

## Article

# Expressing Robot Personality through Talking Body Language

Unai Zabala <sup>†</sup>, Igor Rodriguez <sup>†</sup>, José María Martínez-Otzeta <sup>†</sup> and Elena Lazkano <sup>\*,†</sup>

Department of Computer Science and Artificial Intelligence, University of the Basque Country, Manuel Lardizabal 1, 20018 Donostia-San Sebastián, Spain; unai.zabalac@ehu.eus (U.Z.); igor.rodriguez@ehu.eus (I.R.); josemaria.martinez@ehu.eus (J.M.M.-O.)

\* Correspondence: e.lazkano@ehu.eus

† These authors contributed equally to this work.

**Abstract:** Social robots must master the nuances of human communication as a mean to convey an effective message and generate trust. It is well-known that non-verbal cues are very important in human interactions, and therefore a social robot should produce a body language coherent with its discourse. In this work, we report on a system that endows a humanoid robot with the ability to adapt its body language according to the sentiment of its speech. A combination of talking beat gestures with emotional cues such as eye lightings, body posture of voice intonation and volume permits a rich variety of behaviors. The developed approach is not purely reactive, and it easily allows to assign a kind of personality to the robot. We present several videos with the robot in two different scenarios, and showing discrete and histrionic personalities.

**Keywords:** social robots; talking gesticulation behavior; emotion appraisal; robot personality



**Citation:** Zabala, U.; Rodriguez, I.; Martínez-Otzeta, J.M.; Lazkano, E. Expressing Robot Personality through Talking Body Language. *Appl. Sci.* **2021**, *11*, 4639. <https://doi.org/10.3390/app11104639>

Academic Editor: Alessandro Gasparetto

Received: 4 March 2021

Accepted: 17 May 2021

Published: 19 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Human-robot interaction (HRI) is the study dedicated to understand, design and evaluate robotics systems to be used by or with humans [1,2]. HRI is a multidisciplinary field with contributions from multiple fields such as human-computer interaction, artificial intelligence, robotics, natural language understanding, or social sciences among others.

Social robots have emerged as a class of robots that require a highly evolved type of human robot interaction. These robots cannot be merely teleoperated and must possess skills that are beyond those present in cooperative robots, due to the challenges faced when developing social intelligence; robots that interact with humans should behave as humans do.

Socially interacting robots must have abilities such as communicating using verbal (natural language) or non-verbal modalities (lights, movements or sound); expressing affection or perceiving human emotions; possessing distinctive personality; modelling human social aspects; learning; establishing social relationships [3,4]. Robots able to interact in such manners are being sketched in many applications such as caregivers of the elderly or of people with physical or emotional disabilities, in education, entertainment, and even in domestic scenarios [5–11].

Verbal and body expression in robots are thus of main concern while developing social interaction. This paper aims to link these two aspects of social intelligence by creating a system that coordinates the robot's body language with the nature of its discourse. By "nature" we mean the emotional aspect of the speech, as extracted by a text sentiment analyzer. The numerical output of the sentiment analyzer is then reflected in some degree in the movement or change of state of different parts of the robot body.

Therefore, the contribution of this paper is twofold:

(1) The development of a talking behavior as a combination of talking beat gestures with emotional cues such as eye lightings, body posture of voice intonation and volume. At this stage, the sentiment analysis extracted by the text being vocalized by the robot is used

as input to the talking behavior and each feature is affected by the polarity extracted by the sentiment analyzer. (2) Instead of being purely reactive, the developed approach easily affords modulating to the intensity or type of the actions which accompany the speech depending on the personality we would like to assign to the robot.

The rest of the paper is structured as follows. Section 2 reviews the literature of emotional and affective robotics. Section 3 describes the conducted approach, how beat gestures are generated using a deep generative approach, how different features are used to convey the sentiment extracted by a text sentiment analyzer; and how the robot state is affected by each reaction is explained. Robot personality adjustment is described in Section 4 and the resulting robot behavior is presented and discussed by means of videos in Section 5. Finally, Section 6 outlines the improvements proposed as further work.

## 2. Emotion Expression in Robots

Citing [12], “Creating and sustaining closed-loop dynamic and social interactions with humans requires robots to continually adapt towards their users’ behaviors, their affective states and moods while keeping them engaged in the task they are performing”. In this vein, the Affective Loop is defined as the interactive process in which the user of the system first expresses her/his emotions through some physical interaction involving her body and the system responds by generating affective expression which in turn affects the user making the user respond and step-by-step feel more and more involved with the system [13–16]. Thus, perceiving and showing emotions is essential to convey interaction.

Verbal communication is a natural way of interaction among humans. It is one of the communication channel most used by human beings, it is dynamic and is learned from childhood, even when we do not know how to write or read, we are able to communicate with others through words. Oral language allows us to transmit a message to the receiver, whether it be an opinion, an order, a feeling, etc. We express them through the articulation of a sound or group of sounds with different types of intonation, which can give a greater or lesser emotional charge to what is expressed. The way in which human voice can be modulated plays an important role in the communication of emotions. This process can be very complex and of the utmost importance in human-robot interaction, according to Crumpton and Bethel [17], who highlight the importance of vocal prosody.

However, non-verbal expression is key to understand sociability [18,19].

