UBT International Conference

2013 UBT International Conference

Nov 2nd, 3:15 PM - 3:30 PM

Comparison of emotion evaluation perception using human voice signals of robots and humans

Vesna Kirandziska
Rugjer Boshkovikj, vesna.kirandziska@finki.ukim.mk

Nevena Ackovska
Rugjer Boshkovikj, nevena.ackovska@finki.ukim.mk

Comparison of emotion evaluation perception using human voice signals of robots and humans

Vesna Kirandziska[1], Nevena Ackovska[1]

[1] Faculty of Computer Science and Engineering, „Rugjer Boshkovikj" 16, 1000 Skopje, Republic of Macedonia
{vesna.kirandziska,nevena.ackovska}@finki.ukim.mk

Abstract. Emotion perception is the process of perceiving other people's emotions. It can be based on their facial expression, movement, voice and other biosignals people emit. The evaluation of human's emotion is one characteristic of emotions. One of the research areas in Robotics is adapting humanistic behavior in robots. Today many robots are constructed. Some of them can even perceive emotions. In this paper a custom built emotion aware robot that perceives emotion evaluation is used to investigate the similarity and differences of the robot's and human's emotion perception. Voice signals from real human were recorded and the information for the emotion evaluation was obtained from our robot, but also from a set of human evaluators. This paper presents the results of the experiments done. The experimental results show the difficulty of the problem of emotion evaluation perception in general. The significance of human voice signals in emotion evaluation is also investigated.

Keywords: emotion, evaluation, robots, human

1 Introduction

Human's interaction with robots in the future should be based on natural conventions like natural language or social rules (facial expression, mimic and body movements). So, robots would no longer be just indifferent logical machines. They could become capable of understanding human's feelings, needs and desires. In this paper we are considering emotions as one of the most important information robots should understand. In human-to-robot interaction emotions are especially significant because emotion perception can facilitate communication between these two subjects. Machine learning and more precisely classification algorithms are implemented in robots that perceive emotions. Examples of applications that use emotion classification are video and computer games, human-to-robot interaction systems and even more call centers ([2], [7], [4], [13]). In general, information from speech ([1], [3], [9], [12]), facial expression [14] and brain activity has been used for emotion classification so far. Here, only sound signals are considered. In [7] the justification for using sound features for emotion classification is found. Further, the results in [6] show that certain emotions are correlated with the values of some sound features extracted from human speech.Frameworks for automatic emotion classification are created in [4] and [14]. In [4] online emotion classification is also considered. These can be used in real life applications. Some researches aim at finding new sound features [9] or extracting the best combination of existing sound features [6]. Many different classification algorithms like decision trees [10] neural networks [11] and SVM [16] are being used. Still, there isn't a satisfactory accurate classifier built and there is not a standard set of sound features that is used.

Today, robots that perceive emotions have been constructed. But, unfortunately there is not an attempt made yet to compare the robot's perception of emotion with the human's perception. This is exactly the final goal of this paper. We would like to explore the natural and artificial emotion preceptor system. The research presented in this paper will be exposed in the following sections. The following section describes the robotic system built and used for this research. In Section 3 we present the experimental results for comparison of human's and robot's emotion perception. In the end, the future work and the conclusion for this paper are given.

2 Presentation of the human-robot interaction system used for research

In [15] and [17] a robot that perceives emotion evaluation based on human speech was presented. Its intelligence was based on a custom built emotion classification algorithm. The robot used for this application is the robotic arm, Lynx 5, developed by Lynxmotion.

As in many cultures, when someone is in need of help i.e. has negatively evaluated emotion, the robotic arm gives him/her a hand. The robot should straighten the hand, go in a handshaking position and put the arm back in the starting position when the speaker has negative emotions expressed by his speech. This actually simulates the giving hand gesture.

The automatic emotion evaluation system built in the robotic arm gives the most probable emotion evaluation (positive or negative) from the features extracted. As explained in [17] sound features used in our research here are found important in some psychological researches. In the next section, the model for our human-robot interaction system will be explained in more detail.

2.1 "Wordless call for help" application

The robotic system built for the "Wordless call for help" application is schematically shown on Figure 1. It can be divided in two bigger parts. One concerning the emotion perception (upper part of Fig. 1) and other concerning the robot's behavior in human-robot interaction (lower part of Fig. 1). We created a human-robot interaction application called "Wordless call for help".

The procedure for emotion evaluation perception is conducted in a few smaller steps. First the sound signal is captured and digitalized (Fig. 1A). Next, the sound signal is analyzed with an algorithm for preprocessing and sound transformation (Fig. 1B). In the next step, form the preprocessed sound signal the sound features are extracted (Fig. 1C). Feature selection is done next. It is the process of selection valuable features for a given classification problem. This step is only conducted during the training phase. We introduced a novel approach for feature selection in [17]. It is used in the research presented in this paper. With this process, the most valuable features are extracted, so in the test phase only few features are calculated (Fig. 1D). The final step is the classification of emotion evaluation from the selected sound attributes. It is shown on Fig. 1E. These 5 steps finalize the emotion classification model.
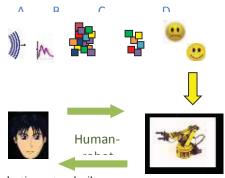

Fig. 1. Schematic view of the robotic system built.

