Toward Automatic Recognition of Children's Affective State Using Physiological Parameters and Fuzzy Model of Emotions

Ovidiu-Andrei SCHIPOR, Ştefan-Gheorghe PENTIUC, Maria-Doina SCHIPOR
Stefan cel Mare University of Suceava, 720229, Romania
schipor@eed.usv.ro

***Abstract*—Affective computing – the ability of a system to recognize, understand and simulate human emotional intelligence – is one of the most dynamic fields of HCI – Human Computer Interaction. These characteristics find their applicability in those areas where it is necessary to extend traditional cognitive communication with emotional features. That is why, Computer Based Speech Therapy Systems (CBST), and especially those involving children with speech disorders, require this qualitative shift. So in this paper we propose an original emotional framework recognition as an extension for our previous developed system – Logomon. A fuzzy model is used in order to interpret the values of specific physiological parameters and to obtain the emotional state of the subject. Moreover, an experiment that indicates the emotion pattern (average fuzzy sets) for each therapeutic sequence is also presented. The obtained results encourage us to continue working on automatic emotion recognition and provide important clues regarding the future development of our CBST.**

***Index Terms*— assisted speech therapy, emotion recognition, fuzzy model, physiological parameters.**

I. Introduction

Emotional intelligence encompasses the ability to assess, understand, and control the affective states of oneself and others. From an informatics system point of view, the most valuable "emotional requirement" is the skill to evaluate user's emotion and to react accordingly [1].

Speech therapy of preschoolers and young schoolchildren can be considered half-blind if it does not take into account the affective state of the subjects. In the next decade, the difference between "good" and "inappropriate" CBST (i.e. Computer Based Speech Therapy System), will be made by the high quality emotional skills [2].

Since 2006 our team has focused on designing and implementation of Logomon, the first CBST for Romanian language. Although it was originally designed for dyslalia therapy (mispronunciation of sounds), this system has been also successfully used for the treatment of dysarthria, dysphonia, and aphasia. The development of this system has been achieved in four steps:

Implementation of a classical CBST architecture that consists of three modules: *Monitor Program* (centralized management of information about subjects, audio records, exercises, and therapy timetable), *3D Articulator Model* (real-time presentation of the correct positioning of exterior pronunciation organs such as lips, teeth, and language for each affected phoneme), and *Exercises Player* (over 35 interactive exercises types); The second step was to put up a *Personal Portable Device* and, in consequence, to continue speech therapy in familial environment; This gadget is an extension of Logomon system and act as a a virtual interface between SLT (i.e. Speech and Language Therapist) and child [3]; The third step consisted in integration of all information about children using fuzzy approach [4] [5]; This approach helped us to extract new knowledge regarding therapeutic process such as frequency, content, and length of treatment sessions.

As a consequence of SLT's and parents' recommendations [6] and in accord with our experience through more than three years of system utilization, the fourth stage has to be concentrated on endowing the Logomon software with emotion recognition skills. Speech quality depends of the subject emotional condition. Some paraverbal parameters such as speed, intensity, timbre and fluctuation of the speech provide evidence concerning the subject's affective state. On the other hand an emotional extended CBST that detects emotions can take several actions in order to maintain positive affective states and avoid unwanted emotions.

In this article we study the utilization of physiological parameters (e.g. respiration and heart rate, skin temperature and conductance) as input data in assessment of affective state. A fuzzy emotion model is used in order to increase the representation accuracy, especially in the case of ambiguous situations. In addition, we present an experiment whose purpose is to identify the specific emotional patterns (probabilistic coefficients matrix) for each therapy sequence. Since this experiment involves human experts – psychologists and SLTs – the obtained results can be considered a reference point (a benchmark) used to measure the performance of automatic emotion recognition.

II. Literature Review

Automatic recognition of affective state has become an increasingly dynamic field with new and emerging applications. As a consequence, there is a variety of researches concerning this topic.

Since emotion is a complex with both internal (biochemical) and external (behavioral) implications [7], there are several channels (modalities) that can be used for acquire valuable information about affective state: facial

This paper was supported by the project " Progress and development through post-doctoral research and innovation in engineering and applied sciences– PRiDE - Contract no. POSDRU/89/1.5/S/57083", project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

Digital Object Identifier 10.4316/AECE.2012.02008

expression, body movement and gestures, speech and paralanguage parameters, physiological signals [8] [9]. Different acquisition techniques, different strategies for features extraction and classification and different advantages and drawbacks – all these aspects indicate what modality or what combination is expected to provide best results for a specific area of use [10] [11] [12].