Some authors working with virtual agents and computer graphics have obtained impressive realistic animations of human characters. They are able to perform an accurate synchronization of the gesture behavior with the synthesized speech [20–23]. The demand for robots to behave in a sophisticated manner requires

the implementation of capabilities similar to those typical of humans: sensing, processing, action and interaction. All of them have to take into account the underlying cognitive functions: motivations, emotions and intentions. Recently effort has been made in the search of behaviors that are able to convey sentiment. As the main mechanism to communicate emotions, facial expression plays a predominant role [24,25].

The robotic head Kismet [26] represents itself a milestone as how the human voice and facial features affect expressiveness. Furhat is another robotic head that shows similar characteristics, but it uses a back projected facial animation system for face to face interaction [27].

The advent of humanoid robots has encouraged researchers to investigate and develop body language expression in robots. The body language uses the gestures, postures and movements of the body and face to convey information about the emotions and thoughts of the sender, and can disclose as much information as words. For instance, Anki’s Cozmo [28] is a tiny robot with impressive body expression [29]. Face emotion is shown in a small LCD screen. It moans and laughs. A kind of shovel that it uses for manipulation purposes adds arm level expression in a wheeled platform. At a different scale, Shimi [30], a smart-phone enabled robotic musical companion far from human morphology, expresses emotion rather differently, using a (faceless) body with a notably small number of DoFs.

Humans can also learn to associate colors with emotions and therefore could be used as another possible channel of communication when in conjunction with adequate cognitive models [31–33]. Color and light patterns can be modulated in a dynamic way to evoke happiness, sadness, anger or fear [34,35].

Different cultures or even individuals could interpret in a different way non-verbal behavior, but it is highly relevant for social interaction for all of them. When engaged in verbal communication with a robot, a person's trust is higher when the robot's gaze is in her direction [36]. In [37], authors propose a system in which the robot expresses itself through gestures in addition to speech, and in which the robot takes into account the human's reactions to adapt its own behavior. They then assess the perception of the person when compared with the speech-only behavior.

In [38] the same authors add facial expressions to their system. They report on an experiment where participants discuss with the robot videos chosen to induce some particular emotion, and the robot tries to adjust its behavior to the emotional content of that video.

Huang and Mutlu [39] conclude that all types of gestures affect user perceptions of the robot's performance as a narrator. Therefore, an important goal is to create a coherent gesture-speech narrative. In [40] a system is trained on a single-speaker audio and motion-capture data and it is able to generate both speech and full-body gestures from that input. A framework for speech-driven gesture production, designed to be applied to virtual agents with the aim of enhancing human-computer interaction is presented in [41].

The main contribution of this paper is the novel combination of all of the previous aspects in a socially interacting robot. In the following sections we introduce our approach to address the goal of generating an adaptive expression behavior for social robots.

### 3. Sentiment to Expression Conversion

As mentioned before, human talking expression can be affected by many factors. The mood, the interactional cues perceived or the character/content of what is being said is reflected in both face and body features. As we are concerned on the development of a natural talking robot behavior, at this stage we opted by analysing the effect of the nature of the verbosity, i.e., the text being pronounced by the robot. Thus, this nature of the verbosity, as a measure of the sentiment, will constitute the affective input of the emotion system of the robot and the measure used to modulate the robot gesturing features.

#### 3.1. Affective Input

The literature reveals two types of emotional models. On the one hand, the theory of basic emotions divides emotions into discrete and independent categories, such as the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) identified by Ekman [42].

Other affective models regard those experiences as a continuum of highly ambiguous states with also a high degree of interrelationship. They describe emotions as linear combinations of *Valence-Arousal-Dominance* (VAD) values. *Valence* is a measure of positivity-negativity of the stimulus, *Arousal* is related to the energy level and *Dominance* addresses the approachability of the stimulus.

These models allow for a wider range of emotions [43,44].

The goal of the research field called *sentiment analysis* is the analysis, from pieces of written language, of writer's attitudes, evaluations, emotions, sentiments, and opinions [45]. Its main purpose is to associate a given text with its polarity: positive, negative or neutral.

A brief review of the sentiment analyzers tested in this work are listed below.

VADER (Valence aware dictionary for sentiment reasoning) sentiment analyzer [46], is a tool for sentiment analysis first designed for social media, but also applicable in different domains. It is based on rules derived from dimensional affective models, but also uses a lexicon. It analyzes the intensity and polarity of the written text and gives as output

the proportion of text for each category and a compound score from the *Valence* scores of each word, according to the lexicon.

VADER is optimized for social media data. The wellknown NLTK library [47] also uses VADER as a sentiment analyzer tool.

TextBlob <https://textblob.readthedocs.io/en/dev/> (accessed on 18 May 2021) is another tool for text analysis, written in Python. Its API provides sentiment analysis and another usual natural language processing (NLP) capabilities, as noun phrase extraction or part-of-speech tagging.

Similar to VADER, it is a rule-based system but slightly more limited in the sense that it does not take into account the punctuation marks. The sentiment analyzer outputs the polarity of the sentence together with the level of subjectivity. The advantage is that it can evaluate pieces of text instead of individual sentences.

The main weakness of the rule-based approaches is that the context is not taken into account, and only single words are analyzed. In order to address this issue, word representation in the form of embeddings places two words with similar meaning close to each other in a *n*-dimensional space. This approach is used in the flair framework [48].

