Proceedings of the IASTED International Conference
Robotics (Robo 2011)
November 7 - 9, 2011 Pittsburgh, USA

SYNTHESIS OF EMOTIONS ON A *HUMAN-ROBOT-INTERACTIVE* PLATFORM

José Augusto Soares Prado
Institute of Systems and Robotics
University of Coimbra
Coimbra, Portugal
email: jaugusto@isr.uc.pt

Lakmal D. Seneviratne(*)
Division of Engineering
King's College
London, UK
email: lakmal.seneviratne@kcl.ac.uk

Jorge Manuel Miranda Dias(*)
Institute of Systems and Robotics
University of Coimbra
Coimbra, Portugal
email: jorge@deec.uc.pt

ABSTRACT
This work explores our developments over a framework and over a platform for human-robot-interaction with emotions. The focus of our framework is on visuo-auditory perception and response. In other words, perception and response can be called analysis and synthesis; the analysis is responsible for the classification of human emotion and the synthesis is responsible for the synthetic expression that the robot must show. This paper is focused on the synthesis and also on how the synthesis can affect human engagement during an interactive conversation with the robot.

KEY WORDS
Emotion Synthesis, Bayesian Approach, Human-Robot Interaction.

1 Introduction

The development of robots that interact with humans implies better interfaces between humans and robots, and capabilities to detect and express emotions. Current studies are mainly focused on telling how to classify an emotion and how to synthesize an artificial emotion. It is common in this area to associate emotions with facial expressions and also with voice features. Most state-of-the-art in the area of emotive robots do not run in real-time, being then still inapplicable to real cases of human-robot-interaction applications. This paper builds on the work of [19, 15] where real-time Bayesian classifiers were presented for visual signal and to auditory signal; independently and respectively. Both of these classifiers give output among the scope *{happy, sad, fear, neutral, anger}*. This classifiers solve the analysis part, however when the robot responds to the human there is the synthesis problem which will be detailed on this paper.

Moreover, according to [4], the presence of the emotions are personality traits which characterize a Social Behavior Profile(SBP). This assertion is valid in both ways: when analyzing the human behavior, and when synthesizing an artificial SBP. Our focus on this paper is to explore how the robot's SBP can influence on the engagement of the human during a dialog with a machine.

The authors gratefully acknowledge support from Institute of Systems and Robotics at University of Coimbra (ISR-UC), from Portuguese Foundation for Science and Technology (FCT). (*) on sabbatical leave at Khalifa University of Science, Technology and Research (KUSTAR), Abu Dhabi, UAE

2 Related Work

2.1 Proposed Approaches to Automatic Emotion Recognition from Audio Signal

Although the field of spoken language processing had recently significant advances [9],[11],[1]; the affect recognition, or in other words, emotional speech area has not progressed that much. Since our focus is to improve the interaction between human and machine by exploring the non-verbal cues, namely facial and vocal expressions; we realized that there is still a lack of satisfactory solutions to be proposed and experimented in these particular fields.

It is known that the problem of vocal expressions analysis include two sub-problem areas: from the input audio it is necessary to specify which signal features to be extracted, from the extracted data it is necessary to classify into some emotion categories. Progresses were made in [23] and [10] however they claim to have a high percentage of accuracy, they are not real-time. In this paper we are using our own model previously proposed on [15] witch is a real time system for auditory emotional classification, though it is well suitable for our purposes.

2.2 Proposed Approaches to Automatic Emotion Recognition from Face Images

Automatic emotion recognition from images clearly includes three sub-problems: finding faces, detecting features, and classification. Many of the current systems assumes the presence of a face in the scene and do not automatically find faces [12],[25]. However, for example in [13, 10] a camera is fixed pointing to the human face, so they do not really need to find faces. In HRI area, the camera is always on the robot and not on the human, most systems assumes good illumination, a clean background and usually they do not provide any automatic or even manual tool to deal with illumination problems. Several improvements have been done in the area of detecting faces [26][24]. In our case for finding faces we are using the

DOI: 10.2316/P.2011.752-065

OpenCV haarlike features [24] this method is well known as being independent of illumination problems. For detecting features, still many of the current approaches do not automatically extract the features, do not consider time sequence frames, and it is common to divide the image in parts instead of analyzing the whole face image at once. A step forward was done in [18] and [13, 10] were the features were detected automatically.

About the classification techniques, we found on literature: template-based classification [25], fuzzy classification, ANN based classification [12], HMM based classification and also Bayesian classification [3][17] and [18].

