Finding key emotional states to be recognized in a computer based speech therapy system

Ovidiu-Andrei Schipor*, Doina-Maria Schipor, Emilia Crismariu, Stefan Gheorghe Pentiuc

Ștefan cel Mare University of Suceava, Universității street, no. 13, Suceava 720229, Romania
Regional Speech and Language Therapy Center, Marasesti street, no. 57, Suceava, 720173, Romania

Abstract

Emotion recognition has become a “must have” for all systems that want to inspire user’s confidence and to interact in a friendly and familiar way. This is why speech therapy, especially for young children, can be considered half-blind if it does not take into account the emotional state of the subject. Most research on affective state identification so far has focused on adult subjects, with a good pronunciation. However, little research has been conducted on adapting classical emotion recognition techniques in “narrow areas” such as children speech therapy, where emotions play a key role. In this paper we investigate which emotional states should be recognized by a CBST (Computer Based Speech Therapy System) in order to reduce the gap between classical and computer assisted therapy. A brief literature review is presented, exploring the recent work in the area. New hypothesis are tested using results from a focus group and from an experiment involving children. Appropriate affective states and measure model are identified. These results encourage us to develop an emotion recognition extension for our CBST named Logomon.

While the vision of a CBST that can replace human SLT (Speech and Language Therapist) is beyond the horizon of the next decade, recent advances has proved that computer can be a valuable “primary assistant” in the therapy process.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the 2nd World Conference on Psychology, Counselling and Guidance.

Keywords: Emotions models, affective computing, computer based speech therapy, emotion recognition;

1. Introduction

The last decade has seen an increasing interest in the role emotion plays in human cognition and social interaction. Several “artificial extensions” of the emotional intelligence were proposed starting with the automatic emotion recognition and simulation. Various models of the emotions have been applied and have proved being very promising in a wide range of areas.

Since 2005, our team has begun the development of Logomon – the first CBST (Computer Based Speech Therapy System) for Romanian language. We started implementing three classical modules: the main program installed on SLT’s PC, the child monitor program installed on a mobile device such as a PDA, and an animated 3D Model of the phono-articulatory system (Pentiuc, Schipor, & Danubianu, 2008). Then we extended this architecture with a fuzzy expert system whose role is to determine the optimal exercises set for each child (Schipor, Pentiuc, & Schipor, 2010). Next, in order to reduce the gap between the classical therapy and the computer assisted therapy, we...
want to add “affective computing” capabilities (Schipor, Pentiuc, & Schipor, 2011).

This article can be seen as a preliminary step for implementing the emotions recognition framework. It defines the requirements that the module must accomplish from the end-user point of view. So in this paper we focus on finding affective states to be considered in order to develop this extension. Because previous researches (Schipor & Schipor, 2009), (Schipor, Pentiuc, & Schipor, 2010) have shown that the most important drawback of actual CBSTs is the lack of empathy, integrating emotion understanding in such systems would be an important qualitative shift. There are two promising consequences of this new approach, both of them referring to the behavior of the computer (Pentiuc, Tobolea, & Schipor, 2010):

- Emotionally reactive behavior - The computer could identify some undesirable affective states that may occur in therapeutic lesson and could modify “on the fly” the next sequences of therapy;
- Emotionally proactive behavior – Knowing the child’s previous emotional behavior (i.e. emotional behavior that is likely to occur in a particular sequence of therapy), the computer could generate personalized lessons.

Positive affective states should be encouraged and maintained while the negative ones should be minimized. Anyway, the system must keep an optimal balance between creating and maintaining positive moods and still going through the required steps.

2. State of the art

Assisted speech therapy is a research and application field that demonstrates the need for close cooperation between humans (i.e. SLT – speech and language therapist, children and parents) and computers. The utilization of computer in speech therapy, initiated in 80s, has been proved extremely beneficial (Bunnel, Engwal, & Polikoff, 2000), (Balter, Engwal, Ster, & Kjellstrom, 2005). Some advantages such as objectivity, patience, and interactivity have contributed to the widespread deployment of such systems. The capacity of computers to provide ever more finely-grained detailed traces of children progress continues to be a major pull in therapy. Having access to a child’s digital trace may also aid SLT in diagnosing learning difficulties that may have previously been overlooked. As the trend towards developing more sophisticated technologies to record and assess a pupil’s output continue, the way computers empathize with subjects must be reconsidered.

The emotions are short time affective states with variable intensity determined by precise stimuli (Salovey & Mayer, 1990). Many theorists have tried to define a list of basic emotions and, consequently, a wide range of research on identification of these states could be referred. In their article (Ortony & Turner, 1990), 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 (Parrot, 2001) offers a model on three levels: primary (6), secondary (25) and tertiary emotions (over 130).