Table 1 shows a comparison example of the sentiment scores obtained by the three mentioned tools. The sentences of the table correspond to the text chosen for the experimentation in Section 5.

**Table 1.** From *The Color Monster* by Ana Llenas. Sentiment extraction results.

Sentence	Flair	VADER	TextBlob
Do not you feel much better	−0.963	0.4404	0.5
I see you are feeling something new	0.999	0.128	0.136
Tell me how do you feel now	0.899	0.0	0.0
When you are sad you hide and want to be alone	0.998	−0.6705	−0.5
You do not want to do anything except maybe cry	−0.654	−0.5142	0.0
When you are afraid, you feel tiny	0.815	0.0	−0.3

None of them take into account the semantics of the text as a whole, neither relate each sentence with the previous one.

The three options gave non-expected outputs, i.e., scores that would be very different if the text was manually annotated. The decision to use VADER for our experiments was made on the one hand, by observation and, on the other hand, by the flexibility to attenuate or enhance the nuance given to a sentence by the analyzer by adding emoticons or punctuation marks. This allows us to obtain a robot expression more in harmony with the non captured semantics of the text.

The *Valence* value places the emotion in the scale ranging from sadness to happiness, thus evaluating its positivity. VADER takes all the *Valence* scores from all the words, summarizes them, and returns a normalized sentiment in the range (−1, +1), being −1 the negative extreme and +1 the positive extreme. We then translate this VADER score from sentiment to emotion in the following manner: (Positive: VADER score  $\geq 0.5$ , Neutral:  $-0.5 < \text{VADER score} < 0.5$ , Negative: VADER score  $\leq -0.5$ ). In this work the emotion value is obtained by a direct translation from the VADER returned value (sentiment) into the emotion value in the sadness-happiness continuum.

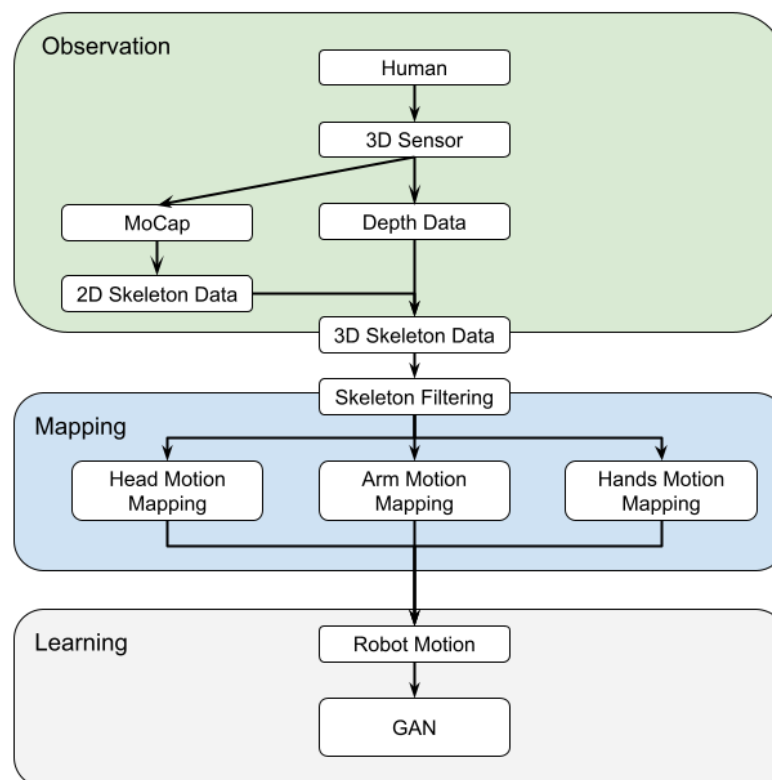
As mentioned before, the sentiment analyzer evaluates each sentence individually. VADER was specifically attuned to sentiments expressed in social media and thereby for short sentences. As a consequence, it is not very sensible, tends to give exaggerated values and the neutral emotion rarely returns a non zero value within the  $-0.5 \cdots 0.5$  range. As an attempt to keep track of the sentiment of the text being verbalized, instead of using the raw compound score, we low pass filter the obtained value, that smooths the compound score over time.

### 3.2. Gesticulation Behavior

Kinesic communication is the technical term for body language, i.e., communicating by body movement. The word kinesics comes from the root word kinesis, which means “movement”, and refers to the study of hand, arm, body, and head movements. Specifically, this section will outline the use of particular gesticulation; we will focus on the generation of beat gestures, i.e., conversational movements of body parts synchronised with the flow of speech but not associated with particular meaning [49]. This group of gestures mainly affects the upper body.

Our approach for talking beats generation consist of the automatic creation of movements that include arm, hand and head’s joint positions in a timeline. Those gestures are generated by a Generative Adversarial Network (GAN) [50] model trained with data captured using a Motion Capturing (MoCap) system, that employs the Intel RealSense RGB-D camera and OpenPose [51] as skeleton tracker. The use of the MoCap to collect data from human speech allows us to capture the naturalness with which we gesticulate when talking and then transfer such properties to the robot. In this way, the gesture generation system allows the robot to generate novel sequences of gestures containing head yaw and pitch positions and arm joints information. A more detailed description about the gesture generation system can be found in [52].

For each sentence to be vocalized by the robot, its duration is calculated and the GAN system is required to generate motion (each movement consists on 4 poses) for that duration. Figure 1 summarizes the process followed to generate movements.