Ekman et al. [5] devoted a lot of attention to the specific subject of emotional states and facial expressions.

2.3 Emotive Robots

Most state-of-the-art in the area of emotive robots do not run in real-time, being then still inapplicable to real cases of human-robot-interaction applications. However recent researches like [20] and [15] shows result about two Bayesian classifiers inside a structure for human-robot-interaction that are applicable to HRI in real-time. These classifiers have both the purpose to classify human emotional state among the scope {anger,fear,sad,neutral and happy}. The first difference between them is clearly the channel of communication, in other words, one of them uses the auditory channel (listen to human voice) and the other uses the visual channel (camera looking to the human face). The second difference is of course the model, since the channels are different, the detection and also the variables are diverse. Thus, each classifier analysis it's respective input signal during a certain period of time and gives a result among the mentioned scope. The synthesis of the facial and vocal expressions is done by mirroring the Bayesian network and from the desired emotion is possible to reach again to the leaves of the Bayesian network.

Therefore, mankind is capable of building robots capable of recognizing and synthesizing emotions, but what to do with these emotions? What is the advantage of having them during a dialog between human and robot? How to measure this advantage? What happens if the *social behavior profile* of the robot varies? Answering these questions was our main motivation in this article.

In our context, Social Behavior Profile (SBP) is the variable which defines the personality of the robot. During [6], an attempt for a social behavior learners analysis during conversations mediated by computers was presented. Four social behavior profiles were analyzed in this work: moderator, valuator, seeker, interdependent. In medicine, for example on [7], we found autism, apathetic and aggressiveness as social behavior profiles. Several other social behavior profiles are listed among the literature, they may vary a lot depending on the author's interpretation and of the context of each problem.

This paper builds on [19, 15] where the used classifiers were defined; and also builds on [4] where the five-factor model personality traits was defined. Based on this mentioned model, we reinterpret the "openness to experience" SBP presented on [4] as Humorous personality. Thus, the selected SBP scope for this work are: *{Neuroticism, Extraversion, Conscientiousness, Agreeableness and Humorous}*.

3 Our approach

Our system structure is composed by two detectors (sensory processing for audio and video), two classifiers (Bayesian classifiers for audio and video), one decision process (Bayesian inference that merges the output of the classifiers). This structure is presented on figure 1. For the detection of features and classification of facial expressions we are using our previously proposed methods from [18]. For the classification of vocal expressions we are using a Bayesian model previously proposed by us in [15]. Both classifiers give a confidence score for emotions among the scope of *{Anger, Fear, Sad, Neutral, Happy}*. Figure 1 shows our structure schematic and in further sections we are going to explain how we did the BMM fusion with social behavior profile.

3.1 Bayesian Mixture Model

The Bayesian mixture model (BMM) technique was used from green level 2 to green level 1 of the Bayesian network presented on figure 3. To be coherent and concise, we present the description of our variables in table 1.

Thus, the mixture is computed by the following equation:

$$P(HE, FE_w, VE_w) = P(HE, FE_w|VE_w).P(HE) =$$

$$P(HE|FE).fw.P(VE|HE).vw.P(HE) \quad (1)$$

last equation is only valid when it is assumed that the variables FE and VE are independent.

The *posterior* can be obtained from the joint distribution, using the Bayes Formula as follow:

$$P(HE|VE_w, FE_w) = \frac{P(VE|HE).vw.P(FE|HE).fw\,.\,P(HE)}{P(VE_w,\, FE_w)} \quad (2)$$

from the summation theorem we can calculate

$$P(VE_w,\, FE_w) = P(VE|HE).vw.P(FE|HE).fw\,.\,P(HE) + P(VE|\sim HE).vw.P(FE|\sim HE).fw\,.\,P(\sim HE) \quad (3)$$

After the BMM takes place, we have the *detected emotion* represented by variable HE, which stands for Human Emotion and vary among the scope *{Anger, Fear, Sad, Happy, Neutral}*. Next step is to compute the probability of response given the human emotion and the robotic SBP that

