A PROTOTYPE OF A MACHINE SPEAKING WITH EMOTIONS

Albert van der Heide[1] **Gracian Trivino**[1]

[1] European Centre for Soft Computing (ECSC), Mieres (Asturias)
{albert.vdheide,gracian.trivino}@softcomputing.es

Abstract

Many computational applications use natural language to communicate with their users. Nevertheless, this communication usually is unnatural due to the absence of emotional content.

This paper presents a prototype of a computational system that, thanks to a simulated model of emotions, is capable of including emotional content in its spoken utterances. The system's emotional content is based on the conditions of light and temperature, and implemented using a Fuzzy Finite State Machine.

Keywords: model of emotions, fuzzy finite state machine, emotional speech.

1 INTRODUCTION

Currently there are quite a number of systems that attend to humans automatically. A number of these systems use voice to transmit content, for example, in the domain of telephony many automated attendants exist, various web applications use speech, and there are many more examples.

However, the interaction with these machines is currently still too artificial. Usually these machines are insensitive to the emotional content being expressed by the communication partner, as well as incapable of modifying the expressed voice. This sensitivity is of importance in human interaction and actually transmits a lot of valuable information.

In order to increase the effectiveness and acceptance of these automated systems these machines would need to respond in a natural way. One way that this can be improved is by making these machines sensitive to the emotional content expressed by a human communication partner, as well as being able to express themselves emotionally as humans do. An emotional aware machine would be able to adjust its behavior in response to the emotional content transmitted by the communication partner.

In this article, we elaborate upon a prototype that was created to simulate an emotional state of a machine. The objective of the prototype is to simulate and express emotional state.

The prototype consists of a virtual being that appears as if it has an emotional state that depends on the weather, such that when the weather is 'bad' the emotional state deteriorates and when the weather is good the emotional state improves. The emotional state is influenced by two parameters: luminosity and temperature during the day. The emotional state is expressed by emotionally modified speech. A depiction of the basic scheme is given in figure 1.

The emotional state of the system is fuzzy, i.e., at any point in time the system can be in various emotional states to certain degree.

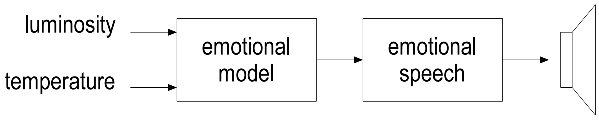

Figure 1: The general scheme of the system.

2 A SIMPLE MODEL TO SIMULATE THE EMOTIONAL STATE

Currently there is no consensus on how emotions should be modeled, nor on what representation to choose. It is clear however that emotions are an integral part of human functioning and are a key part in human cognition [4, 11]. There are general models about the universality of emotions [3, 8] and the expression of emotions in faces [6]. Several researchers have begun to model emotions for conversational agents, for example [13, 5, 20].

We believe that fuzzy logic provides an expressive language that enables our model to represent complex emotional states and behaviors. Fuzzy logic makes it possible to deal with multiple and conflicting emotions and makes it possible to output a mixture of emotions. Fuzzy logic has previously been used in emotion simulation, for example [7].

In previous studies we have learned that Fuzzy Finite State Machines (FFSM) are suitable tools for modeling signals that follow an approximately repetitive pattern [1, 16, 17]. Here, we use a FFSM to model how luminosity and temperature patterns during the day can influence the simulated emotional state of a virtual being.

As part of a first prototype we have developed a FFSM that uses two input variables, namely the average temperature and luminosity during the last 24 hours. These variables are fuzzyfied, subsequently fuzzy rules determine the new state activation based on input and current activation of states. In short, the prototype explores how ambient parameters can influence emotional states of the system.

A FFSM is a tuple $\{S, U, Y, f, g, S_0\}$. We describe each one of its components in the following sections.

2.1 Linguistic Fuzzy States

In the first steps of modeling a system, the designer must define a set $\{S_i\}$ of fuzzy sets that summarize the relevant states of the system. The *linguistic fuzzy state* of the system (S) is a linguistic variable that take its values in the set of linguistic labels $\{S_1, S_2, \ldots, S_n\}$.

We represent the fuzzy state of the system with a *state activation vector* $s(t) = (s_1(t), s_2(t), \ldots, s_n(t))$ where $s_i \in [0, 1]$ and $\sum_1^n s_i = 1$.

We will see in section 2.2 how to obtain $s(t)$ using a set of fuzzy rules.

S_0 is the initial state or *state activation vector* at $(t = 0)$, e.g., $(1, 0, 0, 0)$.