In order to build the model, a training of the emotion classification algorithm is performed. In this research a custom built database was used for training. The database contains 25 sound signals that represent a spoken sentence of only one person. This person, called an actor, spoke the same sentence 25 times, acting 25 different emotions.

The classification model used here is evaluated and trained using the Weka software [8]. For creating the database, sound features were extracted using the PRAAT software [5]. As stated previously, a custom built method for feature selection was used in order to extract the most valuable sound features for emotion perception [17]. More about the classification model and the results are presented in Section 2.2.

Given the emotion classification model the robot acts based on the classification. It can move it servo motors for making the hand shaking gesture if the estimated emotion is evaluated negatively. These elements consists the second part of the model, called the behavior model (lower part of Fig. 1).

The behavior model was built using direct serial manipulator's kinematics. Indeed, the application sends signals to the robot describing the next motor movement. The composition of the movement of all motors, represent the robot's movement.

2.2 Building and evaluating the classification and behavior model

For building the classification model of the robotic system, first a database is build. Some sound features are extracted using the PRAAT software to build a sound feature database. For feature selection first every feature is ranked, and using its rank the best fitting sound features are selected. In [17] a new approach for ranking the sound features based on linear classifier and psychological researches. Using this approach a set of features was selected. These features include pitch range, minimum amplitude and voice brakes. After selecting the features, the classification model was now trained with the new set of features. Linear classifiers were used for the classification model. Indeed neural networks were used as a linear classifier. Using Weka, the model was trained and evaluated. For the evaluation of the classification model the 10-fold method was used. The precision of the classifier built is 85%.

The precision given gives only partial information of weather this classifier is proper for human-robot interaction. On the other side, the classifier should be similar to human perception of emotion evaluation. This presented the need for another research that compares the emotion perception of the robot with human perception. The goal of the next section is to display the results from this research.

3 Experimental results for comparison of robots and humans emotion evaluation perception

The primary goal of this paper was to investigate the dependencies between the human and robot emotion evaluation perception. We would like to check how close their perception for other human emotions is.

In the next sections the results gained in our research are shown.

3.1 Comparison of our robot emotion perception with the perception of humans

More important for robots that can perceive emotion evaluation is weather their evaluation is similar to human evaluation. Indeed, emotion perception is a complex task even for humans.

Usually, a human can perceive emotions on another human more precisely if they know each other. Evermore, no matter how similar emotion expression by humans may be, there are still slight differences between emotion expressions of two different humans. This can be due to different cultural, social and other diversity issues. The idea of this research is to find the correlation between human evaluator's and robot's emotion evaluation detection. Experiments are done with the "Wordless call for help" application. For these experiments a new test database was build. It contains sound signals form one speaker, who is different from the human from which the training of the classifier was done. The idea was to test the classification of the robot when a different person is in question. In the first experiment both the evaluator (who doesn't know the speaker) and the robot were given to classify the sound signals from the test database. The comparison of their evaluation is given on Fig. 2.
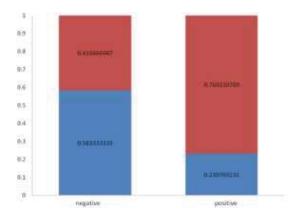
Fig. 2. Emotion evaluation of the evaluator with respect to the robot classifier.

On the x-axis the classification of the robot is given (positive or negative). From the first column in Fig. 2 it can be seen that 58% of the negatively evaluated emotions by the robot were classified the same by the evaluator. This percentage is 76% for the positively evaluated emotions. This implies the similarity of the classifiers: the automatic classification system implemented in our robotic system and the real one by birth obtained by the evaluator. In the next experiment more evaluators were included. These evaluators also did not know the speaker. The general precision of the evaluators given the true emotion evaluations went from 50% to 90%. This describes that also for humans, the emotion evaluation problem, is a difficult problem that has a great imprecision. From the results obtained from this experiment the percentage of equally qualified sound signals by the evaluators and our robot were calculated. The histogram of the percentages is given on Fig. 3. As shown, most evaluators have more than 80% (0.8) coincidence with the robot classification. Only a few evaluators had less than 60% coincidence. This shows the similarity of the evaluators' perception on emotion evaluation with the perception of our robotic system.

This in general shows some evidence that making emotion aware robots that function like humans is possible and applicable.
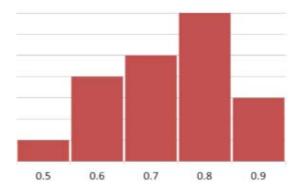

Fig. 3. Histogram of the percentage of equally qualified sound signals by the evaluators, given the robot's classification.

3.2 The significance of the robot's learning phase

In this section the significance of the robots learning phase is stated. When the robot is interacted with a new speaker it uses the classifier built based on a different speaker, so the classifier is not adapted. Indeed, because of the diversity of all people, the emotion evaluation classification must be adjusted.

