.g. Mayer, Salovey, & Caruso's EI test: MSCEIT, Mayer et al., 2003 or Nowicki and Duke's test: DANVA, Nowicki & Duke, 1994). In our opinion, the face and content validity of such tests are highly questionable. One important goal for achieving rapid progress in this field would therefore be to define the standards that tests of emotional sensitivity ought to meet (especially with respect to face and content validity).

The relation of this question to more specific HUMAINE concerns lies in the issue of the assessment of automatic emotion recognition systems. The definition of the performance of an automatic recognition system is, potentially, faced with questions similar to those raised by the assessment of the performance of a human participant in the same matter.

A large number of aspects are involved in this issue. A list of aspects that have been raised during the first reunion of working group 3 is presented in the following outline/summary:

First the issue of the **emotional categories and/or dimensions** that have to be identified/recognized was evoked. This issue is central to the problem of the assessment of emotional sensitivity and, simultaneously, submitted to contingencies. The performance of human participants (in a test of emotional sensitivity) as well as the performance of an automatic recognition system will be totally dependent on the number and the degree of differentiation of the emotion categories/dimensions that have to be discriminated. But it does

not seem sensible to establish a list of states or dimensions that have to be discriminated independently of **contextual factors** (such as interaction type, social situation). This point actually raised another central issue, namely the question of the opportunity of assessing an **overall performance** in emotional sensitivity. It is not unlikely that some human participants could be very competent in identifying emotional expressions in a given context (an interaction type or a social situation involving a certain range of potential emotional reactions), yet very poor in another context (another interaction type or social situation involving other emotional reactions). One pending issue therefore would be to **define the range of contexts – as well as the emotion categories** they entail – in which a person or an automatic recognition system would have to perform (and in which their performance would be assessed). In any case, the context and the emotional categories/dimensions involved would have to be clearly specified in order to interpret the performance reached by a test participant or an automatic recognition system.

A second issue raised was related to the expressive modalities or **communication channels**. In ordinary communication situations, emotions are likely to be communicated audio-visually, but in some contexts they might be communicated via only one channel (e.g. in noisy environments or on the telephone). The decoding performance of a person or an automatic recognition system (provided that it uses more than one channel to recognize the emotions!) might be channel specific. Moreover, evidence to date suggests that some emotions could be preferentially communicated via the audio channel, whereas other emotions could be preferentially communicated via the visual channel. A better recognition performance in one or another channel might therefore result in a better performance for the emotions preferentially communicated in the respective channel. It seems therefore essential **to assess recognition performance for independent as well as for combined channels**.

A third issue concerned the **definition of communication accuracy**. Accuracy is most frequently defined as the **number (or proportion) of expressions correctly discriminated**. As such it is directly dependent on the number of provided response alternatives. Other, less obvious, biases that ought to be controlled for include: unequal representation of the expressed categories combined with answer tendencies (e.g. knowing/guessing that 'anger' is the most frequently expressed emotion and therefore selecting this alternative more frequently than others) and "similarity" of the expressions that have to be discriminated (e.g. does the task imply to differentiate expressions of joy from expressions of anger, or does it imply to differentiate cold anger from hot anger?). The possibility to use **confusion matrices** for the definition of a "degree of accuracy" has been considered. In this case, some answers would be defined as more "wrong" than others, relatively to the distribution of answers of a comparison group.

This leads to the fourth issue discussed during the first reunion of working group 3: accuracy can be defined with reference to various criteria. The most common criteria used in tests appears to be the answers provided by a **comparison group** (or reference group). If groups of people with rather different response patterns can be identified then various comparison groups ought to be used to establish the performance scores. Features of the decoders that might influence the overall response pattern could include: gender, culture (nationality, ethnicity), social class, age and personality.

A central objective of working group 3 will be to gather **data on emotional sensitivity from a variety of samples** (in different countries) in order to assess the extent of variability in emotional sensitivity and the need to identify various comparison groups. Previously mentioned issues such as the consideration of different communication channels and the

examination of confusion matrices will be included in this objective.

The fifth issue raised was related to the **individual characteristics of the encoders**. Individual differences (such as gender or age) are likely to result in systematic variations in emotional expressions. Consequently, it is to be expected that knowledge about some characteristics of the encoder (such as his gender or his age) is needed to identify the emotions he/she expresses. However, it is, so far, unclear to what extent this issue can be included in the objectives of working group 3.

At the conclusion of the first reunion of working group 3, it was decided that the next step of the activities of working group 3 would be postponed to the HUMAINE Summer School on "Data and Databanks" to be held in September 2004. During this Summer School decisions ought to be taken regarding the data (emotional expressions) at hand for the construction of a test of emotional sensitivity.

References:

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication. The PONS test*. Baltimore: The Johns Hopkins University Press.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: the Diagnostic Analysis of NonVerbal Accuracy. *Journal of nonverbal behavior, 18*, 9-35.