2.1.1 Input variables

U is the input vector $(u_1, u_2, ..., u_{nu})$. Typically, U is a set of linguistic variables obtained after fuzzification of numerical measures obtained from sensors.

The emotion-FFSM is sensitive to two linguistic variables [21]; *luminosity* and *temperature*, both obtained from sensors attached to a window in our laboratory. The sensors measure the light intensity coming from the outside of the building as well as the temperature near the window in one minute intervals. In a preprocessing state this signal is smoothed by taking the running average over 10 samples. The basic input signal to the system is the average luminosity and the average temperature during the last 24 hours. Linguistic variables take values ‘high’, ‘medium’, and ‘low’ to represent the current values of luminosity and temperature. The membership functions are depicted in figure 2.

These membership functions have been defined automatically based on the range in input values. (Approximately 35% of the samples are defined as ‘high’, 30% as ‘medium’, and 35% ‘high’).

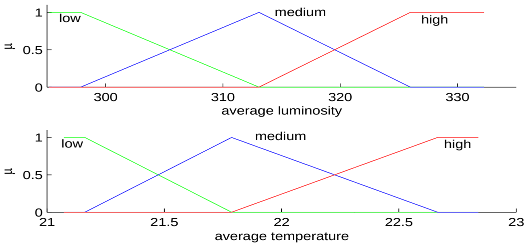

Figure 2: The linguistic variables for luminosity and temperature.

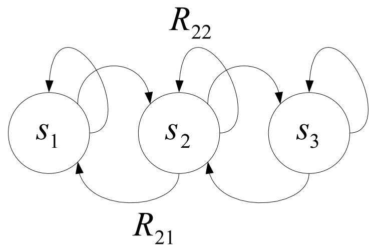

Figure 3: A depiction of the Fuzzy Finite State Machine and the possible transitions within the model.

2.2 Transition function f

We obtain the new *activation vector* using the *transition function* $S[t + 1] = f(U[t], S[t])$.

Once the designer has identified the relevant states in the model, he/she must define the rules that govern the temporal evolution among the states of the system. In our approach these rules are fuzzy rules. In Figure 3 depicts a graph that shows the allowed paths between the states. Here, state s_1 represents a bad or angry mood, state s_2 represents a neutral mood, and state s_3 represents a joyful/happy mood. As personal characteristics of the virtual being we had envisioned that it should be moody when the days are cold and dark, and happy when the days are warm and bright.

The being that we are designing is envisioned as having an pleasant personality, and therefore will be more likely to be in a happy mood than in a depressed mood. Such a personality can be created by modifying the state transition rules, as well as by making use of *TimeToStay* and *TimeToMove* linguistic variables which will be explained in the following.

We distinguish between rules to remain in a state (R_{ii}) and rules to change a state (R_{ij}).

2.2.1 Rules to remain in a state

The designer uses these rules to express the conditions of the system to remain in a specific state. We distinguish between amplitude conditions (C_A) and temporal conditions (C_T).

Using this terminology, the generic expression of these rules is formulated as follows:

R_{ii}: IF ($S(t)$ is S_i) AND (C_A) AND (C_T) THEN ($S(t+1)$ is S_i)

Amplitude conditions. The designer uses the vector of input variables U to write the fuzzy rules that describe the constraints on the input variables to remain in the state S_i. For example, C_A = ((u_1 is High) AND (u_2 is High)).

Temporal conditions. These rules express constraints on the duration of states. The duration of a state (D_i) is calculated as:

$$\text{IF } (S_i > \delta) \text{ THEN } D_i(t+1) = D_i(t) + 1 \\ \text{ELSE } D_i(t+1) = 0;$$

where δ is a threshold that is application dependent.

Typically the temporal constraints are modeled using the linguistic label *TimeToRemain* (TR_{ii}) (see figure 4). The conditions (D_i is TR_{ii}) allows the designer to constrain the maximum time the system is in the state S_i.

In this application, we designed the virtual being as optimistic, e.g. it is not moody indefinitely, although the weather conditions might give it sufficient reason to be so. These personality characteristics have been modeled using temporal conditions.

An example rule to remain in a state is the following:

$$\begin{aligned} R_{11}: \ & \text{IF } (S(t) \text{ is } S_1) \text{ AND } (u_1 \text{ is Low}) \text{ AND} \\ & (u_2 \text{ is Low}) \text{ AND } (D_1 \text{ is } TR_{11}) \\ & \text{THEN } (S(t+1) \text{ is } S_1) \end{aligned}$$

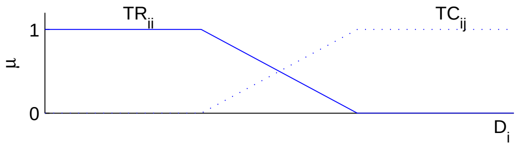

Figure 4: Linguistic labels used for defining temporal constraints, namely, *TimeToRemain* and *TimeToChange*.

