Prototypes of Multimedial Stress Assessment subject to Laboratory Experiments

R.J. van Vark, L.J.M. Rothkrantz, E.J.H. Kerckhoffs

Delft University of Technology
Faculty of Technical Mathematics and Informatics
Section of Knowledge Based Systems
Julianalaan 132
2628 BL Delft
The Netherlands

A research group of Delft University of Technology, Faculty of Technical Mathematics and Computer Science, has been working on automated stress assessment since 1992. Goal of the project is the development of a system for automated stress assessment. The system will use multimedia techniques and is based on the analysis of facial expressions, voice signals and physiological sensor data. These media are analysed in parallel and based on the outcome of these parallel analyses an overall stress score is obtained. Analysis techniques are mostly based on artificial intelligence techniques, like knowledge based systems and artificial neural networks. A detailed description of the system, underlying model and theory are presented in an accompanying separate paper. Several prototypes have been developed for each medium. The prototypes have been tested on a group of 108 computer science students.
In the paper, the developed prototypes will be presented. Secondly, the experimental setting of the test sessions will be described, followed by results of these test sessions. Finally, some conclusions will be presented and future developments concerning automated stress assessment will be covered.

Prototypes
The analysis of multimedial data takes place in several steps. These steps are the same for each medium and as these steps play an important role in the prototypes, they will be described first. The steps are implemented as several processing layers. The first layer is the *filtering & reduction layer*. In this layer several transformations are applied to the raw incoming measurements: noise filtering, outlier correction to eliminate recording errors and data reduction by feature abstraction. The second layer is the *preprocessor layer*. In this layer, measurements are stored for later use. Additionally, simple transformations are applied, like averaging and smoothing. The third layer, the *processor layer*, is very important as it derives hypotheses based on preprocessed features using an intermediate representation. The fourth layer is the *application layer*. This layer consists of several applications of which any subset can be processed simultaneously. The final layer is the *output layer*, which produces two types of feedback. Temporary feedback is used to produce feedback during the session either for the trainer or trainee. Permanent feedback interprets complete sessions and can be used to check progress during successive sessions.

The input of the physiological prototype consists of a number of physiological signals: electrocardiogram, blood pressure, respiration and galvanic skin response. In the filtering & reduction layer the following features are abstracted: heart rate, heart period, blood pressure, respiration rate, depth of respiration and galvanic skin response. The preprocessor layer averages these signals over a short period of time as temporary, cyclic perturbations can be found in physiological signals. The processor layer derives the stress hypotheses using nervous activity as the intermediate representation. Several types of output are built in the prototype. First, physiological behaviour is visualised as two-dimensional paths, in contrast to traditional one-dimensional projection. Points in this 2D-space can be labelled by different types of nervous activity and stress

levels. Another type of feedback is the behavioural compass. This indicates the direction in which behaviour is changing in the near future. Based on this prediction, evasive action can be taken if necessary. Finally, a stress meter is integrated in the prototype. Several areas in the two-dimensional space correspond to certain stress levels and these are visualised as a kind of "econometer".
The sound prototype samples the human voice signal at a rate of 44 kHz and a resolution of 16 bits. From this sampled signal a number of features are extracted: frequency spectrum, frequency spectrogram, fundamental frequency, jitter, zero crossings, energy and average amplitude. Noise filtering is very important as background noise can completely disguise the original voice signal. The first layer also detects if the observant was speaking or not speaking at the time of recording. The abstracted features are averaged over short periods of time. Hypotheses are derived using emotions as intermediate representation. Besides visualisation, only one type of feedback is implemented: a voice stressometer. Using a formula based on a linear combination of vocal features, stress levels can be deducted.
Two prototypes have been developed for facial expressions. The first prototype uses manual extracted features which can be interpreted by neural networks. The neural networks classify these features as different types of emotion. Stress corresponds to certain levels of emotions. Classification of a well known standard set of facial expressions resulted in a correct classification in over 75 percent of the cases. The second prototype concerns the automatic generation of facial expressions. By examining these expressions, more insight is gained in the correspondence between facial expressions and stress.

Experimental setting

The experiments used to test the physiological and voice prototypes were carried out in a laboratory using one computer for stimulus and two computers, one for each medium, for recording and feedback. Voice signals were only recorded during some parts of each session, while physiological signals were recorded during entire sessions. First, procedures during recording and the behavioural tasks were explained to each observant and it was tried to ease the observant as much as possible. After this relaxation period, some basal recordings of rest levels were made. Following these basal recordings which are important to compensate for personal differences, some behavioural tasks had to be performed. A computer game was used at different levels during the first and second task. Each minute the observant had to utter the sentence: "They bought a new car". This sentence contains many voiced signals which are important in voice analysis. The third task was another computer game during which the observant had to perform mental arithmetic. The fourth task was the well-known Stroop Test. This task is known to be extremely stressful. Each task lasted for five minutes and in between tasks was a rest period of two minutes.

Test results

In general, physiological signals indicated higher stress levels during the fourth task than during the first two tasks. For the third task, two groups can be distinguished: one group of people who enjoy the game and one group of people who dislike racing games or are not good at them. The second group did not react significantly, while the first group showed strong reactions. Voice analysis subscribes to the stressful nature of the Stroop Test as higher stress levels were also found using the voice stressometer.
The two-dimensional projection of the physiological prototype was found to be clearer and easier to interpret than the classical one-dimensional presentation. It was easy to classify the observed behaviour and therefore stress without observing the students. It was also possible to predict near future behaviour using the behavioural compass. The voice stressometer performed reasonably well, although the correspondence between vocal features and stress was found not to be completely linear, which presumption was used to implement the voice stressometer.