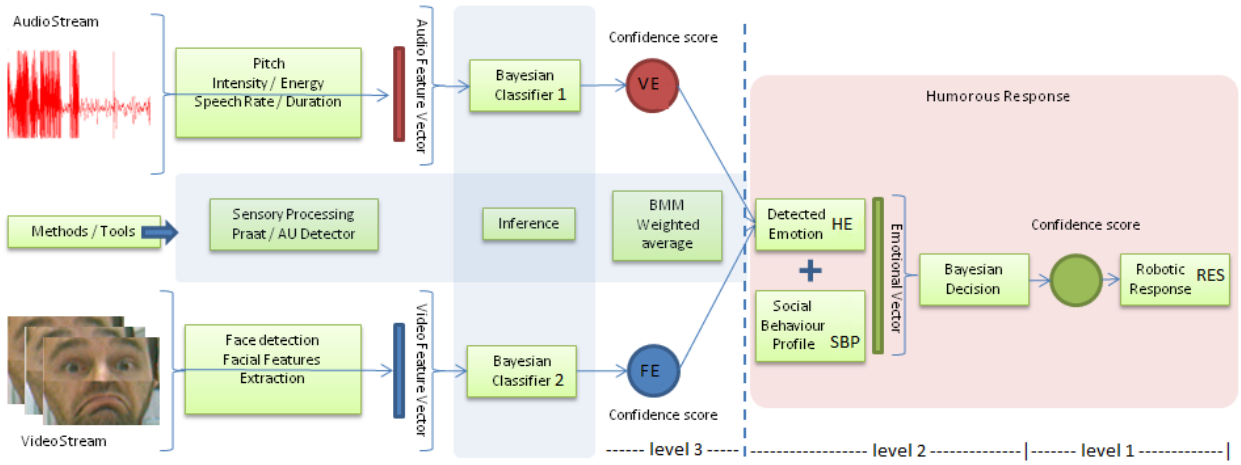


Figure 1: Our system structure: Notice that two modalities signals are processed and then classified, Bayesian Classifier 1 was proposed in [15] and Bayesian Classifier 2 was proposed in [18]. Then, the fusion proposed in this paper takes place to reach to a *fused detected emotion*. Later, by adding a given *social behavior profile* it is inferred the correct response for the humorous profile according to previous learning.

Variable	Description
FE	Stands for Facial Expression, it is a *random variable* among the scope *{anger, sad, fear, happy, neutral};*
VE	Stands for Vocal Expression, it is a *random variable* among the scope *{anger, sad, fear, happy, neutral};*
HE	Stands for Human Emotion, it is a *random variable* among the scope *{anger, sad, fear, happy, neutral};*
SBP	Stands for Social Behavior Profile, it is a *random variable* among the scope *{Sympathetic, Antipathetic, Humorous}*
RES	Stands for Robot Response, it is a *random variable* among the scope *{r1, r2 ... , rn};*
fw	Stands for of *Facial* expressions *Weight*, it is given by the level of confidence assumed for that classifier. $FE_w = P(FE).fw$
vw	Stands for of *Vocal* expressions *Weight*, it is given by the level of confidence assumed for that classifier. $VE_w = P(VE).vw$

Table 1: Description of all the variables used on our Bayesian equations from the input of classified expressions, through the Bayesian Mixture Model for fusion, to the process of decision of robot's response.

vary among the scope *{*Sympathetic, Antipathetic and Humorous*}*. The response (RES) vary among a large scope that contains all possible answers on the database *{r1,r2, ... , rn}*, where n stands for the number of responses on the database. The following equations are for the top of the network (level 1 and level 2), it illustrates the joint distribution associated to the Bayesian Fusion with the social behavior profile:

$$P(RES, H\_E, SBP) = P(HE, SBP|RES).P(RES) =$$

$$P(HE|RES).P(SBP|RES).P(RES) \quad (4)$$

last equation is only valid when it is assumed that the variablesHE, and SBP are independent.

The *posterior* can be obtained from the joint distribution, using the Bayes Formula as follow:

$$P(RES|HE, SBP) = \frac{P(HE|RES).P(SBP|RES).P(RES)}{P(HE, SBP)} \quad (5)$$

from the summation theorem we can calculate

$$P(HE, SBP) = P(HE|RES).P(SBP|RES).P(RES) + P(HE| \sim RES).P(SBP| \sim RES).P(\sim RES) \quad (6)$$

The result of the Bayesian inference is a probability vector with all the probabilities for all the possible responses. Usually what is done is to select the best, and we do no different for the previous presented Bayesian network, we do select the best, more probable result. However, in the very case of this last Bayesian network, not only the best will be considered, a random number generator will help us to add some randomness in the process. At the end, the true response will be given by a random choice

among the 3 most probable responses contained in the result vector RES after the Bayesian inference. We decided to add this randomness in order to increase the *autonomy* and decrease the *predictability* of the robot's responses by the part of the user.