2.2.2 Rules to change of state

The designer uses these rules to express the conditions that provoke the system changing from state S_i to state S_j. Also here, we distinguish between amplitude conditions (C_A) and temporal conditions (C_T).

The generic expression of R_{ij} is formulated as follows:

R_{ij}: IF ($S(t)$ is S_i) AND (C_A) AND (C_T) THEN ($S(t+1)$ is S_j)

Amplitude conditions. In a first approach these conditions coincide with the amplitude conditions to remain in the destination state of the transition. Nevertheless, some tuning could be needed to express a softer condition to change.

Temporal conditions. These conditions are based on the linguistic label *TimeToChange* (TC_{ij}) (Figure 4). The condition (D_i is TC_{ij}) allows the designer to constrain the minimum time the system must be in S_i before to move to S_j.

What follows is an example of rules used for state transitions:

$$\begin{aligned} R_{12}: \ & \text{IF } (S(t) \text{ is } S_1) \text{ AND } (D_1 \text{ is } TC_{12}) \\ & ((u_1 \text{ is Med }) \text{ OR } (u_2 \text{ is Med})) \\ & \text{THEN } (S(t+1) \text{ is } S_2) \end{aligned}$$

2.3 State activation

We calculate the degree of activation of a state using the Takagi-Sugeno-Kang paradigm [14].

In the rule-base a subset of all rules k contribute to the same destination state i. To calculate the activity of state i at time $t+1$ we take into account the rules contributing to the state i, ω_i, weighted by the degree of firing of all rules, ω_k:

$$S_i[t+1] = \begin{cases} \frac{\sum \omega_i}{\sum \omega_k} & \text{if } \sum \omega_k \neq 0 \\ S_i[t] & \text{if } \sum \omega_k = 0 \end{cases}$$

where the degree of firing for each rule ω is calculated using the minimum for the AND operator.

2.4 Output variables

Y is the output vector $(y_1, y_2, ..., y_{no})$. Y is a summary of the characteristics of the system evolution that are relevant for the application. In our case this is directly the emotional state represented by the FFSM. For example, in this application $Y = (0, 0.7, 0.3)$ represents an emotional state between *neutral* and *happy.*

The output vector is calculated by the output function g.

2.5 Output function g

The output function $g(U[t], S[t])$, calculates the values for the output variables $Y[t]$. As a first possible implementation of Y we have chosen the function $Y = s(t) = (s_1(t), s_2(t), \ldots, s_n(t))$. In this case the model output is the current fuzzy state of the system represented as a *state activation vector.*

In the prototype application we are directly interested in the emotional state represented by the model. However, this does always not need to be the case and other output functions could be defined. For example, one could apply the average and the standard deviation to the values of the input variables while the system remained in the considered state. Furthermore, each output variable y_i could represent any parameter which is related to an internal emotional state, and usually more complex functions g are needed.

3 EMOTIONAL STATE EXPRESSION IN XML

The FFSM provides as output directly the state current activation S. We express this emotional state through a voice which states a simple phrase.

For the voice synthesis we use Verbio Text-To-Speech (TTS) [15]. The speech can be annotated using a standard markup language based on and very similar to Speech Synthesis Markup Language (SSML) [19].

In the literature there are various articles describing how emotional speech is different from normal speech and what the parameters are that are modified [12, 18]. It should be noted however that there is not much literature specifically centered on Spanish language [10, 2, 9]. We have used the references specified here and applied these to our speech synthesis method.

For example, when a person is experiencing *joy* a number of parameters are modified. Among these parameters are the fundamental frequency, which is higher than normal, the speech rate, which is faster than normal, etc. Additionally, according to the literature, the fundamental frequency decreases during the phrase expression. We have annotated speech to include these parameters. Table 2 gives an overview of the used parameters. The annotated speech for *joy* is as follows:

```
<prosody pitch="+40%" range="+20%" rate="-30%">Estoy </prosody>
<prosody pitch="+20%" range="+20%" rate="-30%">muy   </prosody>
<prosody              range="+20%" rate="-30%">feliz.</prosody>

<prosody pitch="+40%" range="+20%" rate="-30%">Todo  </prosody>
<prosody pitch="+26%" range="+20%" rate="-30%">ha    </prosody>
<prosody pitch="+13%" range="+20%" rate="-30%">salido</prosody>
<prosody              range="+20%" rate="-30%">bien. </prosody>
```

There are quite a number of parameters that can be modified. Nevertheless, not all parameters we find in the literature can be expressed using Verbio TTS or other TTSs, for example variance in pitch, or mean / range modification of intensity (only one general parameter exists for intensity).