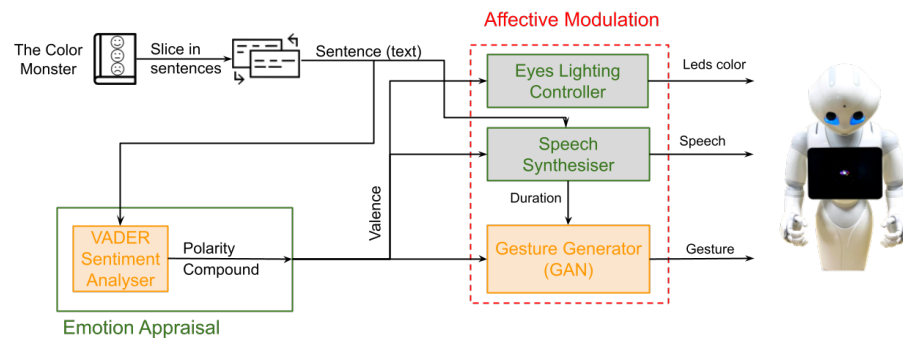


**Figure 1.** Gestures learning process.

### 3.3. Affective Modulation

To appropriately express the emotion obtained from the text analyzer, the sentiment must be mapped into natural body gestures, enriched with facial expressions and voice intonation. This process is performed by the developed “Adaptive Expression” system, which is composed by three main modules: the “Gesture Generator”, the “Eyes Lighting Controller” and the “Speech Synthesiser”. These modules makes the robot adapt its

behavior according to the values obtained from the “Emotion Appraisal” step. Each of the expression features can vary, according to the previous output, in a operational interval experimentally defined for each feature. Figure 2 summarizes the main components that compose the architecture of the system. A more detailed explanation on how this mapping process occurs is described in the following subsections.



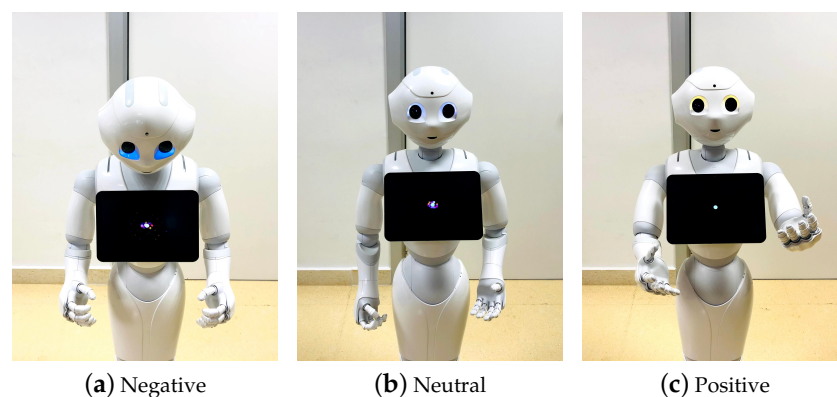
**Figure 2.** The architecture for adaptive talking behavior.

### Body Motion

The “Gestures Generator” module is in charge of the generation of talking gestures, converting the emotion value returned by the “Emotion Selector” module into a collection of gestures. These gestures are generated in such a manner that when executed during the speech, there is a good synchronization.

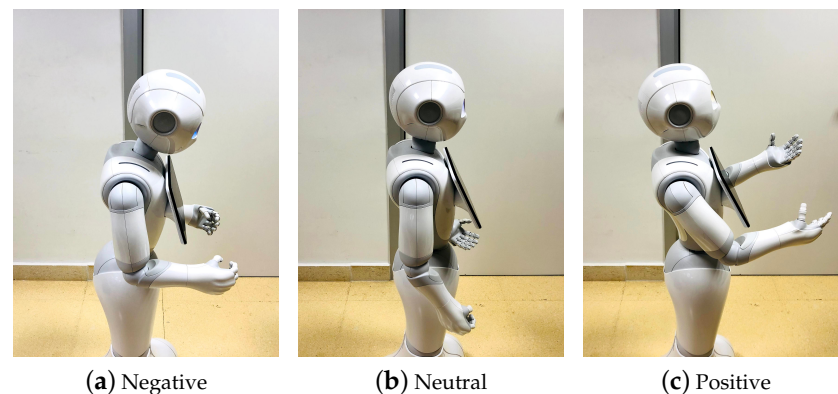
The speed of the gestures execution is adapted to the intended emotion, in order to better convey it. If the emotion is understood as “positive” the gesture will be executed more lively than when the emotion is depicted as “negative”.

The emotion also affects the nod of the head of the robot. When a neutral emotion is portrayed, the robot head will simply look forward. However, the robot will tilt the head in other situations: in case of positive emotions the head will direct upwards, while with negative emotions it will go downwards. In order to obtain valid tilt angles the obtained *Valence* value will be normalized in the head minimum and maximum tilt angle range. In Figure 3 can be seen an example of GAN-generated gestures for three types of emotions: sadness (negative), neutral and happiness (positive).



**Figure 3.** Some examples of generated gestures with emotion.

Finally, the waist is biased to bend down or to stick the chest out, according to the emotion. Figure 4 shows the effect of backbone inclination adjusted by the waist joint.