A. Emotions and Physiological Parameters

The reason why in this paper we have focused on physiological changes is related with the specificities of assisted speech therapy and with age of the subjects. First of all, the quality and quantity of speech uttered by the children make difficult to obtain constant information about affective state. Secondly, since the subjects interact with the program from sitting position, it is difficult to extract affective clues form gestures.

Of course, using physiological channel leads to some disadvantages too such as difficulties in interpretation and correlation of data taken from multiple sources (sensors). The lack of comfort related with “to be connected to sensors” is another drawback. However, using bio-signals provides considerable advantages. The continuous monitoring of subjects and the avoiding of “social masking” are two priceless benefits [9]. Moreover, recent sensor development enables wireless capture of bio-signals, in an unobtrusive way.

The variation in value of each bio-signal can be associated with a variation in affective state [13]. The skin conductance (SC) can be determined by galvanic skin response (GSR) and it is directly proportional with arousal level of a person. Also, negatively charged emotions are correlated with a higher muscle activity which can be established with electromyography (EMG). Higher heart rate variability (HRV) is associated with attention control while a lower HRV has been associated with depression [14]. Blood pressure in the extremities (Blood Volume Pulse - BVP) indicates changes in sympathetic nervous system arousal [15].

In a study involving seven emotional states induced by a computer game, a classification accuracy ranging from 48% (boredom) to 68% (frustration) was obtained [15]. Higher correct classification ratios were archived using electrocardiogram, skin temperature variation, electrodermal activity as input data, short-segment analysis as feature extraction method, and a SVM (i.e. Support Vector Machine) as a pattern classifier: up to 61.8% taking into consideration four emotions and up to 78.4% in the case of only three emotions [16]. The features that are to be extracted for each raw bio-signal come from both time and frequency domain and some of them are particular to specific channels [17] [18].

B. Models of Emotions

Many theorists have tried to define a list of basic emotions and, consequently, a wide range of research could be referred. In their article [19], Ortony and Turner have collated 14 different models. The number of basic emotions considered in these models ranges from 2 to 11 and the total number of distinct emotions is 36. The most invoked six affective states are: fear – included in 9 models, anger – 7, disgust – 6, joy – 5, sadness – 5, and surprise – 5. In addition to these linear models, different hierarchical models have been proposed. For example, Parrot [20] offers a model on three levels: primary (6), secondary (25) and tertiary emotions (over 130).

The emotional state evaluation refers to: selecting a specific emotion from a predefined list (*labeling approaches*), finding a point in plane or space (*dimensional approaches* [21] [22]) or indicating the strength of each basic emotion (*weighted labeling approaches* [23]). The second model assumes that each affective state has a specific location defined by two scales (valence – pleasant/unpleasant and arousal – calm/excited) or by three scales (valence, arousal, and stance – close/open). For example, sad has a negative valence, a low arousal level and, in addition, reflects a closed stance.

The other two models use predefined emotions but, while in the first it is possible to select only one emotion, in the third more emotions are accepted, each of them being weighted with a specific coefficient. This difference become very important in ambiguous situations, when affective state could be described as a combination of basic emotions, each of them with its own intensity.

As far as we know, there is no record of a research endeavor toward establishing relations between affective states and different sequences of children assisted speech therapy and that is why we consider our research being a step forward. The results of this research offer answers to the following questions: What type of emotion model is appropriate for affective state representation? In which therapeutic sequences should automatic recognition system be activated? What is the probability of a certain emotion state to occur in a specific sequence of therapy?

The first question defines the mathematical model that supports the design and implementation of the automatic recognition framework. The second one defines the moments in which our framework should be turned on. The results corresponding to the third question are used for improving the recognition rate and for solving ambiguous situations. All these results can be applied only if they meet consistency condition. In other words, the evaluation of emotions by the different SLTs should lead to similar results (the system could automatically identify an emotion only when human experts can identify that emotion in a consistent manner).

III. METHODS

A. Data Acquisition

In order to acquire physiological signals, we have used Physiology Suite™ – a Thought Technology Ltd. product. This platform has been developed around a multi-modality encoder with five channels, named ProComp5 Infiniti™.

The encoder provides five channels for sensors connection. First two input lines offer a high signal fidelity (2038 samples per second) for very dynamic signals such as BVP (Blood Volume Pulse), while the remaining three input lines are suitable for slower signals such as RSP - Respiration, SC - Skin Conductance, and ST - Skin Temperature (Figure 1).