Another limitation of the current prototype is that it does not directly accept a mixture of emotional states as input. Therefore, for the moment, the prototypical annotations have been constructed for five types of emotional states. These annotations are related to states $\{(1, 0, 0), (0.5, 0.5, 0), (0, 1, 0), (0, 0.5, 0.5), (0, 1, 1)\}$. For the generation of an output sentence at a certain point in time the most similar state's annotation is selected.

Table 1: Emotional states and acoustic features.

	sadness	neutral	joy
pitch mean	$<$		$>$
pitch range	$<$		$>$
pitch contour	$\searrow$		$\nearrow$
intensity	$<$		
speech rate	$>$		$<$

4 EXPERIMENTATION

To test our prototypical system we have applied temperature and luminosity data to it. The data was collected during three weeks. The average values during the last 24 hours has been calculated over the input signal. The membership functions for defining when a temperature or luminosity is 'high', 'medium' or 'low' has been defined automatically based on the range in input values (approximately 35% of the samples are defined as 'high', 30% as 'medium', and 35% 'high').

The activation of the states is evolving in function of the current inputs, as well as on the basis of the current simulated emotional state. An example showing the evolution of the different states is given in figure 5.

The system started in a depressed state due to the initial state activations settings, $(1, 0, 0)$. However, the input to the system did not provide any reason for being in this state, and quickly the system moves from a mainly neutral state towards a joyful state. This state continues for two days, after which the time restrictions for state 3 start to take effect, lowering the state activation. At the same time the temperature and luminosity start to fall, and during day 2-3 the conditions are sufficiently bad to give rise to a depressed state. But later in the week conditions start to improve again, and a neutral state is recovered, followed by a happy state, due to the improved weather conditions.

The speech output at a particular point in time was annotated based on the nearest prototypical state's annotation, as explained at the end of section 3, and displayed in table 2.

Currently the prototype is visualized by an icon on a computer screen with a button next to it. When the button is pushed a voice is heard speaking ‘hello world’ with different emotional intonations.

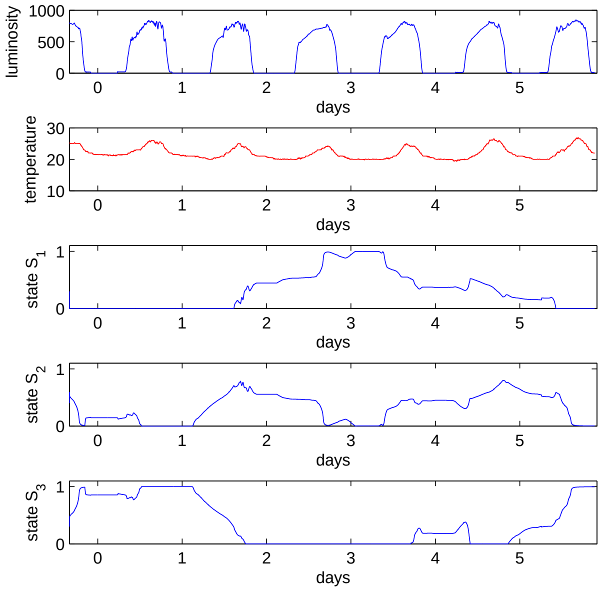

Figure 5: The activity of three states S_1,S_2 and S_3, as well as the input variables *luminosity* and *temperature*.

5 CONCLUSIONS

We have presented a simple prototype as demonstration of how to create a virtual being that seems to have emotional status and that behaves in agreement with these emotions.

Table 2: Different annotations of the words in a sentence based on the emotional state.

	”and	have	a	good	day”
depressed					
pitch mean and contour		-10%	-20%	-30%	-40%
pitch range	-20%	-20%	-20%	-20%	-20%
intensity	-30%	-30%	-30%	-30%	-30%
speech rate	+20%	+20%	+20%	+20%	+20%
depressed / neutral					
pitch mean and contour		-5%	-10%	-15%	-20%
pitch range	-10%	-10%	-10%	-10%	-10%
intensity	-15%	-15%	-15%	-15%	-15%
speech rate	+10%	+10%	+10%	+10%	+10%
joyful / neutral					
pitch mean and contour		+5%	+10%	+15%	+20%
pitch range	+10%	+10%	+10%	+10%	+10%
intensity					
speech rate	-15%	-15%	-15%	-15%	-15%
joyful					
pitch mean and contour		+10%	+20%	+30%	+40%
pitch range	+20%	+20%	+20%	+20%	+20%
intensity					
speech rate	-30%	-30%	-30%	-30%	-30%

In this paper we have contributed to the development of new techniques of human - computer communication. A FFSM has been used to create a model of different simulated emotional states. The computer produces the same spoken utterances with varying emotional content, adding an additional connotation to it's content. The emotional content can be expressed using SSML-like XML annotations.