**Figure 4.** Backbone inclination for different states.

### Facial Expression

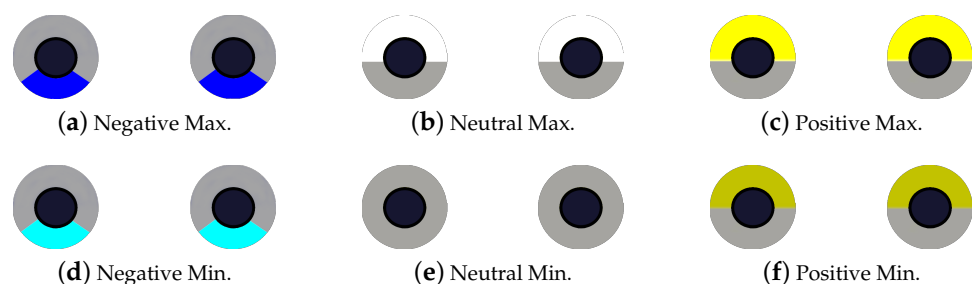
In the field of social robotics, researchers working in the design of robotic eyes have been inspired in the human face and, therefore, have tried to capture the human eyes' movements and shape. However, SoftBank's platform is limited in that sense due to the eyes design structure. In Pepper robots' eyes lie two LED rings which can be programmed in different manners. The color intensity can be changed, there is a choice of different color hues, and they can also be turned on or off during some time span.

Moreover, taking inspiration from facial expressions displayed in cartoons, the LEDs of each eye are grouped in three different patterns as shown in Figure 5. As it can be appreciated, to show different emotions only the half of the eye is used. This pattern showed to be socially better understood by the public than coloring the whole ring of each eye with the same color in the experiments performed in [53].

The conversion from emotion into facial expression is performed by the "Eyes Lighting Controller" module, which changes the color and intensity of the LEDs in the robot eyes. This module codifies the *Valence* value into a color in the form of a set of RGB values suitable to be displayed in the LEDs.

The higher the valence, the more intense is the coloring of the eye.

The colors for each emotion are represented in Figure 5, where the blue-greenish range from RGB(0, 0, 255) to RGB(0, 255, 255) the gray scale from RGB(128,128,128) to RGB(255, 255, 255), and the yellow from RGB(76, 76, 0) to RGB(255, 255, 0).



**Figure 5.** Negative: blue-greenish color. Neutral: gray scale color. Positive: yellow color.

### Voice Intonation

People modulates the intonation of their voices according to the context and also to add emphasis to the speech. Intonation is also correlated with the speaker's mood. In [54] the authors argue that the role of voice intonation in emotion expression is very important and show that when expressing anger, happiness or fear, the speech is uttered at a higher rate and pitch than when sadness or similar emotions are expressed.

We have associated the three emotions that our system implements to intonations portraying positive, neutral and negative moods. One of the limitations of the Pepper's

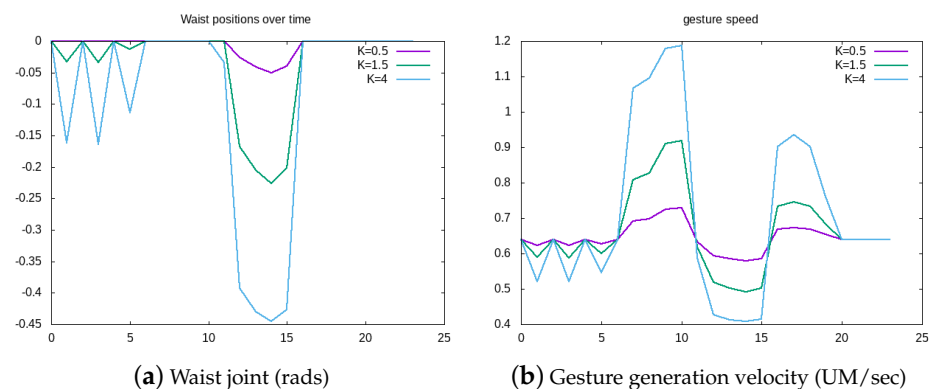
speech synthesizer is that it does not provide a way to directly manage voice intonation. But it is possible to adjust some voice parameters, namely volume, speed and pitch. In our approach we change those parameters according to the emotion value of *Valence*.

#### 4. Adaptive Personality

The behavior produced by the previous steps produce merely reactive expressions, i.e., each sentence's sentiment level is originated at the position where the previous sentence ended.

There is a correspondence to a humanlike manner of expressing text connotations in the sense that each action/perception we take has an impact in our mood or emotional state that, at the same time determines the intensity with which the next action is performed. However, the same text is always narrated with the same voice intonation and body gesturing except for the arms and head yaw motion generated by the GAN.

A straightforward modification allows the robot to show a constrained or exaggerated behavior, depending on the personality we want to assign to the robot. By applying sigmoid functions to each of the expression features on the compound score and adjusting each output to keep the operational interval of each feature, different levels of expression can be achieved, just by adjusting the gain of the exponential. Figure 6 shows the plots of the waist pitch angle and the speed reproduction of the generated arms movements for several exponential ( $K$ ) values. The plots were recorded for the compound score given by VADER over "The Color Monster" tale.



**Figure 6.** An example of the evolution of waist joint position, and gesture generation velocity.

#### 5. Results

In order to show the obtained robot performance, two different scenarios have been defined to emphasize the different aspects of the modulated talking behavior.