There are many studies that reveal the most representative features that are to be extracted for each raw bio-signal, in order to optimize emotion classification [24] [25]: BVP – sub-band spectral powers, spectral entropy, pulse rate variability; SC – mean, standard deviation, values of first and second differences, average number of transient changes; ST – mean value, standard deviation, minimum value, maximum value, max / min ratio; RESP – fundamental frequency, power mean values of sub-bands, sub-band spectral entropy).

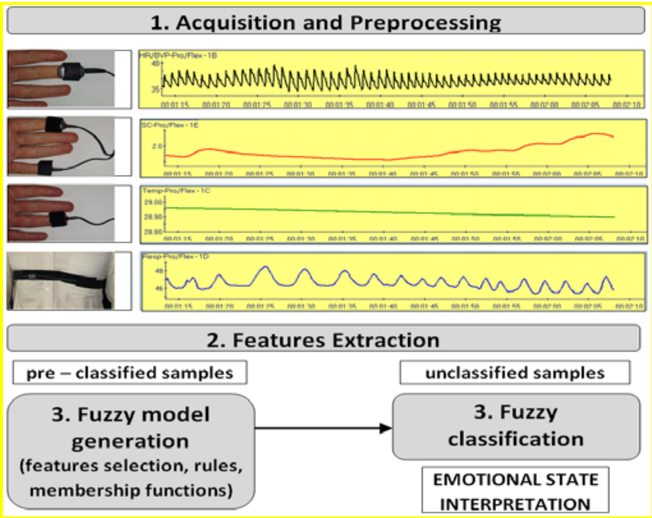

Fig. 1. The top level diagram of Affect Recognition Framework

It is impossible to know a priori which physiological patterns could offer a good discrimination between emotional patterns. Some of the signals' characteristics may have no relevance for our affect recognition module and it is reasonable to eliminate them. Otherwise, these features will act as a garbage input data and they will worsen the recognition rate [26]. That is why the number of initially extracted features is usually greater than the number of used features.

B. Fuzzy Model of Emotions

In this paper we use a flexible representation of emotions based on fuzzy numbers. We start defining E as a finite set of n basic emotions and FM as an infinite set of fuzzy membership functions.

$$E = \{e_1, e_2, \ldots, e_n\}, FM = \{\mu_j : E \to [0,1], j = 1, 2, \ \} \quad (1)$$

Then, we define ES as the set of all emotional states (es_j) that can be represented in this model:

$$ES = \{es_j, j = 1, 2, \ldots \ \}, es_j = \{(e_i, \mu_j(e_i)) \mid e_i \in E \ \} \quad (2)$$

This n-dimensional fuzzy become a representation for an affective state because each element is a 2-uple: the emotion (e_i) and the specific value of fuzzy membership function ($\mu_j(e_i)$).

The cardinal of ES set is the number of affective states that can be described by this model (ε is the precision of representation).

$$|ES| = (1/\varepsilon)^{|E|} \quad (3)$$

The hypercube whose axes are associated with basic emotional states (e_i) become the "recognition space" of this model. A point in this hypercube (n-dimensional vector) could be seen as a graphical representation of a complex affective state.

$$es_j = (\mu_j(e_1), \mu_j(e_2), \ldots, \mu_j(e_n)) \quad (4)$$

According with our previous researches [2], there are five basic emotions related with speech assisted therapy: *happiness, contentment, neutral, tenseness, and nervousness*. In this context, the point (0.3, 0.9, 0.2, 0.1, 0.0) could be associate with a positive state whose fuzzy set is: *{(happiness, 0.2), (contentment, 0.9), (neutral, 0.3), (tenseness, 0.1), (nervousness, 0.0)}*.

IV. The Experiment

In order to identify the opportunity of utilization of above presented fuzzy model in assisted speech therapy, we set up an experiment involving children and psychologists from Regional Speech and language Therapy Center – Suceava, Romania. First, we try to find out if this model can be easily and efficiently manipulated by human experts. Secondly, we want to obtain an average distribution of the five selected basic emotions (i.e. happiness, contentment, neutral, tenseness, nervousness) during three therapeutic stages (i.e. initial speech and language evaluation, exercises for development of phonematic hearing, and the pronunciation of affected sounds using 3D model).

A. Methodology

The subjects (N = 41) were children (from five to nine years old) randomly selected from Regional Speech and Language Therapy Center of Suceava – RSLTCS, Romania. The obtained distribution and magnitude of speech disorders was representative for children from RSLTCS.

Procedure: Two human experts observed the children's emotional state during speech therapy and the individual results were written on an observation sheet. Base on recorded scores, specific fuzzy sets were computed. In addition, physiological parameters were monitored and recorded.

The observation was performed in three distinct sequences of speech therapy (five minutes each): 1. Initial speech and language evaluation; 2. Exercises for development of phonematic hearing; 3. The pronunciation of affected sound using 3D model.