Another key point is the identification and the measurement of the emotions and there is a vast and sometimes contradictory associated literature. A person's emotional state is reflected simultaneously on different, complementary and redundant channels (Salovey & Grewal, 2005), (Mehrabian, 1972). Thus there are verbal (i.e. language and paralanguage parameters), visual (i.e. gestures and face expressivity), physiological (i.e. respiration and heart rate, skin temperature and conductance), and contextual (indirect) channels types. Although physiological parameters provide high accuracy, they have certain drawbacks such as: the need for special equipment, limitations in motion, moral issues and that is why SLT usually do not use this type of data (Pantic & Rothkrantz, 2003), (Takahashi, 2005).

When the humans are placed in the position to assess the emotions of the other people, three different types of approaches are generally used (Power, 2006):

- Labeling approaches – the human experts (e.g. psychologist, SLT) choose a specific emotion from a predefined list (e.g. anger, joy, pleasure, sadness). Although this method is convenient in terms of simplicity, it cannot be used in confusing situations, when emotional state would be described by several words, each of them eventually associated with a certain weight (Jonghwa, 2007);
Dimensional approaches – this method involves identifying affective state as a location on several continuous scales (e.g. pleasant – unpleasant, calm – arousal, attention – rejection). Two-dimensional (valence and arousal) (Russell, 1980) or three-dimensional (valence, arousal, and stance) (Schlossberg, 1954) models are used. The different emotional labels could be plotted at various positions on a plane or into a cube. For example, joy has a positive valence and a high arousal level while disgust has a negative valence and a low arousal level. In addition, joy reflects an open stance while disgust a closed one;

Weighted labeling approaches - instead consider a single emotional state associated with an activity, more emotions are accepted, each of them being related with a specific weight (De Verena & Cooper, 2005). Weighting can be performed, for example, taking into consideration how long the emotion was manifested in the activity (Lawton, Haitsma, & Klapper, 2011). The researchers that performed emotional states assessments through this approach, usually used observation scales and a significance factor was associated with each emotion.

As far as we know, there is no record of a research endeavor toward establishing relations between affective states and different sequences of children’s computer 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 emotions should be recognized by the system? What emotions have therapeutic value?

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 recognition space of the system. The second one defines the moments in which our framework should be turned on. The results corresponding to the third question (i.e. probabilistic values) 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).

3. Methods

The aim of this research is to define, from the psychological point of view, the specifications for the emotion recognition framework. We organize our research in two phases, each with its specific objectives:

A focus group involving two SLTs, two psychologists, and one IT specialist;

○ Finding emotions that have therapeutic value and to be recognized by the framework;

○ Identifying the sequences of the assisted therapy that are suitable for the automatic recognition of emotions;

An experiment involving forty-five children, five emotions and three therapy sequences;

○ Observing children's emotional states during the selected therapeutic sequences (this activity was performed in the same time by two experts);

○ Checking the results in terms of consistency;

○ Obtaining a probabilistic model that indicates the affective states that are likely to occur in a specific therapeutic sequence.

All the experiment’s results were obtained from the interaction between children and our system - Logomon.

3.1. The focus group

In order to perform the focus group, we have selected five persons with expertise in the field of computer assisted speech therapy: two speech and language therapists, two psychologists and one IT specialist. Each of these persons had to choose five emotions from Parrot’s tertiary emotions list, based on practical relevance in speech therapy. Subsequently, the most rated five emotions were selected. The same procedure was used in order to select the therapeutic steps. After two meetings have been accomplished the proposed goals, identifying both emotional states that must be recognized and therapy steps in which automatic emotion recognition system should be activated.

The selected affective and the signs considered to identify them are as follows:
Happiness – participating with great pleasure (laughing, verbal manifestation, and joking);
Contentment – participating with pleasure, with self-confidence (smiling, precise hand movements, relaxed face);
Neutral - participating with some pleasure, relax, and reservation;
Tenseness – participating without pleasure, with reluctance and some anxiety (line between eyebrows or across forehead, looking around, and leg jiggling);
Nervousness – participating with displeasure, with irritation or refusing participation (repeated or agitated movements, hand wringing, and verbal manifestation).

The selected therapy stages are as follows: 1. Speech evaluation by recording with feedback; 2. Exercises for phonematic hearing development; 3. The pronunciation of affected sound using 3D model;

3.2. The experiment

The subjects (N= 41) were children with moderate speech disorders (24 with many sounds mispronunciation, 9 with sound “s” mispronunciation, and 7 with sound “r” mispronunciation), 4 to 9 years old, selected for speech therapy by the SLTs from Regional Speech and Language Therapy Center – Suceava, Romania.

Procedure: The children were observed by two independent human experts during assisted speech therapy and an observation sheet was filled out by each of them (reliability analysis test Alpha - Cronbach α =.83). The observation was performed in all three sequences: 1. Speech evaluation by recording with feedback; 2. Exercises for phonematic hearing development; 3. The pronunciation of affected sound using 3D model. For each child and each stage of therapy five scores (in a Likert type scale ranging to 0 – “absence” to 5 – “maxim intensity”) were obtained. These scores corresponded to the intensity of the selected emotions. Thus, each child received fifteen scores, which means five scores for each of the three sequences.

Research hypotheses
- Each stage of therapy has a specific emotional pattern.
- The human experts made similar assessments.

