

Emotion-Recognition from Speech-based Interaction in AAL Environment

B. De Carolis, S. Ferilli, G. Palestra, D. Redavid

Dipartimento di Informatica, Università di Bari
70126 Bari, Italy
<name.lastname>@uniba.it

Abstract. In Ambient Assisted Living environments assistance and care are delegated to the intelligence embedded in the environment that, in our opinion, should provide not only a task-oriented support but also an interface able to establish a social empathic relation with the user. To this aim social assistive robots are being employed as a mediator interface and, in order to achieve a relation with the user, they should be endowed with the capability of recognizing the user affective state. Since a natural way to interact with a robot is speech, spoken user's input can be used to give to the robot the capability of recognizing the emotions and attitude of the user, thus providing more detail information about the user state. This paper focuses on this topic and proposes an approach based on the dimensional model of emotions in which the valence and arousal of user's spoken input are recognized. The experimental analysis shows the performance in terms of accuracy of the proposed approach on an Italian dataset. In order to show its application in the context of Ambient Assisted Living, an example is provided.

1 Introduction

A Smart Environment should support people in their daily activities by assisting and facilitating users when interacting with environment services in a natural and easy way. The required assistance may be provided to the user through different devices. The choice of an assistive robot agent as an interface is supported by several considerations. First of all, the robot has a physical presence and it may participate in the user's daily life. Assistive robots can move around and perform actions, follow and observe the user in the environment, which is fundamental when designing supportive technologies for elderly people [1]. In addition to typical service-oriented features, assistive robots can be equipped with social and conversational capabilities, thus improving the naturalness and effectiveness of the interaction between users and smart environment services.

Speech is a natural way for humans to interact with robots [2]. Moreover, speech based interaction is seen as an effective interface for smart environments because it is natural, hands-free and it enables different types of users with different capabilities, and disabilities, to interact with systems. Since elderly people are an important user group for smart environments, spoken interaction is of particular benefit for them

since it is natural and does not require particular skills. In addition, spoken user's input can be used not only to issue commands, but also to give to the robot the capability of recognizing the emotions and attitude of the user and this is very important for establishing a social relation and to personalize service execution. Indeed, providing personalized services requires taking into account several factors, which are related to the nature of the service, to user's preferences and to context-related features such as user's emotional state.

In this paper we focus on the latter issue and we will present an acoustic analyzer for the recognition of the emotion. This module is able to extract the prosodic features of user's spoken input and, starting from them is able to recognize the two dimension of emotions: valence and arousal [25]. Then, the module has been used by a social assistive robot embodied in NAO. The robot acts as Interactor Agent in a smart home environment implemented as a Multi Agent System (MAS) [3].

The experimental analysis shows the performance in terms of accuracy of the proposed approach on an Italian dataset. The obtained results also show which combination of features assures a satisfying recognition rate allowing a better understanding of the user's affective state.

The paper is structured as follows. In Section 2 the motivations and technical background for this work are presented. Section 3 briefly describes the MAS architecture implementing the smart home environment. Section 4 describes how VOCE has been developed and Section 5 shows an example on how it can be applied in the context of AAL. Conclusions and future work directions are illustrated in Section 6.

2 Background and Motivations

Interaction with services provided by a smart environment may be provided to the user in a seamless way (i.e. by combining smart home technologies based on sensors and effectors embedded in the appliances of the environment), or using an embodied companion as an interface, or combining both approaches. In all cases, research emphasizes the need of natural and user-friendly interfaces for accessing the services provided by the environment. Moreover, research on social and affective computing suggests that such an assistive environment should provide not only a task-oriented support but also an interface able to establish a social empathic relation with the user.

Several studies report successful results on how social assistive robots can be employed as interface in the assisted living domain. For instance, projects ROBOCARE [4], Nursebot [5], Care-o-bot [6], CompaniAble [7], and Ksera [8] aim at creating assistive intelligent environments for the elderly in which robots offer support to the user at home. However, to be accepted and integrated in the user's daily life, interaction with robots must be spontaneous and natural, and to provide a friendly environment robots must exhibit social capabilities and learn how to react according to the human emotional state. Since speech provides a natural and intuitive way for people to interact with robots, automatic emotional speech recognition will expand the possibilities of interaction.