This prototype is an additional step in the project of building a complex virtual being. This virtual being will be the personal assistant of the user. As humans do, the computerized assistant will be able to communicate information about it's context and emotional state. Of course, to take advantage of this possibility, the emotional states must be influenced by the variables in the context in such a way that the human user will realize the content is modified. For example, the tone of the voice could emphasize a message of warning or the communication of a positive result.

Acknowledgements

This work has been partially funded by the Spanish government (CICYT), under grant: TIN2008-06890-

C02-01, and by the Foundation for the Advancement of Soft Computing.

References

[1] A. Alvarez and G. Trivino. Comprehensible model of a quasi-periodic signal. In *Proc. of the 9th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2009.

[2] P. Boula de Mareüil, P. Celerier, and J. Toen. Generation of emotions by a morphing technique in english, french and spanish. In *Proceedings on Speech Prosody 2002*, pages 187 – 190, 2002.

[3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, and W. Fellenz. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 1:32 – 80, 2001.

[4] A.R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain.* G.P. Putnam, New York, 1994.

[5] E. M. Eisman, V. López, and J. L. Castro. Controlling the emotional state of an embodied conversational agent with a dynamic probabilistic fuzzy rules based system. *Expert Systems with Applications*, 36:9698 – 9708, 2009.

[6] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6:169 – 200, 1992.

[7] M.S. El-Nasr, J. Yen, and Ioerger T.R. Flame fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3):219 – 257, 2000.

[8] N. Fragopanagos and J. G. Taylor. Emotion recognition in human-computer interaction. *Neural Networks*, 18(4):389 – 405, 2005.

[9] I. Iriondo, R. Guaus, A. Rodriguez, P. Lázaro, N. Montoya, J.M. Blanco, D. Bernadas, J.M. Oliver, D. Tena, and L. Longth. Validation of an acoustical modelling of emotional expression in spanish using speech synthesis techniques. In *Proceedings ISCA 2000*, pages 161 – 166, 2000.

[10] J.M. Montero, J. Gutierrez-Arriola, J. Colas, E. Enriquez, and J.M. Pardo. Analysis and modelling of emotional speech in spanish. In *Proceedings of the 14th International Conference on Phonetic*, pages 957 – 960, 1999.

[11] A. Ortony, G.L. Clore, and A. Collins. *The cognitive structure of emotion.* Cambridge University Press, Cambridge, UK, 1998.

[12] M. Schröder. Emotional speech synthesis - a review. In *Proceedings Eurospeech 2001*, volume 1, pages 561 – 564, Aalborg, 2001.

[13] M. Schröder. *Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis.* PhD thesis, Saarland University, 2004.

[14] M. Sugeno. *Industrial Applications of Fuzzy Control.* Elsevier Science Inc., New York, NY, USA, 1985.

[15] Verbio Speech Technologies. Read aloud. http://www.verbio.com/.

[16] G. Trivino and A. van der Heide. An experiment on the description of sequences of fuzzy perceptions. *Proceedings of the 8th International Conference on Hybrid Intelligent Systems (HIS2008), September 10-12th, Barcelona, Spain*, 2008.

[17] G. Trivino and A. van der Heide. Linguistic summarization of the human activity using skin conductivity and accelerometers. *Proceedings of the Conference Information Processing and Management of Uncertainty in Knowledge Based Systems. (IPMU2008), June 22-27, Malaga, Spain*, 2008.

[18] D. Ververidisa and C. Kotropoulos. Emotional speech recognition: Resources, features, and methods speech communication. 48(9):1162 – 1181, 2006.

[19] World Wide Web Consortium (W3C). Speech synthesis markup language 1.0. http://www.w3.org/TR/speech-synthesis/.

[20] I. Wilson. The artificial emotion engine (tm), driving emotional behavior. *AAAI Technical Report*, pages 76 – 80, 2000.

[21] Lotfi A. Zadeh. The concept of linguistic variable and its application to approximate reasoning. *Information sciences*, 1975.