##### 5.1. Scenario 1:

In this scenario the robot pronounces the definition of the word *rainbow* extracted from Wikipedia. This text, which can be considered as neutral, has been manually annotated to force VADER to give negative, neutral or positive sentiment to allow all the sad-neutral-happy continuum to be compared. The main goal of this setup is to show how the adjustment of the exponential gain affects the personality of the robot for the three sentiments extracted by the analyzer. Two different videos allows us to compare the behavior of the robot while keeping the polarity of the text in a concrete state. The first video <https://www.youtube.com/watch?v=8-t-URpHsiQhttp> (accessed on 18 May 2021) corresponds to discrete personality ( $K = 1.5$ ) while the second <https://www.youtube.com/watch?v=1ggN0lstFg4> (accessed on 18 May 2021), recorded with  $K = 4.0$  shows a nervous or even histrionic behavior. Clearly, whatever the nature of the text, the histrionic version shows more extreme expression levels than the discrete one, that will indeed fit with a more natural behavior.



### 5.2. Scenario 2:

A different scenario is set in which the sentiment of the text varies over time and so they do the different talking features of the robot. Several passages from Anna Llenas' story entitled *The Color Monster* have been chosen to demonstrate the robot's performance. In this story, the protagonist is a sweet monster who wants to explain how he feels and uses colour to do so. Allegedly we can not detach the expected expression from what is being said, and this tale has shown to be a perfect tool to evaluate correspondences between meaning and generated expression. Definitively this is a more realistic scenario, appropriate for a storytelling robot. The recorded videos shows how the robot adapts its expression level according to each sentence, showing a progressive modulation of the different features again for the two identities, the discrete <https://www.youtube.com/watch?v=eovp55f1jhs> (accessed on 18 May 2021) and the histrionic <https://www.youtube.com/watch?v=s01noQ1u1jM> (accessed on 18 May 2021) one.

## 6. Conclusions and Further Work

In this work has been presented an adaptive system that implements a expression behavior for humanoid robots. Beat talking gestures are generated, along with other non-verbal clues, as head positioning, modulation of the voice tone, color of the eyes and speed of the movements of the arms. All of these features are related to the sentiment expressed in the sentences spoken by the robot.

On the basis of the obtained results, we consider that the presented approach gives good results and allows to emphasize or fade independent aspects of the robot's expression. It somehow facilitates the adaptation of the robot's personality according to the audience or event. Note that each feature is independent and thus, can be adjusted as such giving raise to multiple different expression behaviors.

VADER has shown to be a valuable tool but rather insensitive to subtle sentences. It behaves as it has been trained to, i.e, to categorize individual short sentences, and thus, it does not take into account either the relation among consecutive sentences or the semantic meaning. Training a more specific classifier should minimize unexpected results and eliminate the text annotation we do using emoticons to modify the values of the compound scores. Moreover, keeping the robot's emotional state as a function of other affective inputs could serve to dynamically adapt  $K$  to modify robot's mood in execution time.

Generated movements are executed by direct kinematics. This does not allow us to shorten the motion area of the arms and instead, the taken approach reduces the number of movements per second generated as a solution. The adjustment of the attainable configuration space would eliminate strange beat gestures when the chosen identity requires it.

Instead of relying solely in the sentimental analysis of the text being verbalized, richer and more insightful signals must be considered to commit with the affective loop requirements. Complying with a dimensional model will also enrich the emotion spectrum and extend the sadness-happiness continuum.

Finally, it would be possible to use multimodal data to train a generative model and produce affectively touched talking features. The comparison between the automatic behavior generated by the model and the one here proposed would be then mandatory. In any case, there is a need for a evaluation in a public performance in order to evidence that observers are affected by the emotional performance of the robot.

**Author Contributions:** Conceptualization, U.Z., I.R., J.M.M.-O. and E.L.; methodology, U.Z., I.R., J.M.M.-O. and E.L.; software, U.Z., I.R., J.M.M.-O. and E.L.; writing—original draft preparation, U.Z., I.R., J.M.M.-O. and E.L.; writing—review and editing, U.Z., I.R., J.M.M.-O. and E.L.; project administration, U.Z., I.R., J.M.M.-O. and E.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has been partially supported by the Basque Government (IT900-16 and Elkartek 2018/00114), the Spanish Ministry of Economy and Competitiveness (RTI 2018-093337-B-100, MINECO/FEDER,EU)