3.2 Relation between SBPs and emotions

Each SBP is related with a way of expressiveness, therefore, there is a strong relationship between the SBP of the robot and which emotion it will express. This relationship is not a direct association, however the robot uses the emotive expressiveness to behave in a certain way that fits the current SBP. In the list below we present the SBP scope and it's relationship with the expression.

- The "Neuroticism" SBP is characterized by being prone to experience feelings that are upsetting; thus the robot express *anger* or *fear*, context dependent.
- The "*Extraversion*" SBP is outgoing, high-spirited, prefer to be around people most of the time; therefore it is associated with the *happy* expression mostlly.
- The "*Conscientiousness*" is well organized, have high standards and always strive to achiece the goals, it is cold, strait forward and thus it is related with the *neutral* expression. This social behavior profile is where robots without emotions fits.
- The *"Agreeableness"* SBP is characterized by good-natured and eager to cooperate and avoid conflict; therefore, all the 10 variables (video plus sound) will be imitated from what was interpreted from the human. The agreebleness SBP is basically *human-imitation* over the two channels, together with the agreebleness phrase defined on the task dependent context (see 3.3).
- The "*Humorous*" SBP is very imaginative and willing to consider new ways of doing things, it is funny and may perform jokes to cheer up the interlocutor. Therefore it is always related to *happy*.

In our case, each emotion that the robot expresses includes the multi-modal channels that we are using, and additionally a pre-defined context of conversation as presented on subsection 3.3.

According to [19], the best Bayesian model for facial expressions encompasses a video feature vector of 7 variables which are: *Eye-Brows, Cheeks, Lower Eyelids, Lips Corners, Chin Boss, Mouth's Form and Mouth's Aperture.* We are not going to detail about these variables in this paper. However, to avoid subjectiveness, it is necessary to understand that when we talk about the robot expressing an emotion, we mean that robot synthesizes these 7 variables according to the learned histogram (likelihoods) presented on [19]. And beyond only facial expressions, robot also synthesizes (during speech of the robot's response) 3 of the variables defined on the audio feature vector of [15], namely *Pitch, Energy (Volume Level) and Speech-Rate* (*Sentence Duration*), according to the learned histogram (likelihoods) presented model.

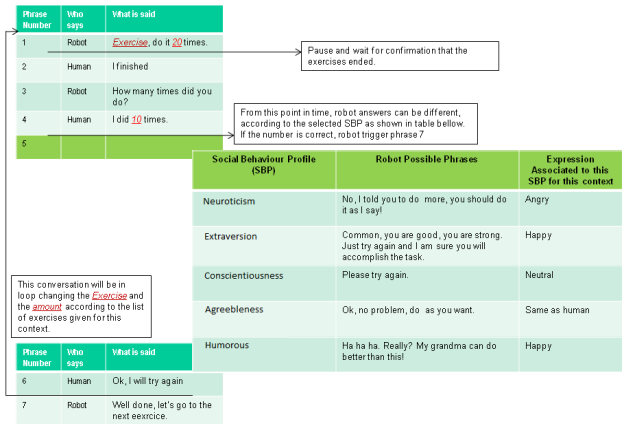


Figure 2: Stimulating exercises — In this study case, the robot is an assistive robot that gives instructions during physiotherapy exercises, each utterance will repeat until the patient confirms that he did the proper amount of exercises.

3.3 Task Dependent Context

The responses *{r1,r2, ... , rn}* are defined by task, in this work we focused on a study case where the robot is an assistive robot to give moral support and instructions during physiotherapy exercises. The robot will thus follow the dialog defined on figure 2. At each cycle, the exercise and the amount of times change according to what the physiotherapist had previously defined (see figure 2).

To be succinct, at each utterance the robot classifies the expression of the human, than it behaves according to the selected SBP, and later the effects of this behavior will be measured according to our proposed assessments presented in section 5.

4 Synthesis

After the Bayesian Mixture Model, response (RES) is already selected. Thus at this point robot knows that emotion it shall present, however, how to present the desired emotion? This is done by the synthesis approach.

4.1 Computational Synthesis

The synthesis can be divided in two layers, computational synthesis and physical synthesis. The first is a computational model where is possible to start from RES (figure 3 pink level 1) and reach to the action *leaf variables* (figure 3 pink level 3). The *leaf variables* are those who belong to the tip of the Bayesian network; these variables are the same that serves as input on the analysis part (figure 3 green level 3 "stimulus"). In the visual channel, our *leaf variables* are all the Action Units. In the auditory channel, the leaf variables are *Pitch*, *Sentence Duration* and *Energy*.

