International Conference on Control, Engineering & Information Technology (CEIT'13)
Proceedings Engineering & Technology - Vol.4, pp.180-182, 2013
Copyright - IPCO

Towards a recommendation system for TV programs based on human behaivior

Mohamed Meddeb Mohamed Adel Alimi
REsearch Group on Intelligent Machines (REGIM)
University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia
Mohamed.meddeb@ieee.org , adel.alimi@ieee.org

Abstract—**In this paper we include the related work of emotional classification and emotional database. Then we propose the architecture of Automatic Emotion Recognition from Speech (AERS). Latter, allows us to introduce the recommendation system for TV programs based on human behavior. Interactivity is understood as a psychological and emotional process, not as a technological one. However, the remote emotional includes unimodal and multimodal approach, and shows the hierarchical recognition emotions steps.** *(Abstract)*

Keywords-component; recommendation system, Multimodal, emotion, ITV, speech, behaviour. **(key words)**

I. Introduction

The speech signal conveys the message not only informative but also repository of information on personality and emotional state of the speaker.

The EMOTION is a motivational and adaptive response of an organism to the social environment. It is part of everyday life in humans. Despite the long history of study of speech and that of emotion, relatively little work has been devoted to the analysis of EMOTIONAL SPEECH. This can be attributed to the formalism of modern science and the methodological problems in the study of vocal emotion. In formal linguistics, researchers have been concerned to describe the regularity of binary language, rather than the variability of the multipurpose floor, and they consider the change due to voice emotion as a random variable, non-routine, which does not deserve scientific analysis. In the psychology of emotion, researchers are primarily interested in demonstrating the nature of emotion (either pre-cognitive or cognitive-post) rather than to describe the vocal expression of emotion, facial or body. Moreover, most psychological studies of emotion are based on data of facial emotion rather than emotion voice data because the latter are more difficult to acquire than first. The formalism is also in the field of speech technology.

The recognition of emotions in speech has many useful applications. In the man-machine interfaces, robots can learn to interact with humans and to recognize human emotions. Robotic Pets, home lighting automation and interactive television (ITV), for example, should be able to understand not only the voice commands, but also other information, such as emotional health humans and modify their actions accordingly. The advent of digital television via satellite, the redundant devices, the proliferation of channels of redundant programs, often complicates the choice of television programs that meet user preferences.

II. PROBLEMATIC

Recently, researchers increasingly realize the importance of knowledge on the emotional speech for the development of research in practice and in the theoretical. The overall mechanism of speech communication can be better understood by understanding the influence of emotion on the production and speech perception. The nature of emotion can be better understood by studying the relationship between the expressions voice, facial and bodily emotion. In speech technology, the addition of personal traits and emotional is essential to increase the naturalness of synthetic speech. The automatic speech recognition can greatly benefit from the development of the system that can recognize speech from different speakers in their various emotional states.

Our works is a study of emotional speech on the basis of data acquired from real situations. It aims to demonstrate how the emotion of the speaker is expressed in its natural speech in terms of acoustic cues and how the listener perceives it in different conditions of hearing. The analysis data consists of spontaneous speech excerpts, TV shows and reality, expressing the joy and sadness (whining voice) Tunisian speakers. Emotions predestined in this work are considered as the true emotions experienced by the speaker, as opposed to stylized emotions imitated by an actor. The choice of emotions actually experienced (not stylized) is based on the following consideration. Since the recording of vocal emotions from natural situations because of methodological problems and mental trauma, most studies are based on data from the emotions expressed by an actor following the experimenter's instructions about how voice. Although these theatrical emotions are supposed to be representative of those we experience in everyday life, both kinds of emotions are different yet at the emotional, motivational and expressive. The expression of emotion by the actor often takes an exaggerated form and involves a theatrical style, whereas the expression of emotion by ordinary people in everyday social interaction takes rather a discrete and follows rules regulating exposure. Considering the difference between emotion and the emotion experienced stylized and the scarcity of studies of emotion experienced, we decided to examine data emotional expressions natural, improvised from interviews. Although these talks have been broadcast through television, the emotions expressed in them are considered authentic compared to what the person felt.

This essay consists of a series of acoustic analysis on the emotional and neutral statements and a series of experiments on perceptual identification of emotion by listeners Tunisians

III. Related Work

A. *TV program recommendation*

Recommender systems for TV program have been studied for the realization of personalized TV Electronic Program Guides. To recommend appropriate TV programs, a user profile that reflects its preferences on the selection of TV programs should be estimated. The typical characteristic used to generate the user profile is the time to look, speech expressions, face and gesture detection and attributes of television programs. The characteristics of television programs, such as genres and performers (actors) are used to estimate the common characteristics of television programs, these attributes can be obtained from metadata, such as TV-Anytime, and IEPG (Internet Electronic Program Guide). The behavior of users to watch TV programs have also been studied.