Thus, for each child and each therapy step, five scores (in a Likert type scale ranging from *0 – "absence"* to *5 – "maxim intensity"*) were obtained.

In order to identify emotional patterns (i.e. fuzzy sets es_{seq1}, es_{seq2}, es_{seq3}) associated with each therapy sequence the average scores were calculated and converted into probabilistic coefficients.

B. Results and Discussions

Base on scores collected from observation sheets we compute the following average fuzzy sets:

$$es_{1st\ sequence} = \{.33, .73, .44, .09, .00\} \quad (5)$$

$$es_{2nd\ sequence} = \{.37, .63, .24, .08, .02\} \quad (6)$$

$$es_{3rd\ sequence} = \{.46, .73, .34, .07, .04\} \quad (7)$$

These fuzzy sets show that the assisted therapy commonly induces an open and positive affective state.

The following two questions arise naturally:

1) There are significant differences between emotional patterns of therapy sequences?

In order to identify specific differences we used the

Paired Sample t Test and the results are presented below:

TABLE I. DIFFERENCES BETWEEN EMOTIONAL PATTERNS

pairs	t value	sign. level
neutral seq. 1 – neutral seq. 2	5.445	.000
neutral seq. 2 – neutral seq. 3	-2.844	.006
neutral seq. 1 – neutral seq. 3	3059	.003
contentment seq. 1 – contentment seq. 2	3507	.001
contentment seq. 2 – contentment seq. 3	-3043	.001
happiness seq. 1 – happiness seq. 3	-2.697	.009
happiness seq. 2 – happiness seq. 3	-2.079	.047

While there are specific patterns for neutral, contentment, and happiness, for the other two basic emotions we identify but a general pattern.

2) Is the fuzzy model appropriate for affect representation?

Even if we didn't make quantitative investigations regarding the easy of usage of the fuzzy model, the psychologists' feedback was a positive one. They find the fuzzy scores as an appropriate and natural way for modeling a complex affective state. On the other hand, we have interested about the similarity between the experts' assessment. This point is very important due the fact that this experiment is the first step of training of emotion recognition framework. A good correlation (Kendall tau rank correlation coefficient - Table II) between fuzzy sets indicated by psychologists is "a must" in order to use them as a reference point.

TABLE II. CORRELATIONS BETWEEN EXPERTS' ASSESSMENTS

	nervousness	*tenseness*	*neutral*	*contentment*	*happiness*
seq. 1	.513**	.333*	.640**	.513**	439**
seq. 2	1000*	-.334*	.683**	.400*	.304*
seq. 3	-.315*	582**	742**	633**	359*

*Note. *p<.05, **p<.01*

These strong correlations ensure that we are on the right way. If human experts wouldn't be able to offer similar assessments then we wouldn't have the possibility to obtain pre – classified samples and to evaluate the performances of recognition framework.

V. Conclusion

In this paper we have proposed an improved, "emotion extended" CBST architecture. We have indicated the advantages and drawbacks of the utilization of physiological parameters in such a task. In addition, a fuzzy model for representation of emotions was presented. We find this approach as being an appropriate and natural way for modeling a complex affective state within assisted speech therapy. The experiment has involved SLTs, psychologists, and children with speech disorders and the results indicated the opportunity of fuzzy approach utilization. The acquired physiological features together with strong correlated fuzzy sets provided by human experts are necessary for framework training. Furthermore, the resulted probabilistic pattern will help the emotion recognition framework to deal with ambiguous situations.

References

[1] S. G. Pentiuc, I. Tobolcea, O. A. Schipor, M. Danubianu, M. D. Schipor, "Translation of the Speech Therapy Programs in the Logomon Assisted Therapy System," Advances in Electrical and Computer Engineering, vol. 10, no. 2, pp. 48-52, 2010. [Online]. Available: http://dx.doi.org/10.4316/AECE.2010.02008

[2] O. A. Schipor, S. G. Pentiuc, M. D. Schipor, "Using a Fuzzy Emotion Model in Computer Assisted Speech Therapy," in *Proc. 3rd International Conference on Software, Services and Semantic Technologies*, Bourgas, Bulgaria, 2011, pp. 189–194.

[3] M. Cerlinca, A. Graur, S. G. Pentiuc, T. I. Cerlinca, "Developing a Logopaedic Mobile Device Using a FPGA," in *Proc. of SACI '07*, Romania, pp. 89-92, 2007.