Emotions are expressed through various communicative signals in humans: facial expressions [9], vocal features [10], body movements and postures [10,11], or a com-

bination of some of them [13,14,15]. In this paper we focus on speech features and how it is possible to use them to recognize emotions in communication with humans.

Recognizing emotions in speech through several features has been a key research issue in robotics, because by recognizing emotional factors the robot can handle social situations. In emotional classification from speech, a multitude of different features denoting prosodic cues have been used. Prosodic features, like pitch, loudness, speaking rate, durations, pause and rhythm were proven to have strong correlations between them, providing emotional information. In the case of the analysis of an entire segment of voice, statistical functions like mean, median, minimum, maximum, standard deviation are applied to the fundamental frequency (F0) base contour [16]. Taking advantage of research work in Music Information Retrieval, Mel Frequency Cepstral Coefficients (MFCCs) are also used with great accuracy in emotion recognition [17]. These features can be used to train a classifier and the learned model can be used to detect emotion in real-time situations.

Several classifiers have been used in this field. Each of them has advantages and disadvantages in order to deal with the speech emotion recognition problem. The more common group includes Hidden Markov Models (HMM) [18, 19] regarded as the simplest dynamic Bayesian networks, artificial neural networks (ANN) [20], support vector machines (SVM) [21], k-NN [22] and Decision Trees [23].

The majority of emotion recognition systems from speech have employed a high-dimensional speech grouped in a big vector of features. In this paper, the most commonly used features in several researches for capturing emotional speech characteristics in time and frequency were selected. The performance of different well known classifiers was compared in order to select the best result to predict the emotion, based on speech emotional data.

3. Overview of the MAS

In [24] we propose an approach based on software agents able to provide what we call Smart Services. A smart service can be defined as an integrated, interoperable and personalized service, accessible through several interfaces available on various devices present in the environment in the optic of pervasive computing.

The objective of the proposed approach is to recognize the users goal starting from percepts (sensors data, user actions, etc.) and provide them with a smart service that integrate elementary services according to the situation. In order to achieve this aim, the environment has to be able to reason on the situation of the user so as to understand which are his/her needs and goals through the composition of the most appropriate smart service. The idea underlying our approach is the metaphor of the butler in grand houses, who can be seen as an household affairs manager with duties of a personal assistant, able to organize the housestaff in order to satisfy the needs of the house inhabitants. To this aim, taking into account the results of a previous project, we have developed a MAS in which the butler agent has to recognize the situation of the user, based on interaction with Sensor Agent, in order to infer possible user's goals. The recognized goals are then used to select the most suitable workflow among a set of available candidates representing a smart service. Such a selection is made by mat-

ching semantically the goals, the current situation features and the effects expected by the execution of the workflow. Once a workflow has been selected, its actions are executed by the effector agents.

One important feature of this architecture is the presence of an agent designed to take care of the interaction with the users. In completely proactive approach, in fact, users may feel a loss of control over the system actions. Therefore we adopt a semi-automatic approach composition of services. The butler proactively propose smart services and leaves, at the same time, the control over proposed service composition to the user to select alternative services, to provide more preference information in order to get a better personalization, to ask for explanation about the proposed services and so on.

The MAS is constituted by the following classes of agents:

- **Sensor Agents (SA)** are used for providing information about context parameters and features (e.g., temperature, light level, humidity, etc.) at a higher abstraction level than sensor data.
- **Butler Agent (BA)** reasons on the user's goals and devises the workflow to satisfy them (see Figure 1).
- **Effector Agents (EA)** each appliance and device is controlled by an EA that reasons on the opportunity of performing an action instead of another in the current context.
- **Interactor Agent (IA)** is in charge of handling interaction with the user in order to carry on communicative tasks. In this case the IA is embodied in the NAO Robot.
- **Housekeeper Agent (HA)** acts as a facilitator since it knows all the agents that are active in the house and also the goal they are able to fulfill.