B. *Emotions classification Algorithms*

As for the choice of emotions classifier there is no uniform a priori answer to the question of which classifier constitutes the best choice. The criterion for selecting an emotion classifier should be related to the task, in order to take into account the regularities of the problem, or of the geometry of the input feature space. Some classifiers are more efficient with certain type of class distributions, and some are better at dealing with many irrelevant features or with structured feature sets. One of the ways to compare the choice is to test the classifiers on the same large and representative database. Many classifiers have been tried for SER[1], and after Weka 1 has appeared it has become easy and straightforward. The most frequently used are Support Vector Machines SVM, Gaussian mixture models GMM, k nearest neighbor KNN, Mel-frequency Cepstral coefficients MFCC and Neural Networks NN.

A rapidly evolving area in pattern recognition research is the combination of classifiers to build the so-called classifier ensembles. For a number of reasons (ranging from statistical to computational and representational aspects) ensembles tend to outperform single classifiers. In the field of speech processing classifier ensembles have also proven to be adequate for the SER task.

C. *Emotional Databases*

There are six databases available for the present study: two publicly available ones, the Danish Emotional Speech corpus (DES) and Berlin Emotional Database (EMO-DB), and four databases from the Interface project with Spanish, Slovenian, French and English emotional speech. All of these databases contain acted emotional speech. With respect to authenticity, there seems to be three types of databases used in the SER research:

Type1: Databases of acted emotions, is obtained by asking an actor to speak with a predefined emotion. In [18] some experiments focusing on the production and perception of real and acted emotional speech supported the opinion that acted emotional speech is not felt when spoken, and is perceived more strongly that real emotional speech.

Type2: Is databases coming from real-life systems as used in: [09].

Type3: Is a database with elicited emotions, where emotions are provoked and self-report is used for labeling control. In this case, no manual labeling is needed.

All of these databases do not address the Arab emotional specificity and behavior, Thus we propose to create a database in Arabic. Sound recording studios will be in professional audiovisual productions.

IV. Proposed Architecture

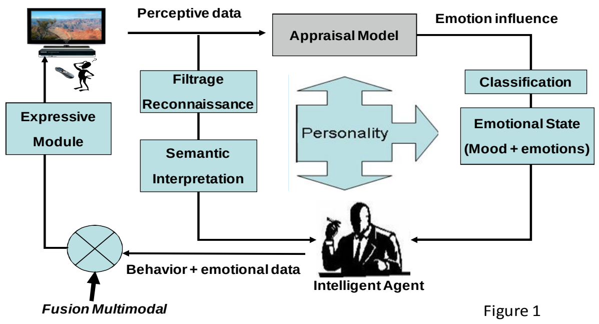

Figure 1

Figure.1 depicts how we view the role of personality (user profile) and emotion as a glue between perception, dialogue and expression. Perceptive data is interpreted on an emotional level by an appraisal model. This results in an emotion influence that determines, together with the personality what the new emotional state and mood will be. An intelligent agent uses the emotional state, mood and the personality to create behavior.

A. *Description*

A system of automatic recognition of emotions based on classical four main phases (Figure.1):

1) *The extraction of acoustic descriptors Figur.2* consisting of an analysis module transforming the speech signal into a sequence of acoustic vectors containing the values of various descriptors (parameters) selected. The objective of this

[1] Weka: software tools for building *SVM classifiers*.

step parameterization is to obtain a compact representation of the main character acoustics of the speech signal that is relevant to discriminate between them different emotional expressions. The intensity of the speech signal or the Central frequency formants are two examples of such acoustic *descriptors.*

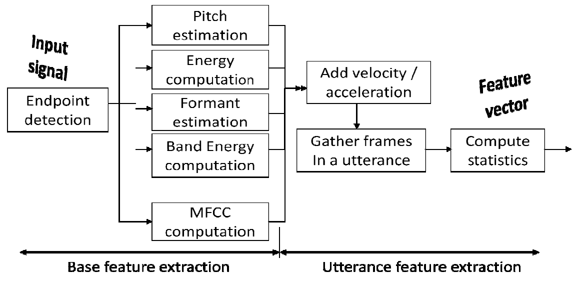

Figure.2

2) During the learning phase, several acoustic vectors corresponding the sounds of the same class are used to create a representative or model characteristic of this class. The representative can be obtained such as centroid vector acoustic characteristics of the grade. A model often will better characterize the distribution of values acoustic vectors for each class (here an emotional class). This representative model or what is traditionally obtained from a database (section4) that has been previously annotated.