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Goodrich, M.A.; Schultz, A.C. Human-robot interaction: A survey. *Found. Trends Hum. Comput. Interact.* **2007**, *1*, 203–275. [CrossRef]
2. Sheridan, T.B. Human–Robot interaction: Status and challenges. *Hum. Factors* **2016**, *58*, 525–532. [CrossRef] [PubMed]
3. Fong, T.; Nourbakhsh, I.; Dautenhahn, K. A survey of socially interactive robots. *Robot. Auton. Syst.* **2003**, *42*, 143–166. [CrossRef]
4. Robert, L.; Alahmad, R.; Esterwood, C.; Kim, S.; You, S.; Zhang, Q. A Review of Personality in Human–Robot Interactions. *Found. Trends Inf. Syst.* **2020**, *4*, 107–212. [CrossRef]
5. Yale University. Socially Assistive Robotics. Available online: <http://robotshelpingkids.yale.edu/> (accessed on 18 May 2021).
6. Feil-Seifer, D.; Mataric, M.J. Defining Socially Assistive Robotics. In Proceedings of the International Conference on Rehabilitation Robotics, Chicago, IL, USA, 28 June–1 July 2005; pp. 465–468.
7. Groopman, J. Robots that Care. Advances in technological therapy. *The New Yorker*, 2 November 2009.
8. Cañamero, L.; Lewis, M. Making New “New AI” Friends: Designing a Social Robot for Diabetic Children from an Embodied AI Perspective. *Soc. Robot.* **2016**, *8*, 523–537. [CrossRef]
9. Čaić, M.; Mahr, D.; Oderkerken-Schröder, G. Value of social robots in services: Social cognition perspective. *J. Serv. Mark.* **2019**, *33*, 463–478. [CrossRef]
10. Søraa, R.A.; Nyvoll, P.; Tøndel, G.; Fosch-Villaronga, E.; Serrano, J.A. The social dimension of domesticating technology: Interactions between older adults, caregivers, and robots in the home. *Technol. Forecast. Soc. Chang.* **2021**, *167*, 120678. [CrossRef]
11. Lia, S.; Wynsberghe, A.V.; Roeser, S. The Complexity of Autonomy: A Consideration of the Impacts of Care Robots on the Autonomy of Elderly Care Receivers. *Cult. Sustain. Soc. Robot. Proc. Robophilosophy 2020* **2021**, 335, 316.
12. Churamani, N.; Kalkan, S.; Gune, H. Continual Learning for Affective Robotics: Why, What and How? In Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Online, 31 August–4 September 2020; pp. 209–223.
13. Paiva, A.; Leite, I.; Ribeiro, T. Emotion Modelling for Social Robots. In *The Oxford Handbook of Affective Computing*; Oxford University Press: New York, NY, USA, 2015; pp. 296–308.
14. Höök, K. Affective loop experiences: Designing for interactional embodiment. *Philos. Trans. R. Soc. B* **2009**, *364*, 3585–3595. [CrossRef]
15. Damiano, L.; Dumouchel, P.; Lehmann, H. Towards human–robot affective co-evolution overcoming oppositions in constructing emotions and empathy. *Int. J. Soc. Robot.* **2015**, *7*, 7–18. [CrossRef]
16. Damiano, L.; Dumouchel, P.G. Emotions in Relation. Epistemological and Ethical Scaffolding for Mixed Human-Robot Social Ecologies. *Humana Mente J. Philos. Stud.* **2020**, *13*, 181–206.
17. Crumpton, J.; Bethel, C.L. A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech. *Int. J. Soc. Robot.* **2016**, *8*, 271–285. [CrossRef]
18. Knight, H. Eight Lessons learned about Non-verbal Interactions through Robot Theater. In *Social Robotics, Proceedings of the Third International Conference, ICSR 2011, Amsterdam, The Netherlands, 24–25 November 2011*; Lecture Notes in Computer Science; Mutlu B., Bartneck, C., Ham, J., Evers, V., Kanda, T., Eds.; Springer: Berlin, Germany, 2011; Volume 7072, pp. 42–51. [CrossRef]
19. Ritschel, H.; Kiderle, T.; Weber, K.; Lingenfeller, F.; Baur, T.; André, E. Multimodal Joke Generation and Paralinguistic Personalization for a Socially-Aware Robot. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness, Proceedings of the PAAMS Collection—18th International Conference, PAAMS 2020, L’Aquila, Italy, 7–9 October 2020*; Lecture Notes in Computer Science; Demazeau, Y., Holvoet, T., Corchado, J.M., Costantini, S., Eds.; Springer: Berlin, Germany, 2020; Volume 12092, pp. 278–290. [CrossRef]
20. Neff, M.; Kipp, M.; Albrecht, I.; Seidel, H.P. Gesture Modeling and Animation Based on a Probabilistic Re-creation of Speaker Style. *ACM Trans. Graph.* **2008**, *27*, 5:1–5:24. [CrossRef]
21. Cassell, J.; Vilhjálmsón, H.H.; Bickmore, T. Beat: The behavior expression animation toolkit. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; pp. 477–486.
22. Ahuja, C.; Lee, D.W.; Nakano, Y.I.; Morency, L.P. Style transfer for co-speech gesture animation: A multi-speaker conditional-mixture approach. In Proceedings of the 28th European Conference on Computer Vision, Online, 23–28 August 2020; pp. 248–265.
23. Bozkurt, E.; Yemez, Y.; Erzin, E. Affective synthesis and animation of arm gestures from speech prosody. *Speech Commun.* **2020**, *119*, 1–11. [CrossRef]
24. Hanson, D. Hanson Robotics. Available online: <http://www.hansonrobotics.com> (accessed on 18 May 2021).
25. Ishiguro, H. Hiroshi Ishiguro Laboratories (ATR). Available online: <http://www.geminoid.jp/en/index.html> (accessed on 18 May 2021).
26. Breazeal, C. Emotion and sociable humanoid robots. *Int. J. Hum. Comput. Stud.* **2003**, *59*, 119–155. [CrossRef]
27. Al Moubayed, S.; Skantze, G.; Beskow, J. The furhat back-projected humanoid head-lip reading, gaze and multi-party interaction. *Int. J. Humanoid Robot.* **2013**, *10*. [CrossRef]
28. Anki. Cozmo. Available online: <https://www.digitaldreamlabs.com/pages/cozmo> (accessed on 18 May 2021).