The independent variables considered for the present study were: the sequence of assisted therapy and the human expert. The dependent variables were the intensity of the five emotional statements observed (i.e. happiness, contentment, neutral, tenseness, nervousness).

4. Results and Discussion

In order to analyze the differences between emotional patterns for each therapeutically sequence we performed the Paired Sample t Test. The results have shown us some differences between the intensity of the emotional states in the three sequences, but the effect of the independent variable was different. Table 1 shows the significant results.

<table>
<thead>
<tr>
<th>pairs</th>
<th>t value</th>
<th>significance level (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral sequence 1 – neutral sequence 2</td>
<td>5.445</td>
<td>.000**</td>
</tr>
<tr>
<td>neutral sequence 2 – neutral sequence 3</td>
<td>-2.844</td>
<td>.006*</td>
</tr>
<tr>
<td>neutral sequence 1 – neutral sequence 3</td>
<td>3.059</td>
<td>.003*</td>
</tr>
<tr>
<td>contentment sequence 1 – contentment sequence 2</td>
<td>3.507</td>
<td>.001**</td>
</tr>
<tr>
<td>contentment sequence 2 – contentment sequence 3</td>
<td>-3.043</td>
<td>.001**</td>
</tr>
<tr>
<td>happiness sequence 1 – happiness sequence 3</td>
<td>-2.697</td>
<td>.009*</td>
</tr>
<tr>
<td>happiness sequence 2 – happiness sequence 3</td>
<td>-2.079</td>
<td>.047*</td>
</tr>
</tbody>
</table>

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

Our first hypothesis is confirmed in a nuanced manner depending on the type of emotion. So, we identified
distinct emotional patterns for neutral (for each of three sequences), contentment (for 1 to 2 and 2 to 3 sequences), and happiness (for 1 to 3 and 2 to 3). For the others two emotional states (nervousness and tenseness) we identified but a global pattern, common to the all three therapeutically sequences.

Even if we could not identify a distinct emotional pattern for each therapeutic sequence, the obtained distributions provide vital information that will help us to build an improved automatic recognition module. That is why, even if our first hypothesis was not fully verified, we see these results as being an important step forward.

Figure 1 shows the average scores matrix for each therapeutic step and emotional state. Based on this matrix, a probabilistic model can be generated. This probabilistic model is to improve the performance of automatic recognition framework and to allow better discrimination between ambiguous situations.

As seen in Figure 1, the positive emotions prevailed in all three therapeutic sequences. This could be due to the children's positive attitudes towards computer or due to the emotional support provided by SLT. However, these fine discriminations could be performed only after the implementation of the framework.

In order to test the similarity between assessments performed by the two experts we used the Kendall tau rank correlation coefficient. The obtained results (see Table 2) revealed a positive relation between the two expert’s assessments of emotional states in the three therapeutically sequences proposed. These results fully confirm our second hypothesis.

Table 2. Significant correlation between experts’ assessments of the emotional states (N=41)

<table>
<thead>
<tr>
<th>therapeutically sequences</th>
<th>nervousness</th>
<th>tenseness</th>
<th>neutral</th>
<th>contentment</th>
<th>happiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Speech evaluation</td>
<td>.513</td>
<td>.333</td>
<td>.640</td>
<td>.513</td>
<td>.439</td>
</tr>
<tr>
<td></td>
<td>.000**</td>
<td>.033*</td>
<td>.000**</td>
<td>.000**</td>
<td>.001**</td>
</tr>
<tr>
<td>2. Phonematic exercises</td>
<td>1.000</td>
<td>-.334</td>
<td>.683</td>
<td>.400</td>
<td>.304</td>
</tr>
<tr>
<td></td>
<td>.000**</td>
<td>.022*</td>
<td>.000**</td>
<td>.002*</td>
<td>.021*</td>
</tr>
<tr>
<td>3. 3D model</td>
<td>-.315</td>
<td>.582</td>
<td>.742</td>
<td>.633</td>
<td>.359</td>
</tr>
<tr>
<td></td>
<td>.020*</td>
<td>.000**</td>
<td>.000**</td>
<td>.000**</td>
<td>.007*</td>
</tr>
</tbody>
</table>

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

Verification of the second hypothesis encourages us to extend the therapeutic system. If human experts would not be able to offer similar assessments then we wouldn’t know how to evaluate system recognition performances.
5. Conclusions

The aim of this paper is to define, from the psychological point of view, the specifications for an emotion recognition framework to be added to our CBST. Therefore we have analyzed different existing models of emotions and several emotion assessment methods have been discussed. The first phase of our research (a focus group) has shown us what emotions and what therapy stages are appropriate for automatic recognition. The second phase (an experiment) has revealed us that there is a specific emotional pattern for some therapeutic step and that human experts offer a consistent emotions evaluation. A probabilistic model that will help framework to deal with ambiguous situations was also obtained. These results encourage us to implement the emotion recognition extension for Logomon system.

Acknowledgements

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.

References