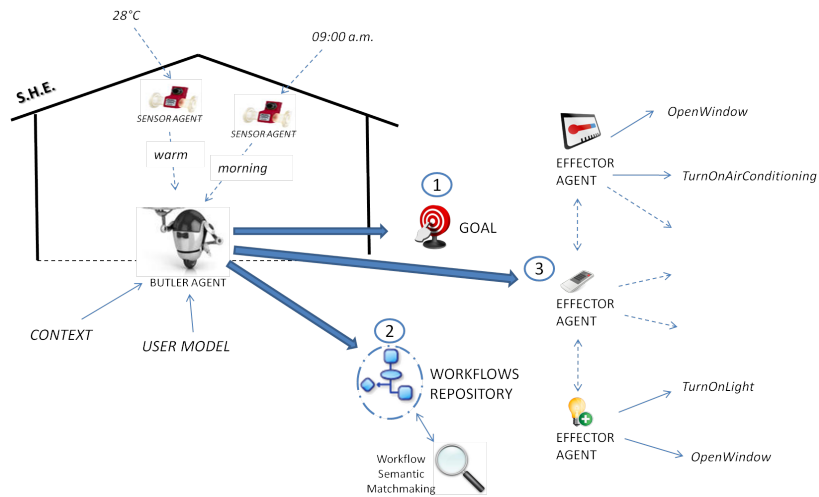


Fig. 1. The MAS architecture

3. VOCE: VOice Classifier of Emotions

Emotions can be classified using two main approaches. Discrete emotion models focus on a defined set of labels denoting emotions (e.g. anger, fear, disgust, happiness, surprise and sadness to name the most common ones). The discrete emotion model has the advantage of clearly distinguishing categories of emotions, however the labels and their number differ a lot from one model to another. By contrast, dimensional models describe the affective space within a limited amount of dimensions. For instance in the circumplex model of emotions [25] only two dimensions are used to represent an emotional state: the valence (from positive to negative or pleasant vs. unpleasant) and the arousal (from high to low or aroused vs. relaxed) dimensions. In contrast to discrete emotions, each emotion can be mapped within this space and this model can be used to determine mixtures of different emotions because they are represented by points in a space.

In VOCE we decided to adopt the dimensional model to classify emotions. Therefore, the analysis of the prosody *user's spoken utterance* is made by two classifiers: one for recognizing the valence dimension and the other for the arousal one.

To this aim we developed a web-service called VOCE 2.0 (VOICE Classifier of Emotions ver. 2.0) that classifies the valence and arousal of the voice prosody with an approach very similar to the one described in [26]. The major steps in speech emotion recognition are audio segmentation, feature extraction and the actual classification of the feature vectors into valence and arousal values.

VOCE can be used in two ways: offline for creating and analysing the emotional speech corpus (Figure 2) and, being a web-service, online for tracking the affect in voice in real-time. While the off-line version allows to build the classifier, the online emotion recognition just outputs the recognised emotions valence and arousal and maps the combination of these values into one of the basic emotions by providing the emotion label.

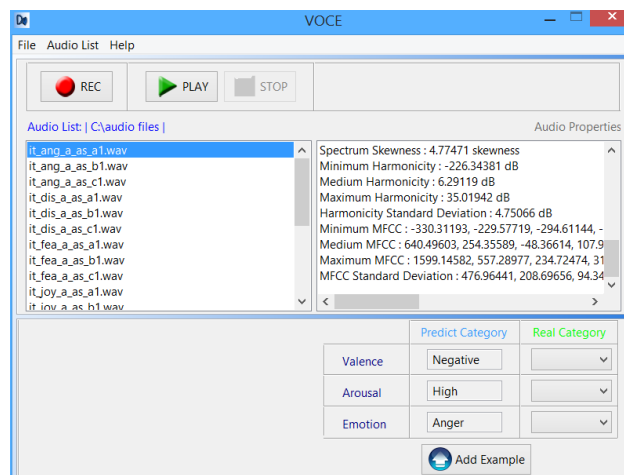


Fig. 2. The interface of the off-line version of VOCE 2.0

Let's see now how the classifier has been trained and how we use it in real-time.