29. Pelikan, H.R.; Broth, M.; Keevallik, L. “Are you sad, Cozmo?” How humans make sense of a home robot’s emotion displays. In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, 23–26 March 2020; pp. 461–470.
30. Bretan, M.; Hoffman, G.; Weinberg, G. Emotionally expressive dynamic physical behaviors in robots. *Int. J. Hum. Comput. Stud.* **2015**, *78*, 1–16. [[CrossRef](#)]
31. Augello, A.; Infantino, I.; Pilato, G.; Rizzo, R.; Vella, F. Binding representational spaces of colors and emotions for creativity. *Biol. Inspired Cogn. Archit.* **2013**, *5*, 64–71. [[CrossRef](#)]
32. Infantino, I.; Pilato, G.; Rizzo, R.; Vella, F. I Feel Blue: Robots and Humans Sharing Color Representation for Emotional Cognitive Interaction. In Proceedings of the Biologically Inspired Cognitive Architectures 2012, Palermo, Italy, 31 October–3 November 2012; Advances in Intelligent Systems and Computing; Chella, A., Pirrone, R., Sorbello, R., Jóhannsdóttir, K., Eds.; Springer: Berlin, Germany, 2011; Volume 196, pp. 161–166. [[CrossRef](#)]
33. Song, S.; Yamada, S. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In Proceedings of the 2017 12th ACM/IEEE International Conference on Human-Robot Interaction HRI, Vienna, Austria, 6–9 March 2017; pp. 2–11.
34. Feldmaier, J.; Marmat, T.; Kuhn, J.; Diepold, K. Evaluation of a RGB-LED-based Emotion Display for Affective Agents. *arXiv* **2016**, arXiv:1612.07303.
35. Johnson, D.O.; Cuijpers, R.H.; van der Pol, D. Imitating human emotions with artificial facial expressions. *Int. J. Soc. Robot.* **2013**, *5*, 503–513. [[CrossRef](#)]
36. Paradedá, R.B.; Hashemian, M.; Rodrigues, R.A.; Paiva, A. How Facial Expressions and Small Talk May Influence Trust in a Robot. In *Social Robotics, Proceedings of the 8th International Conference, ICSR 2016, Kansas City, MO, USA, 1–3 November 2016*; Lecture Notes in Computer Science; Agah, A., Cabibihan, J.J., Howard, A., Salichs, M., He, H., Eds.; Springer: Berlin, Germany, 2016; Volume 9979, pp. 169–178. [[CrossRef](#)]
37. Aly, A.; Tapus, A. Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction. *Auton. Robot.* **2016**, *40*, 193–209. [[CrossRef](#)]
38. Aly, A.; Tapus, A. On designing expressive robot behavior: The effect of affective cues on interaction. *SN Comput. Sci.* **2020**, *1*, 1–17. [[CrossRef](#)]
39. Huang, C.M.; Mutlu, B. Modeling and Evaluating Narrative Gestures for Humanlike Robots. In Proceedings of the Robotics: Science and Systems, Berlin, Germany, 24–28 June 2013. [[CrossRef](#)]
40. Alexanderson, S.; Székely, É.; Henter, G.E.; Kucherenko, T.; Beskow, J. Generating coherent spontaneous speech and gesture from text. In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, Online, 20–22 October 2020; pp. 1–3.
41. Kucherenko, T.; Hasegawa, D.; Kaneko, N.; Henter, G.E.; Kjellström, H. Moving fast and slow: Analysis of representations and post-processing in speech-driven automatic gesture generation. *Int. J. Hum. Comput. Interact.* **2021**, 1–17. [[CrossRef](#)]
42. Ekman, P. Are there basic emotions? *Psychol. Rev.* **1992**, *99*, 550–553. [[CrossRef](#)]
43. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [[CrossRef](#)]
44. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. [[CrossRef](#)] [[PubMed](#)]
45. Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [[CrossRef](#)]
46. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June, 2014; Volume 8.
47. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O’Reilly Media, Inc.: Sebastopol, CA, USA, 2009.
48. Akbik, A.; Blythe, D.; Vollgraf, R. Contextual String Embeddings for Sequence Labeling. In Proceedings of the COLING 2018, 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 1638–1649.
49. McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*; University of Chicago Press: Chicago, IL, USA, 1992.
50. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
51. Cao, Z.; Hidalgo Martinez, G.; Simon, T.; Wei, S.; Sheikh, Y.A. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186. [[CrossRef](#)]
52. Zabala, U.; Rodriguez, I.; Martínez-Otzeta, J.M.; Irigoien, I.; Lazkano, E. Quantitative analysis of robot gesticulation behavior. *Auton. Robot.* **2021**, *45*, 175–189. [[CrossRef](#)]
53. Rodriguez, I.; Martínez-Otzeta, J.M.; Lazkano, E.; Ruiz, T. Adaptive Emotional Chatting Behavior to Increase the Sociability of Robots. In Proceedings of the International Conference on Social Robotics, Tsukuba, Japan, 22–24 November 2017; pp. 666–675.
54. Bänziger, T.; Scherer, K.R. The role of intonation in emotional expressions. *Speech Commun.* **2005**, *46*, 252–267. [[CrossRef](#)]