A Voice-Based Gender and Internal State Combined Detection Model

Amir Aly, Adriana Tapus

Towards an Online Voice-Based Gender and Internal State Combined Detection Model

Amir Aly and Adriana Tapus
Cognitive Robotics Lab/UEI Computer Science Department
Ecole Nationale Superieure de Techniques Avancees (ENSTA-ParisTech)
32 Blvd Victor, 75015, Paris, France
{ami.aly; adriana.tapus}@ensta-paristech.fr

***Abstract*—In human-robot interaction, gender and internal state detection play an important role in making the robot reacting in an appropriate manner. This research focuses on the important features to extract from a voice signal in order to construct successful gender and internal state detection systems, and shows the benefits of combining both systems together on the total average recognition score. Moreover, it consists a foundation on an ongoing approach to estimate the human internal state online via unsupervised clustering algorithms.**

I. Introduction

The automatic recognition of emotions has recently received much attention in order to build more intuitive human robot interfaces. Relevant acoustic features from voice signals such as pitch and energy are needed ao as to reliably recognize emotions and human internal states. These features are, also, highly dependant on the gender; therefore, this study demonstrates how the extracted information of the gender could affect the accuracy of the internal state detection.

II. Database

The database used in this research is the German emotional speech database (http://database.syntheticspeech.de/) [5] with more than 520 voice samples for 5 men and 5 women, discussing 7 emotional states.

III. Gender Detection

The characteristic vector is based on the statistical measuring (e.g. mean, variance, max, min, range) of each of the extracted features from the voice signal (pitch, energy, formants, and mel-frequency cepstral coefficients) [1] [2].

The results of the gender detection recognition (using Support Vector Machine (SVM)) showed that female voices (299 voice sample) are recognized with 93.3% and the percentage of recognition of male voices (224 voice sample) is of 87.5%, making an average score of 90.4% (see Figure 1).

IV. Internal State Detection

The characteristic vector detecting the internal states is similar to the characteristic vector used in gender detection, but it depends only on pitch and energy [3].

The investigated internal states are: anger, boredom, disgust, fear, joy, neutral, and sadness, which are varying in the recognition score as indicated in Figure 2. The seven tested internal states achieved an average recognition score of 86.4% using the SVM algorithm.

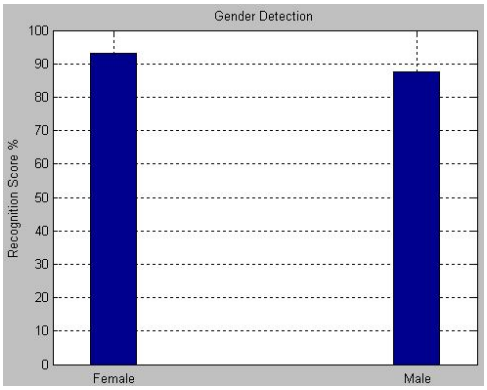

Fig. 1: Female and Male Voice Recognition Score

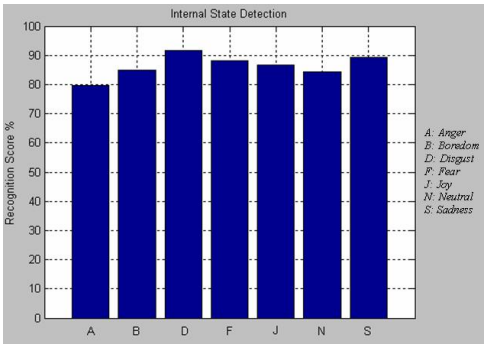

Fig. 2: Internal States Recognition Score

V. Combined Gender and Internal State Detection

The dependency between gender and internal state detection indicates that the recognition score of internal states could be ameliorated if the gender's recognition score is ameliorated in a similar way.

The main idea is to first classify the tested voice sample as being male or female and then to detect the internal state of the person as illustrated in Figure 3.

The total average recognition score is about 93%; however, this recognition system assumed an ideal gender detection (i.e. 100%), so as to reach a total average score of about 93% for the combined model (which is the average value between the ideal gender's recognition score 100% and the internal states' recognition score which is 86.4%). In the real case, the

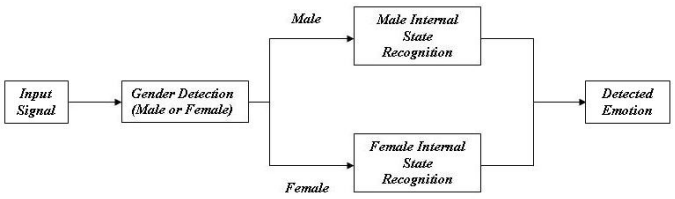

Fig. 3: Combined Gender and Internal States Recognition Architecture

accuracy of the gender detection system had achieved only 90.4% and tests' results revealed a total average recognition score for the combined model of about 88% (which is approximately the average value between the "real" gender's recognition score 90.4% and the internal states' recognition score which is 86.4%).