3) It was during the classification (section3) phase the acoustic vectors voice signal to be analyzed are compared to models or representatives of each class. After this phase, a probability of belonging to each class can be obtained for each acoustic vector.

4) Finally, the decision phase involves a class to a speech segment by exploiting the probabilities of belonging successive acoustic vectors.

B. Example of added functions

The power of TV will be turned on automatically while the user sits in front of it.

A gentle reminding message will show on TV while the user leans against the sofa or tilts his/her head.

A gentle reminding message will show on TV while the user gets too close to it.

The program will be paused automatically while the user leaves and it will be resumed while he/she comes back.

Automatically volume adjusts.

V. CONCLUSION AND FUTURE WORK

This initial work lays first brick of the study and evaluation of emotions in human behavior for the project EmotiTV. Several projects have focused on the case of motor disabilities, children and the elderly, for which there is no way of distraction, no affective reaction (feedback) and no help to make easy for user to choose TV programs. This led us to propose system architecture of a recommendation of TV programs, following the preferences of the viewer any sort. Then for this study we'll create the emotional Arabic database (Tunisian Emotional databases). Allowing us to study the emotional behavior not yet exploited based on the Arabic language. This work is not completed and the study will be extended and will be another article.

REFERENCES

[1] B. Schuller, R. Zaccarelli, N. Rollet, and L. Devillers. 2010. Cinemo a french spoken language resource for complex emotions: Facts and baselines. In Proceedings of the Seventh International Conference on Language Resources and Evaluation.

[2] S. Steidl. 2009. Automatic classification of emotion related user states in spontaneous children's speech. In Studien zur Mustererkennung, volume 28. Logos Verlag, Berlin.

[3] N. Rollet, A. Delaborde, and L. Devillers. 2009. Protocol cinemo: The use of fiction for collecting motional data in naturalistic controlled oriented context,. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, 2009.

[4] L. Devillers and O. Vidrascu, L. end Layachi. 2009. Automatic detection of emotion from vocal xpression. In A blueprint for an affectively competent agent: Crossfertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing. Oxford University Press, Oxford.

[5] L. Devillers and L. Vidrascu. 2007. Emotion recognition, Speaker characterization. Springer-Verlag.

[6] Alexis BONDU Apprentissage Actif par Modèles Locaux 2008.

[7] Scherer, K. R., Johnstone, T., & Sangsue, J. (1998), L.état émotionnel du locuteur: facteur négligé mais non négligeable pour la technologie de la parole, Actes des XXIIème Journées d'Etudes sur la Parole, 249-257, Matigny, Suisse.Scherer, 2003.[8] Ai H., Litman D.J., Forbes-Riley K., Rotaru M., Tetreaut J., Purandare A. Using system and user performance features to improve emotion detection in spoken tutoring dialogs. Proc.INTERSPEECH' 2006, Pittsburgh, 2006.

[8] Engberg, I.S., Hansen, A. V. Documentaion of the Danish Emotional

[9] Speech Database DES. Aalborg, Denmark, 1996.

[10] Kumar R., Rose C. P., Litman D.J. Identification of confusion and surprise in spoken dialog using prosodic features. Proc. INTERSPEECH' 2006, Pittsburgh, 2006.

[11] Liscombe j., Riccardi G., Hakkani-Tuer. Using context to improve emotion detection in spoken dialog systems. Proc. Interspeech 2005, ISCA, pp 1517-1520, Lisbon, Portugal, 2005.

[12] Montero J. M., Gutierrez-Arriola J., Cordoba R., Enriquez E., Pardo J.M. Spanish emotional speech: towards concatenative synthesis COST 258, 1998.

[13] Neiberg, D., elenius K., Laskowski K. Emotion recognition in spontaneous speech using GMM. Proc. INTERSPEECH'2006, Pittsburgh, 2006.

[14] Schuller B., Reiter S., Mueller R., Al-Hames M., Lang M., Rigoll G.Speaker-independent speech emotion recognition by ensemble classification. Proc. ICME 2005, Amsterdam, Netherlands, 2005.

[15] Vogt, T. Andre, E. Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. Proc. ICME 2005, Amsterdam, Netherlands, 2005.

[16] Wilting J., Krahmer E., Swerts M. Real vs acted emotional speech. Proc. INTERSPEECH'2006, Pittsburgh, 2006.

[17] EmotiRob Sébastien Saint-Aimé Université de Bretagne Sud interaction model. In IEEE ROBIO 2009, International Conference on Robotics and Biomimetics.