During the classification, as mentioned in [15] and [18], a likelihood table is filled out

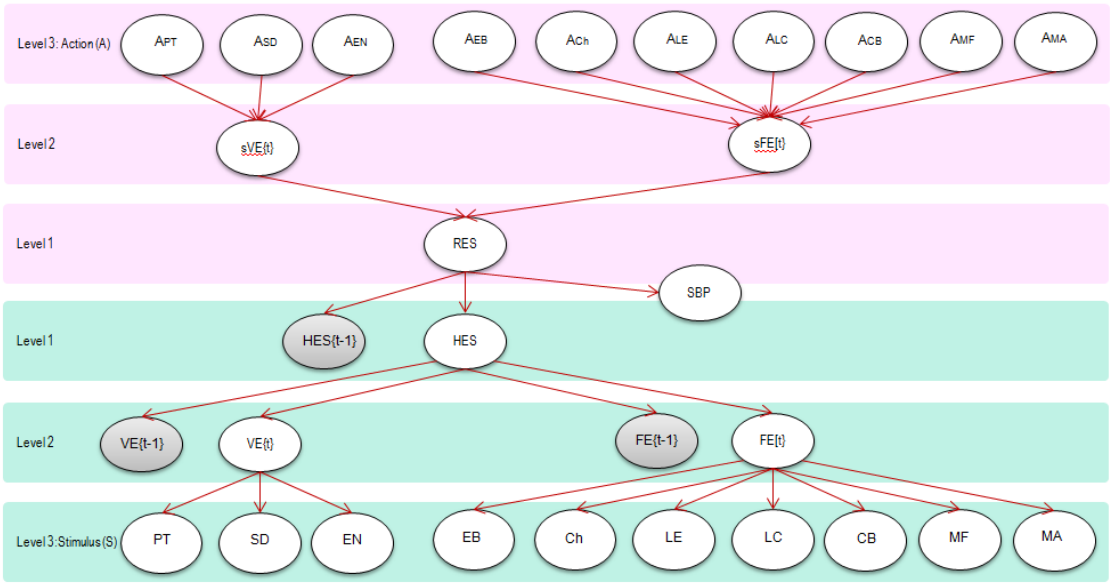


Figure 3: Analysis and synthesis of emotions: analysis starts at green level 3 (bottom-est level in picture) where the 10 *leaf variables* appear. Between the green level 2 and green level 1 the BMM takes place. After RES is defined, the inverse model works based on the learning tables to infer the action, with the leaf variables that are going to be presented by the robot.

with the probabilities of leaf variables for a specific expression. Thus, $P(PT, SD, EN|VE)$ and $P(EB, Ch, LE, LC, CB, MF, MA|FE)$ are stored on the learned table. During the synthesis, this is exactly what is necessary to revert the process. While during analysis a Bayesian inference is needed to reach the correct classification, during the synthesis the likelihood determines directly it's output. In another words, the robot will perform an expression according to what it learned that expression is.

4.2 Physical Synthesis

Our platform was designed in a Scout platform and a head with 2 degrees of freedom and a support for a screen was added to show the expressions. Furthermore a retro-projectable mask was built for a better interactive interface (see figure 4 and 5a).

In both cases the robotic technology used as the experimental platform has an active vision system. This feature allows the robot to move its head towards the tracked face before starting to move its body, thus, the robot will avoid unnecessary movements with the body structure.

The structure in the back of the head allows a simple calibration of the projection. This calibration is possible due to the adjustable distances and mirror angles. The possible adjusts are:

1. distance between projector and first mirror,
2. distance between the two mirrors,
3. angle of the first mirror
4. angle of the second mirror

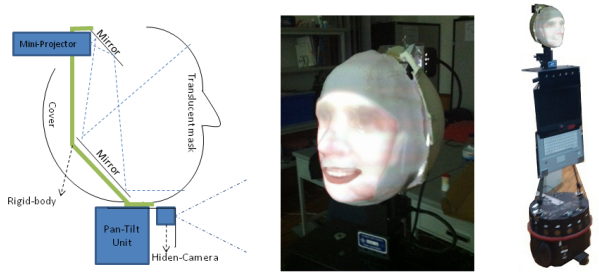


Figure 4: The head is composed by a retro-projected translucent mask which is attached to a rigid body. Two mirrors are attached to this same structure in order to deviate the beam of light projected. This setup was conceived to reduce the necessary distance for projection and thus close the head in a form more close to what would be a humanoid head.