[4] O. A. Schipor, S. G. Pentiuc, M. D. Schipor, "Improving Computer Based Speech Therapy Using a Fuzzy Expert System," Computing and Informatics, vol. 29, no. 2, pp. 303-318, 2010.

[5] M. H. Zaharia, F. Leon, "Speech Therapy Based on Expert System," Advances in Electrical and Computer Engineering, vol. 9, no. 1, pp. 74-77, 2009. . [Online]. Available: http://dx.doi.org/10.4316/AECE.2009.01014

[6] M. D. Schipor, S. G. Pentiuc, O. A. Schipor, "End-User Recommendations on LOGOMON - a Computer Based Speech Therapy System for Romanian Language," Advances in Electrical and Computer Engineering, vol. 10, no. 4, pp. 57-60, 2010. [Online]. Available: http://dx.doi.org/10.4316/AECE.2010.04009

[7] A. Mehrabian. Nonverbal Communication. Aldine-Atherton, 1972.

[8] S.M. Lajevardi, Z. M. Hussain, "Feature Extraction for Facial Expression Recognition based on Hybrid Face Regions," Advances in Electrical and Computer Engineering, vol. 9, no. 3, pp. 63-67, 2009. [Online]. Available: http://dx.doi.org/10.4316/AECE.2009.03012

[9] M. Paleari, B. Huet, R. Chellali, "Towards multimodal emotion recognition : A new approach," in *Proc. of ACM International Conference on Image and Video Retrieval on Medical & Biological Engineering & Computing*, Xi'an, China, 2010, pp. 174-181.

[10] E. Mower, A. Metallinou, C. C. Lee, A. Kazemzadeh, C. Busso, S. Lee, S. Narayanan, "Interpreting Ambiguous Emotional Expressions," in *Proc. of ACII*, Amsterdam, 2009.

[11] G. Castellano, L. Kessous, G. Caridakis, "Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech," Affect and Emotion in Human-Computer, vol. 4868, pp. 92-103, 2008.

[12] M. S. Hussain, R. A. Calvo, "A Framework for Multimodal Affect Recognition," in *Proc. of HCSNet Perception and Action Workshop*, Brisbane, Australia, 2009.

[13] J. Kim, E. André, "Emotion Recognition Based on Physiological Changes in Listening Music," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30, pp. 2067-2083, 2008.

[14] C. L. Lisetti, F. Nasoz, "MAUI: A Multimodal Affective User Interface," in *Proc. of the ACM Multimedia International Conference*, Juan les Pins, France, 2002.

[15] G. N. Yannakakis, H. P. Martinez, A. Jhala, "Towards affective camera control in games," User Modeling and User-Adapted Interaction, vol. 20, pp. 313-340, 2010.

[16] K. H. Kim, S. W. Bang, S. R. Kim, "Emotion Recognition System Using Short-Term Monitoring of Physiological Signals," in *Proc. on Medical & Biological Engineering & Computing*, 2004, pp. 419–427.

[17] O. Grigore, V. Velican, "Self-Organizing Maps For Identifying Impaired Speech," Advances in Electrical and Computer Engineering, vol. 11, no. 3, pp. 41-48, 2011. [Online]. Available: http://dx.doi.org/10.4316/AECE.2011.03007

[18] T. Polzehl, S. Sundaram, H. Ketabdar, M. Wagner, F. Metze, "Emotion Classification in Children's Speech Using Fusion of Acoustic and Linguistic Features," *in Proc. of INTERSPEECH-2009*, Brighton, UK, pp. 340-343, 2009.

[19] A. Ortony, T. J. Turner, "What's basic about basic emotions?" Psychological Review, vol. 97, pp. 315-331, 1990.

[20] W. Parrot. Emotions in Social Psychology. Philadelphia: Psychology, 2001.

[21] J. A. Russell, "A circumplex model of affect," vol. 39, pp. 1161–1178, 1980.

[22] H. Schlossberg, "Three dimensions of emotion," Psychological review, vol. 61, pp. 81-88, 1954.

[23] K. De Verena, L. C. Cooper. Managing emotions in mergers and acquisition. Edward Elgar, 2011.

[24] P. Petta, C. Pelachaud, R. Cowie. Emotion-Oriented Systems: The HUMAINE Handbook. Springer, 2011.

[25] H. Leng, Y. Lin, L. A. Zanzi, "An experimental study on physiological parameters toward driver emotion recognition," *in Proc. of EHAWC*, Beijing, pp. 237-246, 2007.

[26] F. Nasoz, C. L. Lissetti, K. Alvarez, N. Finkelstein, "Emotion Recognition from Physiological Signals for User Modeling of Affect," *in Proc. of M'2003*, Johnstown, 2003.