3.1 Dataset

Albeit our approach is based on the dimensional model, since we could not find any corpus, among the few available for Italian, in which emotions were annotated according to their valence and arousal we used the Emotion dataset [27]. Among the three available ones the Emotion dataset has been used for EVALITA ERT challenge¹ and therefore we could compare our results with other research works in this domain.

The emotional speech characteristics were extracted from the Italian subset of Emotion contains 220 audio files corresponding to sentences for the 6 basic emotions (joy, anger, surprise, sadness, disgust, fear) and the neutral one recorded by professional actors. In order to use the dimensional approach on this dataset we mapped each emotion to the correspondent valence (negative, neutral and positive) and arousal (low, medium, high) using the approach explained in [25]. For instance, "anger" is mapped into negative valence and high arousal, while "sadness" is mapped into negative valence and low arousal.

3.1.2 Features extraction and classification

In developing VOCE we exploited different combinations of features and several classification algorithms. For this task we used Praat [28]. In particular, besides pitch and energy related features, we extracted features related to the spectrum, harmonicity and the Mel-Frequency Cepstral Coefficients (MFCCs) that is used to describe a spectrum frame, its first and second derivative in time are used to reflect dynamic changes.

Table 1 shows the features extracted from each audio file using Praat functions.

In order to find the best set of features we tested three conditions with several classification algorithms:

- Support Vector Machines (SVM), which offers robust classification to a very large number of variables and small samples.
- Decision trees, that work with simple classification rules that are easy to understand. The rules represent the information in a tree based on a set of features.
- Artificial Neural Network (ANN), and in particular the Multilayer Perceptron algorithm.
- k -Nearest Neighbors (kNN) is one of the simplest of classification algorithms available for supervised learning. It classifies unlabeled examples based on their similarity with examples in the training set.

The three set of features were:

- **ALL**: all the attributes in Table 1;
- **MFCC**: MFCC features only;
- **No MFCC**: all the features except MFCC.

¹ <http://www.evalita.it/2014/tasks/emotion>

Feature	Description
Pitch	
PitchMin	Minimum value
PitchMed	Average value
PitchMax	Maximum value
PitchMinMaxDiffLog	Logarithmic differentiation
PitchMinLog	Minimum Logarithmic
PitchMedLog	Average logarithmic
PitchMaxLog	Maximum logarithmic
PitchDevSta	Standard Deviation
PitchSlope	Slope
Energy	
EnergyMin	Minimum value
EnergyMed	Average value
EnergyMax	Maximum value
EnergyMinMaxDiff	Logarithmic differentiation
EnergyDevSta	Standard Deviation
Spectrum	
SpectrumCentralMoment	Central moment
SpectrumDevSta	Standard Deviation
SpectrumGravityCentre	Central tendency
SpectrumKurtosis	Degree of centralization
SpectrumSkewness	Degree of asymmetry
Harmonicity	
HarmonicityMin	Minimum value
HarmonicityMed	Average value
HarmonicityMax	Maximum value
HarmonicityDevSta	Standard Deviation
MFCC	
MFCCnMin	Minimum of nth MFCC
MFCCnMed	Average of nth MFCC
MFCCnMax	Maximum of nth MFCC
MFCCnDevSta	Standard Deviation of nth MFCC

Table 1. The set of features extracted from speech for emotion recognition

From the analysis of the performance of the most commonly used algorithms for classification starting from numeric features the most accurate one were MLP (Multi-Layer Perceptron) and SMO (Sequential Minimal Optimization) algorithm for training a support vector classifier in Weka².

The accuracy was validated using a *10 Fold Cross Validation* technique. A k-fold cross-validation with k = 10 was used to make validations over the classifiers. This technique allowed the evaluation of the model facing an unknown dataset. Results of the classification of valence, arousal and derived emotion labels are shown in Table 2.