5. zoom screen is done with Linux built in desktop zoom
6. focus of the projector

These six parameter are adjusted in the beginning of the projection setup until the eyes and the nose fits to the correct place on the mask. This calibration is done manually and the error is visible by the displacement of the projection size, position or angle over the mask.

Furthermore, as another platform; we developed a 3D virtual world as a "Blender game", where the same core of interactions can be used both over the real robot and/or inside the virtual world; see figure 5b. We generated 14 meshes of heads from 14 persons of our lab, so that we use the face of the real person on the avatar that mimics the person. Stereo vision systems are also an option that we consider. According to [14], the segmentation can be used

(a) (b)

Figure 5: a)One version of our robot: Scout based platform and a head with 2 degrees of freedom.
b)Our virtual world is another option of interaction instead of the robot. A real person can look to the camera, and speak at the microphone; where an avatar mimics this person and the other avatar simulates the robot.

to improve the selection of which user to interact with.

4.2.1 Lips Synchronization

To improve the user-friendly interface we added the capability of lips synchronization on the robotic responses.Since the response phrases comes from a finite database, it was possible to prepare each sentence to be spoken with lips synchronization by using nine visemes that can be seen in figure 6. These nine visemes are associated to nine phonemes and they are used according to what the avatar will speak. Since we are not doing phoneme recognition, this lips synchronization is only possible on the avatar responses where the phrases are known and not on the avatar which is mimetizing the human.

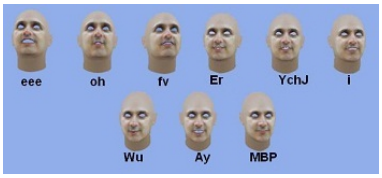


Figure 6: Nine visemes, each viseme is associated to a phoneme. In this figure, from top-left corner to bottom-right corner, the associated phonemes are: eee, oh, fv, er, YchJ, i, Wu, Ay and MBP.

5 Engagement Assessment

It is difficult to measure how the interaction is improving. On literature, [21, 22] claims that a interaction is better when the person consider the system to be funny. Specially those from the European project called "Hahacronym", we found descriptions of results but no detailed descriptions of assessments. However it is understandable that they performed experiment with several persons, while an external agent do a manual classification of how happy was the person with the performance of that system. In [16] description of assessments are more clear where the system was shown to children and what was consider as a joke was also manually measured (by questionnaires after the dialog), they follow an assessment protocol for measure the "jokiness" of each response proposed on [2]. Previously on [2] it was measured the mean of "jokiness", "funniness" and also "heard before" possible classifications for each text, according to their defined assessments. The "jokiness" could be scored from 0 to 1. For "funniness" the range was defined from 1 to 5. For "heard before" the range of score was from 0 to 1.

5.1 Our proposed assessments

Considering the state of the art, there is no common benchmark for this type of system. However, there are existent ideas for assessments that we reinterpret, by defining our own assessments.

The desired measurements must be related to the engagement of the human during the conversation. Touchless interfaces are one of our constraints, though we selected the variables listed bellow:

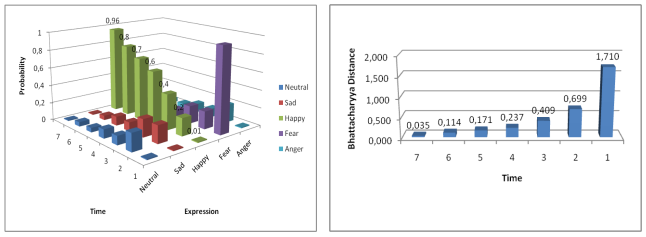
1. Time between phrases (TBP): This variable is measured in seconds and it is annotated along the time.
2. Total Dialog Time (TT): This variable is measured in seconds and it is annotated after the entire dialog.
3. The amount of Happiness (AH): This variable is an integer and represents the amount of times that human expressed happiness during the dialog.

The TBP, TT and AH are annotated during the experiments by an external agent that observes the dialog. This external agent is called engagemeter and currently the annotations for TT and TBP are done manually based on a recorded video of the dialog. The AH and the Error (see 5.1.1 for Error) are automatically calculated. We expect in near feature to have a fully automatic engagemeter.