Therefore, we investigated the change in the robots classification precision during the learning phase. The learning phase includes the adjustment of the classification model previously built.

On Fig. 4 the learning phases given in increasing order are given on the x-axis. On the y-axis the percentage of the precision of the classification model is shown.
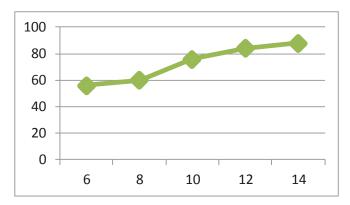

Fig. 4. The precision of the robot classification model during the learning phases.

In the 14th learning phase the precision of the classifier goes up to 90%. This states the importance of adaptation on the classifier model. This adaptation is similar to learning or getting to know the speaker the robot is "talking to". Humans also would give better classification on the emotion evaluation for people they know. In the next section the future work is given along with the conclusion of our work.

4 Future Work and Conclusion

Proper differentiation on emotions is very important. As a consequence, people with psychological background knowledge should be included in feature research. In the future, emotions differentiation in more categories would be essential for getting more precise information. Important sound features for the new categories of emotions can be investigated as well.

The training and testing phase should be done with databases built from voices of different speakers. Also, more evaluators should be included in the research. In this research the problem of emotion classification based on sound features of a human voice is considered. Emotions, based on evaluation, are classified in two classes: positive and negative emotions. We also used information from psychological studies to make a classification algorithm with more humanoid properties. The robot's emotion perception deduced from sound signal processing that is biologically driven showed great precision of 85%. Results showed great similarity of the ability of both robot and human to perceive positive or negative emotions. As a conclusion, biologically driven classification algorithm implemented in human-robot interaction, has a positive influence of the similarity of the robot's and human's perception on emotion evaluation. This result can be used in the future for building classification models used in human-robot interaction.

References

1. V. Kirandziska and N. Ackovska: Sound features used in emotion classification for human interaction with robotic systems. Proc. Ninth International Conference for Informatics and Information Technologies. (2012) 91-95
2. N. Ackovska and S. Bozinovski: Biped Robots: From Inverted Pendulum to Programming 12dof Dancing Postures. Proc. Seventh International Conference for Informatics and Information Technologies. (2010) 3-7

3. T. Vogt, E. Andr´e, and J. Wagner: Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realization. Affect and Emotion in HCI. Lecture Notes in Computer Science: Springer-Verlag. Berlin Heidelberg. vol. 4868. (2008) 75–91

4. T. Vogt, E. Andre and N. Bee: EmoVoice – A framework for online recognition of emotions from voice. Proc. of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems: Perception in Multimodal Dialogue Systems. (2008) 188 – 199

5. P. Boersma and d. Weenink: PRAAT: doing photetics by computer (Version 5.1.05). DOI= http://www.praat.org/ (2006)

6. R. W. Picard: Affective computing. Massachusetts Institute of Technology (MIT), Massachusetts, United States, Media Laboratory Perceptual Computing Section Technical Report No. 321. (1995)

7. K. R. Scherer, R. Banse, H. G. Wallbott, and T. Goldbeck: Vocal Cues in Emotion Encoding and Decoding. Motivation and Emotion. vol. 15. (1991) 123-148

8. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten: The WEKA Data Mining Software, vol. 11. DOI = http://www.cs.waikato.ac.nz/ml/weka/ (2011)

9. L. Caponetti, C. Buscicchio and G. Castellano: Biologically inspired emotion recognition from speech. EURASIP journal on Advances in Signal Processing. vol. 2011. (2011)

10. J. Cichosz and K. Ślot: Emotion recognition in speech signal using emotion extracting binary decision trees. Institute of Electronics, Technical University of Lodz. Poland. Technical Report. (2010)

11. K. Dai, H. J. Fell, and J. MacAuslan: Recognizing Emotion in speech using neural networks. Proc. of the IASTED International Conference on Telehealth and Assistive Technologies. vol. 619. (2008) 31-36

12. O. Kwon, K. Chan, J. Hao and T. Lee: Emotion Recognition by Speech Signals. Proc. of Eurospeech, Genewa. (2003) 125-128

13. Y. Huang, G. Zhang, F. Da: Speech Emotion Recognition with New Spectral Features and Its Application to Pet Robot. Journal of Computational Information Systems, vol. 13. (2011) 4915-4922

14. Mower, M. J. Mataric and S. Narayanan: A framework for automatic human emotion classification using emotion profiles. IEEE Transactions on audio, speech and language processing, vol. 19, no. 5. (2011) 1057-1070

15. V. Kirandziska and N. Ackovska: Human-robot interaction based on human emotions extracted from speech. Proc. Of the 20th Telecommunications Forum - TELFOR. (2012)

16. S. Yacoub, S. Simske, X. Lin and J. Burns: Recognition of Emotions in Interactive Voice Response Systems. HPL Technical Report 136 (2003)

17. V. Kirandziska and N. Ackovska: Finding Important Sound Features for Emotion Evaluation Classification. Proc. EUROCON. (2012) 1637-1643