VI. Online Detection of Internal States

The need for an effective online internal state detection system comes from the fast growing human robot interaction applications which require the robot to be able to deal appropriately with different and varying internal states. Traditional approaches for detecting internal states are based (as illustrated in the first part of this research) on constructing a database knowing in advance the number of clusters. However, it is not possible to construct a good and big database including samples for all human internal states. Thus, we can expect that the robot will not be able to deal with different interactional situations in an appropriate way due to an error in classifying a new or unknown internal state.

Hence, this research tries to integrate an online clustering algorithm to the previously constructed gender and internal state combined detection system. The German speech database with its 7 emotional states will be the origin of the new system, and then new emotional states will be added to the database. The online clustering algorithm (e.g., subtractive clustering) will deal with the new updated database and will try to update the number of clusters to adapt the new coming elements to the database [4]. Subtractive clustering uses data points as candidates for the cluster's centers, and then it calculates for each of the proposed cluster's centers a density function which indicates how the proposed cluster's center is affected by the surrounding points in the dataset according to the following equation (where r is the neighborhood radius, X_i represents the cluster point under test, and X_j represents data points):

$$D_i = \sum_{j=1}^{n} exp(\frac{-(\|x_i - x_j\|)^2}{(r/2)^2}) \quad (1)$$

The first cluster center is chosen as the point that has the maximum density function. Afterwards, this process is repeated until a sufficient number of centers is attained. While repeating this process to find new cluster's centers, all data points are considered as candidates for the new cluster's centers, except the previously final calculated cluster's center and its surrounding (see Figure 4).

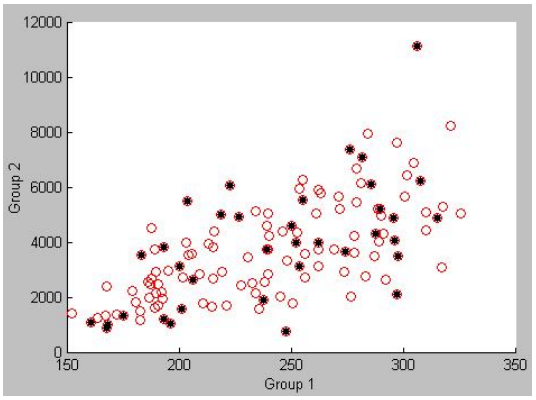

Fig. 4: 2-D subtractive clustering for the extracted features of 2 elements of the database of emotional type: Anger. The clustering centers presented by the black points could be averaged to have a final cluster center for the 2 groups' data.

Supposing a new emotional state is added to the database which will be in this case a N-D clustering process, where N is the number of elements in the database, the subtractive clustering algorithm can work efficiently to precise the whole number of clustering centers, however, the main challenge now is to collect similar centers in terms of the possible internal state they present into separate groups, which will facilitate the detection of the new coming internal states to the database if any. More results will be available by the time of the conference, as this paper reports ongoing work in progress.

VII. Conclusion

This research illustrates a combined model for gender and internal state detection. The total average recognition score of the combined model is about 88% which could be considered as a good score not only because of the large processed database (i.e., all voice samples of the German speech database with more than 520 sample), but also, because of the accuracy of the gender detection system. The whole combined system could be considered as a successful system.

Acknowledgements

This work is supported by the French National Research Agency (ANR) through Chaire d'Excellence program 2009 (Human-Robot Interaction for Assistive Applications).

References

[1] D. Talkin, A Robust Algorithm for Pitch Tracking, *in Speech Coding and Synthesis*, W B Kleijn, K Paliwal eds, Elsevier 1995.

[2] H. Traunmuller and A. Eriksson, A Method of Measuring Formant Frequencies at High Fundamental Frequencies *in the Fifth European Conference on Speech*, Greece, 1997.

[3] C. Breazeal and L. Aryananda, Recognition of Affective Communicative Intent in Robot-Directed Speech *in the Autonomous Robots Journal*, Volume 12 Issue 1, January 2002.

[4] N.Pal and D. Chakraborty, Mountain and Subtractive clustering method: improvements and generalizations *in the international journal of intelligent systems*, volume 15, issue 4, pages 329- 341- 2000

[5] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier and B. Weiss, A Database of German Emotional Speech, *In Proc. Interspeech*, 2005