5.1.1 Controlling Emotion Feedback

It is expected that the robot presents an emotion according to the given SBP (as shown in field "Expression associated to this SBP for this context" of figure 2), thus it is needed a measure to know if this emotion was expressed correctly. Thus, we state error (E) as being the distance from the "output emotion" from the robot and the "expected emotion".

Let's call A the output emotion vector composed by 10 elements (3 variables from sound plus 7 from image, see 3.2 for description of those elements). When the robot synthesizes A, it does it according to the likelihood tables previously filled out. Later A is classified using the same Bayesian networks proposed for human emotion classification as refered on section 1.



(a) (b)

Figure 7: Bhattacharyya Distance reduces as $P(A)$ get closer to $P(R)$

Therefore we have $P(A)$ which is a discrete probability distribution, of A, in the scope X=*{happy, sad, fear, neutral, anger}*. Let's call $P(R)$ the expected discrete probability distribution. Since it is expected one of the five expressions, $P(R)$ will always has 0.96 probability on the expected expression and 0.01 at the other four.

According to [8] the Bhattacharyya distance D_B is the best metric to compare histograms. Thus we selected this metric to compare our histograms and our error function is given by:

$$E = D_{B_{(A,R)}} = -ln \sum_{x \in X} \sqrt{P(A)_x P(R)_x}$$

6 Results

Bhattacharyya Distance reduces as the histogram from the presented output of robot's expression $P(A)$ get closer to the expected expression $P(R)$. Notice, in figure 7, that at time 1 a wrong expression was presented, thus the error increases to $1,71$. Any value of D_B higher than $0,699$ indicates that a wrong expression is being shown. As the expression converges to the expected one, D_B decreases up to near zero.

In our study case presented in section 3.3, a battery of tests were run with several male subjects. The time between phrases (TBP) and the total time (TT) was measured and it is clear that it runs in real-time. However let's be clear that the TBP may have different meaning according to the phrase. Analyzing figure 2, notice that: the time between phrase 1 and 2 ($TBP(1,2)$) is not relevant for interaction purposes, because it is the time the user takes to perform the exercise. The $TBP(2,3)$ is dependent of our silence detector, since phrase 3 is an answer of the robot. $TBP(3,4)$ is relevant, it depends only of the human reaction. $TBP(4,5)$is again dependent of our silence detector. $TBP(5,6)$ is relevant, it depends only of the human reaction. Finally we have also $TBP(4,7)$ when the human did the exercise correctly and it lies in the same reason as $TBP(2,3)$ and $TBP(4,5)$.

The silence detector is a technical feature that detects the end of human speech and will not be discussed in this paper.

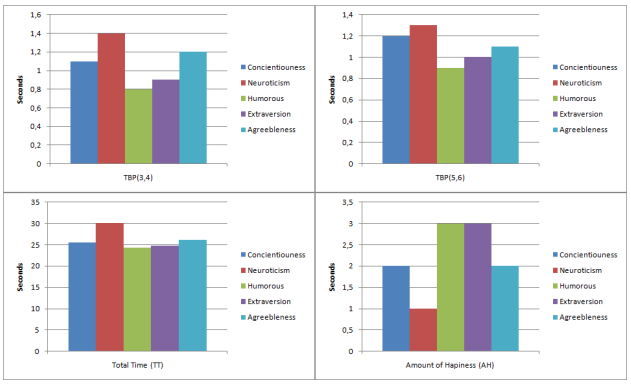


Figure 8: The average of $TBP_{3,4}$, $TBP_{5,6}$, TT and AH for the different Social Behavior Profile of the robot.

Considering this, we present figure 8 with the average of $TBP(3,4)$, $TBP(5,6)$, TT and AH for the different Social Behavior Profile of the robot.

7 Conclusion

The robot's ability of expression has a major effect on the human reaction time during a dialog. By using the assessments defined in section 5.1, we conclude that the human answer faster when when SBP is set to *"humorous"* and *"extraversion"*. The "conscientiousness" SBP is the same as a robot without emotions and results show that in this configuration human response time and total time of the dialog is higher than for *"humorous"* and "*extraversion*" SBP. Moreover we conclude that people enjoy more talking to a funny robot and these results encourage to continue research in humorous robots.

References

[1] Special issue on spoken language processing. *Proceedings of IEEE*, 88:1139–1366, August 2000.

[2] Kim Binsted, Helen Pain, and Graeme Ritchie. Children's evaluation of computer-generated punning riddles. *Department of Artificial Intelligence, University of Edinburgh*, 1997.