Results show that, for both algorithms, using the complete set of features improves accuracy, however using only MFCC related features we get an accuracy comparable

² <http://www.cs.waikato.ac.nz/ml/weka/>

with the one obtained using the all set of features. The worst setting is when MFCC is not considered. As far as the choice of algorithm is concerned, even if MLP had a slight better accuracy, the time to create the model and classify a vocal input is higher (100:1). Since VOCE has to be employed in real-time classification tasks we selected SMO.

Since the arousal dimension is related to the importance of the goal and the valence dimension is related to the achievement vs. the threatening of the goal, our speech classifier performs well in recognizing negative states, like those related to anger, and allows us to distinguish positive from negative attitudes. However, as expected, some emotions are easier to recognize than others. For example, humans are much better at recognizing anger than happiness; therefore, our results can be considered acceptable under this view.

Features	MLP	SMO
ALL		
Valence	70,45	71,36
Arousal	80,90	77,27
Emotion	71,36	68,63
MFCC		
Valence	69,09	64,54
Arousal	80,00	75,00
Emotion	70,45	68,18
No_MFCC		
Valence	64,09	55,45
Arousal	74,09	69,09
Emotion	53,18	53,18

Table 2. Accuracy of the two classifiers for valence arousal and derived emotion labels.

Comparing our results with other works based on the same dataset [29] we can say that our approach has a comparable accuracy over the set of basic emotions.

4 An Example of Application in the Context of AAL

VOCE has been used in real-time as a web service with the NAO Robot for enabling emotion recognition during speech-based interaction (see Figure 3).

We have designed this architecture to endow the Aldebaran NAO robot with this capability. The system is composed by two fundamental units: the NAO humanoid robot and the workstation connected with NAO robot. Audio files in wav format, recorded from 4 microphones located in the head of the NAO, are collected by the Application Programming Interface (API) provided with NAO Software Development Kit (SDK). Captured audio files are sent to the Speech-based Interface module in order to allow the understanding of vocal commands and to recognize the valence and arousal. An Automatic Speech Recognition Module (ASR) performs the first task

whereas the second task is accomplished by Voice Classifier for Emotions (VOCE) Module. Then, the results are sent to Behavior Decision Module that choose the appropriate behavior and send it to the robot to be executed. Communication between the robot and the workstation has been performed using the NAOqi API.

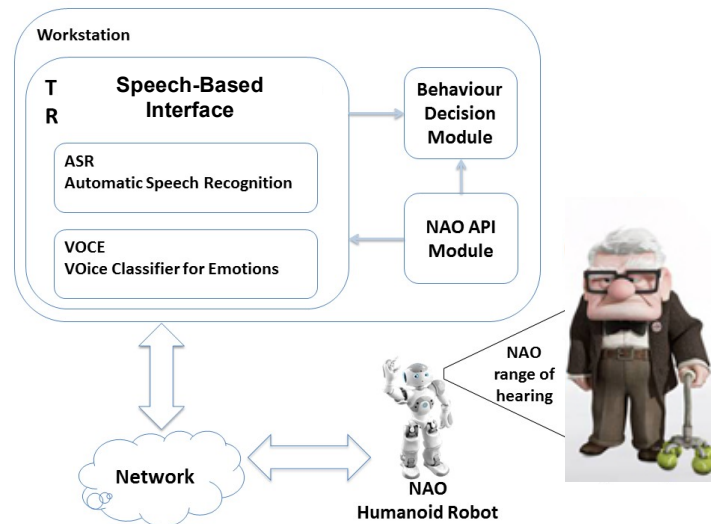


Fig. 3. Overview of the proposed system

Let us now provide an example of how the proposed approach can be integrated in an ambient assisted living environment.

The scenario depicts a situation in which an old man lives in a house equipped with some sensors (to gather data about the house situation) and some effectors (to control appliances in the environments). Moreover the house is equipped with the NAO robot acting as a natural and social interface between the user and the smart home environment, by implementing an easy conversational access to the (digital or physical) services of the environment.