[3] I. Cohen, N. Sebe, A.Garg, M.S. Lew, and T.S Huang. Facial expression recognition from video sequences. *in proc. ICME*, pages 121–124, 2002.

[4] Ian J. Deary, Alexander Weiss, and G. David Batty. Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Journal of Psychological Science in the Public Interest intelligence, personality, and health outcomes*, 11-2:53–79, 2010.

[5] Paul Ekman and E.L. Rosenberg. *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press. Second expanded edition, 2004.

[6] Sebastien George and Pascal Leroux. An approach to automatic analysis of learners social behavior during computer-mediated synchronous conversations. In Stefano Cerri, Guy Gouarderes, and Fabio Paraguacu, editors, *Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, pages 630–640. Springer Berlin / Heidelberg, 2002.

[7] Alice S.M Kau, Elaine Tierney, Irena Bukelis, Mariah H Stump, Wendy R KAtes, William H Trescher, and Walter E Kaufmann. Social behaviour profile in young males with fragile x synfrome: Characteristics and specificity. *American Journal of Medical Genetics*, 126:9–17, 2004.

[8] P. Menezes, F. Lerasle, and J. Dias. Towards human motion capture from a camera mounted on a mobile robot. *Image and Vision Computing*, 29-6:382–393, 2011.

[9] Microsoft. Windows seven buil-in speech recognition: a review.

[10] Mihalis A. Nicolaou, Hatice Gunes, and Maja Pantic. Audio-visual classification and fusion of spontaneous affective data in likelihood space. In *ICPR*, pages 3695–3699, 2010.

[11] Nuance. Draggon naturally speaking. http://www.nuance.com/naturallyspeaking/, 2010.

[12] Gayatri Paknikar. *FACIAL IMAGE BASED EXPRESSION CLASSIFICATION SYSTEM USING COMMITTEE NEURAL NETWORKS*. PhD thesis, The Graduate Faculty of The University of Akron, 2008.

[13] Maja Pantic. Facial expression recognition. In *Encyclopedia of Biometrics*, pages 400–406, 2009.

[14] Jose Prado, Luis Santos, and Jorge Dias. Horopter based dynamic background segmentation applied to an interactive mobile robot. *14th International Conference on Advanced Robotics, ICAR09, Munich, Germany*, 2009.

[15] Jose Augusto Prado, Carlos Simplicio, and Jorge Dias. Robot emotional state through bayesian visuo-auditory perception: focus on auditory perception. *In proceedings of - DOCEIS 2011 - Doctoral Conference on Computing Electrical and Industrial Systems*, 2011.

[16] G. Ritchie. Prospects for computational humor. *In Proceedings of 7th IEEE International Workshop on Robot and Human Communication*, pages 283–291, 1998.

[17] N. Sebe, M.S. Lew, I. Cohen, A.Garg, and T.S Huang. Emotion recognition using a cauchy naive bayes classifier. *in proc. ICPR*, 1:17–20, 2002.

[18] Carlos Simplicio, Jose Prado, and Jorge Dias. A face attention technique for a robot able to interpret facial expressions. In *DOCEIS 2010 - Doctoral Conference on Computing Electrical and Industrial Systems*, 2010.

[19] Carlos Simplicio, Jose Augusto Prado, and Jorge Dias. Comparing bayesian networks to classify facial expressions. *in Proceedings of RA-IASTED, The 15th IASTED International Conference on Robotics and Applications, Cambridge, Massachusetts, USA.*, 2010.

[20] Carlos Simplicio, Jose Augusto Prado, and Jorge Dias. A face attention technique for a robot able to interpret facial expressions. *in Proceedings of the DoCEIS'10 - Doctoral Conference on Computing, Electrical and Industrial Systems. Lisbon.*, Springer - ISBN 978-3-642-11627-8, 2010.

[21] O. Stock and C. Strapparava. Getting serious about the development of computational humor. *In proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 59–64, 2003.

[22] Oliviero Stock and Carlo Strapparava. The act of creating humorous acronyms. *Journal of Applied Artificial Intelligence*, 19:137–151, 2005.

[23] Dimitrios Ververidis and Constantine Kotropoulos. Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48(9):1162 – 1181, 2006.

[24] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[25] Wuhan. Facial expression recognition based on local binary patterns and coarse-to-fine classification. *Fourth International Conference on Computer and Information Technology (CIT'04)*, 16, 2004.

[26] M. H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Machine Intelligence*, 24:34–58, 2002.