It's friday evening and Nicola, a 73 y.o. man, is at home alone. He has an appointment with his friends downtown to play cards like he does almost every friday. Nicola is sitting on the bench in the living room that is equipped with sensors and effectors for controlling appliances in the room and with the social robot that acts as a mediator interface between the environment services and the user (see Figure 4). Nicola received a message saying that his daughter cannot accompany him downtown and this makes him a bit angry. Nicola calls NAO to try to find a solution.



Fig. 4 A simulation of the scenario

In the following we provide an example of the interaction.

R: 'Hi Nicola, what can I do for you?'

Nicola: 'Damn ... (with high arousal and negative valence) I need to go downtown to play cards with my friends and my daughter cannot come to bring me there ...I cannot miss it tonight there is a tournament !' (with high arousal and negative valence).

R: 'Don't be angry for this ... we will find a solution' Do you want me to call your daughter to ask for the permission to call a taxi to bring you there and take you back at a certain time?''

Nicola: 'Yes ... but you know I need a bit of assistance in walking from the car to the bar' (with medium arousal and negative valence).

Nicola: I will not play card with my friends tonight, I feel so lonely. (with low arousal and negative valence).

R: 'Oh, I'm sorry to hear that you are sad. I will ask the taxi driver to help you. OK?'

Nicola: OK.

R: the robot send a message to the daughter that accepts and then calls the taxi.

In this scenario initially the voice classifier recognizes a *negative* valence with a *high* arousal from the prosody of the spoken utterance (Figure 2). This is interpreted as anger and the robot besides expressing empathy (it show to understand the user's feelings) asks the reasons for it. When the robot understands that the user's goal is to go downtown it finds a workflow satisfying this goal by matching the constraints of the situation (the daughter is not available). According to the selected workflow the dialog goes on in order to get some information that are necessary (preconditions) to

execute some of its step (like the permission to call the taxi given from Nicola's daughter).

5 Conclusions and Future Work

We presented a preliminary work towards implementing a speech-based interface between a Social Robot and a user for handling interaction in a smart environment. In our opinion, besides assisting the elderly user in performing tasks, the robot has to establish a social long-term relationship with the user so as to enforce trust and confidence. The underlying idea of our work, in fact, is that the analysis of user's spoken utterances can be used for both issuing requests to the robot and understanding his emotional state, which is important when the interaction happens in everyday life environments. To this aim we developed VOCE 2.0 a module to classify emotions from features extracted from the speech signal according to the dimensional model (valence and arousal). The recognition accuracy results are comparable with other research works in which the same dataset was used [29]. We are aware that an improvement is needed and to this aim we plan to collect a new dataset possibly with example of elderly voices, which may have different range of features from those used in the Emotion dataset. Information on user's emotions coupled to context and activity recognition may give the robot the capability to infer which is or will be the most probable user's mood in a given context.

References

1. S. Thrun, Towards a framework for human-robot interaction, *Human Computer Interaction*, 19 (1&2), pp. 9-24, 2004.
2. Drygajlo, P.J. Prodanov, G. Ramel G., M. Meisser, and R. Siegwart, "On developing a voice-enabled interface for interactive tour-guide robots". *Journal of Advanced Robotics*, vol.17, nr. 7,p.p. 599-616, 2003.
3. B. De Carolis , G. Cozzolongo , S. Pizzutilo , V. L. Plantamura, Agent-Based home simulation and control, Proceedings of the 15th international conference on Foundations of Intelligent Systems, May 25-28, 2005, Saratoga Springs, NY
4. Cesta, G. Cortellessa, F. Pecora and R. Rasconi, Supporting Interaction in the RoboCare Intelligent Assistive Environment, AAAI 2007 Spring Symposium, 2007.
5. J. Pineau, M. Montemerlo, M. Pollack, N. Roy and S. Thrun, Towards Robotic Assistants in Nursing Homes: Challenges and Results, *Robotics and Autonomous Systems* 42(3-4), pp. 271-281, 2003.
6. Graf B, Hans M, Schraft RD (2004) Care-O-bot II – development of a next generation robotics home assistant. *Auton. Robots* 16, 193-205.
7. CompanionAble project (2011) <http://www.companionable.net>
8. ksera.ieis.tue.nl
9. M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, 2000.

10. R. Cowie and E. Douglas-Cowie, "Automatic statistical analysis of the signal and prosodic signs of emotion in speech," In Proc. International Conf. on Spoken Language Processing, pp. 1989–1992, 1996. [
11. N. Bianchi-Berthouze and A. Kleinsmith, "A categorical approach to affective gesture recognition," *Connection Science*, vol. 15, no. 4, pp. 259–269, 2003.
12. G. Castellano, S.D. Villalba and A. Camurri, "Recognising Human Emotions from Body Movement and Gesture Dynamics," In Proc. of 2nd International Conference on Affective Computing and Intelligent Interaction, Berlin, Heidelberg, 2007.
13. H. K. M. Meeren, C. van Heijnsbergen and B. de Gelder, "Rapid perceptual integration of facial expression and emotional body language," *Proc. National Academy of Sciences of the USA*, vol. 102, no. 45, pp. 16518–16523, 2005. [
14. A. Metallinou, A. Katsamanis and S. Narayanan, "Tracking changes in continuous emotion states using body language and prosodic cues," In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2288–2291, 2011.
15. C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzaeh, S. Lee, U. Neumann and S. Narayanan, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal information," In Proc. of ACM 6th int'l Conf. on Multimodal Interfaces (ICMI2004), State College, PA, pp. 205–211, 2004.
16. D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech Communication*, pp. 1162–1181, 2006
17. B. Bogert, M. Healy, and J. Tukey, "The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe-cracking," *Proceedings of the Symposium on Time Series Analysis*, Wiley, 1963.
18. D. Le and E. M. Provost, "Emotion recognition from spontaneous speech using Hidden Markov models with deep belief networks," in *Automatic Speech Recognition and Understanding (ASRU)*, 2013 IEEE Workshop on, pp. 216–221, 2013.
19. J. Wagner, T. Vogt, and E. Andre, "A systematic comparison of different HMM designs for emotion recognition from acted and spontaneous speech," in *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII)*, Lisbon, Portugal, pp. 114–125, 2007.
20. S.A. Firoz, S.A. Raj and A.P. Babu, "Automatic Emotion Recognition from Speech Using Artificial Neural Networks with Gender-Dependent Databases," in *Advances in Computing, Control and Telecommunication Technologies, ACT '09*, pp. 162–164, 2009.
21. C. Yu, Q. Tian, F. Cheng and S. Zhang, "Speech Emotion Recognition Using Support Vector Machines," in *Advanced Research on Computer Science and Information Engineering*, vol. 152, G. Shen and X. Huang, Eds., ed: Springer Berlin Heidelberg, pp. 215–220, 2011.
22. M. Feraru and M. Zbancioc, "Speech emotion recognition for SROL database using weighted KNN algorithm," in *Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1–4, 2013.
23. C.-C. Lee, E. Mower, C. Busso, S. Lee and S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," *Speech Commun*, vol. 53, pp. 1162–1171, 2011.
24. B. De Carolis and S. Ferilli. A multiagent system providing situation-aware services in a smart environment. *Workshop on Ambient Intelligence Infrastructures (WAmI)*. November 13, 2012, Pisa, Italy. In conjunction with International Joint Conference on Ambient Intelligence (AmI 2012).
25. Russell, James (1980). "A circumplex model of affect". *Journal of Personality and Social Psychology* 39: 1161–1178.

26. W.E. Bosma and E. André, "Exploiting Emotions to Disambiguate Dialogue Acts", in Proc. 2004 Conference on Intelligent User Interfaces, January 13 2004, N.J. Nunes and C. Rich (eds), Funchal, Portugal, pp. 85-92, 2004.
27. Vincenzo Galata'. 2010. Production and perception of vocal emotions: a cross-linguistic and cross-cultural study. Ph.D. thesis, University of Calabria
28. www.praat.com
29. Antonio Origlia, Vincenzo Galatà e Bogdan Ludusan. Automatic classification of emotions via global and local prosodic features on a multilingual emotional database. In: Speech Prosody. 2010.